

Doctoral Dissertation

**A Study on Impact Estimation of Hosting Mega-Sport
Events through Machine Learning Models**

機械学習モデルを用いたメガスポーツイベント開催の
影響推定に関する研究

March 2026

Napitiporn Manoli

Graduate School of Sciences and Technology for Innovation
Yamaguchi University

ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest gratitude to my academic advisor, Professor Dr. Masakazu Takahashi, whose guidance extended far beyond this thesis. His unwavering support and mentorship throughout my three years at the Graduate School of Sciences and Technology for Innovation, Yamaguchi University, and his assistance in navigating life in Japan have been invaluable. I am equally indebted to Professor Yoshiyuki Matsuura, whose profound contributions have shaped me into a capable researcher, prepared to embrace future challenges in the international academic arena.

I also extend my sincere appreciation to my examining committee for their insightful feedback and constructive critique, which significantly enhanced the quality of this dissertation. Additionally, I am grateful to all faculty and staff members of the Graduate School of Innovation and Technology Management for fostering an environment of excellence that has enriched my academic journey.

I am deeply grateful to the Support for Pioneering Research Initiated by the Next Generation (SPRING) program of the Japan Science and Technology Agency (JST) for providing a scholarship that facilitated my doctoral studies at Yamaguchi University. Beyond financial support, SPRING's commitment to fostering international academic exchange and interdisciplinary collaboration has been transformative, equipping me with invaluable experiences and professional connections that will serve as the cornerstones of my future academic endeavors.

I must also express my profound gratitude to Japan, a country that has offered me the most remarkable chapter of my life. Japan has profoundly transformed my perspective and enriched my life. The connections I have forged here have not only shaped my professional trajectory but will continue to influence my personal growth for the rest of my life.

Finally, I extend my heartfelt appreciation to my family, whose steadfast support and encouragement have been a constant source of strength throughout this journey.

TABLE OF CONTENTS

ABSTRACT	iii
List of Figures	v
List of Tables	vi
List of Abbreviations	vii
Chapter 1: Introduction	1
Chapter 2: Literature Review	7
2.1 Introduction	7
2.2 Related Work	8
2.3 Research Methodology Review	13
2.4 Research Gap	19
2.5 Conceptual framework	20
2.6 Concluding Remarks	20
Chapter 3: Machine Learning Models for Estimating the Economic and Innovation Impacts	23
3.1 Introduction	23
3.2 Related Work	24
3.3 Data Summary	27
3.4. Analytical Result	28
3.5 Discussion	36
3.6 Concluding Remarks	38
Chapter 4: Machine Learning Models for Estimating Digital Development Impacts	39
4.1 Introduction	39
4.2 Related Work	40
4.3 Data Summary	44
4.4 Analytical Result	44
4.5 Discussion	55
4.6 Concluding Remarks	57
Chapter 5: Traditional Statistical Techniques and Machine Learning Models for Estimating Economic and Innovation Impacts	59
5.1 Introduction	59
5.2 Related Work	60
5.3 Data Summary	63
5.4 Analytical Result	65
5.5 Discussion	84
5.6 Concluding Remarks	86
Chapter 6: Concluding Remarks	89
References	97
Appendix A- Glossary of Terms	105
Appendix B- List of Key Publications	107
Appendix C- Conference Presentations	108

ABSTRACT

Mega-sport events (MSEs) play a significant role in the global landscape by contributing to multiple socioeconomic domains. Prestigious competitions, such as the Olympic Games and FIFA World Cup, are frequently leveraged by host nations to elevate their international standing while generating substantial local and global economic benefits. Simultaneously, the economic activities surrounding MSEs often align with a neoliberal urban agenda, enhancing competitiveness through extensive public and private investments and transforming urban governance into an entrepreneurial model. Another notable effect of hosting MSEs is the substantial benefit of digital advancements that enhance athletic performance, spectator experiences, operational efficiency, and stakeholder engagement. Furthermore, the digital evolution spurred by these events has strengthened connectivity across both sports-related and non-sports sectors. Beyond immediate benefits, innovations foster regeneration and reflect the host nation's capacity to leverage MSEs for long-term development.

A multi-decade panel dataset covering several host nations and multiple MSEs was employed. The methodological framework combined traditional statistical techniques (t-tests, fixed-effects regression, and negative binomial regression) with a suite of machine learning (ML) approaches, including decision trees, stepwise regression, support vector machines, boosted trees, random forests, neural boosting, k-nearest neighbors, and time-series forecasting. Explainable AI methods such as shapley additive explanations (SHAP), partial dependence plots (PDP), and individual conditional expectation (ICE) were applied to enhance interpretability and to identify threshold effects and variable interactions. This integrated design provides a replicable toolkit for assessing multidimensional, nonlinear, and context-specific impacts.

This dissertation is organized into six chapters. Chapter 1 outlines the research background, objectives, and significance, and Chapter 2 reviews the relevant literature, highlights methodological limitations, and identifies research gaps. Chapter 3 applies ML models to estimate the economic and innovation impacts of hosting MSEs in a region. Chapter 4 examines their influence on digital development in depth. Chapter 5 integrates traditional statistical techniques and ML approaches to analyze the effects of ESG parameters and financial factors on economic growth and innovation capability in host countries. Finally, Chapter 6 presents the conclusions, highlighting the key findings, contributions, limitations, and future research directions.

The empirical findings were synthesized across three thematic domains. Economically, MSEs influence outcomes both directly and indirectly, with population size, GDP change, and event-day attendance emerging as the strongest predictors of value-added. Environmental and social factors contribute significantly to GDP per capita growth through nonlinear effects. In the digital domain, high digital evolution scores (DES) are strongly associated with hosting status, confirming that host nations tend to achieve higher DES levels than non-host nations. These findings suggest that the digital legacy of MSEs depends not only on infrastructure investments but also on the host nation's capacity to leverage global knowledge flows and strengthen its competitive position in the long run. Economic factors, such as export volumes, charges for the use of intellectual property, and trade in goods and services, emerged as critical contributors to digital advancement, underscoring the role of trade integration and intangible assets in accelerating the digital transformation.

Innovation capability demonstrated consistent improvement, as evidenced by increasing patent activity, with notable disruptions only during periods of global crisis. Social indices, particularly inclusivity, have consistently emerged as drivers of innovation capability, as reflected in the regression models. Environmental and governmental factors contributed significantly to the ML models, whereas population size and operational expenditure emerged as dominant predictors across the approaches. Comparative analysis revealed that ML models outperformed traditional statistical approaches, particularly in terms of predictive accuracy and the ability to model nonlinear relationships.

This study addresses the limitations of traditional qualitative and statistical approaches by introducing a replicable framework that integrates statistical techniques, advanced machine learning, and XAI to comprehensively evaluate the impact of MSEs. The framework provides a methodological pathway for capturing the complex and long-lasting economic, digital, and innovation outcomes of hosting, thereby offering both academic and policy relevance for future research. Beyond scholarly contributions, it equips international organizers and host nations with the tools to embed sustainability, digital readiness, and inclusivity benchmarks into event planning. It further guides sponsors to align investments with technology, renewable energy, and community development, while enabling policymakers to redirect resources toward R&D, digital infrastructure, and green energy. By integrating ESG principles with advanced analytics, MSEs can be strategically leveraged as catalysts for innovation, digital transformation, and sustainable development at the national and global levels.

List of Figures

1.1	A Dissertation's Conceptual Framework	6
2.1	A Typology of Sports Events by Parent and Chappelet	7
2.2	The Number of Large-Scale, One-Off Sports Events Hosted by Each Country	9
3.1	Plot of The Number of Patent Applications In 1981-2023	33
3.2	Comparison of AIC Values Across Forecasting Models	35
3.3	Comparison of SBC Values Across Forecasting Models	35
3.4	Annual Forecasting Using the Seasonal Exponential Smoothing Model	36
3.5	Semi-Annual Forecasting Using the Moving Average Model	36
4.1	One-way Analysis of Digital Evolution Score by Non-Host and Host Country	45
4.2	Neural Boosting Model Diagram	47
4.3 (a)	Prediction Profiler with Low Digital Evolution Score	47
4.3 (b)	Prediction Profiler with High Digital Evolution Score	47
4.4 (a)	The Residual by Predicted Plot of The KNN Model Applied to Non-Hosted Nations	50
4.4 (b)	The Residual by Predicted Plot of The KNN Model Applied to Hosted Nations	50
4.5	One-way Analysis of Non-Host Nations, Before and After 2016	51
4.6	One-way Analysis of Host Nations Before and After 2016	51
4.7	Average Digital Evolution Score (DES) by Host Countries	53
4.8	Actual vs. Predicted Plot Before and After 2016	54
4.9	Comparison of Predictor Contributions	55
5.1	Predictor Screening of ESG on Patent Number of Host and Non-Host Countries	67
5.2	Actual Vs Predicted of GDP Growth	69
5.3	Leverage Plots of GDP Growth for Environmental, Social and Government	69
5.4 (a)	Actual vs. Predicted Plots of Environmental Factor for GDP Growth	73
5.4 (b)	Actual vs. Predicted Plots of Social Factor for GDP Growth	73
5.4 (c)	Actual vs. Predicted Plots of Governance Factor for GDP Growth	73
5.5 (a)	Actual vs. Predicted Plots of Environmental factors for Patent Application	77
5.5 (b)	Actual vs. Predicted Plots of Social Factors for Patent Application	77
5.5 (c)	Actual vs. Predicted Plots of Governance Factors for Patent Application	77
5.6	The Actual Predicted Plot of Patent Applications	80
5.7	The Shapley Additive Explanations (SHAP) Plot	82
5.8 (a)	The PDP and ICE Plot of the Operational Expenses	82
5.8 (b)	The PDP and ICE Plot of the Other Expenses	82

List of Tables

2.1 Comparison of Traditional vs. Machine Learning-Based Impact Estimation Models	15
3.1 Economic and Sport-Related Variable Description	27
3.2 Importance Value Range	28
3.3 Predictor Screening and Variable Importance Analysis Statistic Results	29
3.4 Nonlinear Regression Analysis Statistic Results	30
3.5 Comparative Analysis of ML Models' Performance	31
3.6 Search Strategy of Patent Applications	32
3.7 Patent Trend Directions and Influencing Factors for FIFA World Cups (2006–2022)	33
3.8 Statistical Data of The Time Series Analysis	34
4.1 Comparison of Prior Literature and The Present Study	42
4.2 Data Description of Host, Non-Host, and DES	44
4.3 Descriptive Statistics of DES for Host and Non-Host Countries	45
4.4 Prediction Performance of The Neural Boosted Model.	47
4.5 Neural Boosting Model's Confusion.	47
4.6 Performance of The Neural Boosted Model For DES Prediction	48
4.7 Performance of The KNN Model for DES Prediction	48
4.8 Descriptive Statistics of DES Before and After 2016	51
4.9 Statistic Description of XGboost Model for DES Prediction Before and After 2016	54
5.1 Sample of Mega-Sport Events Included in the Analysis	63
5.2 Data Summary for Economic Growth and IC Prediction by ESG Factors	63
5.3 Data Summary for IC Prediction by ESG and Financial Factors	64
5.4 Descriptive Statistics of ESG–Patent Relationships in MSE Host and Non-Host Countries	65
5.5 Regression Results of ESG Dimensions on GDP Growth	69
5.6 Regression Results of ESG Dimensions on Innovation Capability	71
5.7 Hypothesis Summary for ESG on Economic Growth and Innovation Capability	72
5.8 Mediation Testing Statistic Result	72
5.9 Bootstrap Forest Prediction Performance for GDP Growth	73
5.10 The Contribution of ESG Factors on GDP Growth Statistic Result	75
5.11 Bootstrap Forest Prediction Performance for Patent Application	75
5.12 The Contribution of ESG Factors on Patent Application Statistic Result	77
5.13 Predictor Screening Statistic Result for IC Prediction	79

List of Abbreviations

DES	Digital Evolution Score
ESG	Environmental, Social, and Governance
FIFA	Fédération Internationale de Football Association
GDP	Gross Domestic Product
IC	Innovation Capability
ICE	Individual Conditional Expectation
IOC	International Olympic Committee
KNN	K-Nearest Neighbors
MA	Moving Average
ML	Machine Learning
MSEs	Mega Sport Events
PDP	Partial Dependence Plots
R&D	Research and Development
SES	Seasonal Exponential Smoothing
SHAP	Shapley Additive Explanations
SVM	Support Vector Machines
UEFA	Union of European Football Associations
XAI	Explainable AI

Chapter 1: Introduction

Mega-sport event (MSE) is defined as a feature of top-tier international athletes, being extensively broadcast, attracting massive spectator interest, and generating wide-ranging benefits for host nations and related industries [1]. Examples include the Olympic Games, Fédération Internationale de Football Association (FIFA) World Cup, and the Union of European Football Associations (UEFA) European Championship. Other prominent competitions that attract global attention include the Tour de France, Super Bowl, and World Championships in athletics, swimming, and skiing. MSEs are recognized for their entertainment value and significant social, cultural, political, economic, and historical importance [2].

Consequently, the structural characteristics of MSEs, whether repeatedly hosted in the same location or rotated across nations, have significant implications. Their exceptional visibility, fixed timelines, and concentrated resource mobilization make them a critical mechanism for examining how countries respond to opportunities and constraints. Mega-sporting events are also defined as ‘ambulatory occasions of a fixed duration that attract a large number of visitors, have a large mediated reach, come with large costs, and have substantial impacts on the built environment and the population [3].

Since the early 2000s, interest in sport event legacies or impacts has increased substantially, and the pursuit of both sport and non-sport legacies from hosting large-scale events has become a central theme in the rhetoric and practices of bid committees, governing bodies, and host governments [4-7]. Solberg and Preuss [8] defined sport event impacts as the ‘planned and unplanned, positive and negative, tangible and intangible structures created for and by a sport event that remain longer than the event itself’. These impacts are typically linked to large-scale events that necessitate substantial investments in infrastructure and urban development.

The continued promotion of sports event legacies by governing bodies and host city proponents has stimulated a growing body of theoretical and empirical research across multiple disciplines, including economics, tourism, physical activity, sociocultural studies, psychology, and political science [9]. Although there is generally a significant financial cost associated with hosting sporting events, there is typically intense competition for the opportunity to host such events because of the perceived anticipated positive benefits to the host. These potential benefits include growth in tourism and employment [10, 11], opportunities for place promotion and marketing of the host community as a destination [12], and improvement in the quality of life of local citizens and their social networks [13].

The literature shows that social and cultural impacts remain the most frequently analyzed (99 studies), with an emphasis on community interaction, cultural promotion, and related dynamics. Economic impacts, examined in 60 studies, highlighted the ripple effects on tourism, employment, and local economies. Environmental impacts, though less studied (28 studies), addressed issues such as carbon emissions and sustainability. Recently, a growing body of research has begun to examine the impact of innovation, particularly technological advancement and digital infrastructure development, positioning MSEs as potential catalysts for broader transformation. A systematic review by Elahi *et al.* [14] confirms that over the past three decades, scholarship has concentrated heavily on economic and social dimensions. Their review identified five primary areas of investigation: economic, social, cultural, political, and environmental impacts. Similarly, the consequences examined in studies on event-attributed structural changes have primarily focused on the economy, accounting for nearly half of the research, followed by social, urban, environmental, and political dimensions [15].

However, the literature reveals a critical imbalance: most studies have narrowly emphasized specific positive dimensions while overlooking broader linkages among economic outcomes, digital evolution, and innovation capability. Equally important, insufficient attention has been given to the consequences of inaccurate estimations, including underestimated financial risks, transportation burdens and long-term infrastructure management challenges.

This dissertation addresses themes in the MSE-hosting literature regarding the host nation's impact on MSE. First, economic impact was the most frequently discussed topic. Generally, studies highlight the economic benefits of event hosting through increased tax revenues, job creation, and ancillary revenues, such as visitor spending on goods and services, which help expand and diversify the local economy. Furthermore, local governments often argue that funds allocated for development projects would not be available without hosting the event [16]. Local governments often promote sporting events and tourism development based on economic benefits. With this focus, scholars examined visitor numbers, additional expenditure through a mega-event and its economic impacts on income and job growth, long-term growth paths, image improvements and host city perceptions [17-19]

It is no longer necessary to attend a mega event in person to experience it. Since the 1980s, the expansion of broadcasting has enabled most audiences to engage with these spectacles through television or digital screens rather than at the venue [20]. The financial value of broadcasting rights has mirrored this shift: for example, revenues for the Summer Olympics increased from USD 34.9 million in Montréal 1976 to USD 2,569 million in London 2012, equivalent to nearly a 23-fold rise in real terms. This dramatic growth reflects both the commercialization of mega-events and the rise of a global attention economy. The International Olympic Committee (IOC, 2014) reported that nearly half of the world's population, around 3.64 billion people, viewed at least one minute of coverage of the 2012 Games. Similarly, the scale of media involvement has grown substantially; from Barcelona 1992 to London 2012, the number of accredited journalists almost doubled, surpassing 24,000, which equates to more than two media representatives per athlete [21]. Together, these developments highlight how mega-events are now primarily consumed through the media rather than in person. However, the reliability of reported economic impacts, both positive and negative, associated with event planning [22-24].

Second, digital evolution impacts the mediated reach of mega-events, which is directly linked to advances in digital technologies such as broadcasting, thereby reflecting their global commercial value as media-driven spectacles [25]. The popularity of mega-events is partly correlated with their ability to attract tourists and large audiences. For instance, the Olympics and the FIFA World Cup not only generate substantial revenues from broadcasting rights but also serve as platforms for showcasing and implementing cutting-edge technologies, which, in turn, influence a wide range of sectors, including tourism, sports management, and urban infrastructure [26-27].

Advancements in wearables, performance analytics, and fan engagement applications are increasingly being adopted to improve athletic performance and enhance the spectator experience. Moreover, event infrastructure often requires extensive technological development and integration, such as transportation (e.g., innovative ticketing solutions and traffic management systems), enhanced urban planning (e.g., innovative city initiatives), and improved connectivity through digital platforms. For example, the 2022 FIFA World Cup in Qatar is highlighted as an instance of substantial infrastructure development, resulting in long-term benefits for urban development and tourism [28]. Qatar benefits from its status as an oil-rich nation with a substantial GDP, enabling large-scale infrastructure investment. However, compared to developed countries such as the United States, Qatar's telecommunications

infrastructure remains limited, with only two television stations and a single internet service provider [29]. This underscores the need to expand and modernize the sector in preparation for hosting the FIFA World Cup.

Third, the impact on innovation capability, technological advancement, and digital development extends beyond immediate improvements during the event; it fosters sustained innovation capability that supports long-term urban renewal and economic growth. MSEs have the potential to spur innovation, particularly in technology and infrastructure [30]. The demand for upgraded facilities compels cities to innovate in project planning and execution, generating advancements that resonate beyond the events [28]. Such developments often lead to improvements in urban spatial structures and enhance the competitiveness of host cities [31]. Hosting MSEs can elevate urban aesthetics and functionality, making previously neglected areas more attractive to investors and residents [32]. These infrastructure projects not only serve the event but also promote sustainable practices and lay the foundation for future research and development, often reflected in increased patent activity [33].

The influx of resources and global attention has accelerated investment in cutting-edge technologies, stimulating patents in event management, infrastructure design, and sustainability solutions [34]. Beyond their immediate applications, these technologies can contribute to the long-term sustainability of host cities by enhancing the efficiency of urban infrastructure, promoting greener construction practices, and improving digital connectivity [29]. However, such investments also pose challenges, including high maintenance costs, the risk of technological obsolescence, and uneven distribution of benefits across local communities [35]. Importantly, collaboration among businesses, universities, and governance bodies during MSE preparation fosters enduring innovation networks. Through joint ventures, research collaborations, and innovation hubs, these partnerships not only embed a culture of innovation but also shape more resilient and adaptive urban economies well beyond such events [36-37].

Inaccurate estimations of the costs associated with hosting MSEs present formidable challenges for organizing bodies, often resulting in resource misallocation, budget overruns, and ineffective legacy planning. Understanding the multifaceted nature of these inaccuracies is essential for assessing their consequences for various stakeholders. Underestimating costs frequently imposes severe financial strain on local governments and their sponsors. Azzali [30] noted that cost overruns are a recurring issue driven by unforeseen expenses, such as venue construction, security enhancements, and operational logistics. These overruns not only burden immediate organizers but may also divert funds from essential community projects, undermining long-term urban development and public service. In some cases, infrastructure constructed for an event remains underutilized after its conclusion, creating lasting economic liabilities rather than benefits [38]-[39]. Conversely, overestimating costs can lead to conservative budgeting, which restricts investment in areas crucial for event success. As Brynildsen and Parent [40] highlight, inflated estimates may reduce the available resources for infrastructure upgrades, security protocols, and hospitality services, thereby diminishing the overall visitor experience and potential legacy outcomes. For beneficiaries, these inaccuracies translate into rising hosting costs, unmet expectations, and growing skepticism regarding whether hosting decisions deliver the promised long-term benefits.

Consequently, the structural characteristics of MSEs, whether repeatedly hosted in the exact location or rotated across nations, have significant implications. With their exceptional visibility, fixed timelines, and concentrated resource mobilization, MSEs provide a critical mechanism for examining how countries respond to opportunities and constraints, particularly

in terms of economic, digital, and innovation impacts. Although extensive research exists on MSEs, significant imbalances remain across the impact domains. Economic studies often rely on short-term models that overlook long-term risks and structural legacies, while the roles of digital evolution and innovation capabilities remain underexplored and largely descriptive topics. Traditional qualitative and statistical approaches struggle to capture the multidimensional and nonlinear dynamics of MSE impacts. Inaccurate estimations create serious challenges for hosting organizations, as under- or overestimating costs can lead to misallocation of resources, budget overruns, and ineffective legacy planning. For instance, Davies and Coleman [41] highlighted the difficulty of accurately measuring attendance at free-to-view events, a challenge that undermines the validity of attendance-based economic models. These constraints highlight the need for more flexible, data-driven methodologies that can integrate diverse indicators and capture the complex, evolving patterns of MSE impacts. For beneficiaries, these inaccuracies translate into rising hosting costs, unmet expectations, and growing skepticism regarding whether hosting decisions deliver the promised long-term benefits.

Although the significance of economic, digital, and innovation evolution is well recognized, studies on MSEs that use predictive or projection-based methods to evaluate their legacy and long-term influence remain limited [42]. Therefore, this dissertation is positioned at the intersection of these two gaps: the underdeveloped methodological domain and the insufficiently examined impact of economic, digital evolution, and innovation capability in the literature. These limitations raise the question of how machine learning methodologies can be applied to more comprehensively estimate the economic, digital, and innovation impacts of MSEs, and whether such approaches provide deeper insights than traditional methods. Accordingly, this study is guided by four key objectives.

- 1) To evaluate the economic impacts of MSEs, with a particular focus on national performance indicators such as the total value added by economic activity and GDP per capita growth.
- 2) To examine the influence of MSEs on national innovation capability, patent applications are used as proxies.
- 3) To investigate the relationship between hosting MSE and national digital development, as measured by digital evolution score (DES).
- 4) To compare the predictive performance of machine learning models with traditional statistical approaches, and to determine their relative accuracy in assessing the multidimensional impacts of MSEs.

Addressing this methodological gap underscores the potential of machine learning (ML) to integrate diverse datasets, uncover hidden interrelations, and generate more precise estimates of how hosting MSEs shapes economic outcomes, digital development and innovation capabilities. Unlike traditional statistical models that rely on restrictive assumptions, ML approaches are inherently data-driven and nonparametric. They excel at detecting nonlinear patterns and high-order interactions across multiple factors, thereby offering a holistic, robust predictive framework. By learning directly from empirical data, ML models can navigate measurement challenges and account for contextual nuances, thereby producing more reliable estimates of the impact of MSE [41].

To bridge this gap, this study applies a suite of ML models to estimate the multifaceted environmental effects of hosting MSEs. This study advances the understanding of the economic, digital, and innovation legacies of such events, while demonstrating the methodological value

of ML as an alternative to descriptive and linear approaches. By delivering evidence-based predictions, this dissertation contributes to scholarly debates and more informed policy and strategic decision-making. This dissertation adopts a hybrid research design that integrates traditional statistical approaches with advanced machine learning and explainable AI (XAI) methods. Regression models, such as fixed-effects and negative binomial, provide benchmarks for linear relationships, whereas ML models, including stepwise regression, support vector machines (SVM), boosted trees, neural boosting, k-nearest neighbors (KNN), random forests, and time-series forecasting, capture nonlinear dynamics, threshold effects, and higher-order interactions. Predictor screening and XAI techniques further enhance interpretability, creating a replicable toolkit for analyzing multidimensional impacts of MSEs.

In line with the research objectives, this study anticipates several outcomes. First, mega-sport events are expected to generate measurable effects on national economic performance, as reflected in shifts in value-added activities and GDP per capita growth. Second, MSEs are projected to act as catalysts for innovation capability, with observable impacts on patenting and technological advancement. Third, the hosting status is expected to play a significant role in accelerating digital evolution, as countries leverage MSEs to enhance infrastructure, trade integration, and knowledge-based assets. Finally, machine learning approaches are expected to outperform traditional regression methods in terms of predictive accuracy by capturing nonlinear dynamics that conventional models often overlook. Taken together, these results suggest that MSEs should be understood not only as temporary events but also as potential drivers of long-term capability building. Their impact appears to be contingent upon governance quality, inclusivity, and the effectiveness of strategic financial allocation, thereby underscoring the conditions required for the realization of sustainable legacies.

This dissertation is organized into six chapters, followed by several appendices addressing the supplementary aspects of the study. Chapter 2 reviews the relevant literature and highlights its importance to this study. Chapter 3 details the application of machine learning and time-series forecasting techniques to estimate the economic and innovation impacts of hosting MSE. Chapter 4 presents the results on the relationship between hosting MSEs, both single- and multi-sport, and their influence on economic indicators and digital development, using machine learning methods such as neural boosting and K-nearest neighbors (KNN). Chapter 5 examines the environmental, social, governance, and MSE-related financial determinants, influence a host nation's economic growth and innovation capability through the application of traditional regression techniques and advanced machine learning methods. Chapter 6 concludes with a synthesis of the findings and a discussion of the limitations and directions for future research.

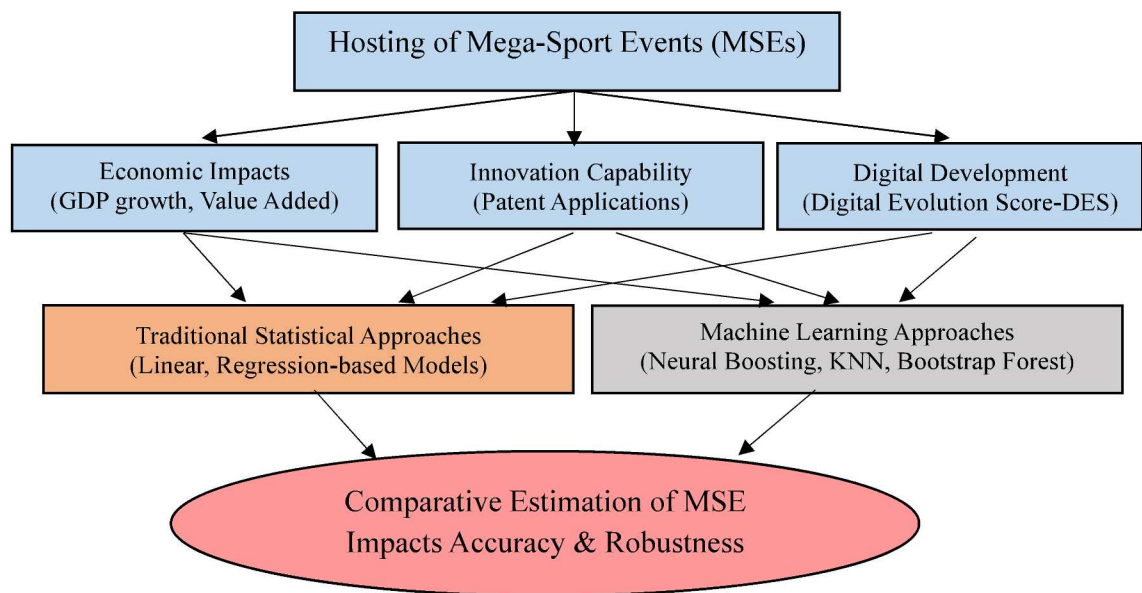


Figure 1.1 A dissertation's Conceptual Framework

Chapter 2: Literature Review

This chapter provides a comprehensive overview of the literature on mega-sport events (MSEs), exploring their definitions, global trends, socio-economic impacts, the role of innovation and technology in MSEs, and existing methodologies for estimating their impacts. This chapter serves as a foundation for understanding the broader area of MES research and the approaches used in impact estimation studies. Moreover, it highlighted some limitations of previous studies on traditional model analysis for forecasting or predicting impacts and how machine learning was employed to overcome these challenges. This chapter is structured into six sections: introduction, related work, research methodology, results from the machine learning literature, discussion, and concluding remarks.

2.1 Introduction

Mega-sport events are often categorized within the broader class of social and cultural mega-events, characterized by their prestige, large scale, and global renown, attracting attention across different social strata [43]. These events are typically one-off, long-term occasions that leave profound positive and negative impacts on host communities [9],[44]. While host applicants frequently emphasize their potential benefits, such claims are often based on optimistic or simplistic assumptions that overlook evidence that hosting can also generate adverse outcomes for local societies. In this regard, Müller noted that not every large-scale event qualifies as a “mega-event” [45]. What distinguishes mega-events from other major events is their sheer scale: in terms of size, investment, and impact, they are so substantial that they shape the entire community’s economy and command extensive global media coverage [46].

As depicted in Figure 2.1, sports events can be broadly categorized into seven types according to Parent and Chappelet's typology [47]. This research specifically focuses on 'one-off' events hosted by different countries on a rotational basis. This includes events organized by both non-profit and for-profit entities, as well as both mono-sport and multi-sport formats.

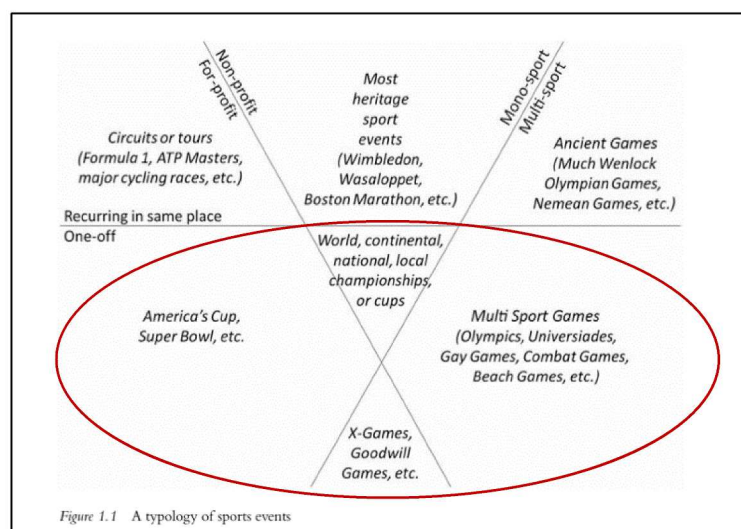


Figure 2.1 A Typology of Sports Events by Parent and Chappelet

Previous studies on the impact and legacy of mega-sport events have employed systematic reviews and empirical research, utilizing a range of methodological approaches. According to an exploratory search by Pickering and Byrne [48], a range of terms, including legacy, outcome, consequence, benefit, development, and transformation, are often used in conjunction with the term impact in the literature. Their review revealed that more than one hundred articles have focused on the impact and legacy outcomes of MSEs. These studies encompass empirical findings and secondary data analyses examining the consequences of event bids, hosting, and legacy initiatives implemented alongside event staging. The remaining articles addressed topics such as impact and legacy delivery, governance guidelines and policies, conceptualization, projection, frameworks, evaluation, and stakeholder perceptions[49]. Although a vast body of global literature addresses the impacts of mega sport events, there is notably scarce research focused on comprehensively predicting all potential impacts. Data-centric methodologies can play a remarkable role in this field.

This literature review section aims to (1) explore the essential challenges in understanding MSE's impact on host countries; (2) explore methodologies used to investigate MSE's impacts; and (3) discover the gaps, especially in MSE's impact estimation and prediction.

2.2 Related Work

This section synthesizes the literature on MSEs, reviewing the field's primary themes, including global trends, challenges, impacts, and legacies. Furthermore, it assesses the state of scholarship by examining standard research methodologies, theoretical frameworks, and prominent gaps in existing studies.

2.2.1 MSE's Global Trends and Challenges for Host Nations

The global proliferation of MSEs has significant economic and geopolitical consequences. However, despite extensive research on the impacts of individual events, a fundamental empirical understanding of their global distribution, frequency, and host characteristics remains underdeveloped.

However, one of the interesting studies in this field is a doctoral thesis by Lee [50], which examined the attribution of 'major' one-off sports events, such as world championships in Olympic sports and the Universal Multi-Sport Games. Lee's research shows that during the period from 1990 to 2012 (24 years), world championships for summer and winter Olympic sports and disciplines, as well as the Olympic Games, the Summer and Winter Universiade, and the World Games, totalling 857 sports events, were attributed to only 60 countries. This study highlights the challenges host countries face, as illustrated in Figure 2.2. Notably, developing countries have historically had limited experience hosting major sports events (fewer than 10). However, since 2000, these countries have increasingly secured such events to capitalize on their associated benefits.

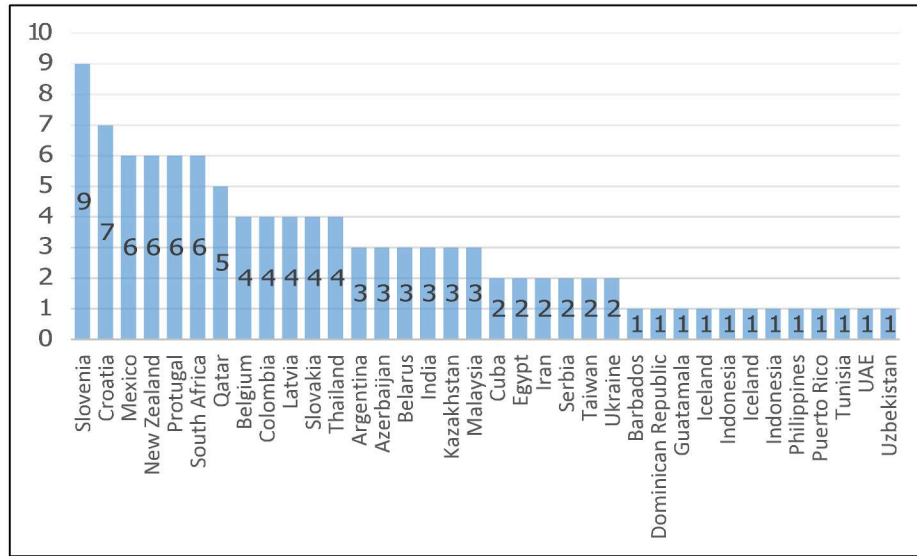


Figure 2.2 The Number of Large-Scale, One-Off Sports Events Hosted by Each Country

In contrast, four of these countries hosted more than 50 major events during the period, including Germany, Canada, the USA, and Italy, averaging more than two per year. Unsurprisingly, the top positions in this ranking are dominated by the wealthier G7 nations, as well as China, Russia, and Switzerland.

In recent years, hosting large-scale, one-off sporting events, hereafter referred to as 'mega sports events, has become increasingly challenging and fraught with risk, often leading to adverse media coverage for the host. Major sporting events, such as the 2013 Confederations Cup, 2014 FIFA World Cup, and 2016 Summer Olympics, all hosted in Rio de Janeiro, were marked by considerable public unrest. Similarly, the 2014 Winter Olympics in Sochi faced widespread criticism over concerns related to security, corruption, and human rights issues. There is a declining interest in bidding for the Summer and Winter Olympic Games, raising concerns for the IOC, which owns the Games, about their manageability and uniqueness[51]. Major world championships are no longer as desirable as they once were in the past.

The primary long-term challenge is to integrate mega-sport events with the evolving needs and ideals of modern society, which necessitates significant innovations. In recent years, the globalization of these events has experienced a major crisis influenced by developments in digitalization and sustainability. This change emphasizes the complicated interplay between mega-sport events and broader socioeconomic and environmental factors.

Therefore, navigating this era of heightened risk and shifting public sentiment demands a fundamental change in how the value of hosting mega-sport events is assessed. Consequently, the ability to accurately estimate and articulate the multidimensional impacts of these events, encompassing the full spectrum of social, environmental, economic, and political consequences, is paramount. A comprehensive, data-based approach is essential not only for managing risks but also for building a credible case for the massive public investment required. Ultimately, such rigorous analysis is crucial for aligning these global spectacles with the long-term strategic goals of a host nation and restoring public trust, which is essential for their future viability.

2.2.2 Impact Investigation in MES Studies

The conceptualization of this study began by addressing a core challenge in the literature: the inconsistent terminology used to define the '**impact**' of MSEs. To ensure a comprehensive review, our literature search utilized a broad range of synonymous terms, including effect, legacy, outcome, consequence, benefit, and transformation.

The most frequently discussed impact type is public life, politics, and culture, with 99 articles highlighting aspects such as community interaction, political outcomes and cultural engagement. This was followed by sports mass participation (65 articles), which emphasized grassroots participation and infrastructure to support increased involvement in sports. Economic legacies are the focus of 60 articles that address the financial impacts of various events. Other impact types include generalized impact discussions (41 articles), which address impact broadly without specific categorization, and built environment non-sporting impacts (39 articles), which focus on urban regeneration and infrastructure development. Environmental legacies, including sustainable design and practices, appeared in 28 articles. Sports' physical infrastructure, such as stadiums and facilities, was covered in 25 articles. Less frequently discussed types included health impacts, information and education, elite sports performance, and symbolic and historical aspects. However, there is no dedicated financial or administrative support for sports articles. This distribution reveals an emphasis on socio-cultural and participatory outcomes over administrative or managerial aspects in impact research [49].

A systematic review by Elahi *et al.* [14] revealed that research over the past three decades has predominantly focused on the economic and social effects of mega-sport events. The review identified five primary areas of investigation: economic, social, cultural, political, and environmental impacts. However, their study also highlighted a critical imbalance in the literature, arguing that this concentration has led to the neglect of several crucial and often adverse effects. Specifically, research gaps have been identified in financial risk, transportation burden, and long-term infrastructure management.

The review of prior studies on the three main impact focuses is offered as follows:

Economic Impact: This is frequently measured using indicators such as tourism influx, job creation, and local business revenues. Mega sports events, such as the FIFA World Cup, typically increase foreign tourist arrivals, stimulating economic growth in the host economy [52]. The measurement variables for assessing the economic impact of mega sports events are multifaceted and encompass a wide range of quantitative data points, including visitor spending, employment generation, changes in local economic activity, tax revenue increases, and long-term infrastructure investment. Visitor Spending is one of the most critical variables for evaluating the economic impact, as it quantifies attendees' expenditures. Tourist arrivals generate substantial revenue for local economies through diverse expenditure patterns, typically including accommodation, food and beverages, transport, retail purchases, and local attractions, which directly contribute to the overall economic dynamism of the host city [53].

Employment Generation assesses the economic impact by calculating employment effects alongside sales and income generated by events, demonstrating how the tourist influx translates into job opportunities across sectors such as hospitality, retail, and services [54]. Changes in Local Economic Activity can be measured by comparing local economic indicators pre- and post-event. Mega sports events play a role in shaping urban policy and economic structures, acting as catalysts for broader economic engagement beyond the immediate timeframe of the event [55]. Tax Revenue enhancement due to increased economic activity is a significant

measurement variable. Taxes generated by increased commercial activity during these events can support improved public services and infrastructure funding.

Infrastructure Investment requires assessing the long-term impacts on local economies, including investments that persist beyond the event duration [48]. It is also essential to consider sponsorship revenues and their associated economic multipliers, which determine how initial spending radiates through the economy. However, various measurement errors can often overstate the economic benefits associated with event-related spending, highlighting the complexities of accurately assessing multipliers [22]. These variables collectively provide a comprehensive framework for analyzing the economic consequences of mega sports events.

Digital Impact: MSEs can serve as powerful catalysts for technological advancement and digital transformation within a host region, often leaving behind a legacy of enhanced digital infrastructure and new business practices [56-57]. The development of applications for sports events is a critical element. For example, the GoPyeongChang Winter Olympic Official application integrated transportation infrastructures (e.g., the high-speed train called KTX, road/rail networks, and sharing car systems) with multiple public transport service providers (e.g., bus companies, subways, and taxis) to provide easier access via integrated digital services [57]. This dimension has become increasingly critical in the wake of global digital acceleration, a trend significantly amplified by the 2020 pandemic, which forced industries worldwide to adopt new technologies [58].

The measurement variables for evaluating the digital impact of mega sports events encompass quantitative metrics and qualitative assessments that reflect audience engagement, content reach, and the overall effectiveness of digital strategies. Social media metrics effectively measure campaign awareness by quantifying outreach impressions, the total number of people who viewed the campaign, and engagement metrics such as likes and shares [59]. Similarly, measuring the reach of digital content, particularly in video campaigns, provides valuable insights into strategic effectiveness [60]. Engagement levels can also be evaluated through click-through rates (CTR) and cost-per-click (CPC) associated with digital marketing campaigns [61]. Improved audience targeting and real-time performance insights can greatly enhance campaign evaluation and user engagement metrics [62].

Furthermore, sentiment analysis through social media provides insights into collective attitudes surrounding significant events, validating the importance of sentiment metrics in assessing the digital impact [63]. Beyond engagement metrics, financial indicators linked to digital productivity, such as increases in ticket sales, sponsorships, and advertising revenues, can signal the economic impact of digital strategies in mega sports events [64]. These measurement variables collectively provide a holistic view of how digital initiatives reinforce audience connections with mega sports events in the short term. However, long-term national measurements remain underexplored. Moreover, the use of digital indicators and big data analytics to measure the digital impact of mega sports events requires further research.

Innovation Capability: The ability to make major improvements and modifications to existing technology and to create new technologies that convert knowledge and resources into new products, processes, and services that create value [65, 66]. This capacity for technological advancement and digital development extends beyond immediate improvements, fostering sustained innovation that supports long-term urban renewal and economic growth.

At the national level, innovative capability represents a country's potential to develop and commercialize a steady stream of new-to-the-world technologies over the long term [65, 67, 68]. This depends significantly on a nation's common innovation infrastructure, whereby the development and commercialization of new technologies occur disproportionately in clusters

of geographic concentrations of interconnected companies and associated institutions in a particular field, linked by commonalities and complementarities [67, 69, 70].

The International Institute for Management Development (IMD) identifies critical factors for measuring the innovative capabilities and performance of nations, grouping competitiveness factors into four major categories: economic performance, government efficiency, business efficiency, and infrastructure [71]. National innovative capability depends on the strength of a nation's common innovation infrastructure, and science, technology, and innovation capability indexes serve as measures of a nation's relative ability to create and foster an enabling environment for economic stability and growth [72]. Research on Thailand's industrial innovation systems demonstrates this multidimensional nature, identifying five levels of innovation capability: (1) organizational innovation capability, (2) process innovation capability, (3) service innovation capability, (4) product innovation capability, and (5) marketing innovation capability. These dimensions provide comprehensive information for assessing the national innovation capability [71].

Mega sporting events (MSEs) serve as important catalysts for developing innovation capabilities in host nations through strategic management and targeted communication initiatives. The organizing committee of Beijing 2022, for instance, published the "Low-carbon Management Pre-Games Report of the Beijing 2022 Games," which led to valuable technological innovations and breakthroughs in venue construction while emphasizing sustainability goals [73]. The innovation capability of MSE host nations can be measured through several observable dimensions. Technological advancement and integration are reflected in the development or adoption of new digital platforms, security systems, and fan engagement technologies during event preparation and delivery [74]. Business model innovation captures how MSEs stimulate new entrepreneurial activity, support micro- and small-enterprise participation, and create adaptive service models within host communities [75].

Innovation ecosystems and cross-sector collaborations, such as partnerships supporting UAV-based security or advanced logistics, further demonstrate a nation's capacity to generate and apply new solutions [76]. Infrastructure legacy measures, including upgrades to transport networks, communication systems, and digital connectivity, highlight long-term innovation outcomes that persist beyond the event [77]. Additional indicators include digital marketing innovations, social-inclusion technologies that enhance accessibility, and operational efficiency improvements resulting from new processes or event-management systems [78].

Patents are widely employed as proxies for innovation because they simultaneously provide legal protection for inventions and quantifiable evidence of technological advancement. They grant exclusive rights that can translate into economic gains while facilitating intra-industry knowledge transfer in contexts of cumulative innovation, where smaller firms contribute to broader technological progress [79]. Beyond their legal function, patent records constitute a rich source of information on inventors, technologies, and institutional networks, enabling the analysis of innovation ecosystems despite biases toward certain technological domains [80]. Empirical studies confirm their validity as indicators: patent statistics reliably forecast technological trajectories, as demonstrated in the hard disk drive industry [81], and a firm's patent portfolio consistently reflects its R&D productivity and technological capabilities [82].

Patents offer important insights into a nation's potential to sustain long-term productivity growth and competitiveness. The performance of each country, as measured by patent criteria, provides guidance for policymakers in allocating resources and improving areas important to the nation. Although many studies provide different factors of innovation capability according to business scale or international perspective, this study considers innovation capability

measurement in the context of hosting MSEs. The number of patents can be used to project technological and innovation performance across different host countries, positioning patents as both economic instruments and analytical tools indispensable for assessing innovation performance at the firm, industry, and national levels. Section 2.3 will devote attention to a specific instrument for measuring impacts.

2.3 Research Methodology Review

This section reviews the methodological approaches used to measure the economic, digital, and innovation capabilities of MSEs, as well as the machine learning models employed for predictive analysis.

2.3.1 Methodological Approach to MES's Impact Studies

To effectively assess the diverse impacts of MSEs, researchers have developed a range of estimation models that can be broadly categorized into qualitative survey-based approaches and traditional quantitative economic models. Qualitative studies typically employ surveys to capture residents' perceptions and attitudes toward events such as the 2022 Winter Olympics [83]. Quantitative approaches utilize economic models, including Input-Output (I-O), Computable General Equilibrium (CGE), and Cost-Benefit Analyses, to estimate economic effects [84]. However, these traditional quantitative models often rely on restrictive assumptions, such as constant returns to scale or static consumer behavior, which limit their accuracy in dynamic, real-world environments.

Despite providing valuable insights, both conventional approaches face significant challenges in clinical practice. Traditional economic models struggle to adapt to real-time data owing to their static assumptions, whereas qualitative approaches often lack predictive accuracy and may introduce subjectivity. Furthermore, research design limitations frequently complicate impact assessments, including difficulties in measuring causal effects and validating underlying model assumptions [41].

Machine learning methodologies offer a powerful solution to address these limitations. Unlike traditional statistical models, ML approaches are inherently data-driven and largely nonparametric, freeing them from restrictive theoretical assumptions. They excel at identifying complex nonlinear patterns and high-order interactions across diverse variables, providing a more holistic and accurate predictive framework. By learning directly from data, ML models navigate measurement challenges and real-world complexities more effectively, thereby building robust and reliable impact estimations [41]. This methodological evolution enables researchers to capture the dynamic nature of MSE's impact more comprehensively.

Superiority of ML is particularly evident in predictive analytics. A systematic review found that machine learning methods improved hospital readmission models, with ML models achieving slightly higher average AUC scores than traditional parametric approaches [85]. The adaptability of ML algorithms enables them to capture nonlinear relationships and intricate interactions among predictor variables that traditional methods often overlook, moving beyond the oversimplifications inherent in conventional logistic regression models. Advanced analytical techniques employing big data and machine learning address environmental, economic, and policy implications across sectors [78]. These methods model intricate relationships among economic activities, ecological impacts, and social dynamics, while enabling cross-disciplinary studies that tackle pressing global challenges, such as climate change [86].

The interdisciplinary nature of impact assessment requires tools capable of merging diverse datasets across economic, environmental, and social dimensions, an area where traditional models face significant limitations. Machine learning has emerged as a transformative approach, offering powerful analytical capabilities for integrating and analyzing high-dimensional data. By bridging disciplinary boundaries and uncovering hidden patterns, ML supports a more holistic understanding of the impacts of MSE and facilitates evidence-based decision-making. Table 2.1 provides a comparative overview of traditional impact estimation models and ML-based methods, highlighting their respective strengths, limitations, and relevance to interdisciplinary studies.

The capacity of machine learning to process and analyze complex nonlinear relationships allows researchers to derive insights that are often obscured in traditional analyses. As ML continues to evolve, its applications across diverse fields hold significant promise for fostering interdisciplinary approaches that promote holistic solutions to multifaceted challenges and enable comprehensive strategies that support sustainable development and innovation.

Table 2.1 Comparison of Traditional vs. Machine Learning-Based Impact Estimation Models

Aspect	Traditional Models	Machine Learning Methods
Model Types	- Input-Output (I-O) Models - CGE Models - Cost-Benefit Analysis - Logic Models - Legacy Cube - Regression	- Neural Networks - K-Nearest Neighbors (KNN) - Gradient Boosting - Random Forests - SHAP, PDP, ICE for interpretability
Data Requirements	Require structured, aggregate data; often static	Can handle both structured and unstructured, real-time, multi-source data
Assumptions	Rigid assumptions (e.g., linearity, equilibrium, rational agents)	Minimal assumptions: model adapts based on data patterns
Flexibility	Low flexibility: models need manual re-specification to incorporate new variables	High flexibility: models adapt and learn from new data without full re-specification
Integration Capacity	Often isolated to one domain (economic, social, or environmental)	Can integrate multi-domain datasets for a holistic view
Scalability	Limited by model complexity and data type	Highly scalable with cloud computing and large datasets
Interpretability	Medium to high (depending on model)	Varies: complex models such as deep learning may be opaque, but tools such as SHAP improve the interpretability
Real-Time Application	Rarely suitable for real-time analysis	Suitable for real-time predictions and monitoring
Use in MSE Literature	Widely used in legacy studies and cost-benefit evaluations	Emerging approach, showing promise in predicting innovation outcomes and impact dynamics
Stakeholder Decision Use	More traditional and familiar; easier to present to policymakers	Requires translation of outputs into interpretable forms for stakeholder engagement

2.3.2 Machine Learning Model for Prediction

In this section, this thesis explores various machine learning models for forecasting, starting with a foundational understanding of weak learner algorithms. The models examined included Random Forests, Stepwise Regression, Support Vector Machines (SVM), Neural Boosted models, generalized regression with LASSO, bootstrap forest, K-nearest neighbors (KNN), Explainable Artificial Intelligence (XAI), and Extreme Gradient Boosting (XGBoost).

Weaker Learner Algorithms: The goal of the weak learner algorithm is to find a weak hypothesis with a moderately low weighted error (i.e., a minimal sum of the weights of misclassified examples). Initially, the distribution π is uniform, but the boosting algorithm updates the weights in each round to focus the weak learner on the pairs (examples, labels) that are the hardest to predict. Thus, the main idea is to run the weak-learning algorithm multiple times, each time on a different distribution of instances, to generate several different hypotheses, as mentioned previously [87].

The weak learner algorithm (weak threshold in our case) selects a ranking feature and converts it into a binary classification $\{+1, -1\}$ using the threshold. The threshold feature

evaluates the training data with scores above the threshold as relevant (score of 1) and the training data with scores below the threshold as non-relevant (score of -1), or vice versa. This weak learner algorithm takes a training set of ranking features ($S = ((x_i, y_i), x_i = (x_{i1}, x_{i2}, \dots, x_{ik}))$), and outputs a weak hypothesis of the following form:

$$h_{li}(S, p_l) = \begin{cases} 1 & \text{if } x_{i,j} > \text{threshold} \\ -1 & \text{if } x_{i,j} \leq \text{threshold} \end{cases} \quad (2.1)$$

The philosophical idea behind weak learner algorithms is to quickly identify weak assumptions that produce moderate error rates [88].

Random forests consist of an ensemble of decision trees, each using a distinct random vector sampled independently and uniformly across all trees in the model. As the number of trees in the forest increases, the generalization error tends to converge towards a limit. The overall error of the forest depends on both the efficacy of the individual trees and the degree of correlation between them. Internal metrics are used to track the error, strength, and correlation of these factors, providing insights into how the model responds to increases in the number of features used for splitting [89].

Given input data $x = x_1, x_2, \dots, x_n$ and output response Y , where x_i represents the i th feature in the feature array x and $Y = F(X)$, the random forest is a notable ensemble method in machine learning. Its purpose is to construct a model Y to estimate the response Y from x , while also analyzing the interplay between the input features and the output response. Each tree in the forest selects a random subset of samples. The tree then refines each node by minimizing the square root errors between that node and its two subsequent split nodes, as follows:

$$SE = \text{MIN} [\sum_L (Y - \bar{Y}_L)^2 + \sum_R (Y - \bar{Y}_R)^2] \quad (2.2)$$

In this context, Y_L and Y_R represent the average response values in the left and right subnodes, respectively, following a split. Unlike conventional methods, a random forest model splits each node based on the most optimal feature from a randomly selected subset of features. This approach often yields superior performance compared to other machine learning techniques, such as neural networks, bagging, and support vector machines [89].

In sports analytics, the application of machine learning algorithms provides invaluable insights that inform both promotional strategies and a broader understanding of athletic performance. The random Forest algorithm was used to evaluate the effects of motion, capturing data both before and after athletic training [90]. A comparative study of predictive models, specifically Poisson regression, random forests, and ranking methods, to analyze the scores of FIFA World Cup matches from 2002 to 2014. Their work elucidated the relative strengths of these diverse modelling approaches in terms of predictive accuracy [91]. Beyond sports, the Random Forest algorithm has also demonstrated effectiveness in financial market analysis, including the screening of high-quality stocks within the CSI 300 index [92]. These varied applications underscore the robustness, versatility, and unique advantages of the algorithm for data classification and prediction, reaffirming its status as a mature, widely applicable machine learning technique.

Stepwise: The stepwise procedures are expected to encompass the (sub)set of predictor variables that influence the response variable and best elucidate the response. Different authors have employed various terms to denote stepwise procedures, with selection criteria other than $p \leq 0.05$ also being suggested [93],[94].

However, stepwise methods also fail to include all variables that genuinely affect the dependent variable and often introduce variables that do not [95]. The application of stepwise multiple regression found that the proportion of erroneously significant results exceeded the expected chance level for all stepwise procedures and across all numbers of predictor variables. The likelihood of obtaining a significant result increases with the number of predictor variables included when using a stepwise procedure [96].

Support Vector Machines (SVMs): Support vector machines are well-known for their ability to serve as universal approximators of any multivariate function to any desired degree of accuracy. The learning problem settings for SVMs are as follows: There are unknown nonlinear dependencies (mapping, function) $y = f(x)$ between the high-dimensional input vector x and scalar output y (or vector output y , as in the case of multiclass SVMs). The formula for the training dataset is as follows.

$$D = \{(x_i, y_i) \in X \times Y\} \quad (2.3)$$

Where $i = 1, l$, l represents the number of training data pairs and is therefore equal to the size of the training dataset D . Often, y_i is expressed as d_i , where d signifies a desired (target) value.

Consequently, SVMs are considered to be supervised learning techniques. Thukaram and Agrawal [97] investigated the potential of SVMs in identifying fault types and locations within distribution feeders. They discovered that SVMs can be practical tools for fault location in distribution systems. The authors assessed SVM performance by simulating various fault types under different conditions, including fault impedance and location. The results indicated classification accuracy of 99.39% for fault type and fault location accuracy within a few meters.

Neural Boosted: Boosting with multilayer neural networks for a character recognition task achieved a remarkable performance improvement. Gradient boosting in machine learning involves employing simple, lower-order models as weak learners, which are sequentially enhanced using first- or second-order gradient statistics to construct a more complex and higher-order model [98]. To simplify the computation, a Taylor expansion of the loss function was used. Because second-order optimization techniques generally outperform first-order techniques and converge more quickly, models are trained using Newton-Raphson steps. Furthermore, for various machine learning tasks, individual model parameters are optimized via regression on the second-order gradients of the Grow Net's outputs [99].

Generalized Regression Lasso: Most analyses are conducted using a generalized linear model (GLM). GLMs can be employed to analyze datasets that encompass any combination of continuous and categorical predictor variables, as well as continuous and discrete response variables, provided that the distribution of the residuals meets certain assumptions. Well-known examples of GLMs include multiple (linear, logistic, or Poisson) regressions, multiway ANOVA, and ANCOVA [100]. A frequently applied extension of the GLM is the generalized linear mixed model (GLMM), which enables users to control for random effect factors such as individual subjects.

The evaluation of prediction models using LASSO regression across multiple imputed datasets indicated that the model's discriminative performance was overly optimistic [101]. This was due to suboptimal calibration, which was caused by over-shrinkage. The estimate of optimism is sensitive to the approach used to handle the imputed data during the bootstrap resampling procedure. Resampling completed datasets tends to underestimate optimism, particularly when selected individuals vary across imputed datasets within a bootstrap step.

Incorporating the MI procedure into the validation yields estimates of optimism that are closer to the actual value, though slightly higher.

Bootstrap Forest: Bootstrap was introduced by Efron [102] to simulate given data to estimate the bias and variance of a statistic $Tn(x_1, \dots, x_n)$, as well as to construct confidence intervals or tests. It is typically used in its nonparametric form, though its parametric version is occasionally used. Under suitable assumptions, Bickel and Freedman [103] proved consistency, which means convergence to the correct value as $n \rightarrow \infty$. For a t-statistic, Singh [104] also demonstrated that the bootstrap is consistent and, under stronger assumptions, showed that it is asymptotically as accurate as the one-term Edgeworth expansion, explicitly indicating that the bootstrap is more accurate than the delta method. This suggests that the bootstrap method leads to a more accurate one-sided confidence interval than the delta method. Decrouez and Hall [105] confirmed that this is indeed true, but the bootstrap offers advantages, even in the lattice case, for a two-sample problem, and the statistician can determine the sample sizes.

K-Nearest Neighbors (KNN): K-nearest neighbor classification, commonly known as KNN, operates on the principle that patterns close to a target pattern x' often hold valuable label information. KNN determines the class of x' by the majority class among its K-nearest patterns in the data space, necessitating a defined similarity measure. As a local method, KNN is effective for large, low-dimensional datasets. Its adaptability is evident in its variants, which support multi-label classification, regression, and semi-supervised learning, making it versatile across a wide range of machine learning applications.

Moreover, the analysis of the KNN classifier performance showed that its accuracy is highly dependent on the chosen distance metric. The authors found that a newly introduced non-convex distance outperformed traditional metrics across most datasets. Notably, KNN using this optimal distance exhibited only a 20% reduction in performance even at a 90% noise level, demonstrating strong robustness. The study further revealed that some distance measures were less sensitive to noise than others, highlighting the varying degrees of noise tolerance among KNN distance metrics [106].

Explainable Artificial Intelligence (XAI): XAI encompasses processes and methods that enable users to understand the output of an AI system. This emerging area of research seeks to address the "black box" problem, characterized by the difficulty that many AI algorithms face in providing transparent explanations of their underlying logic and reasoning, particularly for complex models that automatically make decisions or predictions [107, 108]. Compared to traditional AI, XAI offers greater transparency, interpretability, and accountability within the system, thereby enhancing trust in AI applications, particularly in areas where critical decisions are made, such as healthcare, finance, and criminal justice [109]. XAI is essential for evaluating social and economic forecasts. Policymakers and economists can strategize effectively by understanding how AI utilizes its forecasts. The United States and China have implemented XAI applications for AI regulation. In both countries, AI models aid in predicting and analyzing markets and executing efficient government policies to promote equity and stimulate economic activity [110].

Extreme Gradient Boosting (XGBoost): Extreme gradient boosting is a widely recognized machine learning algorithm known for its strong predictive performance and computational efficiency. Operating within a gradient boosting framework, XGBoost combines multiple weak learners, typically decision trees, to form a robust predictive model by iterative error optimization. Its capacity to capture nonlinear relationships, complex feature interactions, and

high-dimensional data has led to its extensive application across disciplines such as medicine, environmental science and engineering.

In the healthcare domain, XGBoost has demonstrated a notable predictive accuracy. For example, its application to ovarian cancer outcome prediction achieved 88.72% accuracy and an AUC of 82.38%, outperforming several alternative machine learning models and highlighting its effectiveness in oncology survival analysis [111]. Similarly, XGBoost achieved superior performance in predicting acute kidney injury among patients with pancreatitis, with an AUC of 91.93%, outperforming logistic regression models [112]. In diabetes research, XGBoost has been combined with bioinformatics techniques to classify patients based on metabolic signatures associated with diabetic retinopathy, achieving an accuracy of 91.25% [113]. Beyond healthcare applications, XGBoost has proven to be effective in environmental and engineering contexts. In urban flood risk assessment, XGBoost was integrated with satellite-image semantic segmentation to predict urban waterlogging events, demonstrating its ability to handle complex spatial data [114]. Similarly, XGBoost has been applied to mining subsidence prediction in combination with genetic algorithms, further illustrating its adaptability to diverse modelling challenges [115].

As an ensemble learning method, XGBoost is frequently benchmarked against other machine learning algorithms, such as support vector machines (SVM) and random forests (RF), and often achieves superior predictive accuracy. For instance, in heart disease prediction, XGBoost attained the highest accuracy of 95.05%, underscoring its effectiveness in accuracy-critical applications [112]. Its utility has also been demonstrated in soil movement prediction, where XGBoost and related boosting methods were employed to address class imbalance issues, resulting in enhanced predictive performance [116].

The integration of various disciplines within the machine learning framework highlights the collaborative nature of this study. By combining machine learning techniques with economic, social, environmental, and innovation-related indicators across multiple datasets, this study emphasizes the strengths of each model and introduces novel forecasting approaches. These insights establish a foundation for more accurate and data-driven decision-making regarding the impact assessment of mega sports events.

2.4 Research Gap

A significant and well-documented gap in the literature underscores the importance and novelty of this study. Specifically, research employing an impact prediction or projection focus remains in a state of low development [49]. Therefore, this study is positioned directly at the intersection of two main critical research gaps, addressing both an underdeveloped methodological approach and a previously unexamined area of impact. Consequently, prospective studies focused on impact estimation and prediction, which are essential for proactive planning and policymaking, are scarce. By leveraging machine learning, this study offers a more dynamic and holistic approach that enables real-time insights into how various event factors translate into long-term impacts. These empirical findings are discussed in the next chapter.

The modelling results of this study are presented in detail in Chapter 3-5: Analytical Results and Interpretation. While this chapter focuses on the theoretical and methodological foundations, the current section serves as a transition point, outlining the scope of the analyses that follow. Chapters 3–5 comprehensively examine the outputs generated by a range of

statistical models, including t-tests, regression, and negative binomial specifications, alongside machine learning techniques such as neural boosting, k-nearest neighbors (KNN), bootstrap forests, XGBoost, and explainable artificial intelligence (XAI). These models were selected for their ability to capture nonlinear relationships and extract insights from high-dimensional data from multiple domains.

2.5 Conceptual frameworks

As illustrated in Figure 1.1, the research is grounded in the concept of hosting mega-sport events (MSEs), which generate measurable impacts reflected across diverse datasets and enable the estimation of three key impact dimensions: economic performance, digital development, and innovation capability. These datasets were then analyzed using advanced predictive models, such as Neural Boosting and K-Nearest Neighbors (KNN), as well as explainable AI techniques. Simultaneously, conventional regression models were also used to conduct comparative evaluations of predictive accuracy and robustness.

The ultimate objective is to quantify and classify impact outcomes, enabling stakeholders to understand how different types of impacts, including economic impacts, digital development, and innovation capabilities, arise from varying event conditions. The following section previews the modelling results, which are elaborated in Chapter 3-5.

The unique contribution of this study lies in the development of a machine learning–based predictive framework that integrates multidomain indicators, including patent activity, digital evolution scores, ESG performance, and economic data, into a unified impact estimation model. This study’s framework not only enables more accurate forecasting but also improves the interpretability for stakeholders through scenario modelling. This study bridges the critical gap between theory and applied decision-making, offering a robust methodology for governments, event planners, and urban policymakers.

2.6 Concluding Remarks

This chapter maps the evolving academic landscape concerning MSEs, providing a multidisciplinary synthesis of how these events influence host societies beyond their immediate spectacle. Rather than merely focusing on economic or celebratory narratives, the literature increasingly views MSEs as strategic tools for facilitating social impact, driving digital transformation, fostering innovative ecosystems, and advancing urban and environmental objectives. These broadened perspectives are essential for reimagining how this study measures the “impacts” of MSE planning and legacy management.

Notably, the reviewed studies highlight that the impacts are neither linear nor uniformly positive. Variability arises based on the event type, host context, and post-event strategies, necessitating a more adaptable and predictive approach. Machine learning models, as explored in this context, present a promising approach for capturing interdependencies across economic, social, environmental, and technological dimensions. However, much of the existing research continues to examine these areas in isolation, lacking integrated frameworks that link MSE impacts to event features in a data-driven manner. To address this gap, the proposed conceptual framework (Figure 1.1) introduces a structured model that connects MSE typologies, multidimensional impact outcomes and advanced ML-based estimation methods. Specifically, this framework incorporates patent activity, ESG metrics, digital evolution indicators, and macroeconomic variables to estimate the impact of innovation using interpretable ML models.

In Chapters 3-5, the content transitions from theory to empirical application. It details the datasets, feature engineering processes, and machine learning models employed to estimate the post-event impacts. Building on the interdisciplinary foundation established in this chapter, the analytical results demonstrate how a machine learning approach can unveil latent patterns in innovation dynamics, providing valuable insights for policymakers, sponsors, and host cities aiming to maximize long-term returns from hosting a mega-sport event.

Chapter 3: Machine Learning Models for Estimating the Economic and Innovation Impacts

This chapter investigates the application of machine learning and time-series forecasting to estimate the economic and innovation impact of hosting MSEs. Building on the conceptual framework of the previous chapter, this chapter transitions to an empirical analysis using structured data from past FIFA World Cups. Several machine learning models have been evaluated for their effectiveness in predicting economic outcomes. In addition to these short- and medium-term impact estimations, this chapter introduces a complementary analysis of the long-term technological legacy by examining global patent trends in the country. Time-series models, including Seasonal Exponential Smoothing (SES) and Moving Averages (MA), were applied to forecast innovation trajectories related to sports technologies.

3.1 Introduction

Mega-sport events are frequently portrayed as catalysts of economic growth and social transformation; however, their actual impact remains subject to considerable debate. While it is often assumed that these events deliver substantial benefits to host nations through increased consumption, large-scale infrastructure investment, and enhanced global visibility, such assumptions risk oversimplification without rigorous empirical validation. Admittedly, attendees can stimulate local economies via expenditures on tickets, hospitality, and transportation, and organizers commonly allocate substantial resources to venue construction and urban infrastructure upgrading [117]. Similarly, broadcasting rights and corporate sponsorship generate significant financial inflows. However, these apparent gains are often counterbalanced by hidden or indirect costs, including budget overruns, urban congestion, and displacement of existing economic activities [118].

Beyond immediate financial flows, MSEs are expected to shape innovation trajectories by encouraging technological upgrades, knowledge transfer, and intellectual multi-domain development. Nevertheless, the extent of these effects remains challenging to measure and is contingent on institutional governance, strategic legacy initiatives, and the capacity of local industries to internalize and capitalize on new opportunities. Without systematic, data-driven investigations, it remains impossible to disentangle short-term economic stimuli from sustainable, long-term impacts [47]. Accordingly, rigorous empirical analysis is indispensable for objectively evaluating the economic and innovation outcomes of MSEs. By employing robust quantitative methods, this study moves beyond speculative claims and anecdotal evidence to rigorously assess whether MSEs enhance economic performance and foster innovation. Such an approach not only clarifies the magnitude and direction of the impacts but also provides critical evidence for policymakers, investors, and host communities to evaluate whether the substantial costs of hosting are justified by enduring benefits [56].

This overview, although not exhaustive, highlights the variety of short-, medium-, and long-term impacts of MSEs. The focus here is on the short- and medium-term leverage of host nations associated with the FIFA World Cup, as well as economic activities and investments. Historically, assessing the impact of mega sports events has been challenging. Conventional methods struggle with the complexity and diversity of these significant events, resulting in difficulties in accurately evaluating them. Prior research utilizing linear regression and random forest algorithms has provided insights into these impacts; however, it has also exposed limitations in comprehensively capturing the effects of mega sports events on tourism.

This study extends the conventional use of linear regression and random forests by examining a comprehensive model that accounts for various effects, including economic, sports, and social parameters. The objectives of this chapter are as follows: 1) to explore the determinants of economic impact prediction; 2) to investigate the potential of machine-learning models and identify an optimal and highly accurate model for predicting the effects of mega sporting events; and 3) to examine the trends of sports innovations based on global patent activities.

This advancement in machine learning could revolutionize how stakeholders measure and respond to the consequences of MSEs, thereby maximizing positive outcomes and minimizing adverse effects. In addition to assessing short- and medium-term impacts using structured economic and social data, this chapter incorporates a complementary time-series analysis that focuses on long-term technological legacies. Specifically, it examines trends in global patent activities related to monitoring and sports technologies to forecast innovation dynamics following the emergence of MSEs. This approach offers a valuable perspective on how MSEs stimulate R&D over time and supports legacy evaluation from a technology and innovation standpoint. The results of this chapter can guide policy-making and decision-making processes related to the management of international sports events.

The structure of this chapter is as follows: Section 3.2 examines the background and related literature; Section 3.3 provides a concise summary of the data collection; Section 3.4 details the data analytics and presents the analytical findings; Section 3.5 discusses future challenges; and finally, Section 3.6 offers concluding remarks.

3.2 Related Work

This section examines prior research on measuring the impacts of sports and the existing machine learning models used to estimate them.

3.2.1 Overview of Economic, Social, Sport-related, and Innovation Determinants

Assessing the macroeconomic repercussions of hosting major sports events, specifically the FIFA World Cup, involves examining both direct and indirect influences. For instance, Ahlert [119] used factors such as GDP impact, investment in sports facilities, and tourism expenditure at both national and regional scales, employing a dynamic interindustry-based macroeconomic simulation model. Kaplanidou *et al.* [120] explored the multifaceted impact of hosting MSEs on economic, social, and environmental aspects. Additionally, they considered factors such as the event's legacy and long-term development, as well as public perception and support. To effectively estimate the impact of these large-scale sports events, it is essential to undertake a thorough analysis that not only addresses the immediate economic effects but also contemplates the enduring consequences and benefits for the host city and its inhabitants.

The relationship between economic growth and sport-related variables, including the number of venues, total attendance, team rankings, and performance, is a complex interplay in sports economics. Sports infrastructure, such as venues, not only supports national sporting success but also generates substantial economic benefits through ticket sales, tourism, and community engagement [121-123]. Attendance reflects broader socioeconomic and cultural factors, where higher engagement drives a virtuous cycle of economic returns and strengthens community support [124]. Team rankings and performance further influence economic outcomes, as successful teams attract sponsorships, broadcasting deals, and fan loyalty, thereby reinforcing financial stability and enabling reinvestment in facilities and youth development [125, 126]. Economic theories of competitive balance and performance incentives highlight that improved rankings enhance financial productivity through merchandise, sponsorships, and ticket revenue, creating sustainable cycles of growth [127]. Moreover, situational factors, such

as match-day success, strongly affect attendance and local economic activity, with winning performances stimulating spectator turnout and boosting local businesses [128].

The economic impact of hosting global sporting events is deeply conditioned by social factors, particularly population density and a nation's level of development. Population density shapes the scale of economic returns by influencing spectator turnout, which, in turn, drives revenue through ticket sales, tourism, hospitality, and local business activity. Densely populated regions not only attract larger audiences but also tend to possess the infrastructure necessary to accommodate surges in visitors, thereby amplifying their economic benefits. Semenova *et al.* [129] emphasize that social integration, often mediated by demographic dynamics, fosters economic growth by enabling interactions across diverse social groups and creating conditions for collaboration. However, these effects are less straightforward in developing countries, where structural limitations, such as weaker social integration, inadequate resources, and fragile institutions, may offset increased global visibility and infrastructure investment. As Zhang *et al.* [130] highlighted, factors such as social security provision, human capital investment, and economic sustainability critically determine the extent to which these nations can translate hosting into tangible benefits. In contexts where population density does not align with economic capacity, the outcomes of mega-events become highly variable, underscoring the importance of strategic planning and governance in realizing the potential gains.

Patents are widely employed as proxies for innovation because they simultaneously provide legal protection for inventions and quantifiable evidence of technological advancement. They grant exclusive rights that can translate into economic gains while facilitating intra-industry knowledge transfer in contexts of cumulative innovation, where smaller firms contribute to broader technological progress [79]. Beyond their legal function, patent records constitute a rich source of information on inventors, technologies, and institutional networks, enabling the analysis of innovation ecosystems despite biases toward certain technological domains [80]. Empirical studies confirm their validity as indicators: patent statistics reliably forecast technological trajectories, as shown in the hard disk drive industry [81], and a firm's patent portfolio consistently reflects its R&D productivity and technological capabilities [82]. Collectively, these findings position patents as both economic instruments and analytical tools, making them indispensable for assessing innovation performance at the firm, industry, and national levels.

To comprehensively assess the impact of hosting mega-sport events, it is essential to combine economic indicators with social and sport-related factors, as these dimensions interact to shape both immediate and long-term outcomes. Economic variables capture macro-level growth dynamics, whereas social factors, such as population density and developmental status, determine the absorptive capacity of benefits. Sports-related metrics, including attendance, venues, and team performance, reflect event-specific mechanisms driving economic activity. Simultaneously, patent trends serve as a robust proxy for innovation, providing quantifiable insights into technological advancement, knowledge diffusion, and R&D productivity. Integrating these dimensions not only strengthens the predictive capacity of empirical models but also ensures a more nuanced understanding of how mega-sport events influence both economic performance and innovation trajectories, thereby capturing their full legacy.

3.2.2 Existing ML Models in Sport Impact Estimation

Machine learning (ML) models are increasingly used in sports to assess various impacts, including player performance, injury risk, tactical effectiveness, and broader business outcomes. This field, often referred to as sports analytics, leverages large amounts of data to

provide actionable insights. Standard machine learning models and their applications in sports events include regression models, such as linear regression, logistic regression, XGBoost, and multiple linear regression (MLR). These models are frequently used to predict continuous outcomes and their probabilities. For instance, the utilization of performance predictions involves estimating player performance metrics, such as points scored, assists, and distance covered, based on historical data, training loads, and other relevant factors [131].

Tree-based models, such as decision trees and random forests, are robust ensemble methods used for classification and regression. These models effectively manage complex relationships and offer valuable insights into the importance of features. In the context of talent identification and scouting, they are used to classify athletes' potential based on their performance and physical attributes. Additionally, in tactical analysis, these models are instrumental in recognizing patterns in the opponent's play, thereby informing strategic adjustments. Support vector machines (SVMs) are widely used for both classification and regression tasks, particularly in high-dimensional data environments. In sports analytics, SVMs are useful for player classification, enabling categorization based on playing style, role, and potential. Although other models may be more commonly used, SVMs also offer valuable applications in predicting discrete outcomes, such as match results or performance milestones, making them a versatile tool in sports prediction tasks.

Despite significant advancements, several challenges persist in applying machine learning to sports impact estimation. Data quality and standardization remain critical issues because consistent, accurate, and reliable data collection is fundamental to the model's performance. Another key challenge is integrating machine learning insights into coaching workflows, where the gap between analytical outputs and their practical applications must be effectively bridged. Nevertheless, the field continues to evolve rapidly, with new models and techniques emerging that offer increasingly precise and actionable insights to help stakeholders in sports ecosystems make data-driven decisions.

3.2.3 Time-Series Forecasting in Sports and Innovation Impact Studies

Time-series analysis, which involves tracking and analyzing the measurements of a particular variable over time, plays a vital role in numerous scientific disciplines, including sports analytics and event forecasting. Fulcher et al. emphasized the importance of selecting appropriate analytical methods based on the structure and properties of the data, demonstrating that the relationship between the time-series characteristics and the applied techniques is a key determinant of the forecasting accuracy. Their work introduced a robust framework for organizing and comparing time-series datasets, along with the analytical operations applied to them. In the field of sports event analysis, models such as the moving average (MA), autoregressive integrated moving average [132], and seasonal ARIMA (SARIMA) have been widely adopted to forecast variables including ticket sales, event attendance, and fan engagement patterns. These models are particularly effective in identifying and predicting seasonal trends and cyclic behaviors in data, providing both interpretability and computational efficiency for decision-making.

In addition to performance metrics, time-series forecasting plays a crucial role in estimating technology and innovation trajectories such as patent activity and R&D trends. Hyndman and Athanasopoulos demonstrated that exponential smoothing techniques and ARIMA-family models can uncover underlying trends in longitudinal data and provide robust forecasts, even in the presence of volatility or sparse data intervals. Few studies have integrated time-series models to examine technological legacies of mega-sport events. This study addresses this gap by applying robust forecasting techniques to patent application data related to sports

monitoring and video technologies, which serve as proxies for the long-term innovation impacts catalyzed by MSEs. This approach allows for a more nuanced analysis of MSE legacies that extends beyond immediate economic indicators.

3.3 Data Summary

This section describes the data-gathering process, parameter selection, and model input preparation. However, owing to the small sample size ($n = 5$), the findings presented in this chapter should be interpreted with caution. The machine learning models applied were primarily intended for exploratory analysis rather than definitive predictions.

This section presents an overview of the dataset used to assess the impact of hosting the FIFA World Cup. The dataset comprises five observations corresponding to the five tournaments: 2006 (Germany), 2010 (South Africa), 2014 (Brazil), 2018 (Russia), and 2022 (Qatar). A range of potential variables was explored to capture the multifaceted outcomes of these events. The dataset comprised covariates collected in each World Cup year. This classification facilitates a systematic evaluation of the multidimensional legacies associated with hosting the World Cup.

During the initial phase of our research, this study sourced data from various repositories, including FIFA's final event reports, statistical data platforms, and the World Bank data portal. A total of 13 variables were collected for each World Cup and its corresponding host countries. These were modelled using 12 input variables (x_0 – x_{11}) and one target variable (y), representing the total value added in the year the event was held (Table 3.1).

From a methodological perspective, the use of domain-relevant covariates selected before or during the event aligns with the variable-selection principle outlined by Guyon and Elisseeff [133]. Their approach emphasized the importance of relevance, non-redundancy, and interpretability in supervised learning contexts. Further support for the selected variables comes from recent machine learning-based studies, which demonstrate the predictive value of metrics such as GDP per capita, tourist expenditure, and team performance in estimating both the economic and reputational effects of mega-events [134].

Table 3.1 Economic and Sport-Related Variable Description

Input variable	Description
x_0	Number of Venues
x_1	GDP per capita before the event
x_2	GDP per capita after the event
x_3	GDP per capita of absolute change
x_4	Relative change in GDP
x_5	Population size
x_6	Total attendance
x_7	Long-term investment
x_8	Tourist's expenditure
x_9	FIFA ranking (Team's Performance)
x_{10}	PTS (Team's Score)
x_{11}	Developed or developing country (1 if developed, and 0 if developing country)
y	Total Value Added (by Economic Activity, Annual Rate of Growth)

Moreover, the inclusion of variables such as GDP changes, long-term investment, and value-added growth is consistent with Baade and Matheson, who analyzed the economic impacts of World Cup tournaments using similar indicators [135]. Overall, integrating domain-specific classification and machine learning feature selection theory provides a comprehensive and robust foundation for constructing datasets and selecting parameters.

The variable selection process in this study is grounded in theoretical frameworks and empirical evidence drawn from sports economics, mega-event legacy research, and machine learning. To capture the multidimensional impact of hosting the World Cup, the grouping of variables into economic, social, and home advantage categories follows the legacy framework proposed by Solberg and Preuss [8], which identifies economic, infrastructural, and social dimensions as core areas of impact in MSE studies.

3.4. Analytical Result

This section presents the analytical methods and results, encompassing predictor screening, variable importance analysis, nonlinear regression, and performance evaluation of each machine learning model applied to economic impact prediction. It concludes with a comparative assessment of the performance of the alternative time-series models.

3.4.1 Predictor Screening and Variable Importance Analysis

This study employed two methods to evaluate predictors of target variables: 1) predictor screening by boosted trees and 2) variable importance by Random Forest. Table 3.2 provides the standard ranges for categorizing the importance values into strong, medium, and low categories in variable importance interpretation.

As shown in Table 3.3, the predictor screening analysis identified the relative change (x_4) as the most significant contributor to the model, followed by population size (x_5) and GDP per capita prior to the event (x_1), with the relative change of GDP (x_4) accounting for approximately 26 percent of the model. Conversely, the importance analysis showed a different ranking, with population size (x_5), relative change in GDP (x_4), and GDP per capita after the event (x_2) as the top predictors.

The findings indicate that variables with consistently high importance across both predictor screening and random forest analyses, namely relative change in GDP (x_4), population size (x_5), GDP per capita prior to the event (x_1), and GDP per capita after the event (x_2), should be prioritized in forecasting models, as their stability across methods underscores their robust predictive value. Differences in the rankings, such as the elevated prominence of GDP per capita after the event (x_2) in the random forest results but not in the screening analysis, highlight the need to account for interaction effects and nonlinear relationships that univariate methods may overlook. Lower-ranked predictors, such as the team score (x_{10}), may still hold situational relevance, particularly when exceptional event performance amplifies economic and social impacts.

Table 3.2 Importance Value Range

Category	Importance Value Range
Strong	>0.0939
Medium	0.0536 - 0.0939
Low	<0.0536

Table 3.3 Predictor Screening and Variable Importance Analysis Statistic Results

Predictor	Contribution Value	Portion	Variable	Importance Value	Meaning of predictor
x_4	1.0652	0.2609	x_5	0.1757	Strong
x_5	0.6238	0.1528	x_4	0.1376	Strong
x_1	0.4461	0.1093	x_2	0.1073	Strong
x_3	0.4392	0.1076	x_7	0.0948	Strong
x_7	0.3327	0.0815	x_{10}	0.0941	Strong
x_8	0.3282	0.0804	x_8	0.0824	Medium
x_{11}	0.3029	0.0742	x_3	0.0720	Medium
x_6	0.2180	0.0534	x_6	0.0713	Medium
x_0	0.2138	0.0524	x_1	0.0684	Medium
x_2	0.1127	0.0276	x_9	0.0485	Low
x_9	0.0005	0.0001	x_{11}	0.0478	Low
x_{10}	0.0000	0.0000	x_0	0.0000	Low

Reporting results from both predictor screening and random forest importance enables a more comprehensive assessment of the determinants. The former provides a univariate perspective on individual explanatory power, whereas the latter captures multivariate and interaction effects. This dual approach improves predictive validity and interpretability, ensuring that critical determinants, especially those supported by theoretical or empirical evidence, are not overlooked because of methodological bias.

3.4.2 Nonlinear Regression Analysis

Table 3.4 presents the findings from a nonlinear regression analysis examining the relationships between the predictors (x_0 to x_{11}) and the dependent variable. Nonlinear regression was applied to capture complex relationships in which the predictors' impact on the outcome was not linear. The first column of the table shows the growth rate, and the prob.> Chi-square values indicate the statistical significance of each predictor's growth rate in the model.

Typically, a value below 0.05, marked with an asterisk (*), is deemed significant. For instance, the number of venues (x_0) and the population size (x_5) showed substantial growth rates, indicating a strong nonlinear influence on the dependent variable. The statistically significant inflection points for these predictors indicate a model-identified point at which the rate of change in the dependent variable undergoes an important shift. In this analysis, the GDP per capita. Absolute change (x_3) and population size (x_5) exhibited significant upper asymptote values, whereas total attendance (x_6) had a notably lower asymptote, indicating potential natural thresholds in their relationships with the outcome. The relative change (x_4) and total attendance (x_6) showed the lowest MSE and RMSE values, suggesting that they were the best fit. The high R-squared values for these variables also imply a strong fit.

This analysis facilitates an understanding of the nonlinear dynamics of the system, identifies variables with significant nonlinear effects, and assesses the accuracy of these nonlinear models in predicting outcomes.

Table 3.4 Nonlinear Regression Analysis Statistic Results

Predictor	Growth Rate	Inflection Point	Lower Asymptote	Upper Asymptote	MSE	RMSE	R-Square
	<i>Prob > Chi-Square</i>						
x_6	1	1	0.82	<.0001*	0.78	0.88	0.89
x_4	0.99	1	1	1	1.14	1.07	0.84
x_3	0.98	0.92	0.33	0.0048*	1.53	1.24	0.78
x_5	<.0001*	<.0001*	<.0001*	0.0002*	2.05	1.43	0.71
x_7	1	1	0.21	0.08	4.06	2.01	0.42
x_{11}	1	1	1	<.0001*	4.06	2.01	0.42
x_2	1	1	0.53	0.07	4.65	2.16	0.33
x_0	<.0001*	<.0001*	0.51	0.07	6.66	2.58	0.04
x_1	1	1	0.51	0.12	6.66	2.58	0.04
x_8	1	<.0001*	0.51	0.12	6.66	2.58	0.04
x_9	1	1	1	0.12	6.66	2.58	0.04
x_{10}	1	1	0.26	0.34	6.82	2.61	0.02

Overall, the convergence of high-ranking variables across all three methods (predictor screening, random forest importance, and nonlinear regression) provides strong evidence for prioritizing predictors, such as population size (x_5) and relative change (x_4), in forecasting models. Divergence in variable rankings, such as the elevated R^2 for total attendance (x_6) in nonlinear regression, but lower importance in screening and Random Forest, highlights the complementary strengths of different analytical approaches in capturing various dimensions of predictor relevance. The presence of pronounced nonlinear patterns among several key variables underscores the limitations of traditional linear regression in modelling the complex economic dynamics associated with mega-sport events.

This provides a clear rationale for employing advanced machine learning techniques, which are inherently capable of capturing such nonlinearities and interaction effects, thereby offering enhanced predictive accuracy and robustness compared to conventional methods. The following sections present the results of the machine learning models applied to the dataset, which estimate and forecast the impacts of these events with greater precision. Building on the nonlinear regression findings, the subsequent analysis compares the predictive performance of multiple modelling approaches to identify the most effective tools for forecasting event-related impacts.

3.4.3 Comparative Analysis of Predictive Model Performance

In the final analysis, a comparative evaluation of the machine learning models was conducted using R-squared and root mean squared error (RMSE) as performance metrics. The corresponding statistical results are presented in Table 3.5. The stepwise regression model achieved an R-squared of 0.9999 and an adjusted R-squared of 0.9995, with minimal error (RMSE = 0.0285; SSE $\approx$ 0) and a notably low BIC of -21.3976, indicating an excellent fit to the data. The selection process initially incorporated relative change (x_4), yielding an R-squared of 0.8089, followed by GDP per capita in the year preceding the event (x_1), which increased the R-squared to 0.9973. The team score (x_{10}) was added in the final stage according to the stepwise criteria. The resulting model identified a parsimonious set of predictors with exceptional accuracy and explanatory power.

The stepwise regression results reaffirmed the patterns observed in the predictor screening, Random Forest importance, and nonlinear regression analyses. Variables such as relative change (x_4) and population size (x_5), which consistently ranked highly across methods, again emerged as key determinants, although in the stepwise model, GDP per capita prior to the event (x_1), and team score (x_{10}) were retained as part of the final parsimonious set. The near-perfect fit statistics highlight the strong predictive relationship between these variables and the value-added outcomes of hosting mega-sport events. However, the high R-squared values combined with the small sample size raise concerns about overfitting, underscoring the importance of evaluating model performance against more flexible approaches. This sets the stage for subsequent machine learning analyses, which build on the insights from all prior methods and are designed to better capture complex nonlinearities, interaction effects, and context-dependent patterns in forecasting event impacts.

However, the support vector machines (SVM) model uses four support vectors to identify four pivotal data points that define the regression hyperplane. An R-squared value of 0.7537 indicates a strong predictive capacity. The best RASE value (0.5850) represents the average prediction error, with lower values indicating superior performance. The results show a significant weight for the relative change (x_4) at 22.7712, indicating a considerable positive impact on predictions. Other variables, such as the number of venues (x_0), GDP per capita of the previous year (x_1), and GDP per capita a year after the event (x_2), were also weighted by the model, but to a lesser extent, suggesting a more moderate influence on predictions. The reliance on a few support vectors and the prominent weighting of the relative change (x_4) indicate that the SVM effectively capitalizes on the inherent structure of the data.

Nonetheless, the nullification of numerous variables suggests potential feature redundancy or irrelevance, underscoring the need for further feature selection to enhance model interpretability and, potentially, predictive performance. While the preceding analyses applied cross-sectional and multi-year models to estimate the short- and medium-term economic impacts of hosting mega-sport events, the next stage of the study shifts its focus to a longitudinal perspective. This transition is necessary to capture the long-term technological effects of such events, which may unfold over decades rather than in the immediate aftermath of hosting. By employing a time-series analysis of sports-related patent applications, this study tracked innovation trends before, during, and after major tournaments, thereby isolating sustained impacts from short-term fluctuations. This approach complements earlier economic models by addressing a distinct but equally critical dimension of mega-sport event impacts on technological development.

Table 3.5 Comparative Analysis of ML Models' Performance

Machine Learning Methods	N	R-Square	RASE
Training			
Stepwise	5	0.9999	0.0127
Support Vector Machines	5	0.7537	0.5850
Neural Boosted	3	0.0000	0.8179
Generalized Regression Lasso	5	0.0000	1.1788
Bootstrap Forest	5	-0.002	1.1800
K Nearest Neighbors	5	-0.208	1.2954
Validation			
Neural Boosted	2	-1.1447	1.7989

3.4.4 Innovation Impact of MSEs by Time-Series Approach

While the preceding sections of previous analyses focused on modelling the short- and medium-term economic impacts of hosting mega-sport events, this section presents a complementary time-series analysis to estimate their long-term technological legacies. Patent applications related to sports technologies are used as a proxy for measuring technological impact. The search strategy is described in Table 3.6.

The primary aim of this analysis is to evaluate the predictive effectiveness of time-series models in forecasting technological innovation trends influenced by hosting MSEs. This supports the broader research objective of understanding how such events contribute not only to economic performance but also to long-term R&D and technology transfer. To achieve this, various time-series forecasting techniques have been applied across multiple time intervals to address seasonal fluctuations and improve accuracy. These models include simple and double exponential smoothing, moving average (MA), ARIMA, and seasonal ARIMA.

In the initial step, the patent database was searched using document titles and abstracts, yielding 2,286 records worldwide. A screening procedure was applied to remove duplicates and unrelated content. Finally, time-series methods were applied, with the number of patent applications per year as the target variable (y) for prediction. The search strategy is summarized in Table 3.6.

The analysis of patent applications in the first month of each year from 1981 to 2023 revealed pronounced fluctuations, ranging from sharp year-to-year increases to steep declines. The moving average trendline indicates a downturn beginning around 2018, suggesting a potential shift in innovation activity. This change may be influenced by broader economic conditions, evolving technological priorities, and modifications to patent regulations. Periods of heightened activity, such as between 2008 and 2016, contrast with earlier phases of modest growth and the recent sharp decline. The monthly data provide detailed insight into short-term volatility while also revealing longer-term structural patterns, such as the sustained decrease in the final years of the series.

Building on these observed trends, subsequent time-series modelling aims to quantify and forecast the long-term technological legacy of mega-sport events using patent applications as a proxy for innovation outcomes.

Table 3.6 Search Strategy of Patent Applications

Retrieve Date	18/12/2023
Data Base	Thomson Reuters Derwent Innovation Index
Technology Category and Synonym or Relevant Words	T1: Monitoring control; vision, display, viewing, view, controlling, control T2: VDO; video, VDO, recording, record, HD, high-definition, high ADJ resolution, high-resolution
Assumption	Each document can fall into multiple categories of technology
Scope	All countries
Period of data	1981-2023
Gathering results	2,286 patent application documents

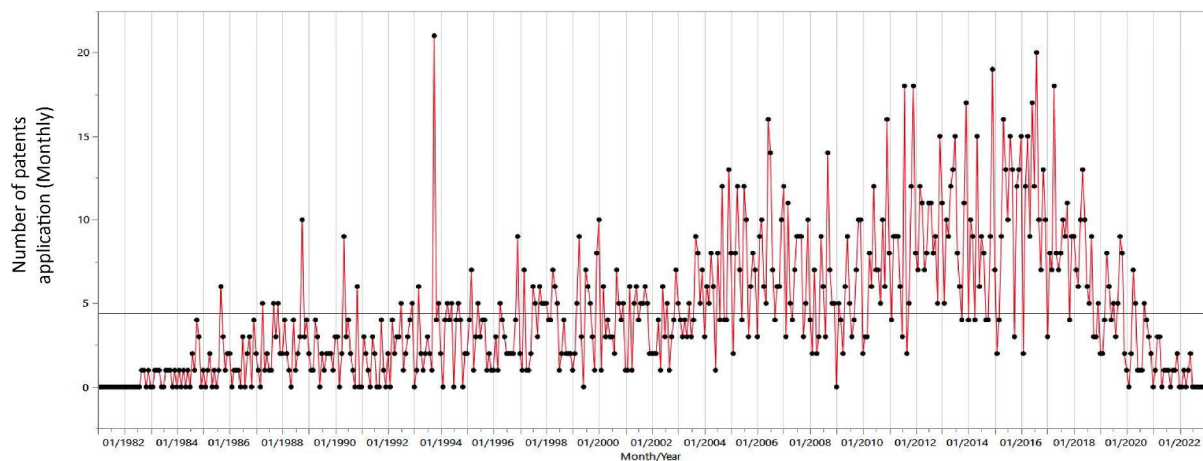


Figure 3.1 Plot of The Number of Patent Applications in 1981-2023

Table 3.7 Patent Trend Directions and Influencing Factors for FIFA World Cups (2006–2022)

Tournament Year	Host Country	Patent Trend Direction	Notable Factors
2006	Germany	Increase	Boost in innovation from event infrastructure and technology upgrades
2010	South Africa	Increase	Increased patent activity linked to global exposure and tech adoption
2014	Brazil	Peak/High	Peak in sports-related patents; strong innovation momentum
2018	Russia	Decline	Start of downturn in global patents; potential economic headwinds
2022	Qatar	Decline	Continued decline; possible pandemic-related disruptions

Table 3.7 summarizes the relationship between five major tournaments, including Germany 2006, South Africa 2010, Brazil 2014, Russia 2018, and Qatar 2022, and the corresponding directions of sports-related patent trends. The 2006 and 2010 events were associated with marked increases in patent activity, potentially driven by infrastructure investment, advancements in broadcasting technologies, and other forms of sport-related technological innovation. Brazil's 2014 alignment with the peak period of patent applications reflects strong innovation momentum within the broader surge identified in the descriptive analysis from 2008 to 2016. In contrast, the 2018 Russia and 2022 Qatar tournaments occurred during a marked decline in patent activity, which was influenced by global economic headwinds, shifting technological priorities, and pandemic-related disruptions. These patterns suggest that while mega-sport events can coincide with innovation surges, their long-term technological legacies are contingent on prevailing macroeconomic and policy conditions. Building on these observations, the subsequent time-series modelling seeks to quantify the relationship between event hosting and patent activity over time and assess whether hosting contributes to sustained technological growth beyond the underlying global trend.

Table 3.8 presents a statistical overview of patent applications gathered across various timeframes. The data reveal an average of approximately 4.413 patent applications per month, quarter, and semi-annual period, and approximately 53.160 per year. The standard deviation of the patent application figures shows consistent patterns on a monthly, quarterly, and semi-annual basis, with values ranging from approximately 3.18 to 4.10. This suggests uniform variability within these time frames. In contrast, the standard deviation of the annual data is much higher, approximately 36.53, owing to the broader dispersion of data points.

A comparison of different time-series forecasting models, as measured by Akaike's Information Criterion [136] and Schwarz's Bayesian Criterion (SBC), is presented in Figures 3.2 and 3.3. Models with lower AIC and SBC values are generally preferred because they penalize the inclusion of excessive parameters, thereby reducing the risk of overfitting. This implies that these models are parsimonious and effective in capturing key trends without adding unnecessary complexity. Consequently, models with the lowest AIC and SBC values are typically chosen as the best fit for time-series analysis.

Table 3.8 Statistical Data of The Time Series Analysis

Statistical Measures	Month	Quarter	Semi-Annual	Annual
Mean	4.413	4.413	4.413	53.163
Standard deviation	4.105	3.378	3.183	36.525
Number of samples	516	516	516	43
Zero Mean ADF	-7.787	-2.713	-1.391	-0.731
Single Mean ADF	-12.309	-4.522	-2.404	-1.212
Trend ADF	-13.851	-5.101	-2.255	-0.117

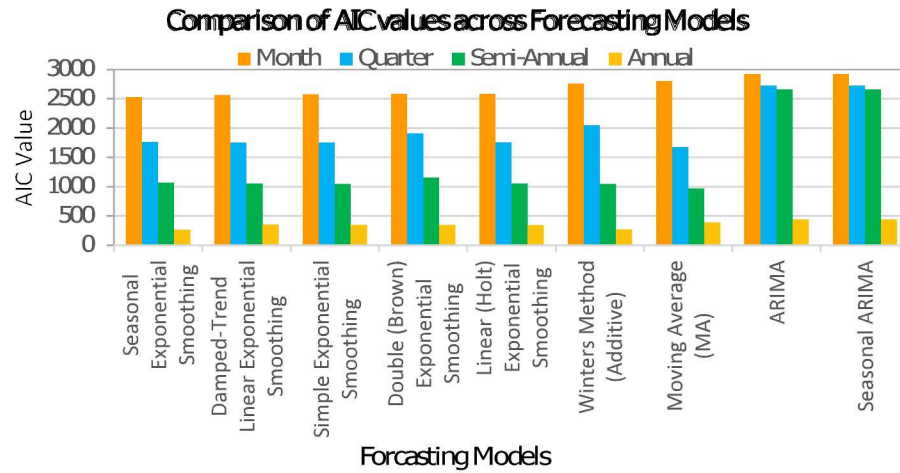


Figure 3.2 Comparison of AIC Values Across Forecasting Models

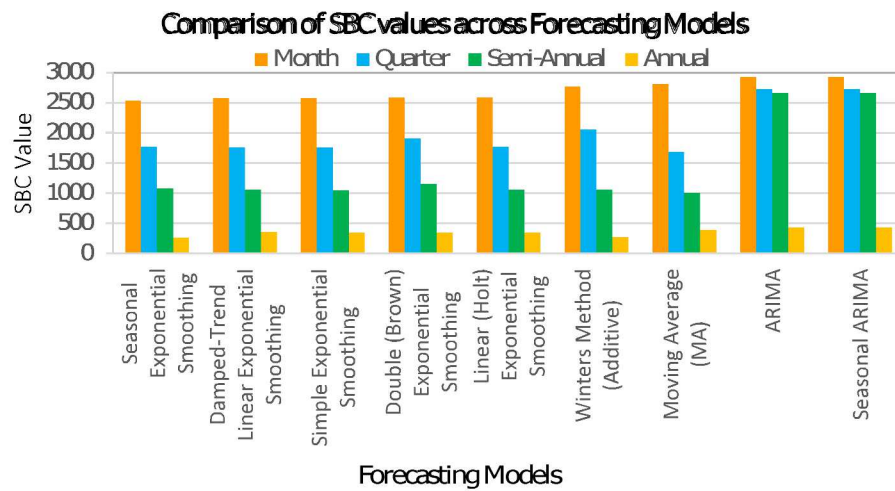


Figure 3.3 Comparison of SBC Values Across Forecasting Models

After a thorough analysis of the performance of each model, it became apparent that the seasonal exponential smoothing (SES) model demonstrated superior predictive accuracy for both monthly and annual trends in patent applications. It achieved the lowest AIC and SBC values of 2529.695 and 2538.136 for monthly trends and 265.122 and 267.924 for annual trends, respectively, as shown in Figure 3.2 and 3.3. This model effectively captures the underlying patterns at different temporal resolutions. For quarterly and semi-annual forecasting, the moving average (MA) model stands out with its robust predictive performance, denoted by an AIC of 970.405 and an SBC of 1000.177 for semiannual intervals. The model further solidified its efficacy for semi-annual intervals, with an AIC of 1,671.666 and an SBC of 1,688.651 for quarterly data.

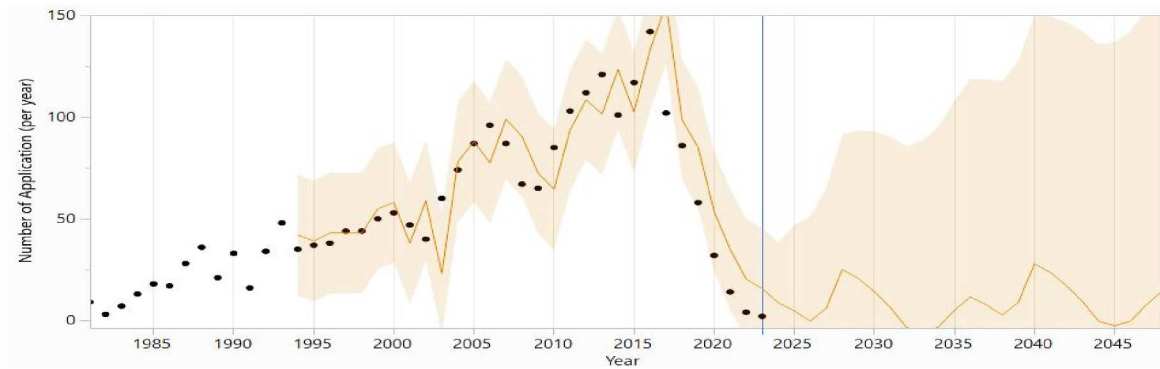


Figure 3.4 Annual Forecasting Using the Seasonal Exponential Smoothing (SES) Model

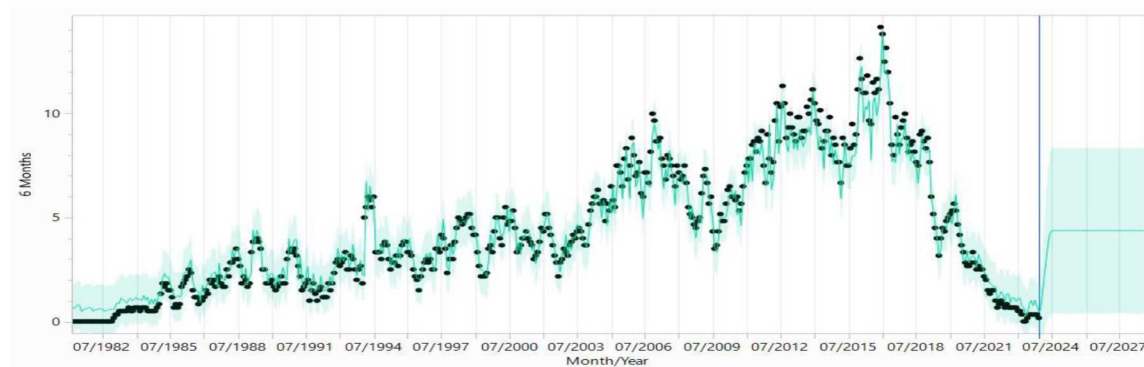


Figure 3.5 Semi-Annual Forecasting Using the Moving Average (MA) Model

These findings are illustrated in Figures 3.4 and 3.5, which show the predicted patent trends and validate the model fit through confidence intervals and the statistical significance of the lag terms. These statistics confirm the model's capacity to represent temporal patterns while maintaining parsimony accurately, a critical consideration in the rigorous process of model selection for time-series analysis.

The findings offer valuable insights for strategic innovation planning within MSE-related sectors, highlighting how patent activity serves as a reliable indicator of technological momentum and legacy. By forecasting trends in relevant technologies, this analysis supports long-term decision-making for R&D investment and policy development. Taken together, machine learning and time-series modelling approaches illustrate the benefits of integrating short-term economic assessments with forward-looking innovation forecasting to enable a more comprehensive and informed approach to MSE impact planning.

3.5 Discussion

A comparative analysis of multiple machine learning models for estimating the impact of mega-sport events (MSEs) reveals a nuanced picture of model effectiveness, variable significance and methodological trade-offs. The findings highlight that no single model is universally optimal; instead, performance is shaped by the dataset characteristics, predictor composition, and specific forecasting objectives. This section interprets the strengths and limitations of each approach, links them to broader applications in event-based economic forecasting and situates them within the long-term technological legacy analysis presented later in this chapter.

This study had some constraints. First, the small dataset, limiting statistical power and increasing the risk of overfitting in ML models [137]. This constraint can also lead to instability in variable importance rankings, making the results sensitive to small data perturbations. Second, full validation results were available only for the neural-boosted model, limiting the ability to systematically evaluate overfitting risk across all approaches. Third, the predictor set excluded institutional, environmental, and governance indicators, thereby narrowing the explanatory scope of the models and potentially omitting important contextual factors in estimating the MSE's impact.

Among the models tested, stepwise regression achieved the highest in-sample performance, with near-perfect fit statistics and minimal error. SVM offers a more favorable balance between accuracy and simplicity, due to its robustness in handling high-dimensional, small-sample datasets. In contrast, neural boosting, KNN, and ensemble methods showed suboptimal performance, possibly due to their greater sensitivity to noise, reliance on larger datasets for stability, and limited feature diversity in this study. These differences underscore the need for context-sensitive model selection that considers data availability, predictor relevance, and stakeholder requirements for interpretability [137].

While the preceding analyses focused on short- and medium-term economic outcomes derived from structured event-level datasets, they did not capture the enduring technological impacts of MSEs. To address this gap, a complementary time-series forecasting analysis was conducted using global patent applications in monitoring and recording technologies as proxies for innovation activity. This long-term perspective is crucial for understanding whether MSEs leave a lasting technological legacy that extends beyond their immediate economic effects.

The time-series results showed that the seasonal exponential smoothing (SES) model delivered the strongest performance for annual and monthly forecasts, effectively capturing both the level and seasonality components. In contrast, the moving average (MA) model outperformed the others for quarterly and semi-annual forecasts, suggesting that its simplicity is advantageous when seasonal patterns are less dominant but short-term fluctuations are essential. These findings highlight the importance of selecting forecasting techniques that align with the temporal structure and periodicity of target data. Time-series analysis is particularly relevant when conventional economic indicators fail to capture innovation-related legacies fully. Patent activity reflects technological advancement and indicates institutional readiness to adopt and commercialize new technologies in sectors linked to MSEs. The observed decline in long-term patent filings since 2018 may indicate market saturation, a shift in innovation priorities, or broader macroeconomic disruption. Such trends have implications for policymakers, sponsors, and R&D managers engaged in legacy planning and strategic investment.

The findings from both analytical streams suggest a complementary methodological strategy: machine learning models are best suited for estimating short- and medium-term economic and social impacts. In contrast, time-series models provide a robust framework for evaluating long-term technological trends. Integrating these approaches creates a more comprehensive, data-driven framework for understanding and forecasting the complex legacies of mega-sport events, thereby enhancing the relevance and utility of this dissertation's contributions to academic research and practical decision-making.

3.6 Concluding Remarks

This study presents a comprehensive investigation of the predictive impacts of MSEs using structured event-level indicators and complementary time-series patent data. Multiple machine learning models, including stepwise regression, SVM, and neural boosting, were applied to estimate the effects of MSEs on economic, sporting, and social outcomes using variables from FIFA reports and international statistical databases. Stepwise regression emerged as the best-performing model on the training data, driven by key predictors such as the relative change in GDP, GDP per capita and population size.

To extend the understanding of MSE legacy impacts, a complementary time-series analysis was conducted using patent applications related to sports innovation technologies as a proxy for innovation. This long-term perspective revealed that the Seasonal Exponential Smoothing (SES) model performed best for annual and monthly forecasts, capturing both the level and seasonal effects. In contrast, the Moving Average (MA) model was superior for quarterly and semiannual trends, where short-term fluctuations were more pronounced. These results underscore the importance of evaluating both immediate and extended legacy dimensions when assessing the outcomes of MSEs.

The integration of machine learning and time-series forecasting constitutes a comprehensive multidimensional analytical framework for evaluating the diverse impacts of mega-sport events. This integrated approach offers novel insights by linking economic indicators with innovation proxies, providing practical guidance for governments, event organizers, and R&D stakeholders involved in strategic legacy planning. Building on the economic, social, and technological legacies examined here, the next chapter shifts focus to the relationship between MSE hosting and national digital impact. While Chapter 3 emphasizes short-term economic impacts and long-term innovation trends, Chapter 4 investigates how MSE-related economic conditions can influence a host country's digital evolution. It applies advanced machine learning models to explore the potential for fostering digital infrastructure benefits.

Chapter 4: Machine Learning Models for Estimating Digital Development Impacts

Chapter 3 examines two key areas: the short-term economic impacts of single-sport events and the long-term innovation trends associated with hosting mega-sport events (MSEs). Building on these findings, this chapter focuses on the relationship between hosting MSEs, both single- and multi-sport types, and their influence on economic indicators and digital development. Using machine learning methods, specifically neural boosting and k-nearest neighbors (KNN), the analysis examines how hosting MSEs affects digital development. By identifying the most influential predictors and comparing model performance, this chapter provides policy-relevant insights into leveraging MSEs for national digital transformation.

4.1 Introduction

As digitalization becomes increasingly central to national competitiveness, MSEs have begun to play a transformative role in physical infrastructure and tourism and in accelerating digital readiness and innovation ecosystems. This chapter shifts the focus from the general economic impacts of hosting major sporting events to specifically examine how these events affect a nation's path of digital development. Digital development is a specialized subset of technology that leverages digital tools, such as computing, the internet, mobile devices, and data, to drive social and economic progress across domains such as healthcare, education, finance, and governance [138]. Whereas technology represents the broad ocean of human invention, digital development is the accelerating current within it and serves as the primary driver of contemporary innovation. This distinction reflects perspectives in technological determinism, which views technology as a transformative force in shaping society [139], and aligns with recent scholarship on the digital economy and digital transformation as catalysts of modern technological advancement [140].

Concurrently, the digital dimensions of MSEs have become increasingly prominent. The integration of advanced digital technologies into event management, fan engagement, and urban services not only enhances the operational efficiency of these events but also contributes to broader national digital development. The high demand for advanced digital integration, encompassing contactless ticketing systems, 5G-enabled broadcasting, innovative mobility solutions, and cybersecurity frameworks, has prompted host nations to accelerate their digital transformation. One significant aspect of this transformation is the implementation of contactless ticketing systems. Various regions have adopted digital ticketing infrastructures that streamline fare collection and enhance customer experience. Relying on innovative technologies such as RFID cards, these systems enable seamless payment processes across various public transport modes, as exemplified by the Octopus card in Hong Kong and similar systems in other urban centers [141]. Furthermore, successful contactless solutions have demonstrated their potential for widespread adoption across various international contexts, enhancing operational efficiency and customer satisfaction in the process. [142].

Moreover, the integration of 5G technology into broadcasting and communication infrastructure signifies a pivotal shift in information dissemination. The proliferation of 5G networks enables rapid data transfer and connectivity, which are crucial for innovative mobility solutions that incorporate autonomous vehicles and connected transport systems. As highlighted by recent studies, the role of digital transformation extends beyond mere efficiency improvements; it also fosters competitiveness and responsiveness in the development of digital economies [143]. Nations are increasingly recognizing the crucial role of robust digital

infrastructure in achieving strategic sustainability goals and enhancing service delivery across sectors [144]. Knott *et al.* noted that MSEs provide a platform for showcasing national digital maturity and attracting international investment [145]. These technological demands often compel host countries to expedite digital initiatives, upgrade their infrastructure, and foster innovation across the public and private sectors. Therefore, a country's economic potential significantly impacts digital investment and development.

Building on the literature linking MSEs to broader economic and technological outcomes, this study seeks to address a critical gap in understanding how hosting influences national digital development. While prior research has examined the economic impacts of digital transformation in general, limited attention has been paid to the specific role of MSEs as catalysts for digital maturity. Moreover, existing studies have primarily relied on traditional statistical approaches, with few using advanced machine learning methods to capture nonlinearities and complex interactions among predictors better.

To bridge these gaps, this study pursues three objectives: (1) to investigate the relationship between hosting mega-sport events and national digital development as measured by digital evolution scores (DES); (2) to identify the economic predictors that contribute most significantly to variations in DES; and (3) to evaluate the predictive accuracy of alternative models for estimating DES in host and non-host nations. This chapter is structured as follows: Section 4.2 reviews the literature on digital legacies and MSEs; Section 4.3 outlines the research design, data, and methodology; Section 4.4 presents the empirical results; Section 4.5 provides analytical findings; and Section 4.6 concludes with implications and future directions.

4.2 Related Work

This section provides a comprehensive overview of the research background and related work on mega-sport events and digital development, the economic drivers of national digital advancement, machine learning for economic and digital evolution predictions, and the selection of digital evolution scores and MSEs.

4.2.1 Mega-Sport Events and Digital Development

The advent of digital technology has significantly reshaped national development trajectories, influencing economic competitiveness, public service delivery and innovation systems. This section synthesizes key findings from the existing literature to explore how the hosting of MSEs is linked to digital evolution and development at the national level.

Barrett argued that digital transformation has become a defining factor in how countries position themselves in the global economy [146]. To enhance its international competitiveness, a nation must demonstrate the ability to adapt and integrate within the global digital sphere. In this context, hosting MSEs serves as a strategic inflection point, prompting governments to accelerate investments in digital public infrastructure, promote digital literacy, and adopt emerging technologies to meet the heightened demands and global visibility associated with such events.

From a practical standpoint, Jukan and Mambretti emphasized the evolution of optical networking as a foundation for enabling nationwide digital services, from real-time media delivery to remote scientific computing [147]. This infrastructure is often rapidly scaled during MSE preparation, developing digital legacies such as broadband access, high-resolution scientific visualization, real-time media production, and integrated service platforms. This demonstrates the broader impact of technological advancements on business capabilities.

According to Kaggwa *et al.*, digital transformation supports equitable and sustainable development, a goal that aligns with the long-term legacy aspirations of MSE hosts [148]. Similarly, Luo *et al.* demonstrated how digital advancement can indirectly stimulate green innovation by reshaping industrial structures, a transition that is often accelerated through infrastructure and regulatory reforms associated with MSE hosting [149].

For instance, the Tokyo 2020 Olympics showcased Japan's capabilities in robotics and artificial intelligence, with key innovations, including 3D athlete tracking systems and assistive robotics, patented by leading institutions such as NTT and Toyota [150, 151]. Similarly, major football tournaments, such as the FIFA World Cup, drive innovation in gameplay and officiating technologies, particularly the Video Assistant Referee (VAR) system, which supports accurate match decisions. These reviews highlight the growing transparency and interconnection between hosting MSEs and national digital development. Digital transformation has shifted from a minor aspect of business to a fundamental component of national policy. By hosting MSEs, both the urgency and visibility required to initiate widespread digital improvements have increased significantly. However, this connection remains largely anecdotal, as there is a distinct lack of empirical evidence systematically comparing host and non-host nations. This study addresses this critical gap by investigating the differential impact of digital evolution on these two groups.

4.2.2 Economic Drivers of National Digital Advancement

The economic implications of hosting MSEs extend beyond short-term gains, offering strategic opportunities that can indirectly foster national digital advancement through increased fiscal capacity, infrastructure investment and broader economic modernization.

The economic momentum generated by MSEs, such as increased investment flows, improved infrastructure, and policy reform, can create a fertile ground for digital transformation. For instance, Wang found that digital innovations improve resource allocation and foster cross-sector collaboration, which are more readily achieved when nations possess fiscal flexibility and investment capital, which is often mobilized during MSE preparation. Georgescu *et al.* demonstrated a strong correlation between digitalization and GDP growth, attributing nearly 70 percent of the observed economic expansion in their study to digital advancements [152]. Their findings underscore the direct contribution of investments in ICT, digital public services, and digital skills to national productivity and global competitiveness in the Philippines.

Similarly, Cruz-Jesus *et al.* highlighted the nonlinear dynamics between digital development and income levels across countries [153]. Their study challenges the notion of consistent digital benefits, demonstrating that in lower-income nations, digital investments can yield significantly greater growth impacts, particularly when combined with infrastructure enhancements and knowledge dissemination. Mshvidobadze further reinforced this connection by arguing that advancements in ICT infrastructure contribute directly to GDP growth and broader digital ecosystem development [154].

While a substantial body of research has explored the economic impact of MSEs, the specific empirical connection between these events and a host nation's digital advancement remains unexplored. Previous studies have highlighted the macroeconomic momentum, investment influx, and infrastructure improvements driven by MSEs; however, none have specifically modelled how these elements contribute to digital progress, particularly using ML techniques. To address this critical gap, the present study employs predictive ML models to

analyze how economic parameters and the hosting of MSEs can be leveraged to strengthen national digital development.

The digital evolution score, as a proxy for national digital development in this study, has emerged as a critical benchmark for assessing national digital progress. It captures a country's capacity to adopt, expand, and sustain digital infrastructure, promoting digital trust, economic participation, and institutional adaptability. These capabilities have become central to national development agendas as nations increasingly perceive digital advancement not only as an end goal but also as a tool for fostering sustainable economic development in host nations.

4.2.3 Machine Learning (ML) for Economic and Digital Evolution Predictions

Machine learning methods provide new insights and forecasting capabilities across various fields, including economics. Kresova and Hess employed machine learning to identify key financial indicators, including raw milk prices, across multiple regions in Russia [155]. These approaches handle multivariate problems, nonlinear regression, and misspecification, thereby enhancing the economic analysis. Zhang and Hamori suggested that integrating ML with traditional economic modelling frameworks can substantially improve predictive performance [156]. Neural-boosted models and KNN have gained recognition for their effectiveness, providing robust tools for forecasting models, as discussed in the following review.

Neural-boosted models have demonstrated strong performance across a range of applications, improving accuracy, efficiency, and overall predictive capability. For instance, Zhang *et al.* employed a neural-boosted model to classify football events, highlighting the value of advanced computational techniques for organizing and analyzing sports data [157]. Similarly, Wen and Liang developed a sports management system using a multilayer neural network to implement a danger warning mechanism, underscoring the importance of neural networks in ensuring event safety and operational security [158]. In economics, neural networks have gained prominence for their robustness in handling complex nonlinear data. The KNN model has been widely used in studies examining the economic impact of mega-sporting events. Researchers have employed the KNN algorithm to analyze the effect of sports tourism on economic growth in developing nations, accounting for factors such as international tourism, arrivals, gross fixed capital formation, and total labor force. Widodo *et al.* focused on residents' quality of life during mega-sport events, using the KNN model to assess perceived satisfaction and support [159].

Table 4.1 Comparison of Prior Literature and the Present Study

Study Focus	Methods Used	Focus on Digital Evolution	ML Application to MSEs	Contribution of the Present Study
Knott et al. (2021); Cruz-Jesus et al. (2016)	Regression; Descriptive Analysis	Indirect/Partial	No	Synthesizes economic and digital variables under the ML framework
Zhang (2020); Widodo et al. (2022)	KNN, Neural Nets in sports data	None	Yes (for sports outcomes)	Applies ML to national-level digital impact from MSEs
Present Study	Neural Boosting, KNN	Direct modeling	Yes (for economic digital link)	Forecasts the digital legacy of MSEs using host and macroeconomic data

Although methodologies such as neural networks and KNN are well established in economic and MSE research, a significant research gap exists in their application to forecasting digital development. The present study bridges this gap by applying neural boosting and KNN models to predict national digital evolution from MSE-influenced economic indicators, thereby advancing methodologically over traditional linear models used for such impact assessments. The novelty of this study lies in applying neural boosting and KNN models to examine how macroeconomic indicators shaped by MSE hosting status contribute to predicting national digital evolution. This represents a methodological advancement over the linear models that are traditionally used in MSE impact assessments.

4.2.4 Digital Evolution Score and MSEs selection

The Digital Evolution Score (DES), drawn from Tufts University's Digital Intelligence Index, served as the dependent variable in this analysis [160]. It offers a comprehensive measure of a nation's digital progress by aggregating data from 90 economies and 160 indicators across four dimensions: supply conditions, demand conditions, the institutional environment, and innovation and change. This multidimensional framework ensures a holistic assessment of nations' digital competitiveness.

To operationalize the DES within the models, a threshold value of 50 was established to distinguish between digitally advanced and emerging nations. This cutoff was selected based on the dataset's statistical distribution, with an average DES of 57.63 (SD = 18.33), a maximum score of 98.82, and a minimum score of 21.11. A threshold slightly below the mean provides a balanced classification that captures both emerging and mature digital economies while avoiding extreme outliers. This approach also aligns with clustering practices in prior digital readiness studies by Chakravorti, ensuring comparability with the established literature and avoiding bias during model training [161].

For the independent variables, the analysis incorporated MSEs that represented both global popularity and rotational hosting patterns. Specifically, the sample included the Summer and Winter Olympic Games, FIFA World Cup, UEFA European Championship (Euro), ICC Cricket World Cup, Tour de France, Rugby World Cup, and NBA games (preseason and regular season). These events were selected following Pedersen and Thibault's criteria: (1) international broadcasting reach, (2) rotation of host locations, and (3) status as one of the world's most popular sports in terms of participation and viewership [162]. These criteria ensure that the events analyzed reflect both global exposure and local hosting impacts, making them suitable for investigating the intersection of MSE hosting and national digital transformation.

In summary, the reviewed literature highlights that while MSEs have been linked to economic growth, infrastructural development, and social outcomes, their role in advancing digital transformation remains underexplored. The Digital Evolution Score provides a comprehensive measure of national-level digital progress, and the selection of major global sporting events ensures comparability across diverse hosting contexts. However, existing studies have primarily relied on linear or descriptive methods, offering limited insight into the complex, nonlinear dynamics that underpin digital development. This gap provides a rationale for employing advanced machine learning approaches in subsequent analyses, enabling a more nuanced examination of how economic factors and hosting status jointly influence digital evolution outcomes.

Table 4.2 Data Description of Host, Non-host, and DES

Variable	Description
x_1	Government expenditure on education, total (% of GDP)
x_2	Total natural resources
x_3	Imports of goods and services (US dollar)
x_4	Imports of goods and services (% of GDP)
x_5	Exports of goods and services (% of GDP)
x_6	Exports of goods and services (US dollar)
x_7	Patent application (resident)
x_8	Patent application (non-resident)
x_9	Total patent application
x_{10}	Human Capital Index (HCI)
x_{11}	Tax revenue (% of GDP)
x_{12}	Foreign direct investment
x_{13}	Charges for use of IP payment
y	Digital Evolution Score (DES): captured through two specifications: (i) a continuous score reflecting the level of digital development, and (ii) a binary dummy variable coded as 1 for observations with DES greater than 50 and 0 for those with DES of 50 or below.
Nation Host	Host MSE Status: operationalized as a binary indicator, taking the value of 1 for nations that hosted a mega-sport event during the observation period and 0 for non-host nations.
Before-After	Before and After MSE Hosting: operationalized as a binary indicator, taking the value of 1 if a nation hosted a mega-sport event prior to 2016 and 0 if hosting occurred after 2016.

4.3 Data Summary

The final dataset comprises 1,875 observations from 89 non-host countries and 24 host countries, spanning 2008 to 2022. Table 4.2 presents the variables used in this analysis. The key explanatory variables, hosting status (MSEs hosting versus non-hosting) and temporal period (pre- versus post-hosting), are used to predict the Digital Evolution Score (DES).

The analytical framework incorporates 13 primary economic predictors to assess their relative contributions to DES variation and predictive performance. Economic indicators were obtained from the World Bank Group (WBG) database, and the Digital Evolution Score (DES), serving as the dependent variable, was sourced from Tufts University's Digital Intelligence Index.

In before-and-after MSE analysis sessions, 2016 was used as the midpoint reference for the typical hosting time of MSEs. This year was selected because it represents the average hosting year across the sampled nations and coincides with the midpoint of the study period (2008–2022). Accordingly, 2016 serves as an appropriate temporal benchmark for distinguishing pre- and post-hosting digital evolution trends while ensuring a balanced temporal coverage across observations.

4.4 Analytical Results

4.4.1 DES Comparison between Host and Non-Host Countries

This analysis examines whether hosting mega-sport events is associated with systematic differences in national digital development outcomes by applying traditional statistical approaches. By comparing the Digital Evolution Scores (DES) of host and non-host countries,

the t-test provides a baseline statistical assessment of whether countries hosting mega-sport events exhibit higher average levels of digital evolution (a continuous score). This comparison serves as an initial, theory-driven validation of the hypothesized link between mega-sport event hosting and digital transformation, complementing subsequent machine-learning analyses that explore nonlinear mechanisms and heterogeneous effects.

Table 4.3 presents the results of the independent-samples t-test comparing Digital Evolution Scores (DES) between countries that hosted mega-sport events and those that did not. The results indicated a statistically significant difference in mean DES scores between the two groups. Host countries exhibit a substantially higher average DES than non-host countries, with the difference remaining robust under unequal-variance assumptions. The narrow confidence intervals around the group means suggest stable estimates despite observable dispersion within each group. The extremely low p-value- provides strong statistical evidence against the null hypothesis of equal means, confirming that host and non-host- countries differ systematically in their levels of digital evolution. The degrees of freedom reflect the adjustment for heteroscedasticity, reinforcing the reliability of inference.

Figure 4.1 visualizes the distribution of the DES for host and non-host countries using a one-way comparison plot. The figure clearly shows an upward shift in the DES distribution for host countries compared to non-hosts-. While both groups display internal variability, the central tendency and interquartile range for host countries are consistently higher. The scatter plot structure further indicates that host countries not only achieve higher average digital evolution outcomes but also exhibit a broader performance spread, suggesting heterogeneous digital trajectories after hosting megasport events. In contrast, non-host- countries cluster more tightly around lower DES values, reflecting comparatively limited digital advancement in these countries.

Table 4.3 Descriptive Statistics of DES for Host and Non-Host Countries

MSEs	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%	p-value	DF
Non-Host	1515	38.3417	20.1316	0.51721	37.3271	39.3562	<.0001*	579.90
Host	360	53.3612	18.4482	0.97230	51.4491	55.2734		72

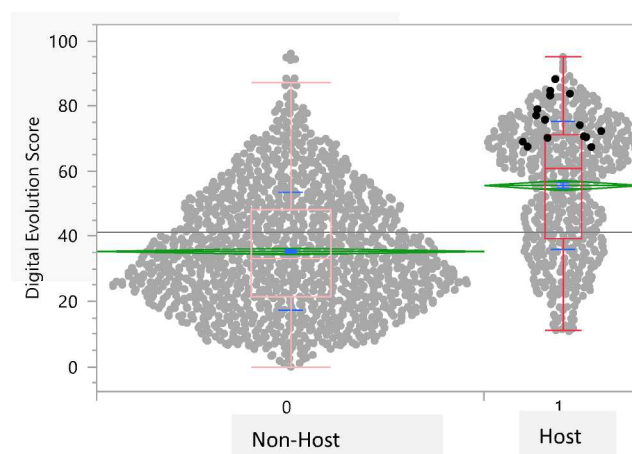


Figure 4.1 One-way Analysis of Digital Evolution Score by Non-Host and Host Country

Collectively, these results provide convergent evidence that hosting MSE is associated with significantly higher levels of digital evolution in the host country. The findings also corroborate the machine learning- results, supporting the argument that MSEs function as catalysts for digital transformation, through accelerated infrastructure investment, institutional modernization, and enhanced global knowledge exchange. Notably, the persistence of variability across host countries implies that hosting alone is insufficient; strategic leveraging and governance quality are crucial for translating event opportunities into sustained digital gains.

Building on these findings, the subsequent analysis employs a neural boosting model to examine further the complex relationship between mega-sport event hosting and digital development outcomes, using the digital evolution score as a binary indicator. The results presented in this section report the model's regression performance and the validation metrics. The evaluation includes R-squared, root mean squared error [98], mean absolute deviation, misclassification rate, negative log-likelihood, and sum of frequencies, as summarized in Table 4.4. The generalized R-squared of 0.199 indicates that hosting status accounts for approximately 20% of the variance in the digital evolution scores. The entropy-based R-squared of 0.120 indicates modest predictive power, suggesting that MSEs play a catalytic yet partial role in shaping a nation's digital trajectory. This supports the view that such events should complement, rather than replace, foundational development policies. These metrics offer a comprehensive view of the model performance. The RASE and mean absolute deviation indicated moderate predictive accuracy.

Furthermore, a misclassification rate of 0.313 means that the model correctly classifies nearly 70 percent of the cases, underscoring its practical utility. The negative log-likelihood value confirmed the model's ability to generate probabilistic estimates that meaningfully aligned with the observed data. Taken together, these results suggest that hosting MSEs exerts a measurable and consistent influence on digital evolution, validating its role as a significant driver of the broader set of factors shaping national digital development.

In machine learning, confusion is a metric used to evaluate the performance of a classification model on data for which the actual values are known. The confusion matrices provide details on the model's accurate and incorrect predictions, classified by type, as shown in Table 4.5. The neural boosting model produced 49 correct predictions for the positive class (high MSE–high DES) and 43 correct predictions for the negative class (non-host–low DES). It also recorded 10 false negatives and 32 false positives, resulting in an overall confusion rate of 0.313. Class-specific performance indicated strong accuracy for the negative class (0.811) and moderate accuracy for the positive class (0.605), with corresponding error rates of 0.189 and 0.395, respectively. These results suggest that while the model more reliably distinguishes non-host nations with lower digital scores, it also demonstrates meaningful predictive power in identifying host nations with higher digital scores. Achieving over 60 percent accuracy for the positive class is noteworthy, given the complexity of capturing digital advancements through MSE hosting, which highlights both the effectiveness of the neural boosting approach and the scope for further refinement with expanded data and features.

Table 4.4 Prediction Performance of the Neural Boosted Model

Validation Y	
Measures	Value
Generalized R-Square	0.1992
Entropy R-Square	0.1186
RASE	0.4506
Mean Abs Dev	0.3847
Misclassification Rate	0.3134
-Log Likelihood	79.267
Sum Freq	134

Table 4.5 Neural Boosting Model's Confusion

Confusion Matrix			Confusion Rates		
Actual	Predicted Count		Actual	Predicted Rate	
Binary Y	Low (0)	High (1)	Binary Y	Low (0)	High (1)
Non-hosted (0)	43	10	Non-hosted (0)	0.811	0.189
Hosted (1)	32	49	Hosted (1)	0.395	0.605

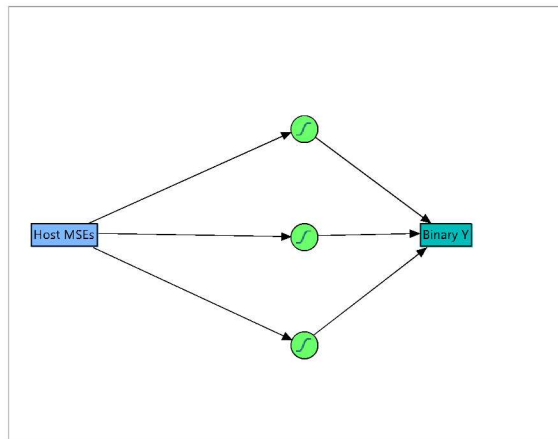


Figure 4.2 Neural Boosting Model Diagram.

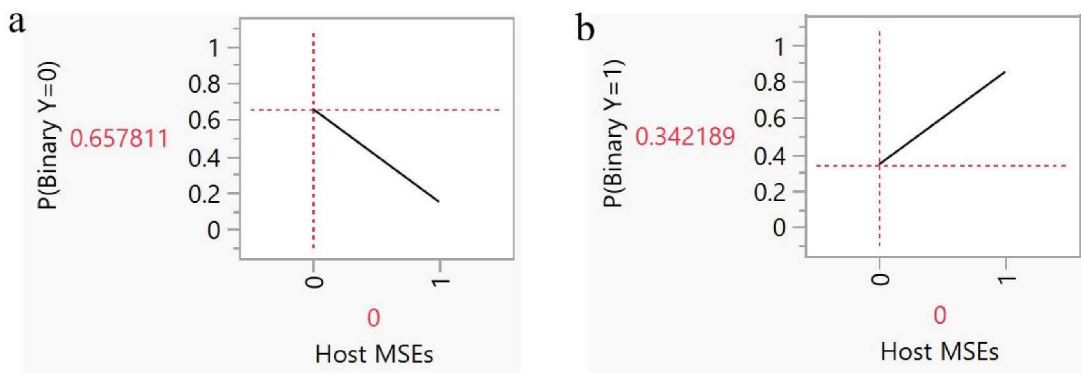


Figure 4.3 (a) Prediction Profiler with Low Digital Evolution Score
(b) Prediction Profiler with High Digital Evolution Score

Figure 4.2 illustrates the neural boosting model, where "Host MSEs" serves as the binary independent variable, indicating whether a nation is hosting a mega sporting event. This input is processed through a series of hidden layers comprising weak learners that utilize weighted connections to transform the data. The objective of the model is to predict the binary outcome "Binary Y," where 1 indicates a high level of digital evolution and zero a low level.

The predictive profiler indicated that the probability of achieving a low digital evolution score for non-hosted nations was approximately 0.658 (Figure 4.3 (a)). Hosting MSE resulted in significant variations in the associated probabilities of the digital evolution scores. The model predicts a decrease in the probability of a low score ($P(\text{Binary } Y=0)$) for the host country, underscoring the potential catalytic impact of MSEs on the host country's digital infrastructure and its capabilities. In contrast, the probability of high digital evolution scores for non-hosted countries was estimated to be 0.342, as shown in Figure 4.3(b). The likelihood of a high score ($P(\text{Binary } Y=1)$) increases when a country actively hosts MSEs, indicating that the responsibilities and opportunities associated with hosting MSEs may contribute to and expedite digital advancement. These predictive insights underscore the transformative potential of hosting MSEs in a country's digital ecosystem.

The findings of this analysis confirm that hosting MSEs is positively associated with national digital evolution. The neural boosting model effectively differentiated between host and non-host countries, achieving an overall classification accuracy of nearly 70% and a positive class accuracy of over 60%. The probabilistic outputs further demonstrate that hosting MSEs substantially increases the likelihood of achieving higher digital evolution scores. Simultaneously, non-host nations are more frequently observed in the lower category. These results provide strong empirical support for the argument that MSE hosting serves as a catalytic driver of digital advancement, reinforcing the view that international visibility, infrastructural investment, and organizational demands linked to MSEs contribute to national digital development trajectories.

Table 4.6 Performance of the Neural Boosted Model for DES Prediction

Non-hosted MSEs		Hosted MSEs	
Measures	Value	Measures	Value
RSquare	0.8158	RSquare	0.8981
RASE	7.3517	RASE	6.0274
Mean Abs Dev	5.3733	Mean Abs Dev	4.3210
-LogLikelihood	303.8343	-LogLikelihood	192.9149
SSE	4810.2126	SSE	2179.7489
Sum Freq	89	Sum Freq	60

Table 4.7 Performance of the KNN Model for DES Prediction

Non-hosted MSEs				Hosted MSEs			
K	R-Square	RASE	SSE	K	R-Square	RASE	SSE
1	0.9227	4.5964	5598.69*	1	0.9740	2.774	1384.70*
2	0.8809	5.7043	8622.97	2	0.8989	5.471	5387.78
3	0.8479	6.4460	11011.10	3	0.8205	7.288	9561.43
4	0.8190	7.0318	13103.10	4	0.7922	7.841	11068.00
5	0.8026	7.3423	14286.00	5	0.7513	8.579	13246.40

In the following step analysis, all economic predictors were used in the neural-boosted model to estimate digital evolution scores for both hosted and non-hosted nations. As shown in Table 4.6, the model performance showed notable differences between the two groups. For non-hosted nations, the model achieved an R-squared value of 0.8158, indicating a reasonably strong fit, with an RASE of 7.3517 and a Mean Absolute Deviation of 5.3733. In contrast, the model performed more effectively for hosted nations, with an R-squared of 0.8981, indicating improved explanatory power. Additionally, the hosted group demonstrated lower error metrics and a Mean Absolute Deviation, suggesting greater predictive precision. These results suggest that the model is better suited to estimating digital evolution in nations that host MSEs, reflecting stronger and more consistent relationships between the predictors and digital outcomes in these contexts.

The performance comparison of the KNN model between non-hosted and hosted mega sports event nations demonstrates notable differences in predictive accuracy and error metrics across varying values of k , as shown in Table 4.7. For non-hosted nations, the model achieved its highest R-squared at $k = 1$, with corresponding RASE and Sum of Squared Errors (SSE). For hosted nations, the model also achieved its highest R-squared at $k=1$, with substantially lower RASE and SSE values, suggesting superior predictive performance and lower error margins compared to non-hosted nations. Moreover, hosted nations consistently exhibit lower SSE values across all ' k ' values than non-hosted nations; this suggests that the KNN model performs better for hosted nations, due to more consistent and impactful data, which provides more accurate and reliable predictions in these scenarios. When comparing model performance, the KNN model, particularly at its optimal k value of 1, consistently outperformed the neural-boosted model in terms of R-squared values in both non-hosted and hosted nations, indicating more accurate and dependable predictions, particularly for the host nations. This comparison demonstrates the potential benefits of the KNN model when the optimal k value is determined and effectively used.

Figure 4.4 shows the residual by predicted plots for the KNN model, illustrating the model's error distribution for non-hosted and hosted mega sporting event nations. The scatter plot shows a spread around the zero line, indicating the difference between the actual and predicted values. For the hosted nations, as shown in Figure 4.4(b), the residuals appear to be mostly centered around the zero line, indicating that the model predictions are generally accurate. This confirms that the KNN model performs exceptionally well for hosted nations, with residuals close to zero; therefore, the model's assumption is valid.

The application of neural-boosted and KNN models demonstrates their value in capturing the nonlinear relationships underlying national digital evolution. Both approaches highlight that predictive performance is consistently stronger for host nations, with higher R-squared values and lower error metrics than for non-host nations. This finding is significant because it suggests that hosting MSEs amplifies the alignment between economic predictors and digital outcomes, creating more structured and explainable patterns for machine learning models to detect. From a practical perspective, the superior accuracy in host contexts highlights the catalytic role of MSEs in driving measurable digital advancements, providing policymakers and event organizers with a reliable predictive tool to anticipate and leverage the legacies of digitalization in the future.

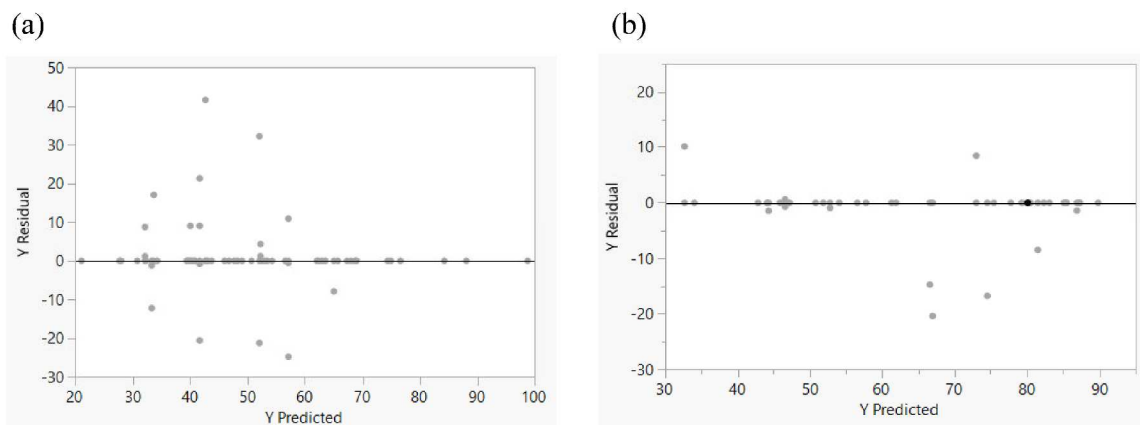


Figure 4.4 The Residual by Predicted Plot of the KNN Model Applied to
(a) Non-Hosted Nations (b) Hosted Nations

4.4.2 The DES Comparison: Before and After MSEs (2016)

This section presents a comparative analysis of digital evolution score (DES) between host and non-host countries across the pre-2016 and post-2016 periods. The subsequent subsections are organized as follows: first, cross-national DES descriptive statistics are reported; second, variations in XGBoost model predictions are analyzed; and third, the predictor screening results for host countries are compared across temporal periods.

The t-test analysis comparing the DES of host and non-host countries across the pre- and post-2016 periods indicates statistically significant improvements in the average DES for both groups, with notable differences in the magnitude and underlying dynamics of digital evolution between host and non-host countries.

Table 4.8 shows statistically significant increases in average DES scores during the post-2016 period for both host and non-host countries, though the patterns differ in important respects. Non-host countries exhibit elevated scores with limited confidence interval overlap, suggesting systematic improvement driven by global digitalization trends, technological diffusion, policy harmonization, and enhanced connectivity, independent of event hosting. Host countries demonstrated similar but more pronounced gains, with substantially higher post-2016 values that reached strong statistical significance. Notably, the wider dispersion among host countries indicates heterogeneous outcomes, suggesting that while hosting correlates with digital advancement, gains vary considerably depending on national strategies, governance capacity, and event-leveraging effectiveness.

The analysis confirms a global upward trend in digital development after 2016, which positively affected both host and non-host countries. However, two critical distinctions have emerged in the literature. First, host countries consistently exhibit higher DES levels in both periods. Second, their post-2016 gains exceeded those of non-host countries. These findings suggest that while global digital transformation establishes a standard baseline trajectory, hosting mega sports events may serve as an accelerator, driving enhanced infrastructure development, institutional modernization, and international knowledge exchange. Crucially, the observed heterogeneity among host countries underscores that hosting alone does not guarantee a uniform digital outcome. This variability highlights the importance of complementary factors, including governance quality, investment prioritization, and strategic event planning, in converting mega-sport events into a sustained digital legacy.

Table 4.8 Descriptive Statistics of DES Before and After 2016

MSEs	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%	p-value	DF
Non-Host								
Before	808	32.9986	18.8297	0.66242	31.6983	34.2989	<.0001*	1513
After	707	44.4480	19.8450	0.74635	42.9827	45.9134		
Host								
Before	192	48.0501	17.6660	1.27493	45.5353	50.5648	<.0001*	352
After	168	59.4311	17.4697	1.34781	56.7702	62.0921		

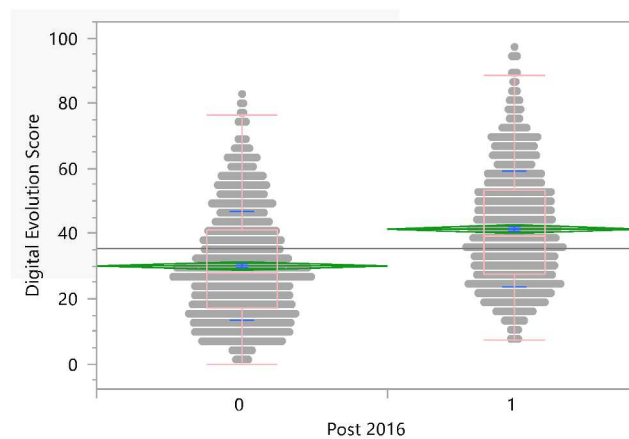


Figure 4.5 One-way Analysis of Non-Host Nations, Before and After 2016

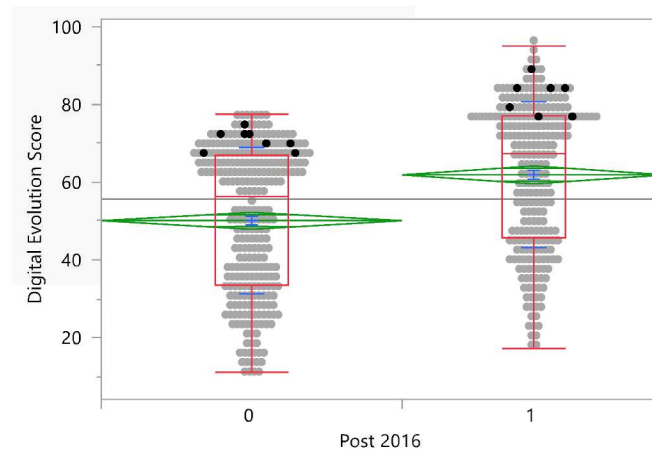


Figure 4.6 One-way Analysis of Host Nations Before and After 2016

Figures 4.5 and 4.6 present distributional analyses of digital evolution scores (DES) across the pre-2016 and post-2016 periods, distinguishing between host and non-host countries. Using scatter plots overlaid with box-and-whisker representations, these visualizations facilitate a comparative assessment of the central tendency, dispersion, and heterogeneity of digital performance over time. Figure 4.5 reveals a clear upward shift in the DES distribution among non-host countries after 2016. The post-2016 period demonstrated an elevated central tendency and an expanded upper range compared with the pre-2016 period. Despite persistent dispersion,

the distributional shift indicates broad-based digital improvements driven by global digitalization dynamics, including technological diffusion and policy convergence rather than event-specific mechanisms.

Figure 4.6 demonstrates a more pronounced temporal pattern in the host countries. The post-2016 distribution exhibited a higher central tendency and greater variance relative to the pre-2016 period. Notably, the upper tail contains several high-performing observations, suggesting that certain host countries have achieved substantial digital gain. Simultaneously, the increased dispersion reveals heterogeneous trajectories, indicating that the magnitude of digital advancement varies considerably across national contexts and strategic frameworks.

Both visualizations confirm a global upward trend in digital development after 2016, affecting host and non-host countries. However, critical distinctions emerged in the comparative pattern. Host countries exhibit consistently higher DES levels and a broader performance range, suggesting that hosting mega-sport events may amplify digital evolution beyond baseline global trends. The wider dispersion among host countries underscores that hosting effects are mediated by complementary institutional and policy factors rather than operating uniformly across all countries. These visual patterns corroborate the statistical findings presented in the corresponding tables, supporting the interpretation that mega-sport events can catalyze digital transformation while highlighting the conditional nature of such effects. The heterogeneity across host countries underscores the importance of governance quality, strategic planning, and investment prioritization in translating event hosting into sustainable digital legacy outcomes.

When examining individual countries, the results revealed a consistent pattern: every host nation exhibited a higher average DES after 2016. For instance, Australia's DES increased from 65.32 to 73.65, Japan's from 65.37 to 76.14, and Germany's from 65.59 to 79.00. Some countries demonstrated particularly sharp increases, such as Bangladesh (13.33 to 23.80), Brazil (30.99 to 40.82), and Russia (34.86 to 49.39), indicating an accelerated digital transformation in nations that initially had lower digital baselines (Figure 4.7).

The t-test results further substantiate the statistical robustness of these observed improvements. All countries reported p-values below 0.001, demonstrating that the differences in DES use before and after 2016 were not only statistically significant but also consistently impactful across the sample. Notably, this significance emerges across a broad spectrum of economic contexts, from technologically advanced nations such as Denmark and Canada to rapidly developing economies such as India and Sri Lanka. This indicates that the post-2016 acceleration in digital infrastructure and technology adoption was broad-based and not confined to specific income groups or to regional clusters. Collectively, these findings strongly indicate that host nations have experienced significantly greater digital evolution since 2016. This systematic increase suggests that the post-2016 global digital environment, combined with country-specific innovation and infrastructure strategies, has contributed to substantial digital advancements. These results offer a compelling foundation for further investigation into how MSE-related investments, national policy shifts, and global technological trends collectively shape the digital trajectory.

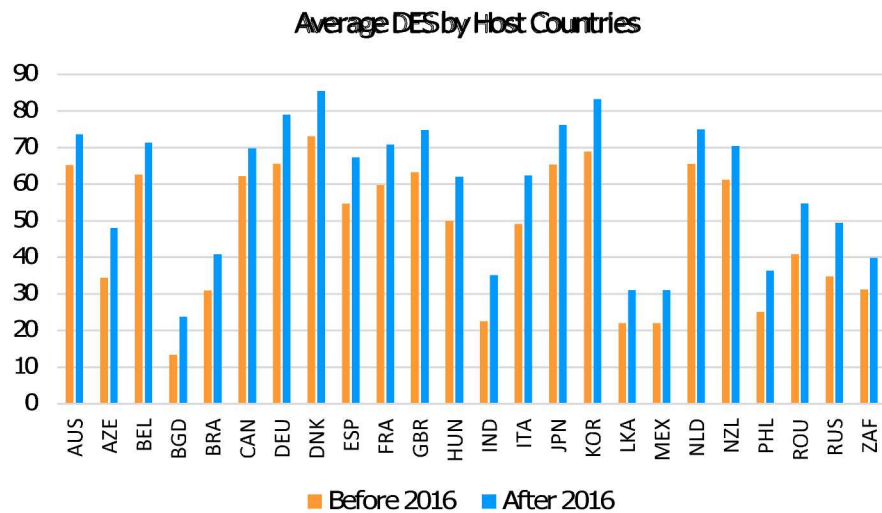


Figure 4.7 Average Digital Evolution Score (DES) by Host Countries

DES Prediction by XGBoost Model: Before and After 2016 Comparison on Host Countries

After the model screening process, XGBoost proved to be the most suitable for DES predictions when the hosting MSE performance before and after 2016 was compared. Table 4.9 summarizes the performance of the XGBoost model in predicting the DES for the pre-2016 and post-2016 periods separately. The results indicate exceptionally high predictive fidelity across both timeframes, reflecting the model's capacity to account for nearly all systematic variations in the DES. The consistently strong fit underscores the stability of the underlying relationships and demonstrates a high degree of concordance between observed and predicted values.

The error metrics provide additional interpretive insights. The model demonstrated a marked reduction in predictive deviation in the post-2016 period, indicating a more stable relationship between the predictors and the DES. This improvement suggests that the digital evolution trajectories have become more systematically structured, enabling a more consistent approximation of the underlying patterns. The accompanying actual-by-predicted plots, as shown in Figure 4.8, illustrate a strong correspondence between the predicted and observed values across both periods, with predictions closely tracking the observed digital evolution trends and maintaining a coherent visual alignment, indicative of reliable performance.

These results demonstrate that XGBoost exhibits superior performance in modeling DES compared to the alternative machine learning techniques examined in prior analyses. The enhanced predictive accuracy observed in the post-2016 period may reflect increasingly structured digital evolution patterns as digital technologies mature, and adoption processes become more systematic across host nations. Furthermore, the gradient boosting architecture of XGBoost appears particularly effective at capturing complex interactions and nonlinear relationships that alternative models fail to represent adequately. Collectively, these findings establish the robustness and precision of XGBoost in predicting DES across temporal periods, thereby justifying its application in subsequent analyses of digital transformation trajectories among mega-sport event host nations.

Table 4.9 Statistic Description of XGboost Model For DES Prediction Before and After 2016

Measure	Before 2016	After 2016	Definition
	Value	Value	
R-Square	0.9996	0.9998	$1 - \text{SSE}(\text{Model})/\text{SSE}(\text{Intercept})$
Correlation	0.9998	0.9999	$\text{Corr}(y, \hat{y})$
Root Mean Square Error	0.3351	0.2158	$\sqrt{\sum (y[j] - \hat{y}[j])^2 / n}$
Mean Absolute Error	0.2507	0.1482	$\sum y[j] - \hat{y}[j] / n$
Mean Response	48.114	59.431	$\text{Mean}(y)$
Mean Prediction	48.108	59.421	$\text{Mean}(\hat{y})$
Sum of Weights	193	168	n

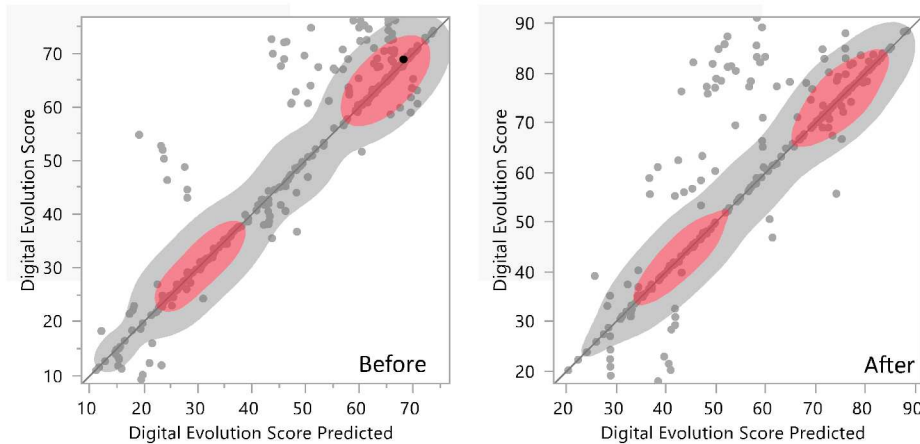


Figure 4.8 Actual vs. Predicted Plot Before and After 2016

Predictor Screening Result: Before and After 2016 Comparison on Host Countries

This subsection presents the comparative results of predictor screening, identifies key variables, and quantifies their respective contributions to model performance before and after 2016.

Figure 4.9 presents the results of the predictor screening process for the pre-2016 and post-2016 periods. The screening results revealed notable differences in the relevance of predictors between the two periods. In the pre-2016 period, payments for the use of intellectual property (IP) made the most significant contribution, followed by tax revenue (% of GDP), total natural resources, and exports of goods and services (US\$). These variables exhibited substantial predictive power, indicating their importance as determinants of digital evolution in the earlier phase.

The post-2016 period exhibited a different pattern of predictor contributions. Charges for IP payments remained the dominant predictor, with a decreased contribution, while patent applications (residents) emerged as a substantially more important variable than in the pre-2016 period. Tax revenue (%GDP) was significant, though its relative contribution has decreased. Government expenditure on education (%GDP) showed enhanced relevance in the post-2016 screening, suggesting a stronger relationship with digital evolution outcomes.

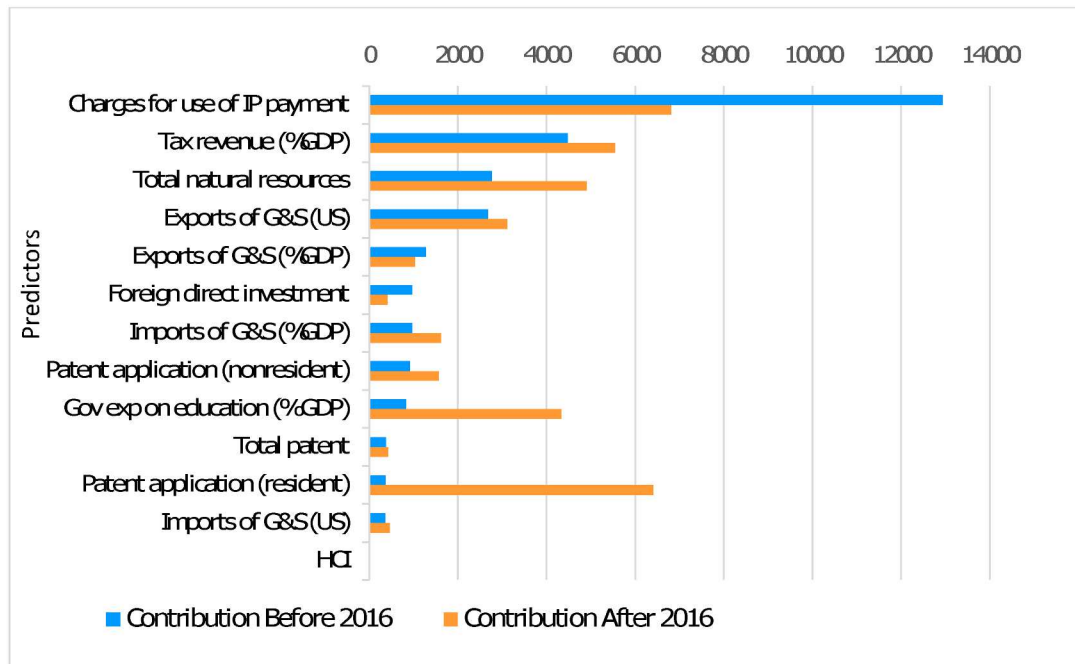


Figure 4.9 Comparison of Predictor Contributions

Several variables consistently demonstrated minimal contributions across both periods, including HCI, total patents, and imports of goods and services (US\$). These variables were identified as weak predictors and may be candidates for exclusion from the final model to improve parsimony without sacrificing predictive performance. The temporal shift in predictor importance suggests that the underlying drivers of digital evolution have evolved between periods, with innovation-related indicators, particularly resident patent applications, gaining prominence after post-2016 period.

4.5 Discussion

The empirical findings demonstrate that hosting MSEs is significantly associated with enhanced national digital development, as measured by the digital evolution score (DES). The machine learning models show that host nations consistently achieve stronger predictive performances than non-host nations. Traditional inferential analyses further corroborate these patterns: independent samples t-tests confirm significantly higher DES levels among host countries, with these differences persisting under unequal variance assumptions.

Temporal analyses centered on 2016 provide further insights into the effects of mega-sport event hosting. While both host and non-host countries exhibit statistically significant improvements in DES after 2016, reflecting broader global digitalization trends, the magnitude of these gains is consistently greater in host nations. Probability profiles and residual analyses further confirm that MSE hosting strengthens the relationship between economic predictors and digital outcomes. Collectively, these findings support the interpretation that MSE hosting functions as an amplifying mechanism for digital advancement, accelerating digital development beyond baseline global trends rather than merely mirroring it.

The temporal shift in predictor contributions between the pre- and post-2016 periods reveal fundamental changes in the mechanisms underlying national digital evolution. While charges for intellectual property (IP) payments remained the dominant predictor across both periods, their reduced relative contribution after 2016 signaled a structural transition in the accumulation of digital capabilities. During the earlier period, elevated IP payments reflected reliance on external technologies and licensed knowledge, consistent with a technology-adoption phase characterized by catch-up strategies.

The post-2016 period demonstrates a pronounced increase in the importance of resident patent applications, indicating a shift toward endogenous innovation and the creation of domestic knowledge. This transition suggests that countries are increasingly moving from technology adoption to technology generation, a pattern more closely associated with sustained digital evolution. The rising contribution of patent activity reflects the maturation of national innovation systems, in which digital capabilities are embedded within domestic R&D ecosystems rather than acquired through external IP licensing. A parallel structural shift in government expenditure on education emerges, with enhanced predictive power in the post-2016 period. Educational investment strengthens human capital, digital skills, and absorptive capacity, all of which are fundamental to an advanced digital transformation. This growing importance suggests that post-2016 digital evolution has become more firmly anchored in long-term capability building rather than short-term technological acquisition.

Although tax revenue (% GDP) remained significant across both periods, its declining relative contribution indicates that fiscal capacity alone became less decisive in explaining digital outcomes once foundational infrastructure was established. Innovation-driven and human capital-based mechanisms have assumed a more central role in shaping digital evolution trajectories. Collectively, the results suggest that the post-2016 digital evolution shifted from externally driven digitalization to an internally driven transformation driven by domestic innovation and education. This transition aligns with the stronger digital performance observed among mega-sport event host countries, where event-related investments and global exposure may accelerate the shift from adoption-based to innovation-led digital development. This pattern reinforces the interpretation that digital evolution increasingly depends on national innovation capabilities and human capital development rather than financial inputs alone, particularly in the post-2016 environment.

Academic Contribution: This chapter advances the academic literature on the impacts of MSE and digital transformation by integrating machine learning into the analysis of legacy outcomes. While prior research has predominantly examined economic or infrastructural impacts using linear regression models, this study contributes by employing nonlinear ML approaches that better capture the complexity of digital transformation processes. The comparative analysis between neural boosting and KNN models also adds methodological depth, demonstrating the importance of model choice and validation in forecasting national-level digital outcomes. Furthermore, evidence that predictive performance improves in host nations provides empirical support for theorizing that MSEs accelerate digital maturity, extending the conceptual boundaries of legacy research into the digital domain.

Practical Contribution: From a practical perspective, the results underscore the utility of ML-based forecasting for event organizers, sponsors, and technology providers. By identifying key predictors, such as trade indicators, intellectual property payments, and export intensity, the models provide actionable insights into which economic levers most strongly support digital development. The superior performance of ML models in host contexts also suggests

that predictive analytics can be used to design targeted interventions that maximize digital legacies, such as by prioritizing digital infrastructure investments or leveraging event-driven visibility to accelerate technological adoption. These insights are valuable for practitioners seeking evidence-based tools to align MSE planning with broader digital transformation goals.

Policy Contribution: The findings provide compelling evidence that hosting MSEs can act as a strategic lever for advancing national digital agendas. The ability of ML models to capture the structured relationships among hosting status, economic factors, and digital evolution outcomes suggests that governments can rely on such approaches to achieve more accurate legacy planning. Policymakers can use these insights to justify investments in digital infrastructure during MSE preparations, design supportive regulatory frameworks for digital adoption, and integrate event legacies into long-term development strategies. The evidence that host nations experience greater predictive alignment also underscores the importance of proactive planning to sustain digital gains after the event.

Despite these contributions, this study has some limitations that should be acknowledged. First, the dataset covered a relatively short time frame and may not fully capture long-term digital legacies beyond the immediate post-event years. Second, although the models included a wide range of economic predictors, other contextual factors, such as governance quality, institutional readiness, and cultural technology adoption, were not incorporated. Third, while the models demonstrated strong performance, particularly for host nations, reliance on secondary data imposes constraints on variable completeness and measurement precision. These limitations suggest opportunities for future research to expand the dataset, integrate additional socio-political predictors, and employ hybrid modelling techniques that combine ML with qualitative legacy assessments.

4.6 Concluding Remarks

This chapter investigated the relationship between mega-sport event hosting and national digital development, using the digital evolution score (DES) as a measure of digital maturity. Through the application of machine learning approaches, including neural boosting, KNN, and XGBoost, alongside traditional inferential analyses, the investigation provided robust evidence that MSE hosting correlates significantly with enhanced digital outcomes. The findings consistently demonstrate superior predictive accuracy, higher R^2 values, and lower error metrics for host nations compared to non-host nations, suggesting that infrastructural investment, technological adoption, and global exposure associated with hosting MSEs produce measurable digital benefits.

A central contribution of this analysis is the confirmation that MSE hosting represents not merely an economic or cultural event but a digital inflection point. Temporal analyses show that although both host and non-host countries experienced significant digital improvements after 2016, reflecting global digitalization trends, host countries achieved consistently larger gains. This pattern indicates that MSEs function as amplifying mechanisms, accelerating digital development beyond baseline global trends through structured investments in connectivity, digital services, and innovation. By distinguishing between host and non-host nations across temporal periods, the analysis reveals that hosting effects extend beyond the immediate event horizon and substantively contribute to national digital positioning in the global landscape.

The chapter further demonstrates a fundamental shift in the mechanisms underlying digital evolution between the pre- and post-2016 periods. The declining relative importance of intellectual property payments and the rising contributions of resident patent applications and educational expenditure signal a qualitative transition from externally driven digitalization to internally driven transformation. This shift indicates that digital maturity now relies more on domestic innovation systems and human capital than on technology adoption and fiscal capacity alone. Notably, this structural transition aligns with the amplified digital performance observed among MSE host countries, where event-related investments and global exposure may accelerate reorientation toward innovation-led development.

This methodological approach underscores the value of combining machine learning with traditional statistical methods in digital transformation research. While conventional regression models often struggle to capture the nonlinear dynamics and complex interactions that characterize digital evolution, machine learning algorithms, particularly gradient boosting, demonstrate a superior capacity to model these relationships. The convergence of findings across multiple analytical approaches strengthens confidence in the robustness of the results and highlights the importance of methodological pluralism in the legacy research. The relative performance differences across models and temporal periods further reveal that digital outcomes among host nations are more systematically structured and predictable, suggesting clearer pathways between economic conditions and digital advancement in the hosting context.

From a policy perspective, these findings underscore the strategic importance of integrating digital legacy planning into MSE hosting frameworks. The evidence that hosting status is associated with enhanced digital outcomes and stronger predictor-outcome relationships suggests that policymakers and organizers can proactively design interventions to maximize digital dividends, such as investing in broadband infrastructure, fostering digital entrepreneurship, embedding advanced technologies into event operations, and strengthening national innovation systems. The predictive insights derived from machine learning models thus serve not only as academic contributions but also as decision-support tools for governments, event committees, and private sector partners seeking to leverage MSEs as catalysts for sustainable digital transformation.

In conclusion, this chapter contributes to the growing literature that positions digital transformation as a critical dimension of MSE legacies. By integrating structured indicator data with advanced predictive techniques and temporal analyses, it establishes an empirical framework for understanding how international sporting events intersect with national digital agendas and innovation trajectories in Brazil. The findings confirm that MSE hosting amplifies digital development, particularly in the post-2016 global digital environment, and that this amplification operates through increasingly innovation-driven and human capital-based mechanisms. While these results are encouraging, they also highlight the need for further investigation into the institutional, governance, and social factors that mediate digital legacy outcomes in diverse national contexts. Building on these foundations, the next chapter extends the analysis to examine how MSE-driven economic dynamics contribute to broader patterns of innovation, providing a deeper exploration of the technological legacies of global sporting events.

Chapter 5: Traditional Statistical Techniques and Machine Learning Models for Estimating Economic and Innovation Impacts

Chapter 4 emphasizes digital evolution as a marker of technological readiness and modernization, and Chapter 5 investigates the economic, social, environmental, and financial determinants that influence a host nation's innovation impact through patent applications as a proxy. By employing multiple machine learning techniques, this chapter provides a nuanced understanding of how environmental, social, and governance (ESG) factors, financial investments, and macroeconomic conditions influence innovation capability in host nations.

5.1 Introduction

The intricate relationships between environmental, social, and governance (ESG) factors, economic performance, and innovation have garnered significant attention in recent academic literature and policy discussions. Although the benefits of ESG investments are often promoted as universally beneficial for sustainable economic development, empirical findings have remained mixed, particularly regarding immediate economic impacts and long-term innovation gains. To address these nuances, this study investigates the differential roles of ESG dimensions using robust statistical and machine learning approaches, providing deeper insights into how ESG components distinctively influence GDP growth and innovation capabilities, proxied by patenting activity, in the host countries of MSEs.

Previous research on MSEs has often conceptualized ESG as a monolithic construct, a simplification that obscures the distinct and sometimes divergent impacts of the Environmental, Social, and Governance pillars. This approach risks overlooking the inherent complexities in the relationships between governance quality, social infrastructure, environmental management, and economic dynamics. Furthermore, a significant gap exists in the literature regarding the use of ESG indicators for predictive modelling, as few studies have focused on forecasting economic and innovation outcomes using a comprehensive set of ESG predictors. This chapter pursues three main aims: (1) to examine the extent to which ESG parameters influence national economic growth (measured by GDP per capita) and innovation capability (proxied by patent applications); (2) to compare the predictive results of fixed-effects, negative binomial, and bootstrap forest models; and (3) to explore how ESG parameters, in combination with financial indicators, shape innovation capability (patent applications).

To enhance clarity and policy relevance, this study disaggregates ESG factors and examines the distinct contributions of each dimension through fixed-effects regression and advanced machine learning models, including bootstrap forest, shapley additive explanations (SHAP), partial dependence plots (PDP), and individual conditional expectation (ICE) analyses. In synthesizing the findings, this chapter redefines the discourse surrounding ESG's role in sustainable economic development, moving beyond simplistic linear assumptions to acknowledge nuanced and multifaceted mechanisms at play. The subsequent sections present detailed analytical results, interpretive discussions, and policy implications, contributing valuable theoretical refinements and practical recommendations for optimizing ESG strategies to cultivate innovation and sustainable growth.

5.2 Related Work

5.2.1 ESG Indicators and Economic Growth

The decision to host an MSE is increasingly aligned with national environmental, social, and governance (ESG) objectives, as these factors are widely believed to influence the long-term economic outcomes of host countries. A substantial body of literature supports this connection at the macro level, demonstrating that strong ESG performance correlates with superior economic results. For instance, firms with high ESG scores attract greater foreign direct investment [163], a key catalyst for GDP expansion, whereas national sustainability reporting has been shown to drive measurable gains in output [164].

Within the ESG framework, the social and governance pillars are often critical drivers of corporate value. Research highlights a long-term causal link between social indicators such as human capital investment and economic growth [165]. At the national level, a robust institutional framework underpins social infrastructure, and transparent governance reduces the overall cost of capital, enhances investor confidence, and strengthens long-term economic growth prospects [166]. However, the literature also reveals that this ESG-growth relationship is not monolithic, with studies finding significant regional heterogeneity and differing effects across developed and developing nations [132]. This context dependency is particularly salient for MSEs, which impose unique and intense pressure on a host's economic and institutional systems. The complexity revealed in prior research motivates the present study, which employs a machine-learning approach to explore nuanced ESG-growth dynamics within six diverse economies that have hosted major sporting events.

5.2.2 ESG performance and Innovation

The relationship between environmental, social, and governance (ESG) performance and innovation development are crucial, particularly in countries that host MSEs. Mega-events provide a unique platform for nations to boost their ESG initiatives while simultaneously enhancing their innovation capacities. Research indicates that strong ESG performance can substantially promote technological and green innovations, which aligns with the increasing demand for sustainable practices in contemporary society. Evidence from firm-level studies suggests that strong ESG performance can be a powerful catalyst, spurring both the quantity and quality of green innovations and enhancing a company's propensity to innovate through improved corporate social responsibility programs [167].

Moreover, Yin and Wang outline how government subsidies can mediate the impact of ESG performance on corporate innovation efficiency. Their findings suggest that firms with exemplary ESG practices are better positioned to secure government financial support, thereby enhancing their innovation capability [168]. This aspect is particularly relevant in the context of mega events, where host nations may provide increased subsidies to align with global sustainability goals, further incentivizing innovation in the sports sector and related industries. Additionally, research has found that nations with higher ESG ratings tend to experience more pronounced green innovation patterns [169]. For instance, He *et al.* highlighted that the interplay between ESG practices and innovation efficiency is notably impactful in state-owned enterprises, especially in developing regions that capitalize on infrastructural improvements linked to mega-events [170]. These events often present opportunities for infrastructural and technological advancements that can significantly elevate a nation's ESG standing while facilitating broader societal and environmental benefits.

Furthermore, the immense investment required for an MSE can create a strategic trade-off in which sustainability agendas are pursued at the expense of the core R&D focus, ultimately diminishing the marginal benefits of ESG performance on innovation [171]. The evidence suggests that while strong ESG practices can unlock substantial innovation advantages, these benefits are contingent on genuine implementation and balanced strategic priorities. Therefore, a critical gap remains in understanding the contextual factors, such as policy pressures, capital market demands, and managerial incentives, unique to MSE host countries that determine whether ESG commitments translate into a genuine, long-term innovative legacy.

5.2.3 Innovation Capability Development in Host Nations

The main drivers of innovation capability in nations hosting mega sporting events (MSEs) are multifaceted and encompass economic, social, and environmental dimensions. These drivers are critical for leveraging the potential benefits of hosting such events, including infrastructure development, national branding, and sustainability. One significant driver is the economic impact of MSEs, which often catalyzes infrastructure development and urban renewal. Hosting a major event, such as the Olympic Games or the FIFA World Cup, requires substantial investment in facilities and transportation systems, which can yield long-term economic benefits for the host nation. For instance, the 2022 FIFA World Cup in Qatar has been highlighted as a pivotal case for understanding the relationship between mega sporting events and infrastructure development, showing how such events can stimulate significant advancements in urban infrastructure [28]. Furthermore, the legacy of these events is often viewed as a strategic component of national development policies, particularly in emerging economies, where MSEs are seen as vehicles for economic growth and social progress [145].

Socially, the involvement of local communities and national sports organizations (NSOs) is crucial to enhancing innovation capability. NSOs are vital for orchestrating the successful delivery of sporting events and fostering a culture of collaboration and engagement among stakeholders [172]. Support from residents, driven by their perceived improvements in quality of life associated with MSEs, further amplifies this effect. Research indicates that when communities recognize the potential benefits of hosting such events, support from key stakeholders increases, leading to greater participation and investment in local sports initiatives [173, 174]. Environmental sustainability has emerged as a critical driver of innovation. The growing emphasis on "green" legacies has prompted organizing committees to integrate sustainability into their planning and execution.

For instance, greening initiatives observed in the Olympic Games have set benchmarks for environmental commitments that can influence future event planning [175]. This focus on sustainability addresses ecological concerns and enhances the host nation's global image, aligning with contemporary demands for responsible event management [176]. Additionally, the branding and marketing opportunities presented by MSEs are significant. Nations often utilize these events to enhance their global standing and attract tourism, resulting in substantial economic benefits for the host country. The strategic positioning of a country as a host can significantly transform its international image, as exemplified by South Africa's branding efforts during the 2010 FIFA World Cup [177]. This branding aspect is closely linked to the innovative strategies employed by host nations to maximize the visibility and impact of events, thereby fostering a competitive edge in the global sports market [178].

Ultimately, a nation's capacity for innovation when hosting mega sporting events depends on a complex blend of economic, social, environmental, and branding factors. By effectively harnessing these drivers, host nations can maximize the immediate benefits of these events

while laying the groundwork for sustainable development and long-term growth. While these macro-level innovation drivers paint a broad strategic picture, it is crucial to examine how direct financial mechanisms influence innovation performance within these events. The following section explores how various financial expenditures can catalyze or constrain innovation outcomes in a mega sporting event environment.

5.2.4 The Influence of Financial Factors on Innovation Capability (IC)

Financial factors play a crucial role in enhancing innovation capability (IC), particularly in mega sporting events (MSEs). The relationship between financial support and innovation is multifaceted, involving direct investment, strategic planning, and resource leveraging to foster an environment conducive to innovation.

One of the primary ways financial factors contribute to innovation performance is through funding allocated by governments and sports administrations to sports federations. Financial backing is essential for developing sports activities, including elite-level sports. Research indicates that sports federations competing for limited resources tend to be more innovative, as they are compelled to explore new avenues for resource generation within their competitive landscapes [179]. This competitive pressure can stimulate creative solutions and innovative practices that may not emerge in a more stable funding environment. Moreover, the financial implications of hosting MSEs extend beyond immediate expenditures. The significant public funds invested in the infrastructure necessary for these events often require demonstrating tangible benefits that exceed the costs. This pressure can drive event organizers and public authorities to innovate their approaches to event management and service delivery, ensuring that the benefits of hosting outweigh the financial burdens [180].

This dynamic is particularly evident in the MSE, where the expectation of a positive return on investment can drive innovative planning and execution. In addition to direct funding, the strategic management of financial resources is critical for maximizing innovation outcomes. This includes aligning venue designs with financial returns and ensuring that the developed infrastructure can serve multiple purposes beyond the event [181].

Failure to do so can result in financial losses and the underutilization of facilities, underscoring the importance of innovative financial planning for MSEs [182]. Furthermore, the interplay between financial investment and innovation is reflected in the broader economic impact. The influx of tourists and associated spending during these events can provide a substantial economic boost to host cities, which can then be reinvested in local infrastructure and community development. Essentially, financial expenditures are a core driver of innovation capability when a nation hosts mega-sporting events. Whether through direct funding, strategic financial management, or the broader economic ripple effects of hosting these resources, it can dramatically boost the innovation capability of all stakeholders.

This study examines a dataset from 1985 to 2022, focusing on six host countries that have hosted two significant global athletic events. This study examines two FIFA World Cups, South Africa (2010) and Brazil (2014), and four Olympic Games: the 2008 Beijing Summer Olympics, the 2010 Vancouver Winter Olympics and Paralympics, the 2012 London Summer Olympics, and the 2014 Sochi Winter Olympics. These events were selected for their availability of updated data, frequent mention in previous studies, and high levels of global fan engagement. Patent data for this study were obtained from the World Intellectual Property Organization database. The search strategy focused on all innovation areas. MSE leveraged data, including environmental, social, and governance (ESG) indicators, obtained from the World Bank database.

The subsequent sections detail the data processing and analytical procedures, organized into two subsections. Section 5.3 investigates the influence of ESG indicators on economic growth and innovation capability using traditional statistical models, including fixed-effects and negative binomial regressions. Section 5.4 extends the analysis by integrating ESG and financial indicators into machine learning models, specifically bootstrap forest, SHAP, PDP, and ICE techniques, to uncover nonlinear relationships and interaction effects in explaining innovation capability.

5.3 Data Summary

This chapter examines the relationship between mega-sport events (MSEs), economic growth, and innovation capability (IC) by incorporating environmental, social, and governance (ESG) indicators and MSE-related financial factors as key predictors. This study employs a longitudinal dataset covering 1985–2020 and comprising 6,732 observations. The MSE-focused subsample shall consist of six host countries that have organized major global sporting events, including the FIFA World Cup and the Olympic Games.

Table 5.1 Sample of Mega-Sport Events Included in the Analysis

Host country	Event name	Event's Year
China	Summer Olympic Game	2008
South Africa	FIFA World Cups	2010
Canada	Winter Olympic and Paralympic Games	2010
United Kingdom	Olympic and Paralympic Winter Games	2012
Brazil	FIFA World Cups	2014
Russia	Winter Olympic Game	2014

Table 5.2 Data Summary for Economic Growth and IC Prediction by ESG Factors

Group of Indicators	Number of Indicators	Description
Environmental dimension	27	Examples are the Food Production Index (2014–2016 = 100), Agriculture, forestry, and fishing value added (% of GDP), Methane emissions per capita, Heating Degree Days, Heat Index 35, and Forest area (% of land area).
Social dimension	20	Examples are the Unemployment rate, Total population, the Economic and Social Rights Performance Score, Fertility rate, the Gini index, and Hospital beds per 1,000 people.
Governance dimension	6	These comprise Government Effectiveness, Regulatory Quality, Voice and Accountability, Political Stability and Absence of Violence/Terrorism, Rule of Law, and Government expenditure on education (% of GDP).
Economic Growth	1	Target Variable 1; GDP growth (annual %)
Innovation capability (IC)	1	Target Variable 2; Number of Patent Applications

Table 5.3 Data Summary for IC Prediction by ESG and Financial Factors

Variable		Description
Financial factors	MS1	Capital expenses: stadiums and precinct development, non-competition infrastructure, civic facilities, neighborhood energy utility, transport, and ports of entry infrastructure.
	MS2	Operational expenses: broadcast and telecommunications, event volunteer training, immigration support, communications, hosting, legacy, and culture.
	MS3	Other expenses: legacy program, other mass mobilization campaigns, government activities, and union commission fees.
	MS4	Total revenue: direct, indirect, and induced economic impact; annual jobs created/sustained; revenue generated by tourists
ESG Indicators	ESG1	Adjusted savings: natural resources depletion (% of GNI)
	ESG2	Agriculture, forestry, and fishing, value added (% of GDP)
	ESG3	CO2 emissions (metric tons per capita)
	ESG4	Control of Corruption: Estimate
	ESG5	Economic and Social Rights Performance Score
	ESG6	Energy imports, net (% of energy use)
	ESG7	Energy intensity level of primary energy (MJ/\$2017 PPP GDP)
	ESG8	Food production index (2014-2016 = 100)
	ESG9	Forest area (% of land area)
	ESG10	Fossil fuel energy consumption (% of total)
	ESG11	Gini index
	ESG12	Hospital beds (per 1,000 people)
	ESG13	Net migration
	ESG14	Population density (people per sq. km of land area)
	ESG15	Prevalence of overweight (% of adults)
	ESG16	Ratio of female to male labor force participation rate (%)
	ESG17	Renewable energy consumption (% of total final energy consumption)
	ESG18	Standardized Precipitation-Evapotranspiration Index
	ESG19	Unemployment, total (% of total labor force) (modelled ILO estimate)
Macroeconomic Variables	Control1	Total GDP
	Control2	Population
	Control3	Avg stock price
	Control4	Exchange rate
Innovation Capability (IC)	Patent	Number of Patent Applications (Target Variable)

Table 5.2 presents the data summary for the analysis of economic growth and innovation capability using ESG indicators. Two dependent variables are examined: annual GDP growth, a measure of economic expansion, and the number of patent applications, a proxy for innovation capability. The explanatory variables were structured according to the World Bank's ESG framework. Before model estimation, all predictor variables were normalized using a min-max transformation to a 0–100 scale to ensure comparability across different measurement units. Missing observations in the time series were addressed using multivariate normal imputation, which preserves the data's underlying covariance structure while minimizing information loss.

Table 5.3 presents the data summary for the IC prediction that incorporates both ESG indicators and MSE-related financial factors. In this stage, innovation capability was modelled using ESG indicators and event-related financial variables, with macroeconomic indicators included as control variables. To improve model parsimony, a preliminary Pearson correlation

analysis was conducted, reducing the number of ESG indicators from 53 (reported in Table 5.2) to 19 for the final analysis (Table 5.3).

Financial data related to MSEs were compiled from official budgetary records and reports issued by event organizers, including the International Olympic Committee (IOC) and Fédération Internationale de Football Association (FIFA), as well as from government archives and open-access sources. These data capture public expenditure, research and development investment, infrastructure development, and funding allocated to innovation programs. For analytical clarity, the financial variables were categorized into capital expenditure, operating expenditure, other spending, and total revenue. ESG indicators and macroeconomic variables were obtained from the World Bank database, and patent application data were sourced from the World Intellectual Property Organization (WIPO).

5.4 Analytical Result

5.4.1 Comparative Analysis of ESG and Patent Relationships in MSE Host and Non-Host Countries

This section examines the relationship between ESG indicators and IC, as measured by patent activity, across MSE-host and non-host countries. Using descriptive statistics, the analysis highlights the systematic differences in the association between ESG dimensions and innovation across the two groups. This comparison provides an initial empirical overview of whether MSE hosting is associated with more substantial alignment between ESG performance and innovation capability.

Table 5.4 Descriptive Statistics of ESG and Patent Relationships in MSE Host and Non-Host Countries

Predictor	Non-Host Countries			Host-Countries		
	Estimate	Std Error	p-value	Estimate	Std Error	p-value
ESG1	-33.1577	223.7456	0.8822	-19244.9	9760.988	0.0501
ESG2	238.5077	134.6887	0.0767	6058.147	5221.605	0.2474
ESG3	774.7439	253.3967	0.0023	45584.47	7836.246	<.0001*
ESG4	6047.004	1400.219	<.0001*	-165993	32335.98	<.0001*
ESG5	-13547.3	7178.012	0.0592	605624.1	108053.2	<.0001*
ESG6	12.91531	7.912805	0.1027	5801.713	654.5493	<.0001*
ESG7	-157.476	338.7227	0.642	-24942.6	10630.31	0.02
ESG8	74.48911	31.08305	0.0166	5511.678	965.2195	<.0001*
ESG9	433.165	44.10397	<.0001*	-4433.42	2152.818	0.0408
ESG10	149.9492	76.06761	0.0488	3723.031	3735.958	0.3203
ESG11	90.38182	150.4995	0.5482	13104.64	3702.229	0.0005
ESG12	684.6818	178.0441	0.0001	9853.126	2409.622	<.0001*
ESG13	6483.858	359.0929	<.0001*	34245.66	10208.33	0.001
ESG14	0.056735	0.003467	<.0001*	-0.3241	0.083771	0.0002
ESG15	-2.0312	0.396274	<.0001*	-1332.28	366.6483	0.0004
EGS16	-704.572	49.93845	<.0001*	13.54154	2200.159	0.9951
ESG17	203.0377	80.24654	0.0115	2456.705	4313.943	0.5697
ESG18	-777.013	629.3191	0.217	15193.79	10456.78	0.1479
ESG19	-1190.39	150.34	<.0001*	-1218.33	5145.748	0.8131

Table 5.4 compares the relationships between ESG indicators and patent activity across MSE host and non-host countries, revealing clear differences in the composition, breadth, and functional roles of ESG determinants in shaping innovation outcomes.

In non-host countries, statistically significant relationships with patent activity are limited to a selective group of ESG indicators, including ESG3 (CO₂ emissions), ESG4 (control of corruption), ESG8 (food production index), ESG9 (forest area), ESG10 (fossil fuel consumption), ESG12 (hospital beds), ESG13 (net migration), ESG14 (population density), ESG15 (prevalence of overweight), ESG16 (female labor participation), and ESG19 (unemployment). These indicators reflect isolated environmental, demographic, and institutional conditions rather than a coherent sustainability framework. The dispersion of significance across heterogeneous ESG dimensions suggests that innovation in non-host countries is influenced by context-specific factors, with ESG effects operating through fragmented and loosely coordinated channels.

By contrast, MSE host countries exhibit a broader and more systematically structured set of significant ESG predictors, including ESG3 (CO₂ emissions), ESG4 (control of corruption), ESG5 (economic and social rights), ESG6 (energy imports), ESG7 (energy intensity), ESG8 (food production index), ESG9 (forest area), ESG11 (income inequality), ESG12 (hospital beds), ESG13 (net migration), ESG14 (population density), and ESG15 (prevalence of overweight), with ESG1 (natural resource depletion) approaching marginal significance. These indicators span all three ESG pillars, indicating that patent activity in host countries is shaped by integrated interactions between governance quality, social infrastructure, energy systems, and environmental constraints. The concentration of significance across multiple dimensions reflects a stronger alignment between sustainability-related conditions and national innovation systems.

Importantly, several ESG indicators that are insignificant in non-host countries, particularly ESG5, ESG6, ESG7, and ESG11, become statistically significant in host contexts. These variables capture institutional inclusiveness, energy dependency, production efficiency, and income distribution, suggesting that MSE hosting strengthens the structural channels through which ESG-related reforms and investments translate into innovation output. Large-scale event preparation and legacy planning appear to reinforce governance capacity, stimulate energy- and infrastructure-related innovation, and enhance social conditions conducive to patent generation.

The results indicate that patent activity in MSE host countries is more systematically and structurally conditioned by ESG performance than in non-host countries, where ESG effects appear more fragmented and driven by isolated contextual factors rather than integrated sustainability mechanisms. This finding supports the interpretation that mega-sport events contribute not only to short-term economic activity but also to longer-term, sustainability-oriented innovation dynamics, in which ESG factors are more deeply embedded within innovation production processes rather than acting as peripheral or isolated influences.

Figure 5.1 illustrates the contribution patterns of ESG indicators to patent activity in MSE host and non-host countries, based on decision-tree-based predictor screening. The results reveal a clear divergence in the influence of ESG dimensions on innovation across the two groups. For non-host countries, innovation outcomes are driven by a relatively limited number of ESG indicators, with contributions concentrated in specific environmental and governance factors. This pattern suggests a more fragmented relationship between sustainability performance and innovation, in which only selected ESG dimensions meaningfully shape patent activity.

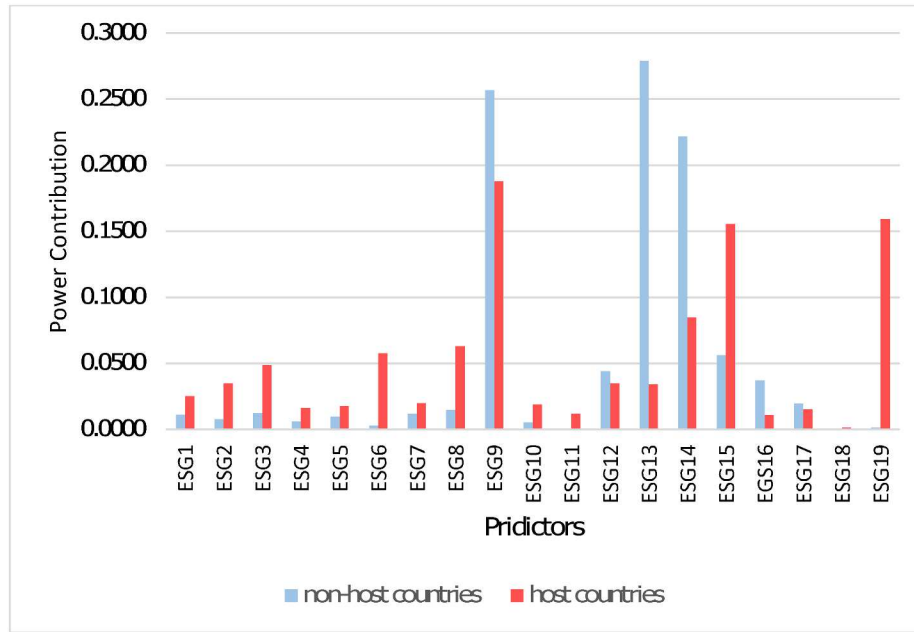


Figure 5.1 Predictor Screening of ESG on Patent Number of Host and Non-Host Countries

In contrast, MSE host countries exhibit broader and more balanced distributions of influential ESG predictors across multiple dimensions, indicating more integrated sustainability–innovation linkages. Innovation capability in these contexts appears to be shaped by the combined effects of environmental stewardship, social development, and governance quality rather than isolated factors. This diversified contribution structure suggests that MSE hosting reinforces institutional coordination and policy alignment, amplifying the influence of ESG on innovation outcomes. The predictor screening results thus reveal that ESG factors in host countries are more structurally embedded within innovation systems, extending beyond simple correlations to occupy constitutive roles in shaping innovation dynamics.

The findings in this section indicate that ESG factors play a more influential and integrated role in shaping innovation outcomes in MSE host countries. Based on this evidence, the next section focuses exclusively on host countries to examine in greater detail how ESG factors affect economic growth and innovation capability, thereby clarifying the internal mechanisms underlying MSE-related legacy effects.

5.4.2 Economic Growth and Innovation Capability (IC) Impacts by ESG Factors on Host Countries

This subsection examines the relationship between Environmental, Social, and Governance (ESG) factors and national economic performance and innovation capability (IC) in host countries. The analytical framework integrated two complementary methodological approaches. First, traditional statistical models, specifically fixed-effects and negative binomial regressions, are employed to test hypotheses regarding the causal impact of ESG factors on economic and innovation outcomes. Second, machine learning (ML) algorithms were applied to assess the predictive power of ESG factors in forecasting economic performance and innovation capabilities. This dual approach enables both causal inference and predictive validation, thereby strengthening the robustness of the empirical findings.

A: Fixed-effects Regression and Negative Binominal Models

The hypotheses for the regression model analysis are as follows.

H1 (Economic):

- H1a: Higher *Environmental* scores → higher GDP growth.
- H1b: Higher *Social* scores → higher GDP growth.
- H1c: Higher *Governance* scores → higher GDP growth.

Control: Population density

Method: Fixed-Effect / Standard Least Squares

The model for interpreting the additive effects of a one-point change in each ESG score on the expected GDP per capita change is as follows:

$$\Delta GDPpc_{it} = \alpha_i + \lambda_t - \beta_E ESG_{E,it} + \beta_S ESG_{S,it} + \beta_G ESG_{G,it} + \gamma PopDens_{it} + \varepsilon_{it} \quad (5.1)$$

PopDen = population density (people per km²) in country *i*, year *t*

H2 (Innovation):

- H2a: Higher *Environmental* scores → greater number of patents.
- H2b: Higher *Social* scores → greater number of patents.
- H2c: Higher *Governance* scores → greater number of patents.

Control: Population density

Method: Fixed-Effects / Negative Binomial

To quantify the distinct contributions of each ESG dimension to national innovation output, a fixed-effects negative binomial regression was estimated, regressing patent counts on three standardized ESG composites and population density while controlling for unobserved country- and year-specific effects. The resulting model allows us to interpret the multiplicative effects of a one-point change in each ESG score on expected patent filings via a log-link specification:

$$\ln(\mathbb{E}[PatentApps_{it}]) = \alpha_i + \lambda_t + \beta_E ESG_{E,it} + \beta_S ESG_{S,it} + \beta_G ESG_{G,it} + \gamma PopDens_{it} \quad (5.2)$$

H3 (Mediation): The positive effect of ESG on GDP is mediated by innovation (ESG→Patent→GDP).

Hypothesis Testing Results:

H1: ESG and GDP Growth

To test H1a–c, a fixed-effects ordinary least squares model was fitted with annual GDP per capita growth regressed on the three ESG composites (ESG_E, ESG_S, and ESG_G) and population density, including country and year dummy variables. As shown in Table 5.5, the model explains a substantial share of the within-country variation ($R^2 = 0.670$; adj. $R^2 = 0.587$). The overall regression is highly significant, F Ratio = 8.00, $p < .001$, indicating that the combined set of predictors reliably forecasts GDP growth.

Contrary to our hypotheses, none of the three ESG pillars had a statistically significant effect on GDP growth: the Environmental ($\beta = 0.260$, $p = 0.801$), Social ($\beta = -1.529$, $p = 0.197$), and Governance ($\beta = -1.037$, $p = 0.121$) composites all fell short of conventional

significance levels. Only population density emerged as a significant control ($\beta = -0.192$, $p = 0.011$), indicating that a higher density is associated with slightly lower GDP growth. Thus, H1a-c are not supported: after controlling for population density and country-year effects, ESG_E, ESG_S, and ESG_G do not significantly predict GDP per capita growth.

Figure 5.2 displays the close alignment between actual and model-predicted GDP growth, with the scatter of points clustering around the 45° line, yielding an RMSE of 10.09 and an R^2 of 0.670 ($p < .001$). This indicates that our specification captures most of the within-country variation in growth rates. Residual dispersion appears homoscedastic across the full range of growth rates, suggesting no obvious nonlinear bias.

Figure 5.3 displays the leverage plots for each ESG composite against the model's residuals from the fixed-effects OLS of the GDP per capita growth. In each panel, the red trend line is essentially flat and fails to achieve significance ($p = 0.8008$ for ESG_E; $p = 0.1970$ for ESG_S; $p = 0.1211$ for ESG_G), indicating no detectable nonlinear pattern or systematic bias in how each predictor relates to the outcome once all other terms are held constant. Moreover, there are no extreme leverage points or outliers that could unduly influence the model's estimates. Together, these diagnostics support the use of simple linear terms for the three ESG composite indices.

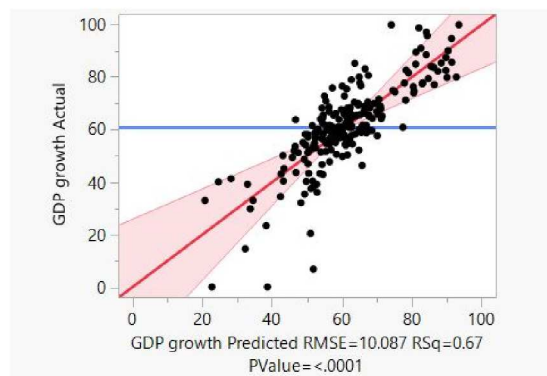


Figure 5.2 Actual vs. Predicted of GDP Growth

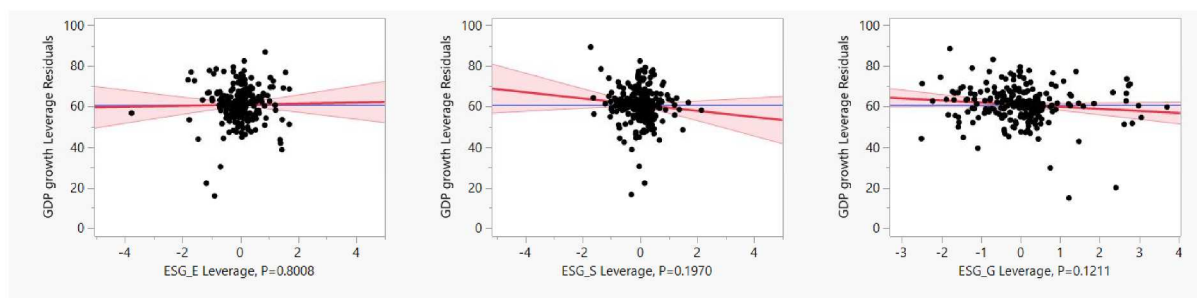


Figure 5.3 Leverage Plots of GDP Growth for Environmental, Social and Government

Table 5.5 Regression Results of ESG Dimensions on GDP Growth

Variable	Estimate	Std. Error	t-Ratio	p-Value	Hypothesis Results
ESG_E (Environment)	0.260	1.030	0.25	0.8008	H1a not supported
ESG_S (Social)	-1.529	1.181	-1.29	0.1970	H1b not supported
ESG_G (Government)	-1.037	0.666	-1.56	0.1211	H1c not supported
Population density	-0.192	0.075	-2.56	0.0113*	control

While our hypothesis tests (inference) indicate that none of the ESG composites (β_E , β_S , β_G) exert a statistically significant effect on GDP per capita growth leading us to reject H1a–c, the model nonetheless achieves strong predictive performance ($R^2 = 0.670$; $RMSE \approx 10.09$), as evidenced by the tight clustering around the 45° line in the Actual-vs-Predicted plot. This apparent paradox reflects the fact that the country and year fixed effects (Log Worths of 24.1 and 7.2, respectively) and population density account for the bulk of the cross-national and temporal variation, leaving little incremental explanatory power for ESG metrics in an inferential sense; however, they deliver a robust overall forecast.

To assess the relationship between ESG dimensions and economic growth, a fixed-effects panel model was estimated, regressing annual GDP per capita growth on the Environmental, Social, and Governance composite scores, alongside population density, while accounting for time-invariant country characteristics and common-year shocks. The resulting specification allows us to isolate the within-country effect of each ESG pillar on growth, net of unobserved heterogeneity, and temporal trends.

$$\Delta GDPpc_{it} = \alpha_i + \lambda_t + 0.260ESG_{E,it} - 1.529ESG_{S,it} - 1.037ESG_{G,it} - 10.192PopDens_{it} + \varepsilon_{it} \quad (5.3)$$

where α_i and λ_t Capture country and year-specific intercepts, respectively, while the negative coefficients on ESG_E ($\beta_E=0.260$, $p=0.801$), ESG_S ($\beta_S=-1.529$, $p=0.197$), and ESG_G ($\beta_G=-1.037$, $p=0.121$) all fall short of statistical significance. In contrast, population density exhibited a modest but statistically significant inverse relationship with growth.

Notwithstanding the absence of statistical support for H1a–c, our fixed-effects model exhibits considerable predictive strength, accounting for 67.0% of the within-country variation in GDP per capita growth ($R^2 = 0.670$) and achieving an in-sample root mean square error (RMSE) of 10.09. This outcome suggests that even if the ESG composites do not register as direct significant drivers once unobserved country- and year-specific heterogeneity are adequately controlled, the integrated components of fixed effects and population density collectively yield highly reliable forecasts of economic growth. Consequently, a model of this calibre offers substantial practical utility by enabling the generation of accurate baseline projections for GDP growth trajectories across diverse national and temporal contexts, against which the marginal effects of targeted policies or institutional reforms can be effectively evaluated.

H2: ESG and Patent Activity

To ensure that the appropriate count model was selected, overdispersion was first tested using a Poisson specification. The Pearson χ^2 statistic was 368.33 on 181 degrees of freedom ($\chi^2/df \approx 2.04$), well above the ideal value of 1, indicating a significant extra-Poisson variation. Consequently, a negative binomial model was adopted, with the estimated dispersion parameter ($\hat{\alpha} = 0.784$, $p < 0.0001$) confirming that this specification appropriately accommodates the observed overdispersion.

A fixed-effects negative binomial model of annual patent counts was estimated using the three ESG composites and population density while controlling for country and year dummies to account for unobserved heterogeneity. As shown in Table 5.6, the Social ESG score is positively and significantly associated with patenting, indicating a 43.6 percent increase in expected filings per 1-point rise in the social composite.

Table 5.6 Regression Results of ESG Dimensions on Innovation Capability

Variable	Estimate	IRR (Exp (Est.))	p-value	Interpretation	Hypothesis Results
ESG_E (Environment)	-0.0696	0.933	0.532	No significant effect; ~7% lower patents per unit $\uparrow$ ESG_E, but not statistically significant	H2a not supported
ESG_S (Social)	0.3616	1.436	0.0046*	Significant positive effect; ~44% more patents per unit $\uparrow$ ESG_S	H2b supported
ESG_G (Government)	-0.2386	0.788	0.0009**	Significant negative effect; ~21% fewer patents per unit $\uparrow$ ESG_G	H2c contradicted
Pop Density	-0.1029	0.902	<.0001**	Significant negative effect; ~10% fewer patents per unit $\uparrow$ density	control

*Notes: IRR = Incidence Rate Ratio. * $p < 0.01$; ** $p < 0.001$.

In contrast, environmental performance does not significantly affect patent counts, whereas governance performance exhibits a significant adverse effect. Population density had a small but highly significant negative association. These findings suggest that national social capabilities, rather than environmental or governance factors, are the strongest drivers of innovation output as measured by patent applications.

Below is the exponentiated specification of our negative binomial model, where each coefficient is presented as an incidence rate ratio (IRR). These IRRs reflect the multiplicative change in expected patent counts associated with a one-point increase in each ESG composite or control variable after accounting for country and year fixed effects:

$$\mathbb{E}[PatentApps_{it}] = \exp(\alpha_i + \lambda_t - 0.0696ESG_{E,it} + 0.3616ESG_{S,it} - 0.2386ESG_{G,it} - 0.1029PopDens_{it}) \quad (5.4)$$

where α_i and λ_t denote country and year fixed effects, respectively. Under this specification, each one-point rise in the social ESG composite (ESG_S) multiplies the expected patent count by $e^{0.3616} \approx 1.436$ (a 46 percent increase), while equivalent increases in the environmental (ESG_E) and governance (ESG_G) composites yield non-significant and adverse multiplicative effects of $e^{-0.0696} \approx 0.933$ (-7%) and $e^{-0.2386} \approx 0.788$ (-22%), respectively. Population density also exerted a modest but significant adverse effect ($e^{-0.1029} \approx 0.902$, 0.8% decrease per unit). By modelling patent counts on the log-link scale, these results succinctly capture the relative impact of each ESG dimension and control while accounting for overdispersion via the negative binomial likelihood.

Table 5.7 Hypothesis Summary for ESG on Economic Growth and Innovation Capability

Hypothesis	Variable	Target	Result
H1a	ESG_E (Environment)	GDP growth	not supported
H1b	ESG_S (Social)	GDP growth	not supported
H1c	ESG_G (Government)	GDP growth	not supported
H2a	ESG_E (Environment)	Patent Application	not supported
H2b	ESG_S (Social)	Patent Application	supported
H2c	ESG_G (Government)	Patent Application	contradicted

*All models include country and year fixed effects (not shown).

Table 5.8 Mediation Testing Statistic Result

Variable	Estimate	Std. Error	t-Ratio	p-Value
Intercept	81.8978	9.7509	8.40	<.0001*
ESG_E	0.5448	1.0267	0.53	0.6068
ESG_S	-0.2973	1.2083	-0.25	0.8041
ESG_G	-1.6199	0.6678	-2.45	0.0154*
Patent Application	-0.09017	0.0551	-1.64	0.1039
Population density	-0.7436	0.3656	-2.03	0.0435*

H3: Mediation via Innovation

In the third step of the mediation analysis, the fixed-effects OLS model of annual GDP per capita growth was re-estimated, including the scaled patent count, the three ESG composites, and population density, while again controlling for country and year dummies. This expanded specification yielded an improved fit ($R^2 = 0.687$, adj. $R^2 = 0.601$, RMSE = 9.56, $F = 7.997$, $p < .001$). Population density remains a significant negative predictor ($\beta = -0.744$, $p = 0.0435$), and the governance composite now exhibits a modest but statistically significant inverse association with growth ($\beta = -1.620$, $p = 0.0154$). In contrast, the scaled patent mediator was not significant ($\beta = -0.090$, $p = 0.1039$), nor were the environmental ($\beta = +0.545$, $p = 0.6068$) or social ($\beta = -0.297$, $p = 0.8041$) ESG dimensions.

Estimate Equation:

$$\Delta GDPpc_{it} = \alpha_i + \lambda_t + 0.5448ESG_{E,it} - 0.2973ESG_{S,it} - 1.6199ESG_{G,it} - 0.09017Patent_{it} - 0.7436PopDens_{it} \quad (5.5)$$

Because the mediator coefficient is not significant, it is concluded that innovation output, measured by patent applications, does not mediate the relationship between ESG performance and GDP growth in the panel.

The results of this study are threefold. For H1, no statistical support was found for the direct influence of any ESG pillar (environmental, social, or governance) on GDP growth, leading to the rejection of H1a–c. Regarding H2, only the social dimension of ESG demonstrated a reliable predictive relationship with patenting, thereby supporting H2b, whereas environmental and governance scores did not yield significant associations (rejecting H2a and H2c). Finally, the mediation test for H3 confirmed that patent activity does not transmit any ESG effect on growth, providing no support for an ESG → Patent → GDP pathway.

B: Bootstrap Forest Model

GDP Growth Prediction

This section reports the results of the bootstrap forest model used to examine the influence of ESG factors on GDP growth. The model's R-squared (R^2) values indicate the predictive power of each group in explaining variations in GDP, while the root mean squared error values reflect the accuracy of the predictions [98].

The results indicate that both the environmental (E) and social (S) groups exhibit strong predictive power, with R-squared values of 0.836 and 0.832, respectively, and relatively low RASE values of 6.348 and 6.420. In contrast, the governance (G) group shows a lower R-squared value of 0.683 and a higher RASE of 8.807, indicating weaker predictive performance.

While the fixed-effects regression did not identify statistically significant linear effects of ESG factors on GDP growth, the bootstrap forest results suggest that these variables still have predictive value when nonlinear patterns and interactions are considered. This difference reflects the methodological distinction: fixed-effects regression tests for average linear associations, whereas bootstrap forests capture more complex relationships that may not manifest as simple coefficient significance. These findings indicate that although ESG may not exert uniform linear effects, it nonetheless contributes meaningfully to GDP growth in nonlinear or context-specific ways.

Table 5.9 Bootstrap Forest Prediction Performance for GDP Growth

Predictors to Target	N	R-Square	RASE
ESG_E to GDP growth	228	0.836	6.348
ESG_S to GDP growth	228	0.832	6.420
ESG_G to GDP growth	288	0.683	8.807

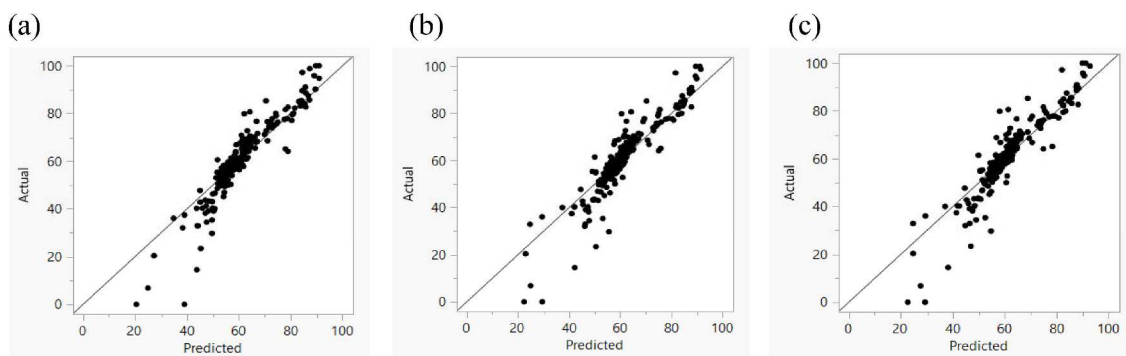


Figure 5.4 Actual vs. Predicted Plots for GDP Growth (a) Environmental Factors, (b) Social Factors, And (c) Governance Factors

Actual Predicted Plot of GDP growth Prediction

This subsection illustrates the predictive performance of the bootstrap forest models through actual-versus-predicted scatterplots for each ESG dimension. By mapping observed GDP growth rates against model-based estimates, these visuals check the model fit and reveal potential biases or heteroscedastic patterns. A side-by-side comparison of the environmental, social, and governance models offers a graphical assessment of their explanatory power, setting the stage for a detailed discussion.

Figure 5.4 displays actual versus predicted plots for GDP growth using environmental (E), social (S), and governance (G) factors from the bootstrap forest model. Figure 5.4(a) shows points tightly clustered along the 45-degree line, indicating a strong agreement between the predicted and actual values, which confirms the high predictive accuracy in Table 5.9, with the highest R-squared (0.836) and the lowest RASE (6.348). Figure 5.4(b) shows strong alignment but more dispersion, corresponding to a lower R-squared value (0.832) and a higher RASE (6.420), suggesting greater variability in the predictive relationships of social indicators.

The 5.4(c) plot shows a wider data spread around the reference line, particularly at the extremes, indicating higher prediction errors. This aligns with the statistical results in Table 5.9, where governance factors have the lowest R-squared value (0.683) and the highest RASE (8.807). The dispersion suggests that governance indicators may not fully capture GDP growth dynamics due to complex effects not represented in the model. These plots confirm that environmental and social variables are stronger predictors of GDP growth than governance variables in this dataset.

The contribution of ESG factors influences GDP Growth

This subsection examines the individual contributions of environmental, social, and governance (ESG) indicators to GDP growth, as estimated by the Bootstrap Forest model. The relative importance of each variable reflects its influence on the predictive output, indicating which ESG dimensions have the most significant impact on economic performance. The top five contributing indicators within each ESG category are presented, highlighting the most influential variables.

As shown in Table 5.10, among the environmental factors, agriculture, forestry, and fishing value added (% of GDP) contributed the most (0.1946), followed by the Heat Index (0.1206) and methane emissions per capita (0.1060). These top variables highlight the links among the environmental dimension, economic productivity, and climate conditions. For Social factors, the most significant contributors were the prevalence of overweight individuals (0.2336), net migration (0.1882), and hospital beds (0.1390), suggesting strong ties between health, demographic changes, and GDP. Under the Governance dimension, Voice and Accountability held the highest proportion (0.3474), followed by government expenditure on education (0.1789) and the Rule of Law (0.1503), indicating that political stability, civic participation, and institutional quality are pivotal governance-related drivers of economic growth.

These results demonstrate that the key contributors to GDP growth span all ESG dimensions, with notable differences. In the environmental category, agricultural output reinforces the relevance of primary sector productivity, particularly in emerging economies. Social indicators, such as the prevalence of overweight individuals and migration, reflect how population health and mobility influence productivity. In governance, Voice and Accountability highlight the importance of democratic institutions in promoting growth. These findings suggest that although all ESG dimensions contribute to GDP growth, their internal drivers differ, offering distinct policy avenues.

Table 5.10 The Contribution of ESG Factors on GDP Growth Statistic Result

Factors	ESG Predictors	Portion
Environmental Factors	Agriculture, forestry, and fishing, value added (% of GDP)	0.1946
	Heat Index	0.1206
	Methane emissions (metric tons of CO2 equivalent per capita)	0.1060
	Land Surface Temperature	0.0575
	Fossil fuel energy consumption (% of total)	0.0550
Social Factors	Prevalence of overweight	0.2336
	Net migration	0.1882
	Hospital beds	0.1390
	Income share held by the lowest 20%	0.0633
	People using safely managed drinking water services	0.0479
Government Factors	Voice and Accountability: Estimate	0.3474
	Government expenditure on education: total	0.1789
	Rule of Law: Estimate	0.1503
	Regulatory Quality: Estimate	0.1486
	Government Effectiveness: Estimate	0.0825

Table 5.11 Bootstrap Forest Prediction Performance for Patent Application

Predictors to Target	N	R-Square	RASE
ESG_E to Patent Application	215	0.785	7.725
ESG_S to Patent Application	215	0.911	4.965
ESG_G to Patent Application	215	0.934	4.282

Patent Application Prediction

This subsection examines how ESG dimensions predict innovation capability, as measured by annual patent applications using the Bootstrap Forest model. Patents serve as a measure of a country's inventive activities. By separately assessing environmental (E), social (S), and governance (G) variables, the analysis identifies which pillar most strongly drives innovation performance in the sample.

Table 5.11 summarizes the model statistics for 215 observations across all three ESG groups. Governance factors yield the highest R-squared (0.934) and the lowest root average squared error (RASE = 4.282), indicating superior predictive accuracy for patent counts. Social factors exhibited a strong correlation with an R-squared value of 0.911 and an RASE of 4.965, whereas environmental factors displayed a more modest correlation with an R-squared value of 0.785 and a higher RASE of 7.725. The uniform sample size ($N = 215$) confirmed that model comparisons were not confounded by differing data coverage.

The results indicate that institutional quality and regulatory effectiveness in the governance pillar are the key drivers of patenting activity. A supportive legal environment and effective governance reduce the transaction costs associated with R&D investments, thereby promoting innovation. Social variables reflecting human capital and demographics also strongly influence innovation, highlighting the importance of an educated workforce. Environmental indicators appear to be less directly linked to short-term patent output, as their effects may unfold over more extended periods. These findings emphasize the importance of governance reform and social development in fostering technological progress.

Whereas the negative binomial regression highlighted only the Social ESG dimension as statistically significant, the bootstrap forest analysis demonstrated that governance, and to a lesser extent environmental factors, also exhibited substantive predictive power in explaining innovation capability. This difference arises because regression isolates average linear effects, whereas the bootstrap forest captures nonlinear patterns and complex interactions. Thus, the two approaches should be considered complementary: regression highlights the most robust linear relationships, whereas the machine learning model reveals additional predictors that enhance forecasting accuracy.

Actual Predicted plots for Patent Application Prediction

This subsection visually assesses the performance of the Bootstrap Forest model in predicting patent applications using ESG factors. Actual versus predicted plots evaluate the model fit and illustrate the predictive strength of each ESG dimension in capturing innovation patterns. Figure 5.5 presents the actual versus predicted plots for patent applications using environmental (a), social (b), and governance (c) factors as predictors estimated using the Bootstrap Forest model. Figure 5.5(a), which represents environmental factors, shows a notable spread of points away from the 45-degree reference line, particularly in the mid-to high-prediction ranges.

This dispersion indicates a relatively weaker model performance, consistent with the lower R-squared (0.785) and higher RASE (7.725). Figure 5.5(b), based on social factors, shows improved clustering along the diagonal, particularly at lower patent levels; however, some deviations remain at higher levels. This pattern aligns with a higher R-squared (0.911) and a lower RASE (4.965), demonstrating that social indicators offer stronger predictive power for patent output. Figure 5.5(c), which models patent applications using governance factors, shows the best fit, with data points tightly aligned along the 45-degree line, indicating strong

agreement between actual and predicted patent counts. This aligns with the highest R-squared (0.934) and the lowest RASE (4.282). The visual evidence confirms that governance indicators are the strongest predictors of innovation performance, followed by social factors. Environmental variables exhibit greater dispersion and lower accuracy, suggesting an indirect influence on innovation. These results demonstrate that effective institutions and a strong social infrastructure enable patenting activities.

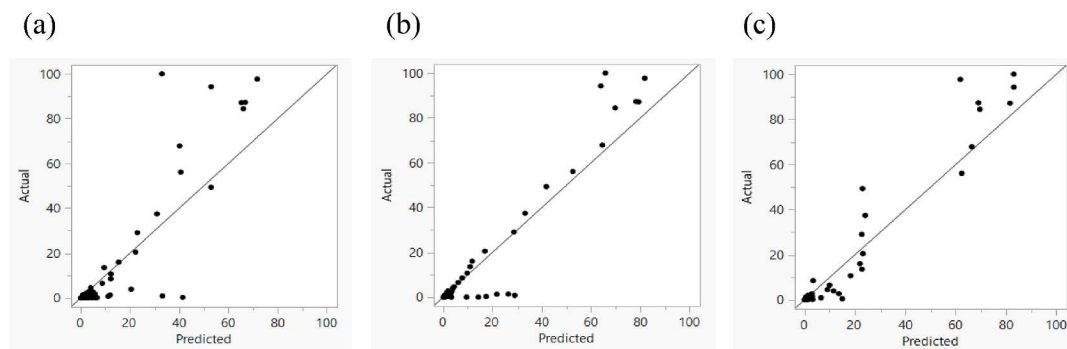


Figure 5.5 Actual vs. Predicted Plots for Patent Application (a) Environmental Factors, (b) Social Factors, and (c) Governance Factors

Table 5.12 The Contribution of ESG Factors on Patent Application Statistic Result

	ESG Predictors	Portion
Environmental factors	Heat Index 35	0.5194
	PM2.5 air pollution, mean annual exposure	0.0763
	Adjusted savings: natural resources depletion (% of GNI)	0.0481
	Food production index (2014-2016 =100)	0.0397
	Level of water stress: freshwater withdrawal as a proportion of available freshwater resource	0.0299
Sical factors	Scientific and technical journal articles	0.5841
	Unemployment, total	0.0623
	Population density	0.0492
	Prevalence of undernourishment	0.0489
	Fertility rate, total	0.0381
Government factors	Voice and Accountability: Estimate	0.5697
	Rule of Law: Estimate	0.2044
	Government Effectiveness: Estimate	0.1683
	Regulatory Quality: Estimate	0.0390
	Government expenditure on education: total	0.0139

The contribution of ESG factors to Patent Applications

This subsection identifies the most influential ESG indicators for patent application outcomes using variable importance scores from the bootstrap forest model. To better understand the drivers of innovation, the top five contributing factors from each of the ESG categories (environmental, social, and governance) are presented and analyzed. This analysis offers more profound insights into which indicators within each pillar are most closely linked to national-level inventive activities.

As shown in Table 5.12, among the environmental factors, heat index 35 has the highest contribution, followed by PM2.5 air pollution, natural resource depletion (% of GNI), food production index, and water stress. In the social dimension, scientific and technical journal articles dominate, far and away the most impactful variable in the table, followed by Unemployment, Population density, prevalence of undernourishment, and fertility rate. As for governance factors, voice and accountability leads the highest score, followed by rule of law, government effectiveness, regulatory quality, and government expenditure on education.

The results reveal distinct patterns of influence across the ESG pillars. In the environmental category, Heat Index 35 far outweighs the contributions of other variables, suggesting a strong link between extreme heat and innovation, possibly driven by climate-related adaptation technologies or policy responses that stimulate inventive activity. In the social domain, the dominant role of scientific and technical journal articles underscores the critical importance of research output and knowledge infrastructure in driving patent applications, reaffirming the role of academia and R&D capacity in the national innovation systems. The governance dimension is similarly led by Voice and Accountability, indicating that institutional openness, civil liberties, and participatory governance are essential for fostering innovative environments. These findings suggest that innovation is influenced not only by economic and technological inputs but also by sociopolitical and environmental conditions, especially those that support freedom, transparency, and scientific exchange.

Taken together, the regression and machine learning results offer complementary insights into the role of ESG factors in driving economic growth and innovation. The fixed-effects and negative binomial regressions emphasize linear average effects, showing that only the social dimension consistently and significantly supports patenting activity. In contrast, the bootstrap forest models emphasize the predictive value of Social, Environmental, and Governance factors, particularly when nonlinearities, thresholds, and interactions are considered. This divergence underscores a methodological trade-off: regression models provide statistical inferences on direct relationships, whereas machine learning reveals more complex dynamics that enhance predictive accuracy. Synthesizing these results suggests that social development is the most robust linear driver of innovation. In contrast, governance and environmental performance play critical roles in shaping long-term trajectories when broader nonlinear effects are considered.

5.4.3 ESG indicators and MSE-Related Financial Factor for IC Prediction

This subsection analyzes the prediction of innovation capability (IC) using ESG indicators and MSE-related financial factors as key predictors, with macroeconomic indicators as control variables. A machine learning (ML) framework was employed, incorporating decision-tree-based predictor screening, actual-by-predicted plots, and explainable AI (XAI) techniques, including shapley additive explanations (SHAP), partial dependence plots (PDP), and individual conditional expectation (ICE) analyses. Predictor screening provides an initial assessment of variable importance and supports feature selection, whereas XAI methods offer deeper interpretability by revealing individual prediction behavior and interaction effects.

Predictor Screening Result for Patent Application Prediction

The predictor screening identifies and ranks the variables that contribute most to predicting patent applications, which serve as proxies for IC. Using a decision tree model, the contribution and proportion of each predictor were assessed to determine their relative importance in explaining the variance in the target variable. As shown in Table 5.13, the results demonstrate that population, operational, and other expenses are the most critical drivers of IC, as measured by patent applications.

This aligns with the existing literature that emphasizes the role of population scale and economic investment in fostering innovation. The contributions of variables such as renewable energy consumption reflect the increasing importance of environmental factors in patent prediction. Interestingly, while total GDP, economic and social rights performance, and Gini Index performance make smaller contributions, their inclusion suggests that socioeconomic inclusivity supports innovation. This aligns with frameworks that advocate equitable and sustainable growth to support long-term innovation.

Table 5.13 Predictor Screening Statistic Result for IC Prediction

Predictor	Contribution	Portion	Rank
Population	26.3274	0.3332	1
Operational Expenses	17.2115	0.2178	2
Other Expenses	8.6319	0.1093	3
Renewable energy consumption	5.4872	0.0695	4
Total GDP	5.2200	0.0661	5
Economic and Social Rights Performance Score	3.8911	0.0492	6
Gini index	2.6511	0.0336	7
Fossil fuel energy consumption (% of total)	2.3298	0.0295	8
Population density	1.3768	0.0174	9
Unemployment	1.2370	0.0157	10

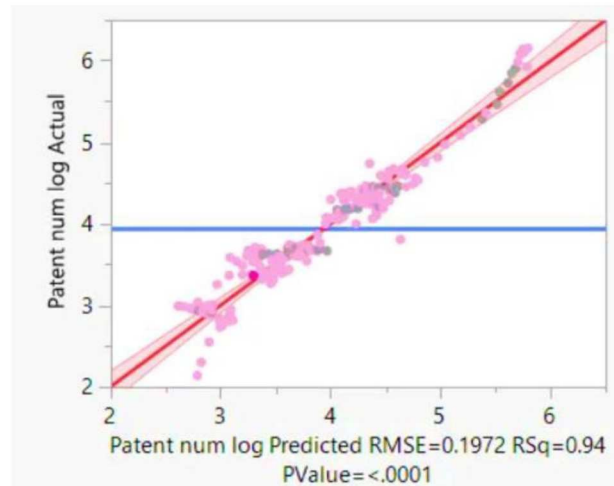


Figure 5.6 The Actual Predicted Plot of Patent Applications

Actual by Predicted Plot

The Actual-Predicted Plot visually shows the model's predictive performance in estimating the logarithm of patent application counts. In Figure 5.6, the plot shows strong alignment between the actual and predicted values, with most data points clustered around the diagonal reference line. An R^2 value of 0.94 indicates that the predictors explain 94% of the variance in patent applications (log-transformed), suggesting a highly accurate model for predicting patent applications. The root mean squared error (RMSE) of 0.1972 also signifies a low prediction error, reinforcing the model's reliability.

The p-value (<0.0001) confirmed the model's statistical significance. The Effect Summary highlights the key predictors contributing to the model. The Total GDP, Fossil Fuel Energy Consumption, Gender Labor Force Ratio, and Hospital Beds (per 1,000 people) exhibit the highest log-worth values, implying their strong influence on the number of patent applications. Predictors with lower log-worth values, such as the Control of Corruption and Gini Index, demonstrated minimal impact. Overall, the model suggests that both economic and environmental indicators play crucial roles in influencing innovation capability, as measured by patent applications.

A comparison of the predictor screening results with the regression model's effect summary revealed both similarities and differences in variable importance. In both analyses, population-related and financial factors of MSE hosting, such as operational and other expenses, were influential, aligning with the substantial impact of total GDP and Fossil Fuel Energy Consumption in the effect summary. However, this contrast suggests that different modelling approaches highlight distinct aspects of the relationship between MSE financial factors, ESG factors, and innovation capability, reinforcing the need for multi-method validation in complex analyses.

The SHAP Plot Result

The SHAP (Shapley additive explanation) plot provides a visual explanation of the contributions of different features to the model's prediction for a specific instance in the dataset. As shown in Figure 5.7, the SHAP plot illustrates how individual features contributed to a model prediction of 3.35, compared to the average prediction of 3.94. Features with positive SHAP values (extending to the right) increased the prediction, whereas those with negative SHAP values (extending to the left) decreased the prediction.

The analysis indicates that several ESG indicators contribute to innovation capability (IC). In particular, the Economic and Social Rights Performance Score (ESG5) and percentage of forested area (ESG9) showed positive effects, suggesting that social inclusivity and environmental sustainability foster innovation. Additional contributors include the prevalence of overweight (ESG19), agriculture, forestry, and fishing value added (ESG2), and adjusted savings from natural resource depletion (ESG1), underscoring the interplay between demographic, economic, and environmental conditions in shaping innovation capabilities. Moreover, Total Revenue (MS4) made a substantial positive contribution, reinforcing the view that broader economic multipliers can stimulate innovation. In contrast, operational expenses (MS2) and other expenses (MS3) exhibited substantial adverse effects, indicating that higher spending in these categories reduced the predicted number of patent applications.

These findings contrast with those from the perspective of predictor screening (Table 5.13). Predictor screening, based on a decision tree framework, evaluates the relative importance of each variable in explaining the variance in patent applications across the sample. From this perspective, operational and other expenses rank highly, reflecting their large share of MSE-related spending and their role in differentiating patenting outcomes across countries and over time. In contrast, the SHAP values focus on the marginal effect and direction of influence, showing that while these financial variables are highly influential, their predominant contribution is negative. This discrepancy can be better understood by considering the composition of these expenses. Operational budgets encompass various elements, including broadcasting and telecommunications, volunteer training, immigration support, hosting, and cultural activities.

Specific components, particularly those related to digital broadcasting and telecommunications, may foster R&D and technology transfer, whereas others provide limited innovation spillovers. Similarly, other expenses include legacy programs and some government activities that could support innovation, as well as mass mobilization campaigns and ceremonial spending that offer little to no long-term benefit. Therefore, the SHAP analysis indicates that although these categories contain innovation-enabling elements, the overall efficiency of their allocation is limited. In other words, disparities in the distribution of these expenditures across innovation-related and non-innovation-related activities reduce their net positive effect on innovation capability.

In contrast, Total Revenue (MS4) shows a consistent positive effect in the SHAP analysis, reflecting its role in capturing broader economic multipliers, such as induced employment and tourism-generated income, which may indirectly support innovation. These results suggest that not all financial flows associated with MSEs are equally beneficial for innovation; their impact depends on how closely expenditures or revenues are connected to R&D and technology-intensive activities.

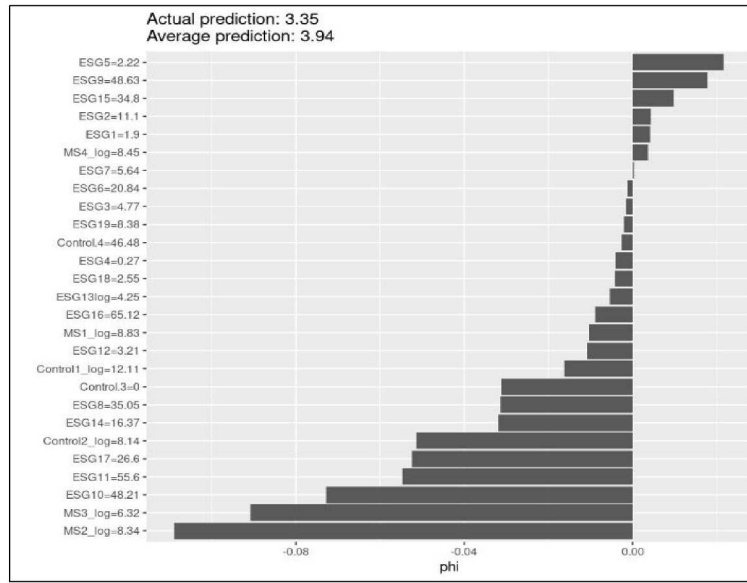


Figure 5.7 The Shapley Additive Explanations (SHAP) Plot

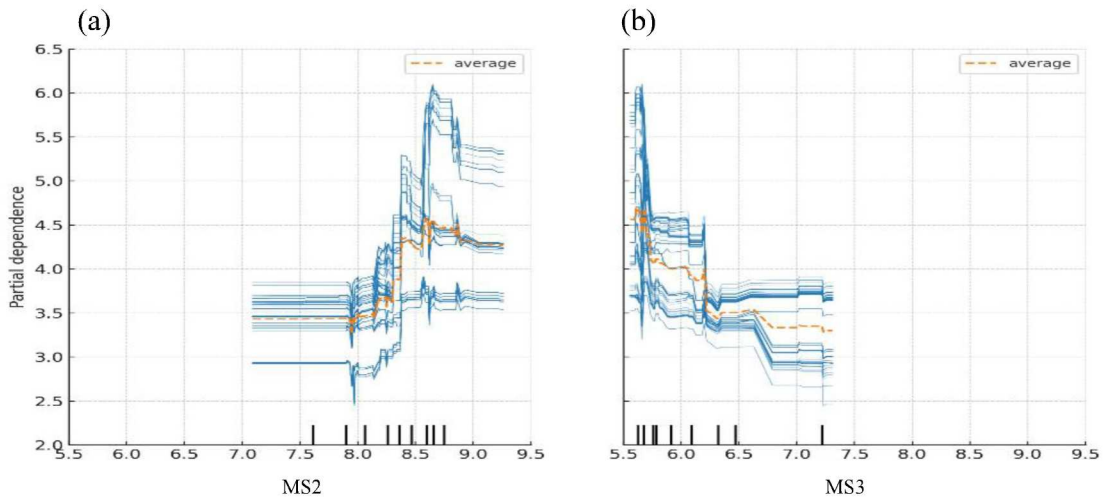


Figure 5.8 The PDP and ICE Plot of (a) The Operational Expense, (b) The Other Expenses

The PDP and ICE Result

This section presents a partial dependence plot (PDP) and individual conditional expectation (ICE) graph for the selected predictor and its impact on the predicted number of patent applications. As shown in Figure 5.8, the orange line represents the PDP, which shows the average marginal effect of the feature value on the predicted outcome for all instances.

The blue lines represent ICE curves, illustrating how individual observations respond to changes in features. These instances included operational expenses (MS2) and other instances (MS3). The analysis of operational expenses (MS2) revealed a nonlinear relationship with the predicted number of patents. The PDP and ICE plots of the operational expense, as shown in Figure 5.8(a), indicate a positive impact up to 8.5, with the predicted value peaking before fluctuating and eventually stabilizing. Notably, between 8.0 and 8.5, most Individual Conditional Expectation (ICE) curves align, indicating a consistent effect across instances. This suggests that increased operational expenses associated with hosting mega-sports events are generally linked to a higher number of patent applications or innovative outcomes. However, beyond a certain threshold, these effects become less predictable.

As shown in Figure 5.8(b), the analysis of other expenses associated with mega-sports events exhibits a negative relationship with the predicted number of patents. The most significant decline in predicted innovation outcomes is observed between 5.5 and 6.5, suggesting that increasing this expenditure category beyond this range tends to diminish innovation capability. Beyond 6.5, the predicted values stabilize, indicating that further increases in other expenses have negligible effects on the outcome. These findings suggest that excessive spending on other expenses (non-operational, non-capital, and event-related) may not contribute positively to innovation capability. This aligns with the existing literature, which emphasizes the need for efficient resource allocation in large-scale events, as unproductive expenditures could divert funds away from investments in technology, infrastructure, and human capital, which are critical drivers of innovation.

The SHAP analysis revealed that operational expenses (MS2) and other expenses (MS3) exert predominantly adverse marginal effects on innovation capability, suggesting that higher spending in these categories tends to reduce patenting outcomes. The PDP and ICE results provide further nuance by demonstrating that these relationships are nonlinear. For operational expenses, the effect is initially positive, with innovation outcomes increasing up to a threshold (approximately 8.0–8.5), after which the impact becomes unstable and less predictable. This pattern suggests that while moderate investment in event-related operations, such as broadcasting or telecommunications, may enhance innovation, excessive spending beyond a critical level yield diminishing returns. In contrast, the PDP/ICE analysis of other expenses confirms the SHAP finding of a negative relationship, with the steepest decline observed between 5.5 and 6.5. Beyond this point, the effect stabilizes, indicating that further increases in such expenditures yield little additional benefit to innovation.

Taken together, the SHAP and PDP/ICE results emphasize that not all financial flows associated with MSEs contribute equally to innovation capability. Operational expenses have some innovation-enhancing potential, but only within an optimal range, whereas other expenses predominantly constrain innovation when they exceed moderate levels. These findings align with the broader literature emphasizing the importance of efficient resource allocation: unproductive spending on ceremonial or administrative event functions may divert resources away from R&D, infrastructure, and human capital, the proper drivers of innovation. Therefore, maximizing the innovation legacy of mega-sport events requires careful budgeting that prioritizes expenditures with clear pathways for technological development and long-term knowledge spillovers.

5.5 Discussion

This chapter synthesizes the statistical and machine learning findings on the role of environmental, social, and governance (ESG) indicators and financial factors in shaping economic growth and innovation capability (IC). By comparing the results from traditional regression models with those derived from ensemble machine learning and explainable AI techniques, the discussion highlights both the methodological trade-offs and the substantive implications for leveraging mega-sport events (MSEs) to enhance long-term IC.

5.5.1 ESG Impacts on Economic Growth: Fixed Effects vs. Bootstrap Forest

The comparison between fixed-effects regression and bootstrap forest models highlights how methodological choices influence conclusions regarding the relationship between ESG and economic growth. In the fixed-effects framework, none of the environmental (ESG_E), social (ESG_S), or governance (ESG_G) dimensions showed statistically significant effects on GDP per capita after controlling for country- and year-level heterogeneity. This suggests that under linear assumptions, ESG indicators do not exert robust within-country effects on growth. In contrast, the bootstrap forest model revealed ESG_E and ESG_S as strong predictors of GDP growth, achieving high R^2 values (0.836 and 0.832, respectively) and low error rates.

This divergence reflects the assumptions embedded in each approach. Fixed-effects regression identifies average linear effects and excels at hypothesis testing under strict assumptions, but it risks overlooking nonlinear or conditional dynamics. Bootstrap forest, a nonparametric ensemble method, flexibly captures threshold effects, interactions, and nonlinearities. For example, environmental investments may contribute more strongly beyond a specific minimum scale, or social development may be more critical in densely populated economies. Such dynamics are obscured in linear models but are captured by the tree-based algorithms. Together, these results suggest that while ESG_E and ESG_S may not exert consistent linear effects, they nonetheless carry predictive value in shaping long-term economic trajectories when nonlinear mechanisms are considered.

5.5.2 ESG Impacts on Innovation Capability: Negative Binomial vs. Bootstrap Forest

A similar contrast emerges when patent applications are analyzed as a proxy for innovation capability. Negative binomial regression, which accounts for overdispersion in count data, identified only the Social ESG pillar (ESG_S) as statistically significant. This finding suggests that improvements in social conditions, including human capital, inclusivity, and demographic development, are consistently and linearly associated with higher levels of patent activity. However, environmental and governance indicators do not exhibit significant effects under this specification.

The bootstrap forest model, designed for predictive accuracy, identified a broader set of contributors. Both Governance (ESG_G) and social (ESG_S) demonstrated high predictive power ($R^2 = 0.934$ and 0.911 , respectively), whereas environmental factors showed weaker predictive power. This suggests that governance quality and institutional structures, although not having a robust linear effect, are influential when nonlinear dynamics are considered. In other words, regression highlights robust linear drivers (primarily social inclusivity), whereas machine learning reveals that governance and institutional quality enhance innovation predictions by capturing context-specific and nonlinear effects. Collectively, the two approaches suggest that, while social development remains the most consistent linear driver, governance plays a significant role in shaping long-term innovation outcomes when broader dynamics are taken into account.

5.5.3 ESG and Financial Factors for IC Prediction: Predictor Screening, SHAP, PDP, and ICE

Predictor Screening

The predictor screening analysis highlights the multifaceted nature of innovation capability, which is shaped by demographic, economic, social and environmental factors. Among these, the population size and operational expenses emerged as the most significant contributors. This aligns with the literature, which emphasizes that larger populations provide broader talent pools and greater demand for innovation. Moreover, investments in event operations can generate spillovers in knowledge sharing, infrastructure, and business activity. Sustainability-related factors, such as renewable energy consumption and fossil fuel dependence, also emerged as important, suggesting that nations that embed green energy policies into their MSE strategies tend to achieve stronger innovative outcomes. Economic stability indicators (e.g., GDP, income equality, and economic rights) played secondary yet meaningful roles, reinforcing the importance of macroeconomic foundations.

SHAP Results

SHAP analysis complements predictor screening by clarifying the direction of the effects. Positive contributions were observed from the economic and social rights performance score (ESG5) and the percentage of forested area (ESG9), demonstrating that social inclusivity and environmental stewardship foster innovation. Additional positive drivers included the prevalence of overweight (ESG19), agriculture, forestry, and fishing value added (ESG2), adjusted savings from natural resource depletion (ESG1), and Total Revenue (MS4). The positive effect of Total Revenue reflects the role of broader economic multipliers, such as induced employment and tourism income, which indirectly support innovation outcomes.

In contrast, operational expenses (MS2) and other expenses (MS3) exhibited substantial adverse marginal effects. This suggests that although these categories account for a significant portion of the variance in patent applications, higher spending in these areas is associated with a reduction in predicted innovation. The divergence from predictor screening reflects methodological distinctions: predictor screening emphasizes explaining variance, whereas SHAP specifies whether the marginal effect is positive or negative. Their composition explains the negative directions of MS2 and MS3. Operational budgets cover both innovation-enabling activities (e.g., broadcasting and telecommunications) and less innovation-intensive functions (e.g., volunteer training, hosting, and cultural activities). Similarly, other expenses include legacy programs with potential innovation benefits and ceremonial or symbolic spending with limited spillover effects. Thus, while these categories are influential overall, the net effect on innovation is adverse because of inefficiencies in allocation.

PDP and ICE Results

The PDP and ICE analyses add further nuance by illustrating the nonlinear thresholds. Operational expenses exhibit a positive effect on innovation up to a threshold (approximately 8.0–8.5), beyond which the marginal benefit declines and becomes unstable. This indicates that moderate investment in operational activities can stimulate innovation, but excessive spending yields diminishing or even negative returns on investment. In contrast, other expenses display a consistently negative association with patenting, with the steepest decline occurring between 5.5 and 6.5 before stabilizing. These patterns suggest that some spending categories may be beneficial only within optimal ranges, while others consistently divert resources away from innovation-intensive uses.

These results underscore the value of combining statistical and machine learning approaches. Regression models, such as fixed-effects and negative binomial, provide inferential clarity and allow hypothesis testing under well-defined assumptions. They are instrumental in identifying robust linear effects, such as the consistent importance of social inclusivity in the model. In contrast, machine learning models prioritize predictive accuracy and capture complex nonlinear dynamics, revealing additional contributors to financial expenditures, such as governance quality and threshold effects. Explainable AI tools (e.g., SHAP, PDP, ICE) further enhance interpretability by illustrating how individual predictors impact outcomes in terms of direction and magnitude. Together, these approaches provide a more comprehensive understanding of how ESG and financial factors shape innovation.

From a policy perspective, these findings highlight the importance of strategic resource allocation for maximizing the innovation legacy of MSEs. While large financial flows accompany mega-events, not all expenditures translate to innovation-enhancing outcomes. Although large in scale, operational and other expenses exert predominantly adverse marginal effects. This suggests the need for stricter budget prioritization, with a focus on redirecting non-essential or ceremonial spending toward research and development (R&D), digital infrastructure, and human capital development. Simultaneously, positive contributors, such as social rights, environmental stewardship, and revenue multipliers, highlight the broader institutional and sustainability dimensions that governments should integrate into event planning.

For event organizers and stakeholders, these findings emphasize that innovation outcomes are not solely a function of total investment but also of efficient allocation. Embedding sustainability-oriented spending, supporting inclusive social policies, and strengthening governance structures can amplify the innovation benefits of MSEs. Future research should investigate the specific subcomponents of operational and other expenses to determine which areas can be reallocated to achieve greater efficiency. Additionally, applying causal inference or experimental approaches could help establish more direct links between expenditure categories and innovation outcomes in the future.

5.6 Concluding Remarks

This study demonstrates that the relationship between ESG performance, financial factors, and national economic and innovation development is far more complex than a direct linear linkage. Comparative analyses using fixed-effects and negative binomial regressions, along with bootstrap forest and explainable AI methods, highlight the importance of a methodological perspective. While regression models reveal few robust linear associations between ESG composites and GDP growth, machine learning approaches uncover the predictive value of Environmental and Social dimensions when nonlinearities and contextual dynamics are considered. Importantly, the social pillar consistently emerges as a robust driver of both economic and innovation outcomes, underscoring the centrality of human capital, inclusivity, and social infrastructure in building a long-term innovative capacity.

The hypothesized relationships between ESG dimensions and innovation capability are based on established theoretical mechanisms. Environmental performance can drive innovation through green technology development and sustainable product design. Social factors, including education investment, labour standards, and human capital development, enhance the quality of the R&D workforce and its capacity to generate knowledge. Governance quality influences innovation through institutional frameworks that protect intellectual property, reduce regulatory uncertainty, and facilitate long-term research and development (R&D)

investments. Collectively, these ESG dimensions shape the enabling environment for innovation, with patent applications serving as quantifiable proxies for national innovation output.

The findings also reveal a differentiated role for governance. Although governance indicators did not appear significant under linear regression models, they proved highly influential in predictive machine learning frameworks, particularly in shaping innovation capability. This suggests that governance quality and institutional strength matter most when accounting for nonlinear dynamics, rather than simple average effects. Conversely, Environmental factors, while less directly associated with innovation capability, contribute positively to economic growth when nonlinear effects are accounted for and support long-term innovation through sustainability and green technology pathways.

The integration of financial indicators provides a further nuance. The predictor screening results underscore the centrality of population size, operational expenses, and revenue flows in shaping national IC. However, SHAP, PDP, and ICE analyses revealed that the directions of these influences were mixed. Operational and other expenses, while important in explaining variance, exert predominantly adverse marginal effects, suggesting that much of this spending is concentrated on event delivery and ceremonial functions, with limited spillovers into innovation. In contrast, Total Revenue shows a consistently positive effect, reflecting broader economic multipliers, such as employment and tourism income, that indirectly strengthen innovation systems. These findings underscore the importance of distinguishing between financial flows that promote R&D and technological transfers and those that divert resources away from innovation-intensive uses.

Methodologically, this study contributes to the literature by combining statistical inference with machine learning prediction and interpretable AI. This triangulated approach demonstrates that, while traditional models clarify robust linear relationships, machine learning models uncover additional dynamics and predictors. Explainable AI tools, such as SHAP and PDP, enhance transparency by showing not only which factors matter but also how and under what conditions they influence outcomes. This integrated framework offers a richer and more actionable basis for both theoretical understanding and policy applications.

The policy implications of this study are significant. For scholars, this research shifts the debate from a simplistic ESG-to-economic and innovation-growth narrative to a more nuanced understanding of ESG and financial flows as mechanisms for building capabilities and driving innovation. For policymakers and event organizers, the evidence underscores the importance of strategically allocating resources. Strengthening the social environment, embedding sustainability initiatives, and optimizing financial expenditure on R&D and infrastructure are more effective pathways to innovation legacies than increasing ceremonial or administrative spending. The findings also suggest the need to set operational budget thresholds and redirect excess resources toward innovation-driven activities.

The limitations of this study provide avenues for future research. The relatively small sample of six host countries limits generalizability and necessitates larger datasets that cover more events and longer time spans than this study. The use of aggregated ESG indicators may obscure the effects of more specific policies, suggesting a need for disaggregated analyses. Future studies could explore subnational or industry-level data, incorporate alternative measures of innovation performance, and examine the causal pathways through which ESG and financial factors influence long-term economic and innovation capability.

In conclusion, this chapter highlights that the development of IC in the host country depends less on the scale of financial or ESG commitments and more on the efficiency and direction of these investments. Social inclusivity, governance quality, and sustainability policies are the most critical enablers of innovation, whereas revenue multipliers reinforce their effects. Conversely, excessive operational and administrative spending undermined innovation outcomes. By integrating statistical models with machine learning prediction and explainable AI insights, this study provides a comprehensive analytical framework for understanding and optimizing the socioeconomic and technological impacts of mega sports events. Chapter 6 synthesizes the findings from across the dissertation to consolidate the theoretical contributions and practical implications of this research.

Chapter 6: Concluding Remarks

This dissertation explores how mega-sport events contribute to national economic growth, digital evolution, and innovation capability through the lens of economic indicators, sport-related factors, environmental, social, and governance indicators, and financial factors using advanced machine learning methodologies. The study's conceptual foundation, established in Chapter 2, addresses key gaps in the existing research. The literature on MSE underscores its wide-ranging sociocultural, economic, environmental, and political impacts. While social and cultural effects have been the most frequently studied, economic and ecological dimensions have also received significant attention, and research on the impacts of innovation is just beginning to emerge. Simultaneously, prior studies often relied on traditional qualitative or linear quantitative methods, which limited their ability to capture nonlinear dynamics.

This gap provides a rationale for applying advanced approaches, such as machine learning, to enable a more comprehensive assessment of how MSEs contribute to economic growth, digital transformation, and innovation capability. Traditional methods, primarily statistical techniques, tend to focus either on immediate financial returns or on descriptive accounts of social and environmental impacts. This leaves a gap in understanding the nuanced and potentially nonlinear pathways through which MSEs influence long-term national innovation trajectories.

The literature review in Chapter 2 highlights several themes. First, although the economic benefits and costs of MSEs are well documented in the short term, their long-term contributions to sustainable development remain complicated and multifaceted. Second, the social and governance dimensions of MSEs are often treated as qualitative narratives, leaving their measurable contribution to national-level innovation capability underexplored. Third, the environmental dimension has largely been framed in terms of sustainability obligations or ecological costs, without clear links to innovation outcomes, such as energy consumption or sustainable infrastructure. Finally, the methodological review shows that most existing studies rely on regression or descriptive methods, which are ill-suited to capturing the nonlinear, interaction-driven dynamics inherent in MSEs.

Based on this review, this dissertation is structured around four central research questions:

1. **Economic Performance:** In what ways do MSEs influence national economic performance, particularly in terms of the total value added by economic activity and GDP per capita growth?
2. **Innovation capability:** How do MSEs affect national innovation capability, as proxied by patent applications and related technological indicators?
3. **Digital Evolution:** To what extent do MSEs shape digital development outcomes, as measured by digital evolution indices?
4. **Methodological Comparison:** Do machine learning models outperform traditional statistical approaches in prediction accuracy when analyzing the impacts of MSE across multidimensions?

These domains were examined using a multidimensional set of parameters. Economic indicators include imports and exports of goods and services, tax revenue, long-term investments, and tourist expenditure. Sport-related factors include variables such as the number of venues and team performance. Furthermore, the analysis incorporates the dual lenses of economic, environmental, social, and governance indicators, complemented by financial factors associated with hosting MSEs. Methodologically, the framework combines

statistical models (fixed effects, negative binomial regression) for inferential testing with machine learning models, such as bootstrap forests, neural boosting, K-nearest neighbours, time series, and explainable AI tools (SHAP, PDP, ICE) for predictive accuracy and interpretability of results. This study particularly emphasizes the methodological significance of employing high-efficiency, high-accuracy prediction models, as such models are essential for identifying the determinants that influence a nation's capacity to leverage mega-sports events as catalysts for sustained economic, innovative, and digital advancement.

The first investigation in Chapter 3 introduces the hypotheses for applying machine learning models to estimate the economic impact of SMEs. The hypotheses posited that machine learning approaches, such as support vector machine (SVM) and stepwise models, would demonstrate strong performance in estimating the dynamics of total value-added (by economic activity). In contrast, time-series techniques would be particularly effective in capturing patterns and predicting patent applications. Specifically, it tested whether incorporating economic, social, and sports-related variables would improve prediction accuracy. These hypotheses posit that MSEs contribute to short-term economic fluctuations and long-term innovation trends.

The results from boosted trees and random forests indicate that population size and relative GDP change are the most influential determinants of total value added, with sports event-day attendance showing the strongest nonlinear relationship. In comparative model testing, stepwise regression and SVM achieved the highest predictive accuracy, demonstrating the effectiveness of machine learning approaches in capturing the nonlinear dynamics of key determinants.

Although stepwise regression reports an exceptionally high R^2 value, the literature review clearly acknowledges that stepwise procedures are prone to variable selection bias and overfitting. However, this apparent tension is methodological rather than logical. The stepwise regression results are not presented as substantively reliable evidence, but rather as a benchmark that illustrates the upper bound of model fit in a small-sample environment. By juxtaposing this result with ML models and explicitly discussing its limitations, the analysis highlights the dangers of overinterpreting goodness-of-fit statistics under small-N conditions. In this sense, the stepwise model functions as a cautionary example rather than a preferred analytical solution.

The time-series analysis of patent activity from 1981 to 2023 reveals patterns closely associated with hosting major events. For example, the 2006 and 2010 FIFA World Cups coincided with marked increases in innovation, while the Brazil 2014 World Cup aligned with the peak of the 2008–2016 patent surge in the country. In contrast, Russia 2018 and Qatar 2022 corresponded with declines, reflecting the combined effects of global economic challenges, shifting technological priorities, and pandemic disruptions. These fluctuations illustrate how MSEs can act as catalysts for technological advancement and innovation capability, particularly when coupled with favorable economic and policy environments.

The time-series forecasting results indicate that the Seasonal Exponential Smoothing (SES) model produced the most reliable predictions for monthly and annual patent trends. In contrast, the Moving Average (MA) model performed better for short-term forecasts. Overall, the analysis confirms that machine learning approaches achieve high accuracy in capturing the impact of MSEs. Specifically, the SVM and stepwise models were the most effective in predicting economic outcomes. In contrast, the SES and MA models provided robust forecasts for long- and short-term patent activity, respectively. These results highlight the dual role of MSEs in stimulating immediate economic activity and fostering long-term innovation.

While this dissertation consistently demonstrates the advantages of machine learning (ML) approaches over traditional statistical methods, particularly in capturing nonlinear relationships, it also explicitly acknowledges that the FIFA World Cup analysis in Chapter 3 is based on an extremely small sample size ($n = 5$). Consequently, the risk of overfitting is recognized and cannot be ignored. This tension does not constitute a contradiction but rather reflects a difference in analytical scope. In this study, the superiority of ML should not be interpreted primarily in terms of predictive accuracy under small-N conditions. Instead, ML is positioned as a methodological framework that is superior in its capacity to uncover structural patterns, nonlinear dynamics, and interaction effects that are difficult to detect using conventional linear models. Under small-sample conditions, ML results are therefore interpreted as exploratory and structure-revealing rather than as definitive predictors.

The second investigation in Chapter 4 hypothesizes that hosting MSEs is positively and significantly associated with national digital development, as measured by digital evolution scores (DES). It further posits that MSE-hosting status influences the predictive accuracy of selected machine learning models in estimating DES. To test these assumptions, advanced machine learning models, specifically neural boosting and K-nearest neighbors (KNN), were employed to capture the nonlinear interactions between the DES, hosting status, and key economic indicators.

The results from the neural boosted model show that hosting MSEs is positively associated with national digital development, as reflected in a higher DES with moderate predictive power. The neural boosting model successfully differentiated host from non-host countries, achieving nearly 70 percent overall classification accuracy and over 60 percent accuracy for the positive class. In contrast, probabilistic outputs confirmed that hosting MSE substantially increases the likelihood of attaining a higher DES. This suggests that the responsibilities and opportunities associated with hosting MSEs may contribute to and expedite digital advancement.

Temporal analyses centered on the 2016 structural break provide additional insights into this relationship. Although both host and non-host countries experienced statistically significant improvements in the DES after 2016, reflecting broader global digitalization, the magnitude of these gains was consistently larger among MSE-hosting nations. These findings operate at different causal levels and are therefore complementary rather than contradictory. The post-2016 increase reflects global digital transformation trends, whereas MSE hosting is best understood as an accelerator rather than the sole driver of digital evolution. Hosting MSEs amplifies and accelerates digital development by mobilizing investment, institutional coordination, and international knowledge flows within an already expanding global digital environment.

A comparison of the predictor contributions before and after 2016 reveals a substantive shift in the drivers of national digital evolution. Charges for the use of intellectual property (IP) remained an influential predictor in both periods; however, their declining relative importance after 2016 suggests a transition away from reliance on externally acquired technology. In the pre-2016 period, higher IP payments indicated technology adoption and catch-up strategies, consistent with an externally driven phase of digitalization. In contrast, the post-2016 period is characterized by a marked increase in the contribution of resident patent applications, signalling a shift toward endogenous innovation and domestic knowledge creation. This transition implies that digital evolution increasingly depends on internally generated technological capabilities rather than imported IP.

A parallel structural change was observed in government education expenditure, which showed stronger predictive power in the post-2016 period. Increased educational investment enhances human capital formation, digital skills, and absorptive capacity, which are core prerequisites for advanced digital transformation. Simultaneously, although tax revenue as a share of GDP remained statistically relevant across both periods, its diminishing relative contribution suggests that fiscal capacity alone becomes less decisive once foundational digital infrastructures are established. These results indicate that post-2016 digital evolution reflects a broader transition from externally driven digitalization to an internally anchored digital transformation grounded in domestic innovation systems and education. This shift is particularly pronounced in countries hosting MSEs, where event-related investments, international exposure, and policy mobilization appear to accelerate the reorientation from adoption-based to innovation-led digital development processes. Methodologically, KNN's superior performance, especially for host nations, underscores the importance of selecting machine learning models that can capture localized, nonlinear patterns in national digital trajectories.

In conclusion, the empirical evidence from Chapter 4 supports the hypothesis that hosting MSEs contributes positively to higher digital evolution, both directly through enhanced national DES and indirectly through trade integration and knowledge-based assets. Hosting MSEs does not merely coincide with global digital progress; rather, it strengthens the structural mechanisms through which economies convert resources and knowledge into sustainable digital advancements. These findings establish a critical bridge to Chapter 5, which investigates how financial and ESG-related factors shape innovation outcomes and long-term legacy formation in MSE-hosting nations.

The final investigation in Chapter 5 presents a set of hypotheses on the relationships between ESG indicators, financial factors, economic growth, and innovation capability. First, it was hypothesized that the three ESG dimensions (environmental, social, and governance) exert differential effects on economic growth and innovation capabilities. This hypothesis was evaluated using traditional statistical approaches, including fixed-effect and negative binomial models. Second, it was hypothesized that financial factors associated with MSEs, such as capital and operational expenditures, would positively influence their innovation capabilities alongside ESG and macroeconomic predictors. This hypothesis was tested using the bootstrap forest prediction model. The third hypothesis is that explainable AI techniques (SHAP, PDP, and ICE) will generate new insights into the direction, magnitude, and thresholds of these effects, thereby enhancing interpretability and informing policy recommendations.

The t-test statistics demonstrate that the relationship between ESG performance and innovation outcomes differs fundamentally between MSE host and non-host countries. Among non-host countries, ESG-patent relationships are limited and fragmented, with only a small subset of sustainability indicators showing meaningful associations with innovative activity. In contrast, MSE host countries exhibit stronger, broader, and more systematic ESG-patent linkages spanning all three ESG dimensions, suggesting that innovation capability is more deeply embedded in sustainability-oriented institutional and policy frameworks.

Decision tree-based predictor screening reinforces this distinction, revealing that ESG factors contribute to innovation in a more diversified and balanced manner among host countries. In contrast, non-host countries exhibit a narrower set of influential ESG predictors. These findings imply that MSE hosting functions as a structural amplifier, strengthening the mechanisms through which ESG-related investments, institutional quality, and social development translate into innovative outcomes. This evidence supports the interpretation that

mega-sport events contribute to longer-term, sustainability-driven innovation dynamics, motivating subsequent focus on host countries.

The estimation of the effects of ESG and fiscal factors on economic growth and innovation capability is highly dependent on the methodological approach and model specification. For economic development, fixed-effects regression reveals no significant within-country effects of ESG dimensions on GDP per capita. In contrast, bootstrap forest models identify environmental and social factors as strong nonlinear predictors, highlighting threshold and interaction effects overlooked in linear specifications.

For innovation capability (proxied by patent applications), negative binomial regression confirms the consistent linear contributions of social factors, while environmental indicators remain insignificant, and governance displays an inverse effect. However, bootstrap forest models reveal the additional predictive importance of ecological and governance factors, underscoring the role of institutional quality once nonlinear dynamics are considered. These divergent results demonstrate that model choice critically shapes inferences regarding ESG effects and that machine learning approaches capture the complexity obscured by traditional statistical methods.

In addition, when ESG parameters are combined with macroeconomic and MSE-related financial factors for IC prediction, predictor screening, and XAI methods (SHAP, PDP, and ICE), these methods further clarify these relationships. The predictor screening results indicate that population size and operating expenses are dominant predictors of patent activity, while green energy policies and social rights are positively associated with innovation.

Conversely, the SHAP results suggest that excessive operating and other expenditures yield diminishing or negative returns. While these variables remain influential, their marginal contributions are predominantly negative, reflecting inefficiencies in resource allocation. Disparities between innovation-oriented and non-innovation-related spending reduce their overall effectiveness in supporting innovation capability. In contrast, total event revenue shows a consistent positive effect, reflecting its role in capturing broader economic multipliers, such as induced employment and tourism-related income, which, in turn, can foster this. Taken together, these results suggest that not all financial flows associated with MSEs are equally beneficial for innovation; their impact depends on the extent to which expenditures or revenues are connected to R&D and technology-driven activities.

The dissertation emphasizes that the application of explainable AI (XAI) techniques, such as SHAP, PDP, and ICE, enhances the interpretability of ML models. At the same time, it is evident that policy-relevant interpretations are derived from a limited subset of variables (e.g., intellectual property payments and patent activity). This does not represent a contradiction, but rather reflects the distinction between theoretical potential and practical implementation. The interpretability achieved in this study is intentionally partial and selective. Rather than aiming for an exhaustive explanation of all predictors, the analysis concentrates on variables with clear institutional, economic, and policy relevance. Accordingly, the contribution of XAI in this dissertation lies in enabling policy-relevant interpretability, rather than delivering full causal transparency across all model components.

The findings can be synthesized across four thematic dimensions: economic, digital development, innovation capability and methodology. Economically, MSEs influence outcomes both directly and indirectly, with population size, relative GDP change, and event-day attendance emerging as the strongest predictors of the total value added. Environmental and social factors also contribute significantly to GDP per capita growth through nonlinear relationships, underscoring that benefits depend less on the event itself and more on national

strategies and the country's macroeconomic context. In the digital domain, high digital evolution scores (DES) are strongly associated with hosting status, confirming that nations hosting MSEs tend to achieve higher DES levels. These findings indicate that the digital legacy of MSEs depends not only on infrastructure investments but also on host nations' capacity to leverage global knowledge flows and strengthen their competitive positions. Simultaneously, economic factors such as export volumes, charges for the use of intellectual property, and trade in goods and services emerged as the strongest contributors to digital advancement, underscoring the critical role of global trade integration and the commercialization of intangible assets in accelerating digital transformation.

Innovation capability was similarly enhanced, with time-series analyses showing an increase in patent activity, except during periods of global crisis. Social indices, particularly inclusivity in host countries, consistently emerge as drivers of innovation capability, as evidenced by the negative binomial model. Environmental and governmental factors were also found to contribute to the prediction of innovation capability, as identified using the bootstrap forest model (nonlinear contribution). In addition, the population size and operational expenditures emerged as dominant predictors of innovation capability. Methodologically, this study advances the systematic integration of statistical techniques with machine learning and XAI. While regression models, such as fixed effects and negative binomial, provide rigour in testing linear relationships, ensemble-based models, including boosted trees, neural boosting, and KNN, capture nonlinearities, thresholds, and context-specific dynamics. The addition of XAI tools (SHAP, PDP, ICE) further enhanced the interpretability by clarifying the variable importance, marginal effects, and critical thresholds. This triangulated framework improves predictive accuracy and establishes a replicable toolkit for evaluating complex socioeconomic phenomena beyond the scope of MSEs.

The title and abstract emphasize impact estimation and prediction, whereas the discussion and conclusion place greater weight on explanation, interpretation, and policy implications. This shift does not indicate inconsistency but rather reflects a deliberate hierarchical research design. Specifically, the primary contribution of this dissertation is explanatory: it seeks to clarify how economic, digital, and innovation outcomes are structurally related to MSE hosting and national characteristics. Predictive modeling serves as a secondary contribution, providing empirical support and methodological robustness to the explanatory framework. This hierarchical positioning—primary explanatory, secondary predictive—allows the study to integrate machine learning into impact research without conflating prediction accuracy with causal inference.

Overall, the apparent tensions identified in this dissertation reflect deliberate methodological choices that distinguish structural explanation from predictive ambition, positioning machine learning as a complementary, structure-revealing framework rather than a definitive causal or forecasting tool under constrained empirical conditions.

This study has five main real-world implications.

Academic Implications: This study introduces a replicable methodological framework that integrates statistical analysis, advanced machine learning, and XAI to evaluate the legacy of MSE. By combining linear and nonlinear perspectives, this framework enhances methodological rigour and broadens the analytical toolkit for impact estimation research. Beyond the MSE, it can be extended to other global platforms, such as world expos, cultural

festivals, and climate summits, thereby broadening the scope of legacy assessment and fostering cross-disciplinary insights.

Event Organizers (IOC, FIFA, and similar bodies): International governing bodies, such as the International Olympic Committee (IOC) and FIFA, can embed sustainability and digital evolution benchmarks into event requirements, making them integral to the evaluation and monitoring of host city performance. By institutionalizing standards for energy consumption, digital infrastructure readiness, and social inclusivity, organizers can create accountability mechanisms that not only elevate legacy outcomes but also set new global benchmarks for responsible mega-event management.

Sponsors and Stakeholders: Corporate sponsors, investors, and partner institutions can strategically align their event-related investments with innovation-driving categories such as technology infrastructure, sustainability initiatives, and community development programs. This alignment not only maximizes positive spillovers for host nations but also strengthens corporate reputation, brand equity and long-term stakeholder trust. The findings suggest that targeted sponsorships in digital innovation, renewable energy, and inclusive social programs generate measurable societal impacts, thereby differentiating sponsors as leaders of responsible, forward-looking engagements.

Communities and Civil Society: Local communities, residents, and civil society organizations stand to benefit directly from more inclusive and innovation-driven MSE planning. By prioritizing social inclusivity, digital access, and sustainability in event design, communities can gain new employment opportunities, expand their digital infrastructure, and improve public services that persist beyond the events. Civil society groups can also shape accountability frameworks by ensuring that investments address local needs, such as education, cultural preservation, and environmental protection. Thus, MSEs can serve as global spectacles and catalysts for community empowerment and for long-term social resilience.

Government and Policy Implications: National and local governments can draw on these findings to design budget frameworks oriented toward economic growth, digitalization, and innovation when bidding for or hosting MSEs. By reallocating resources from non-essential operational and ceremonial expenses to R&D, digital infrastructure, and green energy initiatives, event-related investments can be converted into sustainable technological and social benefits for the future. Simultaneously, the application of advanced analytical tools, such as machine learning and explainable AI, enables decision-makers to evaluate trade-offs, interaction effects, and thresholds in budget allocation with more precision. Embedding ESG principles, particularly social inclusivity and governance quality, into event strategies not only strengthens national innovation ecosystems but also aligns event planning with long-term developmental priorities for digital transformation and sustainability. This integrated approach ensures that MSEs are strategically leveraged as catalysts for economic and social advancement.

Although this dissertation offers valuable insights, it also highlights avenues for future research. First, the dataset was restricted to specific games and host countries, potentially limiting the generalizability of the findings to other countries. Expanding the scope to include a broader set of hosts and non-hosts would enable more robust comparative analyses. Second, reliance on patents as the sole proxy for innovation may not fully capture the multidimensional nature of innovation, particularly in terms of quality and diffusion. Therefore, future studies should incorporate complementary indicators, such as R&D intensity, technology adoption rates, and composite innovation indices, to provide a more comprehensive assessment.

Building on the insights of this study, future research could extend the methodological framework beyond mega-sport events to include annual medium- and small-scale sporting events, thereby enabling the construction of more consistent longitudinal datasets. The framework can also be applied to the rapidly growing domain of e-sports, where patterns of interaction, digital engagement, and impact measurement differ substantially from traditional events. In addition to exploring new event types, future studies should incorporate broader economic indicators, such as employment dynamics, tourism spillovers, infrastructure utilization, and productivity growth, to capture more nuanced effects beyond aggregate GDP measures. Similarly, innovation outcomes should not be limited to patent counts but expanded to include R&D intensity, technology adoption rates, startup formation, and knowledge transfer mechanisms, which may better reflect how events stimulate national and regional innovation ecosystems. Such extensions would not only broaden the scope of legacy assessment but also foster cross-disciplinary insights into how diverse forms of sports and digital competition contribute to sustainable economic development and long-term capability building.

References

1. Chappelet, J.-L. and M.M. Parent, *Routledge handbook of sports event management*. 2017: Routledge.
2. Rookwood, J. and K. Adeosun, *Nation branding and public diplomacy: Examining Japan's 2019 Rugby World Cup and 2020 (21) Olympic Games in the midst of a global economic downturn and the COVID-19 pandemic*. Journal of Global Sport Management, 2023. **8**(3): p. 573-593.
3. Müller, M. and C. Gaffney, *Comparing the Urban Impacts of the FIFA World Cup and Olympic Games From 2010 to 2016*. Journal of Sport and Social Issues, 2018. **42**(4): p. 247-269.
4. Leopkey, B. and M.M. Parent, *The governance of Olympic legacy: Process, actors and mechanisms*. Leisure Studies, 2017. **36**(3): p. 438-451.
5. Preuss, H., B. Seguin, and N. O'Reilly, *Profiling major sport event visitors: The 2002 commonwealth games*. Journal of Sport and Tourism, 2007. **12**(1): p. 5-23.
6. Ritchie, J.B., *Turning 16 days into 16 years through Olympic legacies*. Event management, 2000. **6**(3): p. 155-165.
7. Veal, A.J., K. Toohey, and S. Frawley, *The sport participation legacy of the Sydney 2000 Olympic Games and other international sporting events hosted in Australia*. Journal of policy research in tourism, leisure and events, 2012. **4**(2): p. 155-184.
8. Solberg, H.A. and H. Preuss, *Major sport events and long-term tourism impacts*. Journal of Sport Management, 2007. **21**(2): p. 213-234.
9. Brent Ritchie, J., *Assessing the impact of hallmark events: Conceptual and research issues*. Journal of travel research, 1984. **23**(1): p. 2-11.
10. Hall, C.M., *Hallmark tourist events: impacts, management and planning*. 1992.
11. Getz, D., *Event Management & Event Tourism*. Canada, Cognizant Communication Corporation, 1997.
12. Smith, A., *Reimagining the city: The value of sport initiatives*. Annals of tourism research, 2005. **32**(1): p. 217-236.
13. Misener, L. and D.S. Mason, *Creating community networks: Can sporting events offer meaningful sources of social capital?* Managing leisure, 2006. **11**(1): p. 39-56.
14. Elahi, A., S. Gholampour, and F. Askarian, *The effects of sports mega-events on host communities: A systematic review of studies in three recent decades*. Sports Business Journal, 2021. **1**(1): p. 13-30.
15. Koenigstorfer, J., et al., *Mapping research on legacy of mega sporting events: Structural changes, consequences, and stakeholder evaluations in empirical studies*. Leisure Studies, 2019. **38**(6): p. 729-745.
16. Burbank, M., G. Andranovich, and C. Heying, *Olympic Dreams: The Impact of Mega-Events on Local Politics*. Lynne Rienner, Boulder, CO. 2001.
17. Burgan, B. and T. Mules, *Economic impact of sporting events*. Annals of tourism research, 1992. **19**(4): p. 700-710.
18. Hall, C.M., *The definition and analysis of hallmark tourist events*. GeoJournal, 1989. **19**: p. 263-268.
19. Jago, L.K. and R.N. Shaw, *Special events: A conceptual and definitional framework*. Festival management and event tourism, 1998. **5**(1-2): p. 21-32.
20. Horne, J., *The four 'knowns' of sports mega-events*. Leisure Studies, 2007. **26**(1): p. 81-96.
21. Chappelet, J.-L. *Managing the size of the Olympic Games*. in *The Olympic Games: Meeting New Global Challenges*. 2016. Routledge.
22. Crompton, J.L., *Economic Impact Analysis of Sports Facilities and Events: Eleven Sources of Misapplication*. Journal of Sport Management, 1995. **9**(1): p. 14-35.
23. Crompton, J.L., S. Lee, and T.J. Shuster, *A guide for undertaking economic impact studies: The Springfest example*. Journal of travel research, 2001. **40**(1): p. 79-87.
24. Tyrrell, T.J. and R.J. Johnston, *A framework for assessing direct economic impacts of tourist events: Distinguishing origins, destinations, and causes of expenditures*. Journal of travel Research, 2001. **40**(1): p. 94-100.

25. Müller, M., *What makes an event a mega-event*. Managing Sport Mega-Events; Routledge: London, UK, 2016. **8**.
26. Grix, J. and B. Houlihan, *Sports mega-events as part of a nation's soft power strategy: The cases of Germany (2006) and the UK (2012)*. British Journal of Politics and International Relations, 2014. **16**(4): p. 572-596.
27. Smith, A., *Leveraging sport mega-events: New model or convenient justification?* Journal of Policy Research in Tourism, Leisure and Events, 2014. **6**(1): p. 15-30.
28. Al-Dosari, K., *The significance of mega sporting event on infrastructure development: A case of FIFA 2022 World Cup in Qatar*. Journal of Social Sciences (COES&RJ-JSS), 2020. **9**(3): p. 1295.1319-1295.1319.
29. *Qatar Infrastructure, Power, and Communications*. Available from: <https://www.nationsencyclopedia.com/economies/Asia-and-the-Pacific/Qatar-INFRASTRUCTURE-POWER-AND-COMMUNICATIONS.html>.
30. Azzali, S., *Mega sporting events and their impact on the built environment: Lessons learned from the past*. A/Z ITU Journal of the Faculty of Architecture, 2019. **16**(2): p. 25-37.
31. Fan, J., et al., *Evaluating the impact of mega-sports events on urbanization focusing on land-use changes using a scenario-based model*. Sustainability (Switzerland), 2021. **13**(4): p. 1-19.
32. Wang, M., H.X. Bao, and P.-t. Lin, *Behavioural insights into housing relocation decisions: The effects of the Beijing Olympics*. Habitat International, 2015. **47**: p. 20-28.
33. Chen, Y., *The future of mega-events, post-COVID-19 pandemic*. International review for spatial planning and sustainable development, 2022. **10**(2): p. 168-178.
34. Falla, M., J. Prior, and B. Jacobs, *Assessing the impact of sporting mega-events on the social and physical capital of communities in host cities: the Gold Coast 2018 Commonwealth Games experience*. Commonwealth Journal of Local Governance, 2022(26): p. 5-27.
35. Flyvbjerg, B., A. Budzier, and D. Lunn, *Regression to the tail: Why the Olympics blow up*. Environment and Planning A: Economy and Space, 2021. **53**(2): p. 233-260.
36. Lee, S. and B.D. Krohn, *A study of psychological support from local residents for hosting mega-sporting events: A case of the 2012 Indianapolis super bowl XLVI*. Event Management, 2013. **17**(4): p. 361-376.
37. Deng, Y., *Shaping mega-event flagships: a case study of Expo Center of Expo 2010 Shanghai, China*. Facilities, 2012. **30**(13/14): p. 590-610.
38. Ludvigsen, J., & Petersen-Wagner, R., *From television to YouTube: digitalised sport mega-events in the platform society*. Leisure Studies, 2022. **42**: p. 615 - 632.
39. Trillas, F. and E. Serrano, *The Economics of Major Sporting Events: The Case of the Barcelona Olympic Games. An Overview 25 Years Later*. European Accounting and Management Review, 2016. **2**(2): p. 77-102.
40. Brynildsen, K.B. and M.M. Parent, *Test events as risk management tools in the context of the olympic movement's agenda 2020*. Event Management, 2021. **25**(2): p. 115-133.
41. Davies, L., R. Coleman, and G. Ramchandani, *Measuring attendance: issues and implications for estimating the impact of free-to-view sports events*. International Journal of Sports Marketing and Sponsorship, 2010. **12**(1): p. 6-18.
42. Thomson, A., K. Toohey, and S. Darcy, *The political economy of mass sport participation legacies from large-scale sport events: a conceptual paper*. Journal of Sport Management, 2021. **35**(4): p. 352-363.
43. Roche, M., *Megaevents and modernity: Olympics and expos in the growth of global culture*. 2002: Routledge.
44. Kim, H.J., D. Gursoy, and S.-B. Lee, *The impact of the 2002 World Cup on South Korea: Comparisons of pre-and post-games*. Tourism management, 2006. **27**(1): p. 86-96.
45. Müller, M., *What makes an event a mega-event? Definitions and sizes*, in *Leveraging mega-event legacies*. 2018, Routledge. p. 13-28.
46. Hover, P., et al., *Integrity & sport events*. 2016.
47. Chappelet, J.-L. and M.M. Parent, *Routledge handbook of sports event management*. 2015: Routledge.

48. Pickering, C. and J. Byrne, *The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers*. Higher Education Research & Development, 2014. **33**(3): p. 534-548.
49. Thomson, A., et al., *Sport event legacy: A systematic quantitative review of literature*. Sport management review, 2019. **22**(3): p. 295-321.
50. Lee, C.J., *Effects of sport mega-events on city brand awareness and image: Using the 2009 world games in Kaohsiung as an example*. Quality and Quantity, 2014. **48**(3): p. 1243-1256.
51. Chappelet, J.-L., *Mega sporting event legacies: a multifaceted concept*. 2012.
52. Zawadzki, K., *The economic legacy of mega sporting events. The impact of hosting European Olympic Games on GDP Growth Through Infrastructure Development*. Polish Journal of Sport and Tourism, 2022. **29**: p. 36-42.
53. Hu, C. and S. Cole, *Comparing Expenditure Patterns and Travel Characteristics Among NFL Fans With Different Levels of Team Identification*. Tourism Review International, 2016. **20**(1): p. 3-10.
54. Lee, S.K., J. Harris, and M.R. Lyberger, *The Economic Impact of College Sporting Events: A Case Study of Division I-a Football Games*. Event Management, 2010. **14**(2): p. 157-165.
55. Feng, Q., *Analyzing the Economic Impact of Super Bowls: A Qualitative Framework for Holding Mega Sports Events in Asia*. 2024: p. 10-17.
56. Preuss, H., *A method for calculating the crowdingout effect in sport mega-event impact studies: The 2010 FIFA world cup*. Development Southern Africa, 2011. **28**(3): p. 367-385.
57. Jung, S., et al., *Evaluating Global Integrated Transportation Application for Mega Event: Role of Trust and Exchanging Personal Information in Mobility as a Service (MaaS)*, in *22nd International Conference on Human-Computer Interaction, HCI International 2020*, C. Stephanidis, M. Antona, and S. Ntoa, Editors. 2020, Springer Science and Business Media Deutschland GmbH. p. 575-584.
58. Parnell, D., et al., *Football Worlds: Business and networks during COVID-19*, in *COVID-19 and the Soccer World*. 2022, Routledge. p. 22-29.
59. Ruzlin, A.N.M., et al., *Promoting Mental Health During the COVID-19 Pandemic: A Hybrid, Innovative Approach in Malaysia*. Frontiers in Public Health, 2021. **9**.
60. Chan, L., et al., *Review of Evaluation Metrics Used in Digital and Traditional Tobacco Control Campaigns*. Journal of Medical Internet Research, 2020. **22**(8): p. e17432.
61. Arshanapally, S., et al., *Use of a Paid Digital Marketing Campaign to Promote a Mobile Health App to Encourage Parent-Engaged Developmental Monitoring: Implementation Study*. Jmir Pediatrics and Parenting, 2022. **5**(2): p. e34425.
62. Ezurike, N., *Sequential Explanatory Mixed Method Approach to Understanding the Impact of Digital Media on Media Relations and Audience Engagement*. British Journal of Marketing Studies, 2024. **12**(6): p. 134-149.
63. Szell, M., S. Grauwin, and C. Ratti, *Contraction of Online Response to Major Events*. Plos One, 2014. **9**(2): p. e89052.
64. Feng, J., et al., *The Economic Impact of Digital Productivity on Sports Competitions*. Advances in Economics Management and Political Sciences, 2025. **171**(1): p. 1-12.
65. Furman, J.L., M.E. Porter, and S. Stern, *The determinants of national innovative capacity*. Research policy, 2002. **31**(6): p. 899-933.
66. Albaladejo, M. and H. Romijn, *Determinants of innovation capability in small UK firms*. 2001, ECIS working paper No. 00.13.
67. Porter, M.E., *The competitive advantage of nations London*. Recherche INAISE/DG5, 1990.
68. Porter, M.E. and S. Stern, *The new challenge to America's prosperity: Findings from the innovation index*. (No Title), 1999.
69. Porter, M.E., *Clusters and the new economics of competition*. Vol. 76. 1998: Harvard Business Review Boston.
70. Porter, M.E., *Enhancing the microeconomic foundations of prosperity: the current competitiveness index*. The global competitiveness report, 2001. **2002**: p. 52-73.
71. Wonglimpiyarat, J., *Innovation index and the innovative capacity of nations*. Futures, 2010. **42**(3): p. 247-253.

72. Kao, C., et al., *Measuring the national competitiveness of Southeast Asian countries*. European Journal of Operational Research, 2008. **187**(2): p. 613-628.
73. Xu, Z., C. Wu, and X. Li, *Residents' Perceptions and Behavioral Intentions towards Mega-Sports Events: A Case Study of Beijing 2022 Olympic Winter Games*. Sustainability (Switzerland), 2022. **14**(22).
74. Tjørndal, A., *Sport, Innovation and Strategic Management: A Systematic Literature Review*. Brazilian Business Review, 2016: p. 38-56.
75. Kirby, S.I., M.B. Duignan, and D. McGillivray, *Mega-sport events, micro and small business leveraging: Introducing the "mse-msb leverage model"*. Event Management, 2018. **22**(6): p. 917-931.
76. Al-Dosari, K., et al., *Security Supply Chain Using UAVs: Validation and Development of a UAV-Based Model for Qatar's Mega Sporting Events*. Drones, 2023. **7**(9): p. 555.
77. Ramasamy, B., H. Wu, and M. Yeung, *Hosting Annual International Sporting Events and Tourism: Formula 1, Golf or Tennis?* Tourism Economics, 2021. **28**(8): p. 2082-2098.
78. Trkulja, M., *The PERSPECTIVE OF DIGITAL MARKETING OF THE SPORT MEGA EVENTS IN THE 21st CENTURY*. SMB, 2025. **11**(2): p. 33-46.
79. Mann, R.J. and T.W. Sager, *Patents, venture capital, and software start-ups*. Research Policy, 2007. **36**(2): p. 193-208.
80. Stek, P.E. and M.S. van Geenhuizen, *The influence of international research interaction on national innovation performance: A bibliometric approach*. Technological Forecasting and Social Change, 2016. **110**: p. 61-70.
81. Igami, M. and J. Subrahmanyam, *Patent statistics as an innovation indicator? Evidence from the hard disk drive industry*. The Japanese Economic Review, 2019. **70**(3): p. 308-330.
82. Wu, Y.C.J. and H.P. Liu, *Technological innovation assessment of business-to-business electronic marketplaces*. Journal of the American Society for Information Science and Technology, 2006. **57**(8): p. 1093-1104.
83. Xu, Z., C. Wu, and X. Li, *Residents' perceptions and behavioral intentions towards mega-sports events: A case study of Beijing 2022 Olympic Winter Games*. Sustainability, 2022. **14**(22): p. 14955.
84. Fourie, J. and M. Santana-Gallego, *The impact of mega-sport events on tourist arrivals*. Tourism Management, 2011. **32**(6): p. 1364-1370.
85. Patel, M.S., et al., *Using remotely monitored patient activity patterns after hospital discharge to predict 30 day hospital readmission: a randomized trial*. Scientific reports, 2023. **13**(1): p. 8258.
86. Jiang, H., et al., *Exploring the effects of technology-related informal mathematics learning activities: A structural equation modeling analysis*. International Journal of Science and Mathematics Education, 2025. **23**(1): p. 49-69.
87. Schapire, R.E., *The strength of weak learnability*. Machine learning, 1990. **5**: p. 197-227.
88. Drucker, H., D. Wu, and V.N. Vapnik, *Support vector machines for spam categorization*. IEEE Transactions on Neural networks, 1999. **10**(5): p. 1048-1054.
89. Breiman, L., *Random forests*. Machine learning, 2001. **45**: p. 5-32.
90. Dong, J. and D. Dong, *[Retracted] Application of Motion Effect Evaluation Algorithm Based on Random Forest*. Computational Intelligence and Neuroscience, 2022. **2022**(1): p. 2039423.
91. Groll, A., et al., *Prediction of the fifa world cup 2018-a random forest approach with an emphasis on estimated team ability parameters*. arXiv preprint arXiv:1806.03208, 2018.
92. Dou, C., T. Zhao, and Z. Guo, *Establishment and Analysis of Multi-Factor Stock Selection Model Based on Support Vector Machine in CSI 300 Index Constituent Stocks Market*. in *6th International Conference on Financial Innovation and Economic Development (ICFIED 2021)*. 2021. Atlantis Press.
93. Quinn, G.P. and M.J. Keough, *Experimental design and data analysis for biologists*. 2002: Cambridge university press.
94. Field, A., *Discovering statistics using SPSS Sage publications*. London, 2005.
95. Derksen, S. and H.J. Keselman, *Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables*. British Journal of Mathematical and Statistical Psychology, 1992. **45**(2): p. 265-282.

96. Mundry, R. and C.L. Nunn, *Stepwise model fitting and statistical inference: turning noise into signal pollution*. The American Naturalist, 2009. **173**(1): p. 119-123.
97. Agrawal, R. and D. Thukaram. *Identification of fault location in power distribution system with distributed generation using support vector machines*. in *2013 IEEE PES Innovative Smart Grid Technologies Conference (ISGT)*. 2013. IEEE.
98. ALzubi, J.A., et al., *Boosted neural network ensemble classification for lung cancer disease diagnosis*. Applied Soft Computing, 2019. **80**: p. 579-591.
99. Badirli, S., et al., *Gradient boosting neural networks: Grownnet*. arXiv preprint arXiv:2002.07971, 2020.
100. Dobson, A.J. and A.G. Barnett, *An introduction to generalized linear models*. 2018: Chapman and Hall/CRC.
101. Musoro, J.Z., et al., *Validation of prediction models based on lasso regression with multiply imputed data*. BMC medical research methodology, 2014. **14**: p. 1-13.
102. Efron, B., *Bootstrap methods: another look at the jackknife*, in *Breakthroughs in statistics: Methodology and distribution*. 1992, Springer. p. 569-593.
103. Bickel, P.J. and D.A. Freedman, *Some asymptotic theory for the bootstrap*. The annals of statistics, 1981. **9**(6): p. 1196-1217.
104. Singh, K., *On the asymptotic accuracy of Efron's bootstrap*. The annals of statistics, 1981: p. 1187-1195.
105. Decrouez, G. and P. Hall, *Asymptotic expansions and roth's theorem*. НОСТЕЙ И ЕЕ ПРИЛОЖЕНИЯ, посвященная 100-летию со дня рождения БВ Гнеденко (Москва, 26–30 июня 2012 года). Тезисы докладов/Под редакцией АН Ширяева, АВ Лебедева. М.: ЛЕНАНД, 2012. 400 с. ISBN 978-5-9710-0492-9, 2012. **26**: p. 89.
106. Abu Alfeilat, H.A., et al., *Effects of distance measure choice on k-nearest neighbor classifier performance: a review*. Big data, 2019. **7**(4): p. 221-248.
107. Tiwari, R., *Explainable AI (XAI) and its applications in building trust and understanding in AI decision making*. International J. Sci. Res. Eng. Manag, 2023. **7**: p. 1-13.
108. Linardatos, P., V. Papastefanopoulos, and S. Kotsiantis, *Explainable ai: A review of machine learning interpretability methods*. Entropy, 2020. **23**(1): p. 18.
109. Arrieta, A.B., et al., *Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI*. Information fusion, 2020. **58**: p. 82-115.
110. Tang, H., et al., *Analysis and evaluation of explainable artificial intelligence on suicide risk assessment*. Scientific reports, 2024. **14**(1): p. 6163.
111. Azar, A.S., et al., *Application of Machine Learning Techniques for Predicting Survival in Ovarian Cancer*. BMC Medical Informatics and Decision Making, 2022. **22**(1).
112. Qu, C., et al., *Machine Learning Models of Acute Kidney Injury Prediction in Acute Pancreatitis Patients*. Gastroenterology Research and Practice, 2020. **2020**: p. 1-8.
113. Yağın, F.H., et al., *Pilot-Study to Explore Metabolic Signature of Type 2 Diabetes: A Pipeline of Tree-Based Machine Learning and Bioinformatics Techniques for Biomarkers Discovery*. Nutrients, 2024. **16**(10): p. 1537.
114. Tong, J., et al., *A Study on Identification of Urban Waterlogging Risk Factors Based on Satellite Image Semantic Segmentation and XGBoost*. Sustainability, 2023. **15**(8): p. 6434.
115. Gu, Z., et al., *Research on Mining Maximum Subsidence Prediction Based on Genetic Algorithm Combined With XGBoost Model*. Sustainability, 2022. **14**(16): p. 10421.
116. Kumar, P., et al., *Addressing Class Imbalance in Soil Movement Predictions*. 2023.
117. Zimbalist, A., *Circus maximus: The economic gamble behind hosting the Olympics and the World Cup*. 2020: Brookings Institution Press.
118. Flyvbjerg, B. and A. Stewart, *Olympic proportions: Cost and cost overrun at the Olympics 1960-2012*. 2012.
119. Ahlert, G., *Hosting the FIFA World Cup™ Germany 2006: Macroeconomic and regional economic impacts*. Journal of Convention and Event Tourism, 2007. **8**(2): p. 57-78.
120. Kaplanidou, K., et al., *Qatar World Cup 2022: Awareness of the event and its impact on destination and country character perceptions in the US tourism market*. Tourism Review International, 2016. **20**(2-3): p. 143-153.

121. Congdon-Hohman, J. and V.A. Matheson, *International women's soccer and gender inequality: Revisited*, in *Handbook on the economics of women in sports*. 2013, Edward Elgar Publishing. p. 345-364.
122. Jiang, C.H. and F. Liang, *Research on the sports economy effection in national economic development*. Applied mechanics and materials, 2014. **556**: p. 6519-6521.
123. Cooper, J. and D.H. Alderman, *Cancelling March Madness exposes opportunities for a more sustainable sports tourism economy*, in *Global tourism and COVID-19*. 2021, Routledge. p. 71-81.
124. Kokolakis, T. and F. Lera-Lopez, *Sport promotion through sport mega-events. An analysis for types of olympic sports in London 2012*. International Journal of Environmental Research and Public Health, 2020. **17**(17): p. 1-17.
125. Fernandez-Cortes, J., et al., *Influence of game indicators on the ranking of teams in the Spanish soccer league*. Applied Sciences, 2023. **13**(14): p. 8097.
126. Cho, S.-Y., *A league of their own: Female soccer, male legacy and women's empowerment*. 2013, IAI Discussion Papers.
127. Bandiera, O., I. Barankay, and I. Rasul, *Team incentives: Evidence from a firm level experiment*. Journal of the European Economic Association, 2013. **11**(5): p. 1079-1114.
128. Aquino, R., et al., *Influence of situational variables, team formation, and playing position on match running performance and social network analysis in brazilian professional soccer players*. The Journal of Strength & Conditioning Research, 2020. **34**(3): p. 808-817.
129. Semenova, G.N., et al., *Social integration as a factor of economic growth: experience and perspectives of developing countries*. International Journal of Sociology and Social Policy, 2021. **41**(1/2): p. 37-49.
130. Zhang, M., X. Zou, and L. Sha, *Social security and sustainable economic growth: Based on the perspective of human capital*. Sustainability, 2019. **11**(3): p. 662.
131. Silvino, V.O., et al., *The Use of Machine Learning in Sports Performance: A Systematic Review*. Translational Journal of the American College of Sports Medicine, 2025. **10**(2): p. e000304.
132. Algarhi, A.S. and M.P. Karimazondo, *The impact of ESG on economic growth: Evidence from BRICS+ countries*. Theoretical Economics Letters, 2024. **14**(4).
133. Guyon, I. and A. Elisseeff, *An introduction to variable and feature selection*. Journal of machine learning research, 2003. **3**(Mar): p. 1157-1182.
134. Wang, T. and F. Sun, *Global gridded GDP data set consistent with the shared socioeconomic pathways*. Scientific data, 2022. **9**(1): p. 221.
135. Baade, R.A. and V.A. Matheson, *The quest for the cup: assessing the economic impact of the world cup*. Regional studies, 2004. **38**(4): p. 343-354.
136. Sassenberg, C., C. Prentice, and P. Thaichon, *The Effect of Infrastructure on the Social Legacy of Mega Sporting Events*. Multidisciplinary Business Review, 2025. **18**(1): p. 94-111.
137. Shmueli, G., *To explain or to predict?* Statistical science, 2010: p. 289-310.
138. Development, U.A.f.I., *Digital strategy 2020-2024*. 2020, USAID: Washington, DC.
139. Smith, M.R. and L. Marx, *Does technology drive history?: The dilemma of technological determinism*. 1994: Mit Press.
140. Brynjolfsson, E. and A. McAfee, *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. 2014: WW Norton & company.
141. Polasik, M., T.P. Wisniewski, and G. Lightfoot, *Modelling customers' intentions to use contactless cards*. International Journal of Banking, Accounting and Finance, 2012. **4**(3): p. 203-231.
142. Prayoonphan, F. and X. Xu, *Factors influencing the intention to use the common ticketing system (spider card) in Thailand*. Behavioral Sciences, 2019. **9**(5): p. 46.
143. Camodeca, R. and A. Almici, *Digital transformation and convergence toward the 2030 agenda's sustainability development goals: Evidence from Italian listed firms*. Sustainability, 2021. **13**(21): p. 11831.
144. Karadayi-Usta, S., *Sustainability through digital transformation: EU practices*. Development, 2024. **2**(1): p. 2434.
145. Knott, B. and C. Tinaz, *The legacy of sport events for emerging nations*. Frontiers in sports and active living, 2022. **4**: p. 926334.

146. Barrett, C.R., *The digital evolution*. 2006, Springer.
147. Jukan, A. and J. Mambretti, *Evolution of optical networking toward rich digital media services*. Proceedings of the IEEE, 2012. **100**(4): p. 855-871.
148. Kaggwa, S., et al., *Digital transformation and economic development: a review of emerging technologies' impact on national economies*. World Journal of Advanced Research and Reviews, 2023. **20**(3): p. 888-905.
149. Luo, S., et al., *Digitalization and sustainable development: How could digital economy development improve green innovation in China?* Business strategy and the environment, 2023. **32**(4): p. 1847-1871.
150. *Toyota Robots Help People Experience Their Dreams of Attending the Olympic and Paralympic Games Tokyo 2020*. 2019 [cited 2019; Available from: [https://global.toyota/en/newsroom/corporate/28912712.html#:~:text=500%20seats%20during%20the%20Olympic,16%20rows%20with%2032%20seats\).](https://global.toyota/en/newsroom/corporate/28912712.html#:~:text=500%20seats%20during%20the%20Olympic,16%20rows%20with%2032%20seats).)
151. *NTT's Contribution to Olympic and Paralympic Games Tokyo 2020 - From the perspectives of Telecommunication Services with Cybersecurity -*. 2021 [cited 2021; Available from: [https://group.ntt/en/newsrelease/2021/10/21/211021a.html#:~:text=Presentation%20Slides\(PDF%3A1.80MB,ntt%2Dpr%40ntt.com](https://group.ntt/en/newsrelease/2021/10/21/211021a.html#:~:text=Presentation%20Slides(PDF%3A1.80MB,ntt%2Dpr%40ntt.com).
152. Georgescu, I., et al. *Correlative approach to digitalization and economic growth*. in *Proceedings of the International Conference on Business Excellence*. 2021.
153. Frederico Cruz-Jesus, T.O., F. Bação, Z. Irani, *Assessing the pattern between economic and digital development of countries*. Information Systems Frontiers, 2016. **19**: p. 835 - 854.
154. Mshvidobadze, T., *DIRECT IMPACT OF ICT ON URBAN ECONOMIC GROWTH BY DEMONSTRATING AGENT-BASED MODELING FOR SMART CITIES*. Social Economics, 2025(69): p. 38-44.
155. Kresova, S. and S. Hess, *Identifying the determinants of regional raw milk prices in Russia using machine learning*. Agriculture, 2022. **12**(7): p. 1006.
156. Zhang, Y. and S. Hamori, *The predictability of the exchange rate when combining machine learning and fundamental models*. Journal of Risk and Financial Management, 2020. **13**(3): p. 48.
157. Zhang, Y., et al. *A multi-modal transformer approach for football event classification*. in *2023 IEEE international conference on image processing (ICIP)*. 2023. IEEE.
158. Wen, J. and D. Liang, *Construction of Sports Management System Based on Multi-Layer Neural Network Model*. Applied Mathematics and Nonlinear Sciences, 2023. **8**(2): p. 1809-1818.
159. Widodo, S., H. Brawijaya, and S. Samudi, *Building a Predictive Model for Chronic Kidney Disease: Integrating KNN and PSO*. Paradigma-Jurnal Komputer dan Informatika, 2024. **26**(1): p. 58-64.
160. Mastercard., T.F.S.a.T.U.a. *Digital Evolution Index*. 2020; Available from: <https://digitalevolutionindex.tufts.edu/trajectory>.
161. Chakravorti, B., et al., *Digital in the time of covid*. Trust in the digital economy and its evolution across, 2020. **90**.
162. Pedersen, P.M., *Contemporary sport management*. 2011: Human Kinetics.
163. Tanjung, M., *Can we expect contribution from environmental, social, governance performance to sustainable development?* Business Strategy & Development, 2021. **4**(4): p. 386-398.
164. Biju, A.V.N., et al., *Is national economic growth associated with environmental social and governance performance of firms?* Business Strategy & Development, 2024. **7**(3): p. e70009.
165. Işık, C., et al., *ECON-ESG factors on energy efficiency: Fostering sustainable development in ECON-growth-paradox countries*. Gondwana Research, 2024. **135**: p. 103-115.
166. Zahid, R.A., et al., *ESG, dividend payout policy and the moderating role of audit quality: Empirical evidence from Western Europe*. Borsa Istanbul Review, 2023. **23**(2): p. 350-367.
167. Zhang, C., et al., *Impacts of a mega sporting event on local carbon emissions: A case of the 2014 Nanjing Youth Olympics*. China Economic Review, 2022. **73**.
168. Yin, X. and L. Wang, *Effects and Mechanisms of ESG Performance Affecting Firms' Innovation Efficiency - Based on Regression Analysis Perspective*. Applied Mathematics and Nonlinear Sciences, 2024. **9**(1).

169. Fu, Q., X. Zhao, and C.-P. Chang, *Does ESG performance bring to enterprises' green innovation? Yes, evidence from 118 countries*. *Oeconomia Copernicana*, 2023. **14**(3): p. 795-832.
170. Wan, H., J. Fu, and X. Zhong, *ESG performance and firms' innovation efficiency: the moderating role of state-owned firms and regional market development*. *Business Process Management Journal*, 2024. **30**(1): p. 270-290.
171. Yang, X., et al., *ESG performance and corporate technology innovation: Evidence from China*. *Technological Forecasting and Social Change*, 2024. **206**: p. 123520.
172. Choi, K.H. and B. Leopkey, *The professionalization of a national sport organization as event hosting legacy*. *Sport, Business and Management: An International Journal*, 2024. **14**(5/6): p. 752-773.
173. Kim, C. and K. Kaplanidou, *The effect of sport involvement on support for mega sport events: Why does it matter*. *Sustainability*, 2019. **11**(20): p. 5687.
174. Kaplanidou, K., et al., *Quality of life, event impacts, and mega-event support among South African residents before and after the 2010 FIFA World Cup*. *Journal of travel research*, 2013. **52**(5): p. 631-645.
175. Samuel, S. and W. Stubbs, *Green Olympics, green legacies? An exploration of the environmental legacies of the Olympic Games*. *International Review for the Sociology of Sport*, 2013. **48**(4): p. 485-504.
176. Hayes, G. and J. Horne, *Sustainable development, shock and awe? London 2012 and civil society*. *Sociology*, 2011. **45**(5): p. 749-764.
177. Knott, B., A. Fyall, and I. Jones, *The Nation-Branding Legacy of the 2010 FIFA World Cup for South Africa*. *Journal of Hospitality Marketing and Management*, 2013. **22**(6): p. 569-595.
178. He, D. and X. Chen, *The effect of event prestige on event participation intention: The case of the FISU world university summer games*. *Urban Science*, 2022. **6**(4): p. 90.
179. Winand, M., et al., *Determinants of service innovation: A typology of sports federations*. *International Journal of Sport Management and Marketing*, 2013. **13**(1-2): p. 55-73.
180. Hautbois, C., *From monopoly to competition: The new strategic challenges facing public sports organizations*, in *The Global Sport Economy*. 2019, Routledge. p. 227-273.
181. Sroka, R., *International sporting mega-events and conditionality*. *International Journal of Sport Policy and Politics*, 2021. **13**(3): p. 461-477.
182. Shao, X., X. Ma, and F. Fan, *Network infrastructure, digital technology innovation and enterprise performance: evidence from Chinese listed companies*. *Applied Economics*, 2025: p. 1-15.

Appendix A- Glossary of Terms

Term	Definition
A. Mega-Sport Events & Legacy Concepts	
Digital Legacy	Persistent improvement in national digital infrastructure and capability.
Economic Impact	Sustained economic benefits of event hosting through increased tax revenues, job creation, and ancillary revenues.
Event Hosting Status	Binary indicator identifying whether a country hosted a mega-sport event.
Event Legacy	Long-term outcomes of hosting mega-sport events beyond the event period.
Event Leveraging	Strategic use of MSE investments to maximize long-term benefits.
Innovation Legacy	Long-term enhancement of innovation capability measured through patent activity.
Mega-Sport Event (MSE)	Large-scale international sporting events generating significant economic, social, digital, and innovation impacts.
B. Economic, Social, and Environmental Indicators	
Capital Expenses	Infrastructure investments linked to hosting events.
Environmental, Social, Governance (ESG)	Framework measuring sustainability and institutional quality.
GDP per Capita	GDP divided by population, representing average economic output.
Gross Domestic Product (GDP)	Total value of goods and services produced within a country.
Operational Expenses	Event-related management and operational costs.
Population Size	Demographic variable influencing economic scale and innovation.
Relative GDP Change	Normalized GDP growth indicator before and after events.
Tourist Expenditure	Spending by visitors associated with MSEs.
C. Digital Transformation Measures	
Digital Evolution Score (DES)	Composite index measuring national digital development.
Digital Intelligence Index	Longitudinal dataset assessing global digital maturity.
Digital Readiness	Capacity to adopt and sustain digital technologies.
D. Innovation and Patent-Based Measures	
Innovation Capability	Ability to make major improvements and modifications to existing technology, and to create new technologies that convert knowledge and resources into new products, processes, and services that create value.
Patent Application	Legal request protecting an invention; proxy for innovation.
Research and Development (R&D)	Systematic innovation-oriented research activities.
E. Machine Learning and Statistical Methods	
Bootstrap Forest	Ensemble tree-based ML method using bootstrapping.
K-Nearest Neighbors (KNN)	Distance-based nonparametric ML algorithm.

Term	Definition
Machine Learning (ML)	Algorithms that learn patterns from data.
Neural Boosting	Hybrid neural network and boosting model.
Nonlinear Regression	Regression capturing nonlinear relationships.
Predictor Screening	Method ranking variables by contribution strength.
Stepwise Regression	Automated variable selection regression method.
Supervised Learning	ML approach using labeled data.
Support Vector Machine (SVM)	ML model constructing optimal decision boundaries.
Time-Series Forecasting	Modeling trends and patterns over time.
Extreme Gradient Boosting (XGBoost)	A high-performance ensemble machine learning algorithm that uses gradient-boosted decision trees with regularization to improve prediction accuracy, computational efficiency, and model generalization.
F. Explainable Artificial Intelligence (XAI)	
Explainable AI (XAI)	A set of methods and techniques that make machine learning model outputs transparent and interpretable by explaining how input features contribute to predictions, thereby improving trust, accountability, and decision-making.
Individual Conditional Expectation (ICE)	A visualization method that illustrates how a model's prediction changes for an individual observation as a specific input variable varies, revealing heterogeneity in predictor effects across instances.
Model Interpretability	Degree to which model decisions are understandable.
Partial Dependence Plot (PDP)	A visualization technique that shows the average effect of a selected input variable on a model's predicted outcome while holding all other variables constant.
Shapley Additive Explanations (SHAP)	An explainable AI method based on cooperative game theory that quantifies the contribution of each input feature to a model's prediction by assigning additive importance values.
G. Model Evaluation Metrics	
Generalized R-Squared	Fit measure for nonlinear models.
Mean Absolute Deviation	Average absolute prediction error.
Misclassification Rate	Proportion of incorrect classifications.
Negative Log-Likelihood	Probabilistic model fit metric.
RASE	Root Average Squared Error.
RMSE	Root Mean Squared Error.
R-Squared	Variance explained by a model.
H. Research Design and Data Handling	
Normalization	Scaling data for comparability.
Secondary Data	Data obtained from existing sources.
Zero Imputation	Missing data replacement with zero values.

Appendix B- List of Key Publications

Reviewed Book Chapters

- 1) Manoli, N., Takahashi, M., Matsuura, Y. (2024). A Study on Time Series Approach Applications for Optimized Forecasting of Mega Sporting Events. In : Uden, L., Ting, IH. (eds) Knowledge Management in Organisations. KMO 2024. Communications in Computer and Information Science, vol 2152. Springer, Cham. doi.org/10.1007/978-3-031-63269-3_9

Reviewed Full Conference Paper

- 1) Manoli, N., Takahashi, M., & Matsuura, Y. (2023). Analyzing the Impact of the Sports Event with Machine Learning. In *2023 5th International Conference on Advancements in Computing (ICAC)* pp. 358-363. IEEE. doi: 10.1109/ICAC60630.2023.10417499.
- 2) Manoli, N., Takahashi, M., & Matsuura, Y. (2024). A Study on Machine Learning Algorithms Selection for Estimating the Impact of FIFA World Cups. In *2024 International Research Conference on Smart Computing and Systems Engineering (SCSE)* Vol. 7, pp. 1-6. IEEE. doi: 10.1109/SCSE61872.2024.10550898.
- 3) Manoli, N., Takahashi, M., & Matsuura, Y. (2024). Towards Decoding the Intricate Connection Between Hosting the MSEs and Advancing Digital Infrastructure and Capabilities with Machine Learning. *Procedia Computer Science*, 246, pp. 1670-1679. doi: 10.1016/j.procs.2024.09.650
- 4) Manoli, N., Takahashi, M., & Matsuura, Y. (2025). Optimizing Economic Predictors for Digital Growth in Hosting MSEs with Neural Boosted and KNN Models. In *2025 International Research Conference on Smart Computing and Systems Engineering (SCSE)* pp. 1-6. IEEE. doi: 10.1109/SCSE65633.2025.11030971.
- 5) Manoli, N., Takahashi, M., & Matsuura, Y. (2025). Decoding Innovation Growth: An Explainable AI Analysis of How Mega Sport Events Transform Financial Landscapes and Sustainability Metrics. *Procedia Computer Science*, 270, 2593-2602. doi: 10.1016/j.procs.2025.09.381

Appendix C- Conference Presentations

- 1) Estimating the Impact of Mega-Sport Events from the Bibliographic Database. *2023 Annual Conference of the Electronics, Information and Systems Society*, IEEEJ, Sapporo, Japan, September 29, 2023
- 2) Analyzing the Economic Impact of Sports Events with a Machine Learning Approach. *JSAI Special Interest Group on Business Informatics (23rd SIG-BI)* Tokyo, Japan, September 23-24, 2023
- 3) Exploratory Research on Machine Learning Algorithms for Estimating the Effects of Mega-Scale Sporting Events. *JSAI Special Interest Group on Business Informatics (24th SIG-BI)*, Okinawa, Japan, March 9-11, 2024
- 4) Exploring the Connection Between Economic Leverage and Digital Advancement of Hosting MSEs with Machine Learning. *The 4th AAGBS International Conference on Business Management (AiCoBM)*, Malaysia, August 12-13, 2024
- 5) Measuring Innovation Returns from Mega-Sport Events: A Machine Learning Analysis of Financial and ESG Factors. *JSAI Special Interest Group on Business Informatics (26th SIG-BI)*, Okinawa, Japan, March 15–17, 2025