Cricket Team Selection Based on Complex Dynamics Using Machine Learning

A Thesis Submitted for the Degree of Master of Computer Science

W.M.H.G.T.C.K. Weerakoon

University of Colombo School of Computing

2024

DECLARATION

Name of the student: W.M.H.G.T.C.K. Weerakoon
Registration number: 2019/MCS/094
Name of the Degree Programme: Master of Computer Science
Project/Thesis title: Cricket Team Selection based on Complex Dynamics using Machine Learning

1. The project/thesis is my original work and has not been submitted previously for a degree at this or any other University/Institute. To the best of my knowledge, it does not contain any material published or written by another person, except as acknowledged in the text.

2. I understand what plagiarism is, the various types of plagiarism, how to avoid it, what my resources are, who can help me if I am unsure about a research or plagiarism issue, as well as what the consequences are at University of Colombo School of Computing (UCSC) for plagiarism.

3. I understand that ignorance is not an excuse for plagiarism and that I am responsible for clarifying, asking questions and utilizing all available resources in order to educate myself and prevent myself from plagiarizing.

4. I am also aware of the dangers of using online plagiarism checkers and sites that offer essays for sale. I understand that if I use these resources, I am solely responsible for the consequences of my actions.

5. I assure that any work I submit with my name on it will reflect my own ideas and effort. I will properly cite all material that is not my own.

6. I understand that there is no acceptable excuse for committing plagiarism and that doing so is a violation of the Student Code of Conduct.

Signature of the Student	Date (DD/MM/YYYY)
	20.08.2024

Certified by Supervisor(s)

This is to certify that this project/thesis is based on the work of the above-mentioned student under my/our supervision. The thesis has been prepared according to the format stipulated and is of an acceptable standard.

	Supervisor 1	Supervisor 2	Supervisor 3
Name	Dr. H.K.T.C. Halloluwa		
Signature			
Date	24.08.2024		

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Mr. Saman Jayantha, Head Coach of Sinhalese Sports Club and former Sri Lankan Cricketer, for his invaluable contribution to this research. His expertise, insights, and unwavering support have played a pivotal role in shaping the direction of this study. Mr. Jayantha's willingness to share his wealth of knowledge in cricket, coupled with his guidance on various aspects of team dynamics and player performance, has been instrumental in the success of this research. I am truly thankful for his mentorship and the time he dedicated to providing valuable perspectives that enriched the quality of this work.

ABSTRACT

Cricket is a sport which has a history, well established governing body and an economy built around. The selection of a winning team is a crucial measure, so the selection committees of teams take extensive efforts to formulate the winning team combination in terms of batsmen, bowlers and allrounders. Usually the selection process is meticulous and will be biased. Hence, this study mainly focused to propose a data driven approach where historic match data under complex dynamics being used and analyzed for an optimal team combination selection in cricket.

The data for the study is secondary data scraped from cricket statistics repositories and data pertaining to the One Day International matches of the Sri Lankan cricket team since 2009. The data were feature engineered before conducting multi-output, multi-class classification tasks and Fuzzy Logic inferencing, and during feature engineering the pitch condition was derived using historic pitch report data.

A neural network with 7 dense layers was designed which used 7 input features; stadium, stadium type, minimum and maximum temperature, average wind speed of the stadium, and the winner of each match and classify into three output the number of batsmen, bowlers and allrounders.

The neural network yielded 76% of performance under 80:20 split on training and testing data for 100 epochs. Three Fuzzy Inference Systems were developed to rate and rank players based on historic performances. These inferencing systems have yielded 75%, 67% and 62% accuracies for Batting Performance FIS, Bowling Performance FIS and Allrounder Performance FIS respectively. This domain inherits data imbalance problem and there are unmeasurable attributes such as psychological, physiological and political aspects, deliverables of the study can be considered as a decision support models for cricket team selection. In future, more empirical methodologies need to be carried out to increase the performance of the neural network model and Fuzzy Logic Inference Systems in terms of a more adaptable model for squad selection in cricket.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS 2
ABSTRACT 3
CHAPTER 1 7
INTRODUCTION 7
- **1.1. Cricket and Technology** 7
- **1.2. Statement of the problem** 7
- **1.3. Research Aims and Objectives** 8
  - **1.3.1. Aim** 8
  - **1.3.2. Objectives** 9
- **1.4. Scope** 9
- **1.5. Structure of the Thesis** 10

CHAPTER 2 11
LITERATURE REVIEW 11
- **2.1. Cricket Match Format Comparison** 11
- **2.2. Predictive Analytics in Sports** 12
- **2.3. Team Selection** 13
- **2.4. Machine Learning Techniques for Team Selection** 15
  - **2.4.1 Multi-objective Evolutionary Algorithms** 15
  - **2.4.2 Fuzzy Logic** 16
  - **2.4.3 Other Supervised Learning Algorithms** 17
- **2.5. Summary** 18

CHAPTER 3 19
METHODOLOGY 19
- **3.1. Fundamental Approach** 19
- **3.2. Design Considerations** 20
- **3.3. Model Architecture** 21
  - **3.3.1. Data Acquisition** 22
  - **3.3.2. Data Acquisition from ESPNCricinfo, World Weather Online and Sports F1** 24
  - **3.3.3. Data Preprocessing on Player Performance Data** 33
  - **3.3.4. Feature Engineering on Complex Dynamics Data** 34
  - **3.3.5. Designing the TCS Model** 35
  - **3.3.6. Designing the Player Rating and Ranking System** 37
- **3.4. Summary** 40

CHAPTER 4 41
EVALUATION AND RESULTS 41
- **4.1. Experimental Results for TCS modelling** 41
- **4.2. Experimental Results for Player Rating and Ranking modelling** 44
  - **4.2.1. Batting Performance FIS** 46
  - **4.2.2. Bowling Performance FIS** 50
  - **4.2.3. Allrounder Performance FIS** 54

4.3. Summary **58**
CHAPTER 5 **59**
CONCLUSION AND FUTURE WORK **59**
5.1. Predicting Team Composition **59**
5.2. Player Rating and Ranking Model **60**
5.3. Future Prospects **61**
5.4. Summary **62**
APPENDICES - I **63**
APPENDICES - II **67**
APPENDICES - III **69**
REFERENCES **72**

LIST OF FIGURES

Figure 1. Highlevel system workflow 22
Figure 2. Web Scrapping Overview 23
Figure 3. Sample player profile from ESPNCricinfo. 26
Figure 4. Batting performances listed in the site 27
Figure 5. Bowling performance listed in the site 29
Figure 6. World Weather Online Site 30
Figure 7. Sport F1 website for pitch reports 32
Figure 8. Layer Specifications of the Final TCS Model 37
Figure 9. Developed FIS for Batting, Bowling and Allrounder Performance Ranking System 38
Figure 10. Input Membership Functions Distribution for Initial FISs 44
Figure 11.Output Membership Functions Distribution for Initial FISs 45
Figure 12. Initial Validation Performance Batting Performance FIS 46
Figure 13. Input membership distribution after Betting Performance FIS tuned 48
Figure 14. Output membership distribution after Batting Performance FIS tuned 49
Figure 15. Validation plots of Tuned Batting Performance FIS, RMSE = 0.25 49
Figure 16. Initial Validation Performance Bowling Performance FIS 50
Figure 17. Input membership distribution after Bowling Performance FIS tuned 52
Figure 18. Output membership distribution after Bowling Performance FIS tuned 53
Figure 19. Validation plots of Tuned Bowling Performance FIS, RMSE = 0.33 53
Figure 20. Initial Validation Performance Allrounder Performance FIS 54
Figure 21. Input membership distribution after Allrounder Performance FIS tuned 57
Figure 22. Output membership distribution after Allrounder Performance FIS tuned 57
Figure 23. Validation plots of Tuned Allrounder Performance FIS, RMSE = 0.38 58
Figure 24. Distribution of Team Combinations 60
Figure 25. Correlation Analysis for Batting Performance Data 69
Figure 26. Correlation Analysis for Bowling Performance Data 70
Figure 27. Correlation Analysis for Allrounder Performance Data 71

LIST OF TABLES

Table 1. Considered 34 players for the research 25
Table 2. Attributes extracted from player profiles 26
Table 3. Attribute description of patting performance data 28
Table 4. Attribute description of bowling performance data 29
Table 5. Weather data attributes 31
Table 6. Pitch report attributes 32
Table 7. Input Feature Parameters for FIS 38
Table 8. Output Feature Parameters for FIS 39
Table 9. Performance Accuracies considering individual classification for no. of batsmen, bowlers and allrounders 41
Table 10. Neural Network Model performances for Multi-ouput and Multiclass classification 42
Table 11. FIS specifications of Tuned Batting FIS 46
Table 12. FIS specifications of Tuned Bowling FIS 51
Table 13. FIS specifications of Tuned Allrounder FIS. 55

CHAPTER 1

INTRODUCTION

This chapter provides context on the study and further describes the background pertains to the study, approaches used for the study, the significance of the study and the main motivations for the study to conduct. In addition, this chapter describes the intended objectives to be achieved, research questions and the scope of the study along with an overview of proposed methodology.

1.1. Cricket and Technology

Cricket is a globally popular sport with its origins in England during the 16th century, evolving from a children's game to an international phenomenon. The first international cricket match took place in 1844, and the official International Test match occurred in 1877. The Marylebone Cricket Club (MCC) initially governed the sport's rules, but the International Cricket Council (ICC) was established to oversee the game globally. Cricket is played with two teams of 11 players each, taking turns batting and bowling in innings, aiming to score the most runs. The ICC has standardized rules for cricket, encompassing various formats such as Test matches, One Day Internationals (ODIs), and T20 matches. The game is played on an oval field with a rectangular pitch, and there are specific rules for boundaries. Players are categorized as batsmen, bowlers, and all-rounders based on their skills in batting, bowling, and fielding. Selection committees of cricket boards choose players for national teams.

Technology has significantly influenced cricket, creating an industry around the sport, with leagues like the Indian Premier League (IPL) gaining immense popularity. Stakeholders, including fans and sponsors, are deeply invested in their supported teams, and strategic decisions play a crucial role in team selection. Overall, cricket has a rich history, diverse formats, and a global appeal that continues to captivate audiences worldwide.

1.2. Statement of the problem

The problem statement revolves around leveraging machine learning techniques to enhance the process of selecting cricket teams. Traditional methods of team selection often rely on subjective assessments and historical performance data. However, cricket involves various complex dynamics, including individual player performances, historical match statistics, environmental conditions, and strategic considerations, that contribute to the overall success of a team.

The goal is to develop a sophisticated machine learning model that can analyze and understand these multifaceted dynamics. The model should take into account individual player metrics such as batting averages, bowling figures, and fielding statistics. Additionally, it should consider historical match data, including team and player performances in various conditions, against different opponents, and in different formats such as Tests, ODIs, and T20s.

Moreover, playing conditions such as pitch characteristics, weather conditions, and venue-specific factors should be factored into the model. Strategic elements, like player compatibility, opposition analysis, and team balance, should also be incorporated to ensure a comprehensive understanding of the selection process.

The objective is to create an intelligent system that, based on historical and real-time data, can predict the optimal combination of players for a given match or series. This system should provide insights to selection committees, aiding them in making informed decisions that maximize the team's chances of success.

In summary, the project aims to develop a machine learning-based cricket team selection system that goes beyond traditional approaches, considering a wide range of factors that influence team performance. This holistic approach is expected to contribute to more effective and data-driven decision-making in the dynamic and competitive world of cricket.

1.3. Research Aims and Objectives

1.3.1. Aim

The principal aim of this research is to design and develop a scientific framework that supports cricket officials in Sri Lanka in making informed team selection decisions. The framework aims to utilize sports analytics, machine learning ideas, statistics, and data mining approaches to assess and rank players based on batting, bowling, and wicket-keeping performance. The overarching goal is to enable the optimal selection of a 15-member squad from a pool of 25-30 players for specific cricket series. The selection process will consider various cricket-related factors, including the characteristics of the opposing team, match venue, individual player historical performance, and current form. Additionally, the system aims to categorize players into 'all-rounders' and 'non-all-rounders' and, based on tournament or match characteristics, determine the best team combination (best 11 players) for the upcoming match. This involves considering the ratio of batsmen, all-rounders, and bowlers (including spinners and fast bowlers) required, as well as other factors such as the characteristics of the opponent team. The project acknowledges the evolving nature of cricket, with rule modifications and new features contributing to what is termed as the "Complex Dynamics" of the sport.

1.3.2. Objectives

1. Evaluate and rank players in the pool based on their individual batting, bowling, and wicket-keeping performances.
2. Establish criteria for selecting the best 15-member squad from the player pool, considering factors such as opposing team strength, match venue, historical player performance, and current form.
3. Analyze the characteristics of the given tournament or match to determine the most suitable team combination (best 11 players), factoring in the required ratio of batsmen, all-rounders, and bowlers (including spinners and fast bowlers).

1.4. Scope

This research study focuses on ODI match format out of the 3 match formats defined by ICC. The test match format played for 5 days and the complex dynamics of this context is highly complex due to many uncontrollable changes in some external factors, such as weather. Since, once a team is fixed for the entire 5 days of the match, there will not be change in the player composition. Hence, fundamentally use of predictive analytics can be in vain for test match format in cricket, but it should be noted that team selection using predictive analytics for test matches are viable. In the context of the T20 match format, it is much more dynamic in nature than that of an ODI match format. The game is carried out in a considerably less time period and the rate of change in variables is high. For instance, the match flow can be disrupted when the umpire calls for a free hit. Another reason for choosing the ODI format is that the number of players who are playing ODIs is more than that of T20 or Test. These reasons provide a great opportunity to select the ODI format for this research purpose. Further, in this research it is considered the limited number of pools of Sri Lankan cricketers. The data about their individual batting and bowling performances will be considered and to evaluate the performance in fielding, the number of catches that a player secured will be considered. But it should be noted that the specified variable is fielding position dependent.

1.5. Structure of the Thesis

Chapter 2: Provides the background study on sports analytics and literature review on how predictive analytics techniques have been used in domain of cricket. Further reviews about the use of predictive analytics within the research domain.

Chapter 3: Discusses the research design and the proposed methodology that intended to follow to meet research aims and objectives.

Chapter 4: Presents the obtained results after following the proposed research methodology and discusses the results in summary.

Chapter 5: Concludes the entire dissertation with the concluding remarks and stating the limitations of the study along with the potential future works for this research study.

CHAPTER 2

LITERATURE REVIEW

The purpose of this chapter is to present some of the prior research projects and findings linked to this topic. The first section of this chapter covers cricket history and how it spread around the world, general rules and terminology, Sri Lankan cricket history, and milestones. After then, this chapter moves on to prior methods for delivering various predictive metrics connected to cricket and players. At the end of this chapter, existing machine learning-based algorithms for sports predictive analytics are outlined.

2.1. Cricket Match Format Comparison

Cricket is an elaborated, comprehensive sport governed by ICC internationally. According to ICC there are 12 main countries who participate in international level cricket. The 12 countries are, Australia (1909 - Founding member), England (1909 - Founding member), South Africa (1909 - Founding member), India (1926), New Zealand (1926), West Indies (1926), Pakistan (1952), Sri Lanka (1981), Zimbabwe (1992), Bangladesh (2000), Afghanistan (2017) and Ireland (2017). In addition to these 12 teams there are other associate cricket playing countries having eligibility only for ODI membership in ICC. Cricket has 3 main match formats, such as Test match format, ODI match format and T20 match format. These different formats have a uniquely defined set of rules that cricket teams should adhere to. When considering test match format, the total number of overs is unlimited, typically on average 76-82 overs will be delivered. Test matches consist of 4 innings per match and an inning lasts until all the players of the batting team bowled out. Then the opposition team gets to bat and the same analogy will be repeated for the other two innings. The batting team captain can declare to close an inning before all the players get bowled out. For test match formats, players should have more endurance, as the match is going to last for 5 days. Test match formats have no limitation whatsoever in number of overs per bowler, powerplay, fielding constraints etc. In the context of the ODI match format, there are only 50 overs per team, and do not need much endurance from players [1], [2]. During this match format, a bowler gets to bowl only 10 overs per inning and there are 3 power plays; powerplay 1: for first 10 overs and powerplay 2 and 3 for 5 overs each. During the first 10 overs: 3 fielders outside the 30-yard circle. For the next 30 overs 4 fielders need to be outside the 30-yard circle. For the last 10 overs, 5 fielders should be outside the 30-yard circle. For the T20 match format, there are only 20 overs per inning and was introduced in 2003. This match format is considered to be more dynamic, enjoyable and youth-centric. 4 overs will be granted to each bowler within an inning and there is a powerplay. This match format requires only 3-4 hour time duration to complete an entire match. But the 20th over should be delivered within the first 75 minutes. If not, the batting team gets extra 6 runs. If the match gets tied, a super over will be offered to both teams. The team which scored the most runs is considered as the winner. There can be seen International as well as domestic fixtures like Indian Premier League (IPL), Sri Lanka Premier League (SLPL) etc. When closely examining the match formats, there can be seen some distinctive features/ attributes that affect the victory of the particular match type. Test match format and ODI format require endurance of players due to the long duration of the match, and meanwhile T20 match format requires players with agility and rapid striking rate. The concept of powerplay in ODI and T20 match formats is also a drastic feature which is a critical contributing concept for winning a match. Powerplay was introduced in limited overs match format, i.e., ODI and T20 to increase the strike rate of batsmen, because batsmen find it difficult to get accustomed to collecting most of the runs in a limited number of overs. This is mainly due to their practiced defensive nature for test matches [3]. Silva, Manage and Swartz, [4] in 2015 explain that during the powerplay teams need to consider the batting style: whether to be active or passive.

2.2. Predictive Analytics in Sports

Predictive analytics is a concept that is used in a wide variety of domains. With recent technological advancements, the domain of sports has also been considered for adopting predictive analytics. Hence, researchers have been researching on different aspects how predictive analytics can be applied to sports. With such initiation many national teams in western countries, big sports clubs incurred capital investments to use predictive analytics to gain powerful insights of both the game and players through analyzing data using statistical, machine learning and computational methods. Moreover, some people, as stakeholders, show interest in what are the odds of a victory in a particular game/ match for their supporting team. Investors, team managers, and national sports authorities have long desired a clear picture of their team's performance, as well as the ability to identify amateur players with the potential to become stars in the near future, find optimal player placements within teams, create the best team combinations of players, identify each player's strong and weak areas and improve accordingly, and rank players in the team. The end consequence would be that the individual teams would have more opportunities to win. When predictive analytics are used in sports, it becomes sport analytics and it has numerous potential applications. Many team sports such as football, baseball, cricket are using sport analytics and moreover, can be seen in sports such as athletics, badminton, tennis where individuals are performing. The study conducted by Tavana et al. [5] discusses how a fuzzy inference system is used in player selection and team formation in football. It is stated that the study had conducted in two phase: the first phase uses a fuzzy ranking method to evaluate alternative players and select the best performers for inclusion in the team, while the second phase uses a Fuzzy Inference System (FIS) to evaluate alternative combinations of the selected players and select the best arrangements for a 4-4-2 team formation. The proposed method is demonstrated using real data from the Parsan1 Soccer Club (PSC), an Iranian professional soccer team situated in Tehran. In the study, the linguistic variables used to evaluate the players' performance on numerous criteria are transformed using fuzzy sets. To deal with the difficulty of describing players' skill levels and performance ratings with discrete values, linguistic variables are used. Finally, they offer a team structure of Defenders (5, 2, 7, 1), Mid Elders (12, 14, 13, 8), and Forwards (18, 20) according to their system, which the respective coaching staff of the team approved.

2.3. Team Selection

With time, cricket has created an industry, instead of being a mere sport. With that being the context, there are a variety of stakeholders, who are keen to invest in the sport at many levels. According to Ahmed, Jindal and Deb, in 2011, [6] with the introduction of limited over match formats, many leagues have been introduced. These leagues are highly monetized and merchandised. Furthermore, the investor of a team which plays for a league, aims to win the league. Therefore, basically team selection for a match is considered as a crucial aspect. Not only leagues, but matches fixed by ICC require team selection and usually the responsibility of team selection is delegated to the Selection Committee of a particular country. Selection committee is appointed by the National Cricket Board of a particular country. Furthermore, researchers explain the types of parameters are considered as vital for team selection. It states that parameters related to batting performance, bowling performance and fielding performance should be taken into consideration when selecting players to a team. Moreover, the authors of the article had mentioned a list of performance measurements that needed to be focused for each performance. No. of matches played, total runs made, the strike rate, total no. of wickets taken, no, of maiden overs delivered are some such statistical parameters that researches have used during their study and researchers predominantly considered the team selection for T20 match format. Further, the researchers had defined performance measurement matrices for batting, bowling and fielding. For batting, it is the ratio between total runs that a batsman scored against the number of times that particular player gets bowled. They termed the ratio as batting average. The bowling average is considered as the performance measurement for identifying bowlers and equated as the ratio between total runs conceded against number of wickets taken. The fielding performance is measured using the ratio between total catches to total number of innings that the player played and called the ratio player fielding performance measure. Authors further elaborate that these criteria are defined by the selection committee considering the format of the match, skill levels of the opposition team and ultimate goal of winning the match. Preston and Thomas, 2000 [7] has shown the importance of selecting the strategically optimal team for winning of limited overs match formats. The authors had treated the team selection as a stochastic dynamic programming problem and formulated a scientific approach for team selection in terms of batting skills. Accordingly, authors discussed that for the two innings there should be two different strategic approaches of batting, i.e., if you bat in the first inning, strategy is to score as many runs as possible and if you bat in second inning, be strategic to endure and chase the run. So simply any team selection should be considered to have a batting lineup of high to low striking rates in first innings and for the second innings, a lineup of players with fluctuating strike rates. Further, researchers state that there are several other situational parameters that should be considered, because after all cricket is a highly dynamic and uncertain context to predict. Amin and Sharma, 2014 [8] in their study focus on team selection for T20 match format. In their study they have mentioned that selecting a team is not merely trying to find a combination of players who can do batting, bowling and fielding, but to find the best mix of captaincy, batsmen, bowlers, all-rounders and a wicket keeper. Moreover, according to the literature review, the information coaches use to make team selection decisions: a scoping review and future recommendations, by Fiander et al. [9] review that coaches use age and experience, skills, physical characteristics and stakeholder perspective for a team selection. Also, Trninic et al., [10] state that for a comprehensive team selection, the selection committee (especially coaches) should consider the overall potential and quality of players. Study also explains that adequate sport specific motor skills, optimal body structure, specific cognitive capacity and technical - tactical experience are essential aspects to consider when selecting players for a team. Hence, the factors which are considered for a team collection are very dynamic, comprehensive and complex.

2.4. Machine Learning Techniques for Team Selection

2.4.1 Multi-objective Evolutionary Algorithms

According to Pérez-Toledano et al., [11] in 2019 team selection in the domain of sport is viable and feasible. They have used multi-objective evolutionary algorithms to optimally select a team of players for basketball. Here, the researchers have used Non-dominated Sorting Genetic Algorithm II (NSGA-II), which has stochastic search methodology and imitation of natural biological evolution. This same algorithm has been used by Ahmed, Deb and Jindal, [12] in 2011 to formulate decision making in team selection in cricket. Researchers have used the NSGA - II evolutionary algorithm to optimize decision making in overall batting, bowling and fielding and mainly focused on T20 match format in cricket. Furthermore, researchers claim that this proposed approach of use of evolutionary algorithms has yielded a plethora of high-performing team choices, rather than having a single optimal team. Furthermore, this study states that, because of the use of evolutionary algorithms, the selection authorities could exploit the opportunity of identification of key players within the squad. Moreover, Sarda, Sakaria and Deulkar, 2015 [13] have used the improved version of NSGA-II, the INSGA-II in context of football and researches have concluded that the same improved algorithm can be used for T20 format of cricket with slight modifications. In the study Applying Genetic Algorithm to Select an Optimal Cricket Team by Sathya and Jamal, [14] in 2009 discusses the use of Standard Genetic Algorithm (SGA) for the optimal team selection for cricket. The researchers have considered the natural evolution principles when formulating the fitness function for the algorithm and focused on the fact that optimality changes with the change of gene composition. Here, the researchers have used evolutionary approaches to identify the most effective and diverse player combination for the team. Kusumsiri and Perera, [15] in 2017 have used performance measures, such as batting average, bowling average, number of wickets taken as predominant features to demarcate the optimal team. M.AqilBurney et al., [16] have proposed an approach which consists of consideration of player personal performance, team performance and players. The researchers have opted to use Island Genetic Algorithm, hence, they have claimed that their approach is suitable for any multiplayer sport. The main reason that they state of being their methodology is generic is that the island genetic algorithm has the ability to resolve multiple conflicting constraints using mixed crossover treatment

2.4.2 Fuzzy Logic

Fuzzy logic application in cricket, predominantly revolves around the player performance evaluation [17]. Researchers have conducted formulation of fuzzy systems with eight inputs (runs scored, balls faced, strike rate, out, amount of fours, sixes, team strength, and opponent team strength), and is utilized to produce the output (ranked). All of these input variables have an impact on a player's ranking. For this fuzzy system, they devised 96 rules. Except for out, which has two membership functions, each input has three membership functions. The biggest disadvantage of this technique is that it does not take into account the relative performances of other players in the same match. It just shows the percentage change in performance as well as the influence on ranking. Furthermore, this study exclusively examines a cricket player's performance in terms of batting. According to this study, it should be able to pick the top 15 players from a pool and rank them. In addition, this approach does not address how to choose the best team composition. Features such as the present performance, the venue, and performance history data should all be considered. The concept of using a fuzzy inference system for player rating, on the other hand, can be researched further as a suitable beginning point for a better and new solution. Furthermore, Premkumar, Chakrabarty and Chowdhury, [18] in 2020 explain that fuzzy logic can be applied for player ranking in a more dynamic approach, considering the factor analysis. Here, it discusses the comparative performances within the team, location impact, pitch impact, batting innings impact, opposition impact, team impact, strike rate impact is considered for batting skill assessments. Moreover, for bowler ranking wickets, economy of bowling, innings impact, strike rate impact, pitch impact, opposition impact was considered. As per the researches the main limitation of the study was that not having high dimensionality for the dataset. Accordingly with some variables the player ranking could be robust and with some other set of variables it will not be. Fuzzy logic has been used for team selection in other sports, predominantly in multiplayer sports. According to Sałabun et al., [19], objective fuzzy inference systems can be utilized for team selection in multiplayer sports. Here researchers have conducted the study for football and a multi-criteria model has been developed known as COMET (Characteristic Objects Method), thereby to evaluate position of the player. Further, researchers have mentioned the use of multi-criteria decision analysis (MCDA) model, Technique of Order Preference Similarity (TOPIS) method. Researchers further claim that with minimal changes, this approach can be used for player performance analysis in other multiplayer sports.

2.4.3 Other Supervised Learning Algorithms

Subramanian R. I. and Ramesh S., [20] proposed a method based on artificial neural networks, to rate players and select specific players for a competition. They use neural networks to forecast each cricketer's future performance based on their previous performance. They recommend cricketers to be included in the World Cup 2007 based on the ratings given and by applying heuristic methods. The feature list they utilize to develop NN's is insufficient to include a few critical parameters that affect player performance, which is a notable flaw in this technique. Players' present form, past performances, and the opponent's country are not considered. To validate the applicability of NNs, the suggested approach was examined using real-world performance of cricketers during the 2007 World Cup. The findings demonstrate that NNs can be a beneficial decision-making tool in a team selection process. Shetty et al., [21] in 2020 discuss the use of concepts and techniques of machine learning for team selection. In this study researchers have considered using a data driven approach to compute player performances and then to carry out selection with respect to ODI match format. The study had categorized players with the aid of random forest classifier, support vector machine algorithms. The produced model via the study has yielded accuracies of 76%, 67% and 95% for batsmen, bowlers and all-rounders respectively. Further, the study has considered non player centric parameters such as weather and pitch condition. Moreover, researchers discuss that parameters such as number of maiden overs, number of remaining overs, number of wickets left of previous games could be used to predict the potential team for the future team. In the study of A Survey on Team Selection in Game of Cricket Using Machine Learning by Punjabi et al., [22] uses supervised learning technique to recognize the most effective set of players considering the performance contribution offered by players in different skills of cricket. Moreover, researchers explain that in order to be an effective batsman, bowler and a fielder, a player should score maximum runs, take maximum wickets and should stop scoring runs by the opposition team respectively. This study has used Naive Bayes algorithm to produce predictions. Al-Shboul et al., [23] in 2017 consider the use of neural network approach to select optimal set of players for a sport.

The literature has shown that there are potential limitations within each research. When highly focusing on the variable selection for team selection shows inevitable error-prone nature. Nor researcher has put efforts to identify a scientific approach for determining the factors which contribute to optimal team selection. Further, only few researches have been conducted for the use of fuzzy logic in team selection in cricket.

2.5. Summary

At the beginning of this chapter, it has discussed the nature of cricket with respect to 3 different cricket match formats that had been introduced by ICC. The intention was to make sure that there is enough awareness about the uniqueness of each match format. Hence, based on the match format, the selection of an optimal team changes and there are many factors to consider. These factors were termed as "complex dynamics". Then, it is more elaborated how a typical team gets to select without the intervention of a computer and further discusses how different researchers have considered different factors as team selection criteria. Finally, a review was carried out to understand how machine learning has contributed to sport analytics. There the review clearly elaborated under three topics: multi-objective evolutionary algorithms, Fuzzy logic and other supervised learning algorithms that can be used for team selection. Chapter 3 is about the proposed methodology, architectural view and experimental setup of the research study.

CHAPTER 3

METHODOLOGY

This chapter outlines the suggested approach to answering the research questions by in-depth explaining the study's design elements and architectural perspective. Additionally, the experimental design of this study will be described here. Here, the major goal is to offer a solid technique that can result in accurate, dependable analytics for the Sri Lankan cricket team in one-day internationals. The main underlying presumptions and other design issues are then discussed. A systematic and thorough evaluation of the literature revealed that no previous research has yet completely satisfied the goals of this study. Important elements of those methodologies will be used as a starting point for this study's goals, and when appropriate, fresh but efficient and accurate approaches will be included. The experimental setting, data set, tools, and programming languages used in this study will all be covered in more detail later on in this chapter.

3.1. Fundamental Approach

The primary objective of this study is to give the Sri Lankan cricket team with trustworthy analytics for One Day International cricket matches that will serve as a decision support system for team administration, coaches, captains, etc. Under that, we seek to address the finding a method for identifying complex dynamics that impact the fixture and finding a winning ratio of batsmen: allrounders: bowlers ratio based on the complex dynamics identified. Fuzzy logic, statistical techniques, and neural networks are emphasized as the best ways used to implement these three challenges, respectively, among the examined approaches. As discussed in the previous chapter, the findings in existing literature depicts some limitations on the performance level of existing prediction models. For instance, it was found that evaluators often based on fuzzy logic are constructed to rate players according to scores. These ranking systems differ from one another in the way they choose their fuzzy rules and features. But for a variety of reasons, none of the current ranking systems can fully achieve this goal. Some of them severely diminish overall accuracy by leaving out or neglecting numerous significant elements that should have been taken into account [7,8,16]. It should be noted that these predicting approaches are trained upon old, historic data, which raises the validity and applicability concerns due to new rules and regulations imposed by ICC on ODI match format. Because of "power play" overs, for instance, 280–300 runs are no longer considered a "high score," and seeing a player play a score of 150 or more is not uncommon. The current systems must occasionally be modified because of changes to the rules. Therefore, most of these assessors are currently obsolete. In order to implement the best 15 player selection from a team of players, we are careful to identify the majority of the key influencing elements. We also consider the following alternatives while announcing the regulations. Further, many other external factors determine the victory of a team, predominantly weather and playing pitch. Another difficulty is that the winning team combination problem hasn't been directly addressed in any of the sports literature that has already been published. But ANN was chosen to address the issue after analyzing the body of literature and merging current approaches. Predictive models have frequently been constructed using neural networks. It is the intention to find answers to research issues based on the earlier work with strong foundations and in cooperation with the unique methodologies suggested in this study. In order to create an ideal team using fuzzy logic and other interpretable machine learning and statistical techniques, this study proposes a thorough methodology for identifying complex dynamics. This methodology will be used to identify complex dynamics, apply them to cricket player composition selection, and produce an optimal team. It is thought that the aforementioned methodologies will be well adapted to the goals of this research because there is a considerable amount of data, including player, match, weather, and pitch data. Additionally, based on the information gleaned from the literature research, the proposed methodology makes an effort to include data sources and additional methodologies.

3.2. Design Considerations

Since the domain of the predictive model is in a sports related environment, there are several concerns that can be identified prior to planning the actual architecture of the model. According to the findings in literature and expert ideas, below are some key points that must be considered when designing prediction models for this research.

- Since the deliverable of the study is being a predictive model for team composition in ODI format with player ranking rules, it must be noted that there is feasibility of developing a decision support system for team selector in cricket.
- Identifying the key characteristics of individuals that can influence the best 15-man SL team roster. When creating the appropriate models, each player's significant characteristics that fall under the categories of batting, bowling, and allrounding performances should be taken into consideration.

- Further, optimal team composition needs to be selected considering the factors such as weather condition and pitch condition etc.

3.3. Model Architecture

This section discusses the architectural and functional design of the optimal team selection system which uses complex dynamics. Proposed methodology basically consists of 2 sub-systems namely,

1. Complex dynamics identification module and best composition selection module
2. Player ranking module

The proposed architecture of experiment is depicted in Figure 2.

As shown in Figure 2, scrape data from ESPN Cricket (https://www.espncricinfo.com/), World Weather Online (https://www.worldweatheronline.com/cricket.aspx) and Sport F1 (http://sportsf1.com/) stored inside a temporary database prior to data preprocessing. After data preprocessing, the preprocessed data analyzed for feature engineering using several feature engineering techniques. Then, the finally prepared data is stored in a designated database. From the database the data oriented for player performance with several complex dynamics used to develop the fuzzy system and the rest of the reduced dataset which includes data related complex dynamics utilized for the optimal team composition classifier development. Finally, the overall system produced the optimal player mapping while considering complex dynamics involved in cricket.

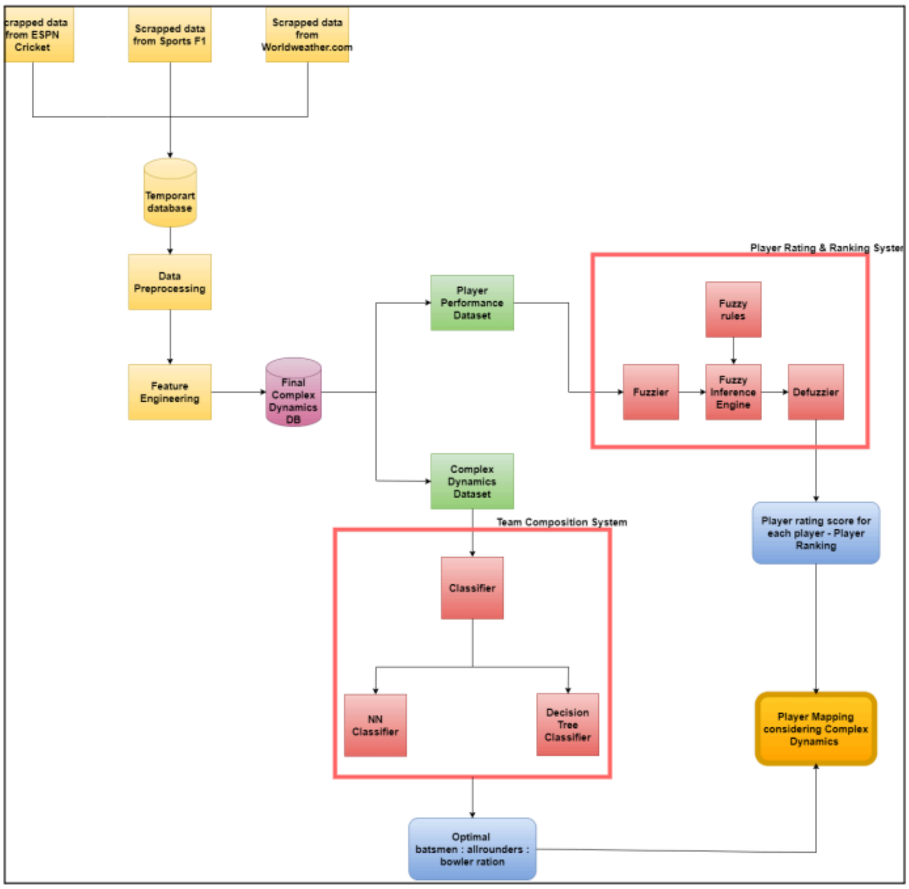


Figure 1. Highlevel system workflow

3.3.1. Data Acquisition

This research study requires a variety of cricket related data from individual player performances for each cricket skill to weather data, pitch condition data and opponent team data. Gathering all this data from a variety of sources, storing data and managing data is a crucial task. Therefore, finding a good and useful technique was vital to making use of these data. Additionally, a number of web scrapers were developed in order to collect pertinent data into one particular database because different data is contained in different URLs. After then, the initial database was updated by adding new associations between entities. The written automated software application was used to collect data, which were then put into a database. The method we utilized to gather the data, known as "Web Scraping Technology," will be briefly explained in this part, along with examples of how it might be applied in actual situation

3.3.1.1 What is Web Scraping?

The World Wide Web (WWW) can be considered as a fountain of data, as it contains and generates huge amounts of data. Even though WWW holds a lot of data, the immediate machine readability capacity is very low. Web scrapers are conventional mechanisms that extract data from web documents. Web scrapers refer to the document structure of a web page and extract data within the web page structure. Hence, web scrapers are such computer programs typically mimic human web browsing by implementing low-level Hypertext Transfer Protocol (HTTP) or by incorporating a fully-featured web browser, like Mozilla Firefox. Web scraping also concentrates more on structuring unstructured data from the web, often in HTML format, so that it may be kept and used for analysis in a central local database or spreadsheet. Online price comparison, contact scraping, weather data monitoring, website change detection, research, web mashup, and web data integration are among the uses of web scraping.

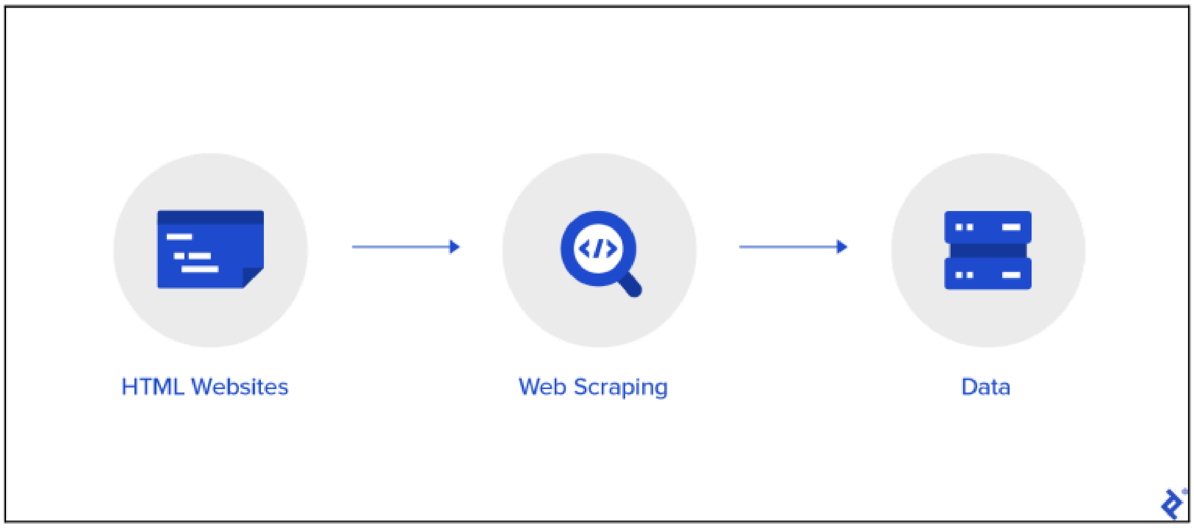


Figure 2. Web Scraping Overview

If a website provides a way for a visitor's browser to download content and render that content in a structured way, then almost by definition, that content can be accessed programmatically. Hence, any content viewable in a web page can be scrapped. Further, web scraping avoids the traditional methods of data fetching from websites such as human copy-and-paste, saving a lot of time and effort. It should be noted that the assistance from a web scraper is very much important when the website does not provide APIs to obtain data. The main issue with web scraping is that there is no universal scraper developed to scrap every detail from every website. Based on the website structure web scrapers should be scripted. There are several methods of web scraping.

- Regular Expression Matching: A simple approach to extract information from web pages based on regular expression-matching facilities of programming languages
- HTTP programming: Static and dynamic web pages can be retrieved by posting HTTP requests to the remote web server
- DOM parsing: Retrieving the dynamic content generated by client-side scripts.
- HTML parsers

3.3.2. Data Acquisition from ESPNCricinfo, World Weather Online and Sports F1

For this study, the data required were scrapped from ESPNCricinfo, World Weather Online and Sport F1 websites respectively. These sites hold data in different aspects that are related to cricket. ESPNCricinfo holds player-oriented data as well as team-centric data. Mainly the data required for player performance, player profiles and team performance were extracted from ESPNCricinfo website. The World Weather Online website disseminates weather data with respect to each playground, stadium situated in each cricket playing country. All real time and historic weather data could be retrieved from this particular site. Sports F1 site assisted in extracting pitch condition and pitch report data. Following section will describe each web scraper developed to extract data from each site.

3.3.2.1 Player Profile Data

In order to rank players and give them a performance score, the first thing that needed to be done was identify each player and obtain their basic information. There were 167 players in Sri Lanka who have played international ODI cricket since 1975 and for this research, an overall pool of 34 players were considered out of them who are still playing cricket after 2009. The list of 34 players is given in Table 1.

Table 1. Considered 34 players for the research

K.I.C. Asalanka	R.A.S. Lakmal
K.N.A. Bandara	D. Lakshan
M. Bhanuka	A.D. Mathews
P.V.D. Chameera	B.K.G. Mendis
L.D. Chandimal	P.H.K.D. Mendis
A. Dananjaya	R.T.M. Mendis
D.M. de Silva	M.D.K.J. Perera
P.W.H. de Silva	N.I.T.C. Perera
N. Dickwella	P.B.B. Rajapaksa
A.M. Fernando	C.A.K. Rajitha
N. Pradeep	P.A.D.I.R. Sandakan
B. Fernando	M.D. Shanaka
B.O.P. Fernando	M. Theekshana
W.I.A. Fernando	I. Udana
C. Gunasekara	J.D.F. Vandersay
M.D. Gunathilaka	P. Nissanka
P. Jaawickrama	F.D.M. Karunaratne
C. Karunaratne	

This site has a comprehensive set of details with respect to players; the name, age, school, major teams, player roles, bowling and batting arm etc.

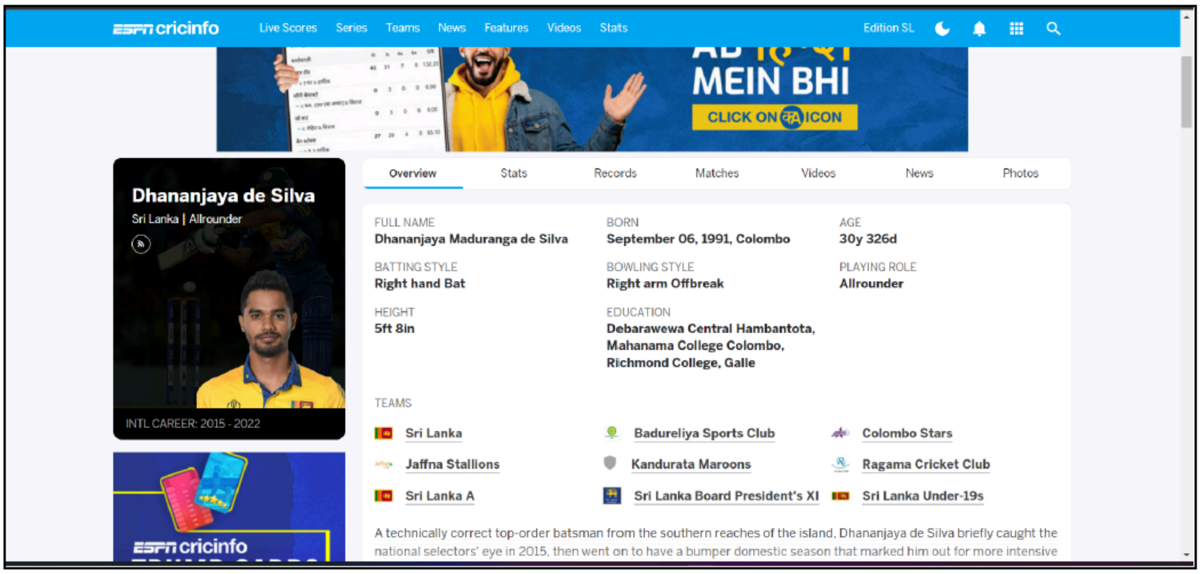


Figure 3. Sample player profile from ESPNCricinfo.

Table 2. shows the attributes extracted from player profiles.

Table 2. Attributes extracted from player profiles

Attribute	**Extracted player profile data**
Name	Name of the player
Born	Date of Birth
Major teams	Prominent teams he has played in
Current age	Current age of the player
Playing role	Player’s playing role in Sri Lankan team
Batting style	Player’s batting style
Bowling style	Player’s bowling style
Fielding position	Player’s fielding position
Height	Player’s height

3.3.2.2 Player Performance Data – Batting Performance Data

Another web scraper implemented to retrieve batting performance of each player from the website. ESPNCricinfo has batting statistics related to each country with respect to the three match formats. As this research is focused on ODI match format, only ODI batting performances were retrieved using the scrapper.

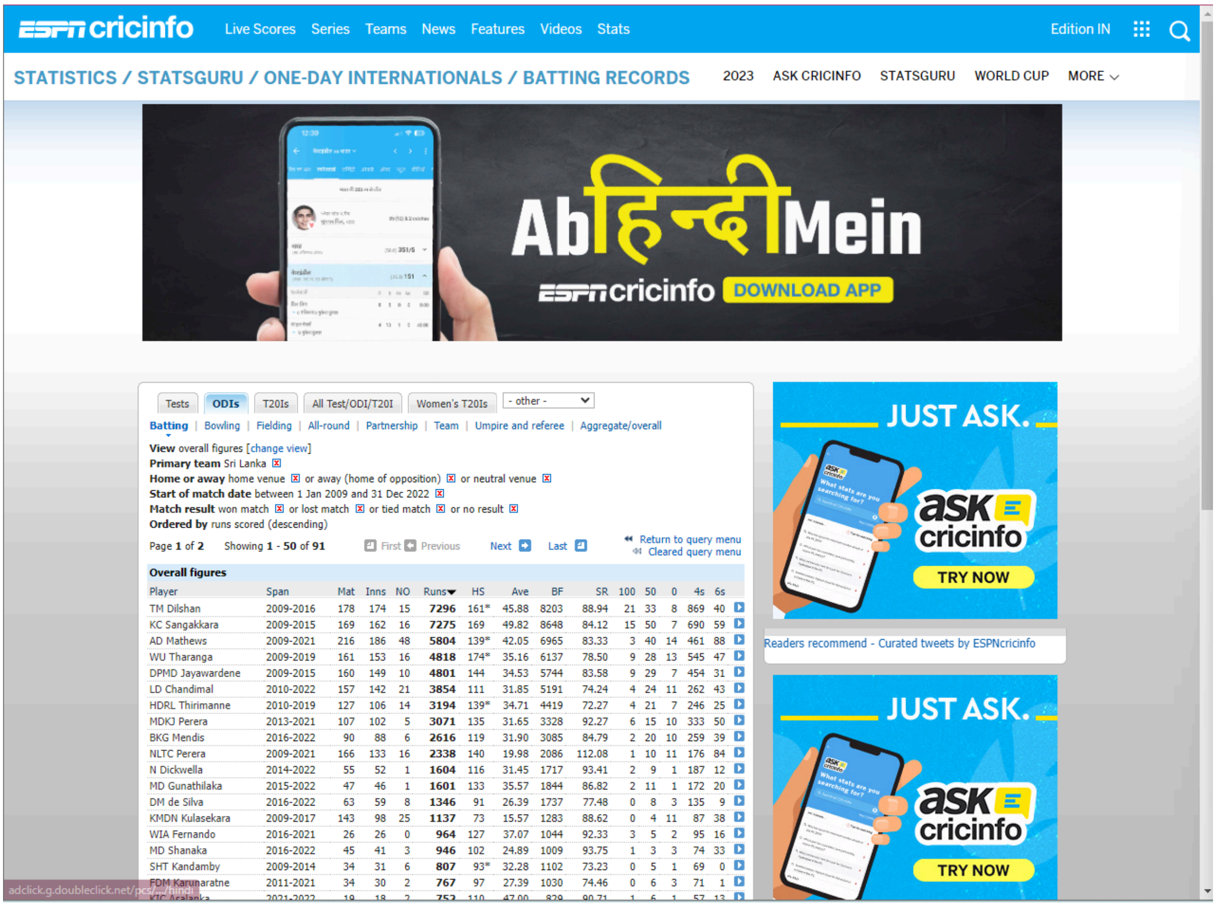


Figure 4. Batting performance listed in the site

This web page consists of all the data of a particular batting inning of each player along with all the related details such as span of playing, number of matches played, number of innings played, number of runs scored, highest score, strike rate etc. Since this web page contains all that data, it was very useful for this research to get insights of batting skills of each player in each match, by scraping this data. A screenshot of a web page which was used to extract data is depicted in Figure 4. above and details about the extracted data is represented in Table 3. Additionally, sample code for batting performance scraper is shown in Appendix A.

Table 3. Attribute description of patting performance data

Attribute	**Description**
Span	Duration of playing
Mat	Number of matches played
Inns	Number of innings played
NO	Number of not outs
Runs	Number of runs scored
HS	Highest score
Ave	Average score
BF	Number of balls faced
SR	Average strike rate
100	Number of centuries scored
50	Number of half centuries played
0	Number of zeros scored
4s	Number of boundary 4 scored
6s	Number of boundary 6 scored

3.3.2.3 Player Performance Data – Bowling Performance Data

Similarly, as in above batting performance data extraction, bowling performance data also extracted. Figure 5. below shows the site screenshot of the data display, Table 4 explains the description of attributes gathered and in Appendix B the web scraper will be attached respectively.

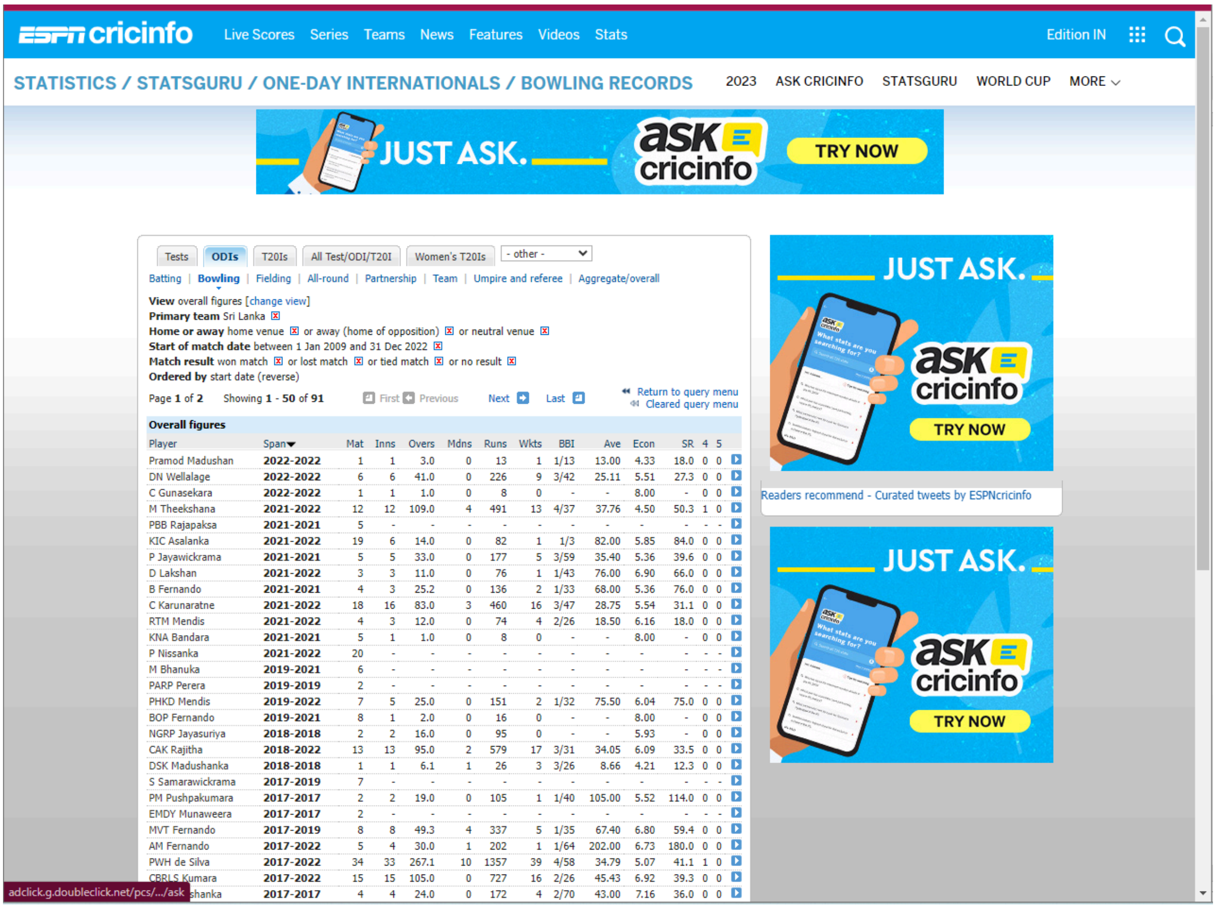


Figure 5. Bowling performance listed in site

Table 4. Attribute description of bowling performance data

Attribute	**Description**
Span	Duration of playing
Mat	Number of matches played
Inns	Number of innings played
Overs	Total number of overs delivered
Mdns	Number of maiden overs
Runs	Number of runs given
Wkts	Number of wickets taken
BBI	Best Bowling Inning
Ave	Bowling average
Econ	The average number of runs conceded per over bowled
SR	Bowling strike rate

4	4 Wicket hauls
5	5 Wicket hauls
Ct	Number of catches
St	Number of stumps taken

3.3.2.4 Play Ground Weather Data

Weather plays a major role in cricket matches. Then different attributes in weather impact on pitch condition and ball spinning. Furthermore, weather is out of the control of the player. The historic weather data pertaining to each ground that the Sri Lankan team has ever played can be retrieved from the World Weather Online website. Figure 6. depicts the screenshot of the site.

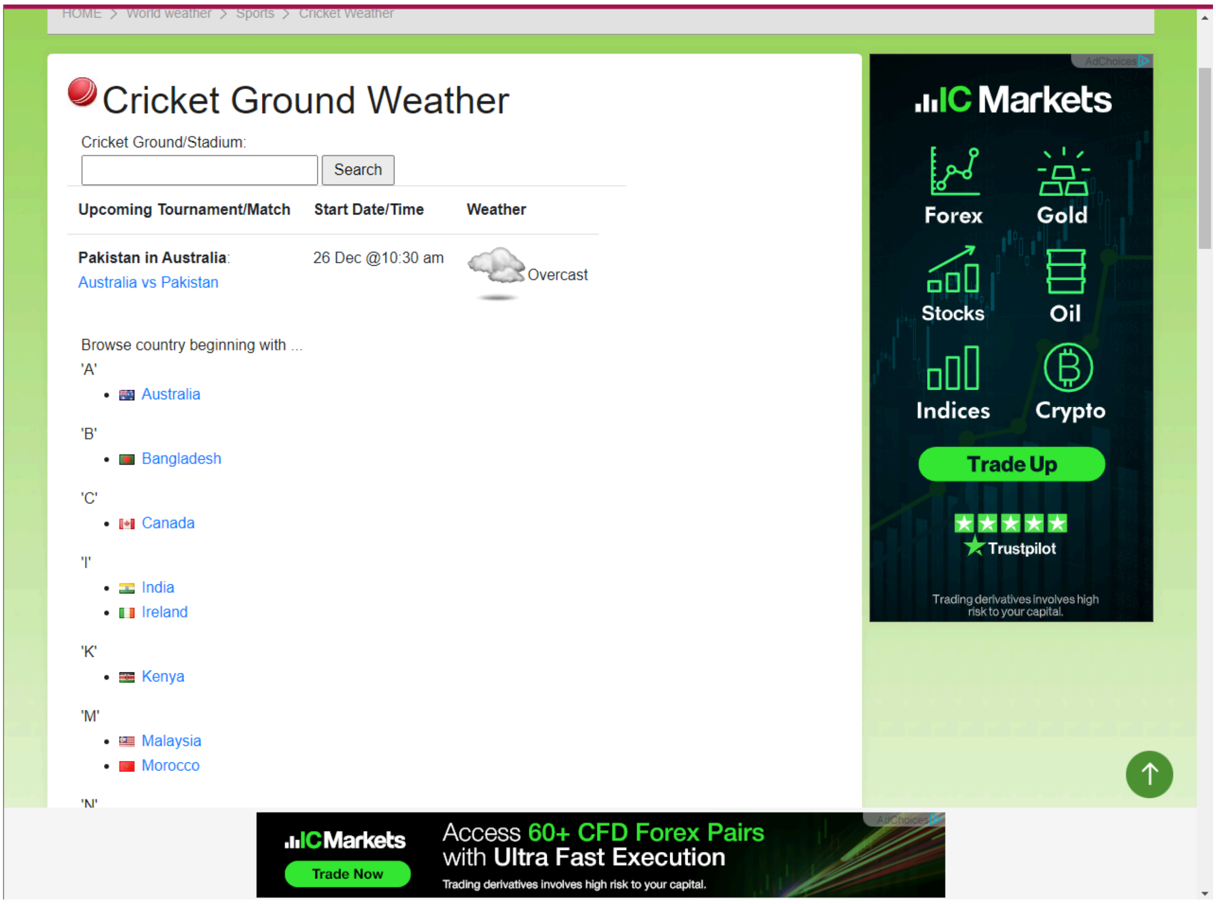


Figure 6. World Weather Online Site

Attributes of the collected weather data are as follows.

Table 5. Weather data attributes

Attribute	Description
Date	Date when the match played
Stadium	The stadium where the match played
Weather	The kind of weather the day had (Sunny, Overcast, Rainy etc.)
TempMin	Minimum Temperature of the day
TempMax	Maximum Temperature of the day
MoonRise	Time of moon rise
MoonSet	Time of the moon set
SunRise	Time of the sun rise
SunSet	Time of the sun set
Average Wind Speed	Average wind speed during the day

3.3.2.5 Pitch Report Data

The pitch report is an important piece of data that many experts use to predict the victory of a match in cricket. Pitch report is a comprehensively gathered data about number of matches played on the pitch, number of winners who won the match after batting first and batting second, margins of scores etc. With the analysis of the pitch report, valuable trends can be mined that thereby can be used to predict the winning of the game. The pitch report data were retrieved from Sports F1 website, which displays pitch reports of all cricket playing stadiums/ grounds. Figure 7. depicts the screenshot of the site.

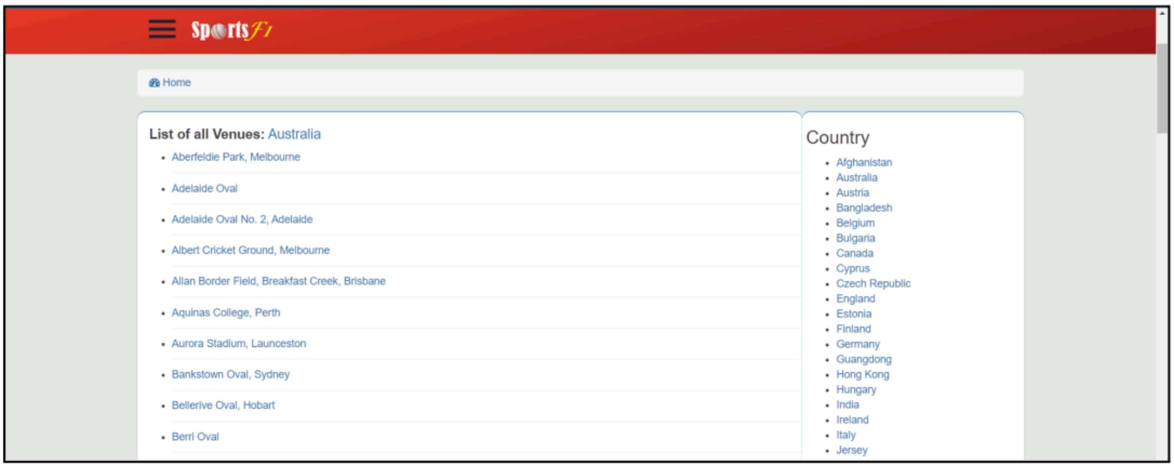


Figure 7. Sports F1 website for pitch reports

Pitch report attributes are as follows.

Table 6. Pitch report attributes

Attribute	**Description**
Ground	The cricket playing ground/ stadium
Country	The country which the ground/ stadium situated
Total match played	Total number of matches played in a particular ground/ stadium
Batting first won	Number of matches that team batted in first innings
Batting second won	Number of matches that team batted in second innings
Tie	Number of matches that draw without the win or loose
Avg. score in 1^{st} bat	Average score which scored by teams batted first innings
Highest score	Highest score scored in the ground/ stadium
Lowest score	Lowest score scored in the ground/ stadium
Below 200 score	Number of times scored less than 200
Score between 200 and 249	Number of times scored 200 and 249
Score between 250 and 299	Number of times scored 250 and 299
Above score 300	Number of times scored more than 300

Data have been carefully gathered, with an emphasis on cricket matches with the following nations, in preparation for future implementations that would rank players according to their results versus particular opponent teams:

1. England
2. Australia
3. South Africa
4. West Indies
5. New Zealand
6. India
7. Pakistan
8. Bangladesh

3.3.3. Data Preprocessing on Player Performance Data

The data accumulated in sports domain tend to have spread of attribute values in larger scale. Carrying out further analysis on such data leads to biased solutions, impacting the decision-making process. Hence, in order to avoid this issue, the normalization technique was carried out to standardize all the attributes across the dataset. Once normalization conducted the numeric attributes were on the same scale (0,1) thereby preventing biasness in the final solution. The following equation was used as the normalization equation.

$$X_{i,new} = \frac{X_{i,old} - X_{min}}{X_{max} - X_{min}}$$

In ODI match format all players have no opportunity to bowl due to the limited overs, hence there are missing values in some specific fields such as strike rate (runs conceded/overs bowled). Fuzzy logic cannot support missing values, hence a method to deal with such missing data was required. Since some players have actually bowled but have not allowed any runs to be scored, it is unfair to fill in such missing data with a value of 0, and also, a player with a low strike rate like 0 will be rated as having excellent bowling performance. Consequently, to lessen this issue, such the greatest value multiplied by 120% was used to impute missing values for the unique trait. A value of 72 was recorded for all the players who had not bowled a single over, for instance, if the lowest bowling strike rate ever recorded was 60 for a specific cumulative match dataset. This figure was obtained by multiplying 60 by 120 percent. It is fairly balanced because a player with a high bowling strike rate is regarded as a poor bowler. As a result, all players who have never bowled a ball in their lives will have a lower normalized strike rate than the player who has had the poorest economy rate performance. If the highest value is not multiplied, a player who has never bowled a ball in his life will have the same economy rate as the player who has fared the worst in economy rate. This is why multiplying by a figure like 120 percent is necessary. Since it can be assumed that a player who has not bowled a single ball should be a worse bowler than a player who has actually bowled a single ball despite having done poorly, this is quite unfair. Also mechanized was this procedure.

3.3.4. Feature Engineering on Complex Dynamics Data

3.3.4.1 Deriving the Stadium Tendency for Batting and Bowling

According to experts, the decision in team combination selection in cricket is highly dependent on the nature of the pitch or stadium that the match is taking place. Since the primary data pertaining to pitch conditions were not available, pitch report data were used to derive the usual stadium condition in terms of batting favorable, bowling favorable or neutral. The provided Python code in the appendix conducts a comprehensive analysis of cricket match data, employing various operations for feature engineering and insights into stadium characteristics. Initially, the script calculates average scores, highest scores, and lowest scores for each cricket stadium, subsequently comparing these averages against the overall mean score. The code further delves into score distribution, categorizing matches into score ranges and determining the percentage of matches falling within each category. Notably, the script labels stadiums based on their characteristics, identifying them as 'Batting Friendly,' 'Bowling Friendly,' or 'Neutral' depending on predefined thresholds. Additionally, the code groups stadiums and computes statistical measures, such as mean scores, providing valuable insights into each stadium's performance. Finally, the script merges the derived stadium labels back into the main dataset and saves the augmented dataset as a CSV file for future reference and analysis. It is worth noting that some variable names, such as 'score_columns,' are referenced but not explicitly defined in the provided code and may need adjustment based on the specific DataFrame and column names in your dataset.

3.3.4.2 Deriving the Temperature Labels

The provided Python code segment focuses on temperature categorization within a DataFrame named 'df.' It sets specific temperature thresholds, designating 30 as the threshold for high temperatures and defining a moderate temperature range with lower and upper bounds set at 20 and 29, respectively. The code then creates two new columns, namely 'Minimum Temp Temperature_Label' and 'Maximum Temp Temperature_Label,' based on these temperature conditions. Using the pandas cut function, the script bins the 'TempMin' and 'TempMax' columns into labeled categories such as 'Minimum Low,' 'Minimum Moderate,' 'Minimum High,' 'Maximum Low,' 'Maximum Moderate,' and 'Maximum High.' Finally, the code merges these newly created temperature label columns back into the original DataFrame, ensuring that each row is now enriched with information about the temperature category it falls into. This categorization can facilitate subsequent exploratory data analysis by allowing researchers to analyze and interpret temperature-related patterns in the dataset.

3.3.5. Designing the TCS Model

In machine learning, the nature of a prediction task is determined by the characteristics of the target variable or variables that the model aims to predict. When dealing with multiple target variables, the task is termed a "multi-output" task. This scenario arises when a single instance is associated with more than one target variable, and the model is expected to make predictions for each of them independently. On the other hand, the term "multi-class classification" is used when the prediction task involves assigning instances to one of three or more classes. Combining these concepts results in a "multi-output, multi-class classification" task, indicating that the model has to predict multiple target variables, and each of these variables can take on one of several classes. For instance, in healthcare applications, a model might predict both the specific type of disease and the severity level, making it a multi-output, multi-class classification problem. Recognizing these characteristics is essential for selecting appropriate algorithms, evaluation metrics, and preprocessing techniques tailored to the complexity of the task at hand.

When analyzing the dataset, the problem domain, and incorporating expert opinions, it became evident that predicting the outcome of a cricket match involves a "multi-output, multi-class classification." In cricket, a team's composition is not a single-dimensional factor but a combination of various aspects, such as the batsman, bowler, and all-rounder roles, each contributing to the team's overall performance. The prediction task extends beyond a binary win/loss scenario, as it requires forecasting the performance of individual players and their specific roles in the match. With multiple players and diverse roles involved, the prediction model needs to output a set of outcomes for each player category, leading to a multi-output setup. Furthermore, the classification aspect emerges from the categorization of outcomes, encompassing various classes such as winning, losing, or potential ties, adding another layer of complexity to the predictive modeling. This multi-output, multi-class classification approach provides a more distinctive understanding of the factors influencing match outcomes, offering valuable insights for team composition and strategic decision-making in the realm of cricket.

Multi-output Decision Trees, Multi-output Random Forests, Multi-output Support Vector Machines, Multi-output Neural Networks, Multi-output Naïve Bayes and Multi-output Gradient Boosting are supervised learning algorithms that facilitate multi-output multi-class classification tasks. Hence, initially several of these algorithms were employed for the classification task and multi-output neural network was the outperforming model for the task. Further in order to come up with the best neural network architecture several network architectures were built as listed in appendix. When considering the high-level architecture, the neural network has the ability to manipulate 10 complex dynamics as the inputs and provide classification for 3 outputs as number of batsmen, bowlers and allrounders. The 10 input complex dynamics for the neural network architecture are as follows.

1. Name of the stadium
2. Nature of the stadium
3. Weather of the stadium for a particular match
4. Minimum Temperature Label of the stadium for a particular match
5. Maximum Temperature Label of the stadium for a particular match
6. Average wind speed label of the stadium for a particular match
7. Team 1 of the match
8. Team 2 of the match
9. Winner of the match
10. Margin of win for the match

The final neural network that was chosen as the best model for predicting team composition in cricket composed of following specifications as listed below Figure 8.

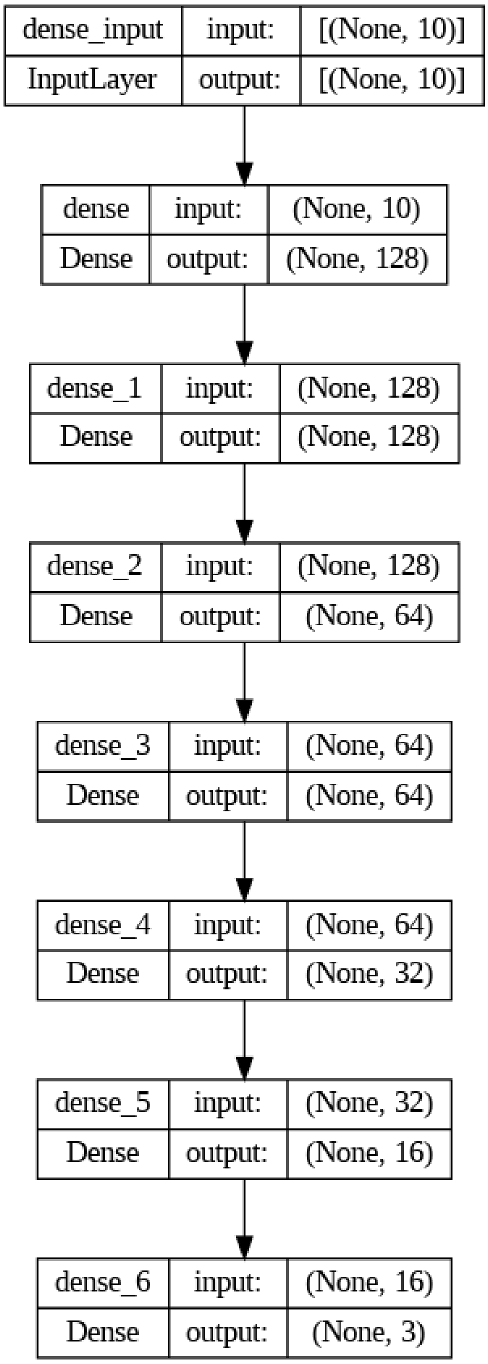


Figure 8. Layer Specifications of the Final TCS Model

3.3.6. Designing the Player Rating and Ranking System

Fuzzy logic inferencing was used derive rules to rank and rate players within their expertise in terms of batsmen, bowlers and allrounders. Three Fuzzy Inference Systems (FIS) termed as Batting Performance FIS, Bowling Performance FIS and Allrounder Performance FIS were built with following specifications.

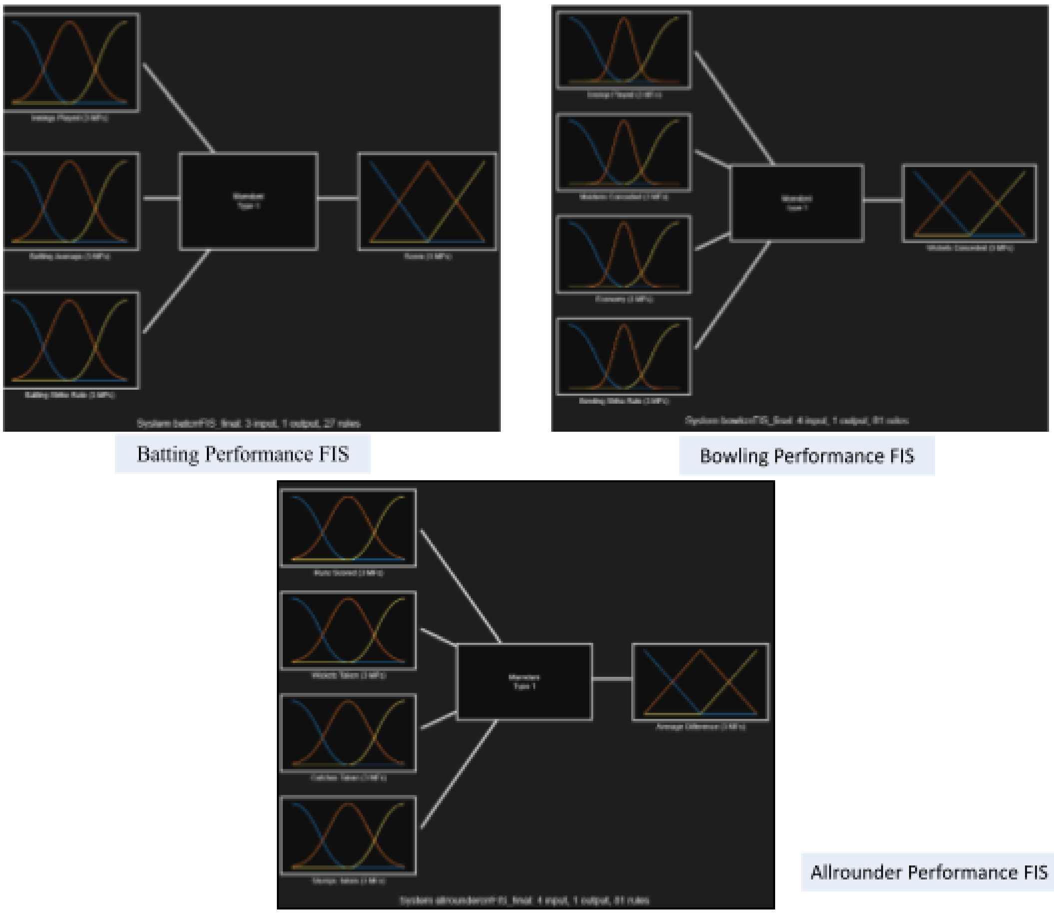


Figure 9. Developed FIS for Batting, Bowling, and Allrounder Performance Ranking System

The following Table 7 and Table 8 tabulate the Input and Output feature membership functions used with input and output level values.

Table 7. Input Feature Parameters for FIS

FIS	**Input Feature**	**Membership Function**
1. Batting Performance FIS	Innings Played	MF1='Low':'zmf',[0 0.5] MF2='Medium':'gaussmf',[0.17 0.5] MF3='High':'smf',[0.5 1]
	Batting Average	MF1='Low':'zmf',[0 0.5] MF2='Medium':'gaussmf',[0.17 0.5] MF3='High':'smf',[0.5 1]
	Batting Strike Rate	MF1='Low':'zmf',[0 0.5] MF2='Medium':'gaussmf',[0.17 0.5] MF3='High':'smf',[0.5 1]
2. Bowling Performance	Innings Played	MF1='Low':'zmf',[0 0.5] MF2='Medium':'gaussmf',[0.09 0.5]

FIS		MF3='High':'smf',[0.5 1]
	Maidens Conceded	MF1='Low':'zmf',[0 0.5] MF2='Medium':'gaussmf',[0.09 0.5] MF3='High':'smf',[0.5 1]
	Economy	MF1='Low':'zmf',[0 0.5] MF2='Medium':'gaussmf',[0.09 0.5] MF3='High':'smf',[0.5 1]
	Bowling Strike Rate	MF1='Low':'zmf',[0 0.5] MF2='Medium':'gaussmf',[0.09 0.5] MF3='High':'smf',[0.5 1]
3. Allrounder Performance FIS	Runs Scored	MF1='Low':'zmf',[0 0.5] MF2='Medium':'gaussmf',[0.1769 0.5] MF3='High':'smf',[0.5 1]
	Wickets Taken	MF1='Low':'zmf',[0 0.5] MF2='Medium':'gaussmf',[0.1769 0.5] MF3='High':'smf',[0.5 1]
	Catches Taken	MF1='Low':'zmf',[0 0.5] MF2='Medium':'gaussmf',[0.1769 0.5] MF3='High':'smf',[0.5 1]
	Stumps Taken	MF1='Low':'zmf',[0 0.5] MF2='Medium':'gaussmf',[0.1769 0.5] MF3='High':'smf',[0.5 1]

Table 8. Output Feature Parameters for FIS

FIS	**Output Feature**	**Membership Function**
1. Batting Performance FIS	Batting Score	MF1='Low':'trimf',[-0.5 0 0.5] MF2='Medium':'trimf',[0 0.5 1] MF3='High':'trimf',[0.5 1 1.5]
2. Bowling Performance FIS	Wickets Conceded	MF1='Low':'trimf',[-0.5 0 0.5] MF2='Medium':'trimf',[0 0.5 1] MF3='High':'trimf',[0.5 1 1.5]
3. Allrounder Performance FIS	Average Differences	MF1='Low':'trimf',[-0.5 0 0.5] MF2='Medium':'trimf',[0 0.5 1] MF3='High':'trimf',[0.5 1 1.5]

3.4. Summary

This research chapter outlines a methodological framework to provide reliable analytics for the Sri Lankan cricket team in one-day internationals, specifically focusing on the intricate task of Team Composition Selection (TCS). The study incorporates complex dynamics such as player roles, match conditions, and opponent teams into a multi-output, multi-class classification approach. The proposed neural network architecture is designed to predict team composition using ten key input variables. Data acquisition involves web scraping from diverse sources, and preprocessing includes normalization and feature engineering. Additionally, the chapter introduces a fuzzy logic-based player rating and ranking system tailored for batting, bowling, and all-round performance. In essence, this research provides a comprehensive strategy for advancing predictive analytics in cricket, addressing the unique challenges of the sports domain and offering valuable insights for decision-making in team selection.

CHAPTER 4

EVALUATION AND RESULTS

This chapter presents the findings derived from experiments conducted on various model performances for Team Composition Selection (TCS) and the development of Fuzzy Inference Systems (FISs) dedicated to Player Rating and Ranking. This examination of experimental outcomes contributes to evaluating the developed models, to identify potentially outperformed models for a decision support system in Cricket team selection. Further, results demonstrated in this chapter are facilitating informed conclusions and offering a valuable resource for future model enhancements.

4.1. Experimental Results for TCS modelling

The tables below show the accuracy performance results for the TCS model with model specifications.

Table 9 provides performance accuracies for the Random Forest Classifier and XGBoost Classifier models developed considering classification for individual figures for number of batsman bowlers and allrounders.

Table 9. Performance Accuracies considering individual classification for no. of batsmen, bowlers and allrounders

Algorithm	Model Name	Precision
Random Forest Classifier	Batsman Classifier	0.613
	Bowler Classifier	0.548
	Allrounder Classifier	0.323
XGBoost Classifier Attempt - I	Batsman Classifier	0.645
	Bowler Classifier	0.452
	Allrounder Classifier	0.323
XGBoost Classifier Attempt - II	Batsman Classifier	0.500
	Bowler Classifier	0.403
	Allrounder Classifier	0.339

Than having three separate models to classify, it is efficient to develop a single model which can classify no. of batsmen, bowlers and allrounders. Hence, with this approach the classification task becomes a multi-output, multiclass classification. Table 10 tabulates the neural network models that have been built to conduct the multi-output, multiclass classification.

Table 10. Neural Network Model performances for Multi-ouput and Multiclass classification

Model Specifications		Performance	
		Accuracy	Loss
Model 1	Layer (type) Output Shape Param # == dense_1 (Dense) (None, 128) 1152 dense_2 (Dense) (None, 64) 8256 dense_3 (Dense) (None, 32) 2080 dense_4 (Dense) (None, 16) 528 dense_5 (Dense) (None, 3) 51 == Total params: 12,067 Trainable params: 12,067 Non-trainable params: 0	0.548	6.858
Model 2	Layer (type) Output Shape Param # == dense_1 (Dense) (None, 128) 1408 dense_2 (Dense) (None, 128) 16512 dense_3 (Dense) (None, 64) 8256 dense_4 (Dense) (None, 64) 4160 dense_5 (Dense) (None, 32) 2080 dense_6 (Dense) (None, 16) 528 dense_7 (Dense) (None, 3) 51 == Total params: 32995 (128.89 KB) Trainable params: 32995 (128.89 KB)	0.677	12.189

	Non-trainable params: 0 (0.00 Byte)		
Model 3	Layer (type) Output Shape Param # == dense_1 (Dense) (None, 128) 1408 dense_2 (Dense) (None, 128) 16512 dense_3 (Dense) (None, 64) 8256 dense_4 (Dense) (None, 64) 4160 dense_5 (Dense) (None, 32) 2080 dense_6 (Dense) (None, 16) 528 dense_6 (Dense) (None, 3) 51 == Total params: 32995 (128.89 KB) Trainable params: 32995 (128.89 KB) Non-trainable params: 0 (0.00 Byte)	0.629	12.124
Model 4	Layer (type) Output Shape Param # == dense_1 (Dense) (None, 128) 1408 dense_2 (Dense) (None, 128) 16512 dense_3 (Dense) (None, 64) 8256 dense_4 (Dense) (None, 64) 4160 dense_5 (Dense) (None, 32) 2080 dense_6 (Dense) (None, 16) 528 dense_7 (Dense) (None, 3) 51 == Total params: 32995 (128.89 KB) Trainable params: 32995 (128.89 KB) Non-trainable params: 0 (0.00 Byte)	0.758	12.396

4.2. Experimental Results for Player Rating and Ranking modelling

Initial FIS design for batting performance, bowling performance and allrounder performance are shown in the Table 7 above. Based on the Table 7 specification Figure 10 and Figure 11 respectively illustrate the initial membership function distribution for the FISs

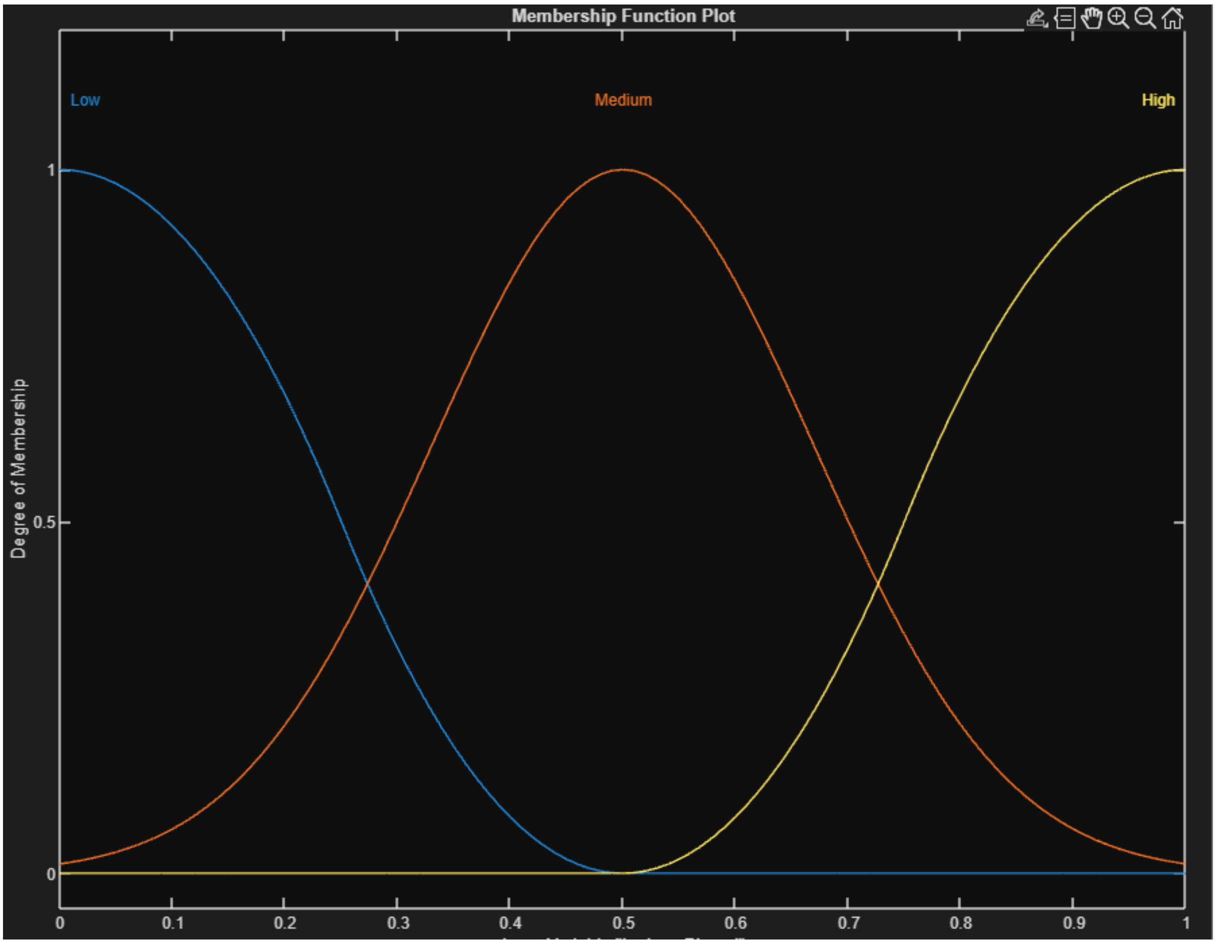


Figure 10. Input Membership Functions Distribution for Initial FISs

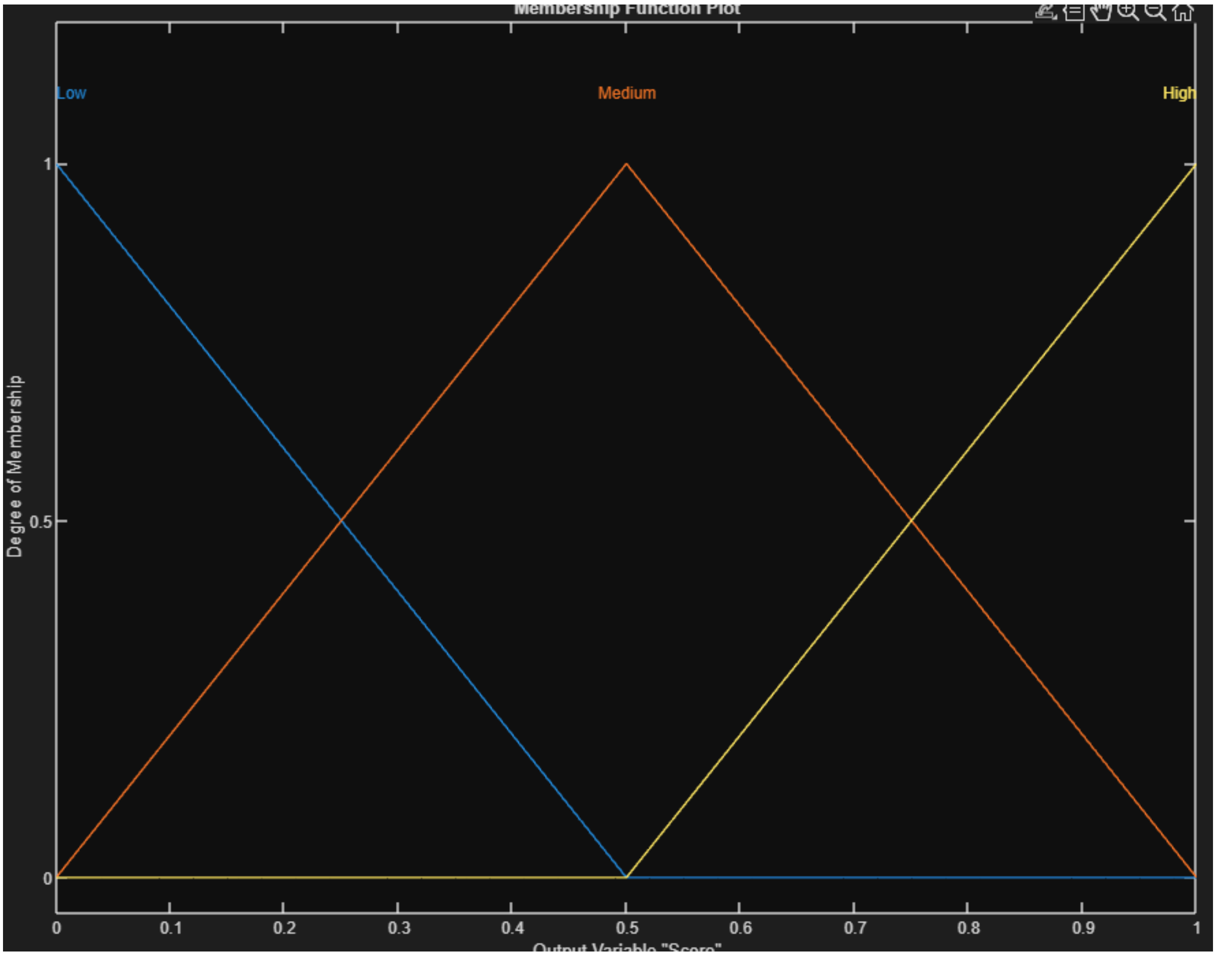


Figure 11. Output Membership Functions Distribution for Initial FISs

4.2.1. Batting Performance FIS

The FIS validating performance of the initial Batting Performance FIS is depicted in the below Figure 12.

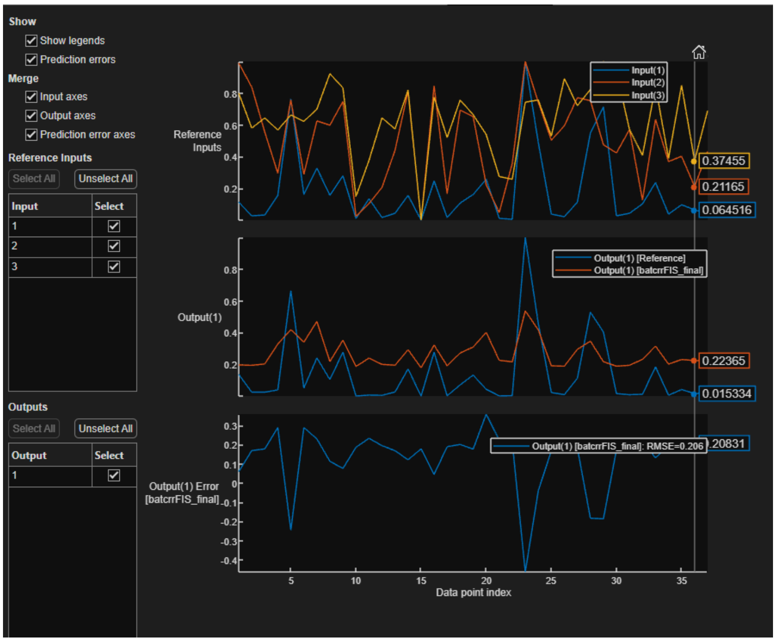


Figure 12. Initial Validation Performance Batting Performance FIS

According to Figure 12 above, the Root Mean Square Error (RMSE) is 20%, which has yielded nearly 80% accuracy when ranking batsmen. Yet, in order to tune the rules, a Genetic Algorithmic approach was considered (All FISs tuned using Genetic Algorithm). Table 11 tabulates specifications of the tuned batting FIS as follows.

Table 11. FIS specifications of Tuned Batting FIS

Input Name	**Class**	**Membership Function Type**	**Parameters**
Innings Played	Low	Z-shaped	[0.547216 0.261871]
	Medium	Gaussian	[0.341125 0.696667]
	High	S-shaped	[0.451739 0.217802]
Batting Average	Low	Z-shaped	[0.864148 0.516997]
	Medium	Gaussian	[0.039185 0.536788]

	High	S-shaped	[0.414429 0.775028]
Batting Strike Rate	Low	Z-shaped	[0.255962 0.123381]
	Medium	Gaussian	[0.141336 0.937450]
	High	S-shaped	[0.959875 0.399067]
Output Name	**Class**	**Membership Function Type**	**Parameters**
Score	Low	Triangular	[-0.5 0 0.5]
	Medium	Triangular	[0 0.5 1]
	High	Triangular	[0.5 1 1.5]

Figure 13 and Figure 14 show the membership function distribution for tuned batting performance FIS.

Figure 13. Input membership distribution after Batting Performance FIS tuned

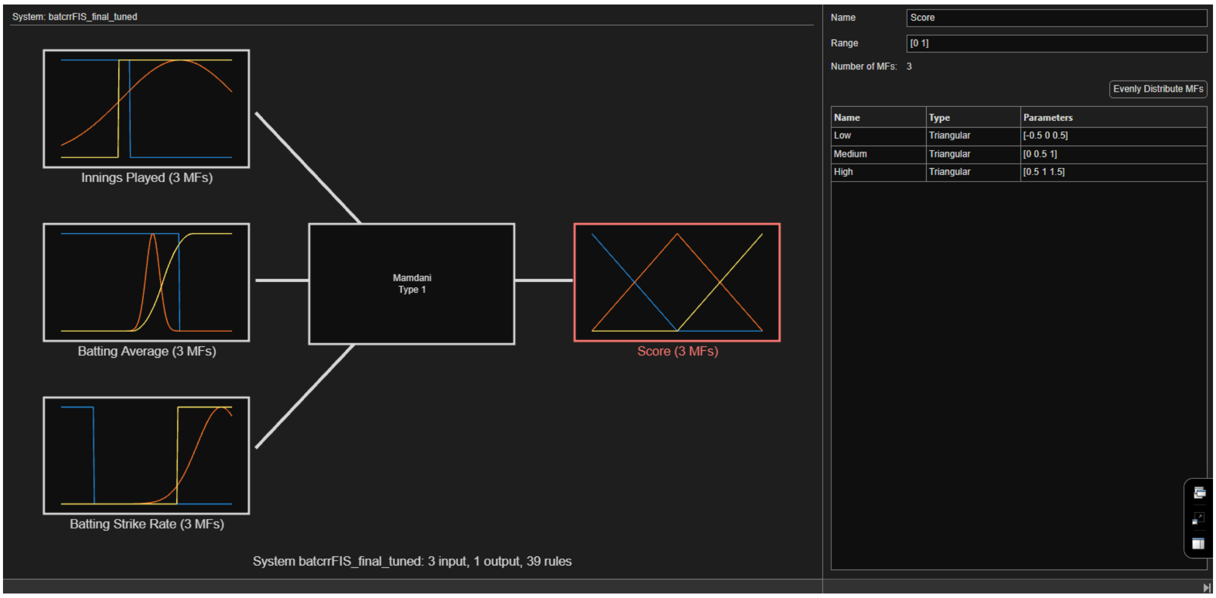


Figure 14. Output membership distribution after Batting Performance FIS tuned

The validation accuracy fluctuations are shown in the Figure 15 below.

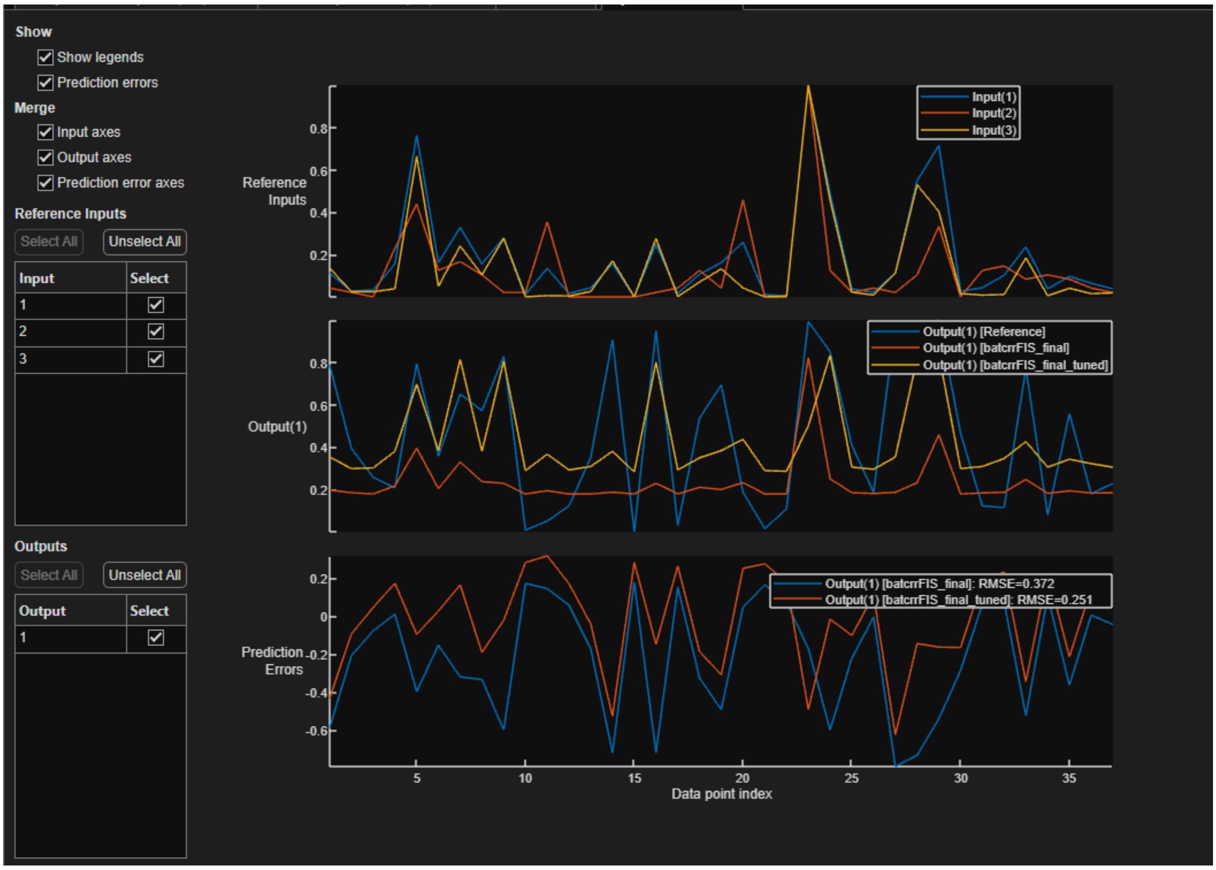


Figure 15. Validation plots of Tuned Batting Performance FIS, RMSE = 0.25

4.2.2. Bowling Performance FIS

The FIS validating performance of the initial Bowling Performance FIS is depicted in the below Figure 16.

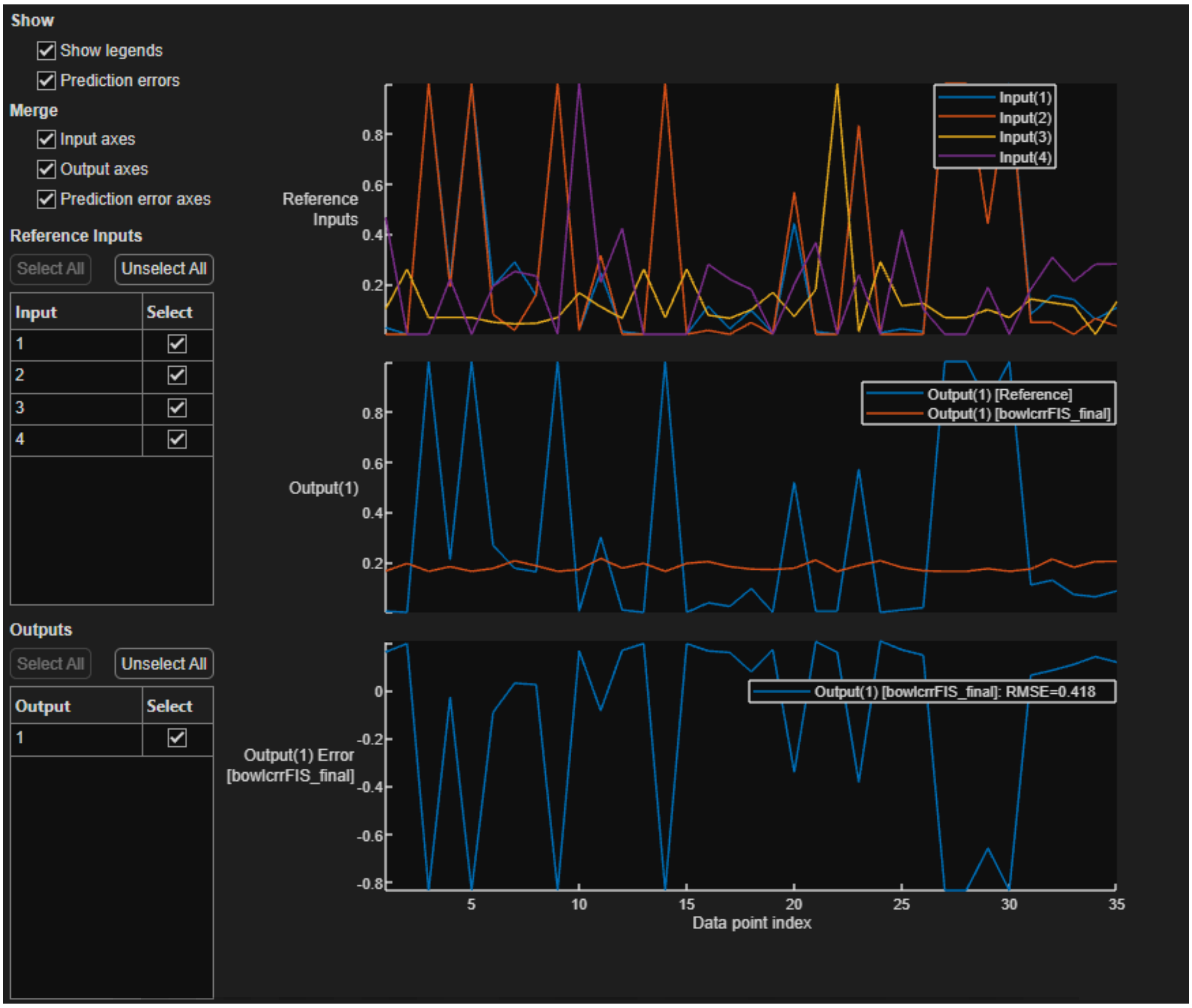


Figure 16. Initial Validation Performance of Bowling Performance FIS

According to Figure 16 above, the Root Mean Square Error (RMSE) is nearly 42%, which has yielded 58% accuracy when ranking bowlers. Yet, in order to tune the rules, a Genetic Algorithmic approach was considered (All FISs tuned using Genetic Algorithm). Table 12 tabulates specifications of the tuned bowling FIS as follows.

Table 12. FIS specifications of Tuned Bowling FIS

Input Name	Class	Membership Function Type	Parameters
Innings Played	Low	Z-shaped	[0.0811258 0.547871]
	Medium	Gaussian	[0.192028 0607866]
	High	S-shaped	[0.10592 0.590609]
Maidens Conceded	Low	Z-shaped	[0.944986 0.705951]
	Medium	Gaussian	[0.089999 0.748509]
	High	S-shaped	[0.16869 0.427793]
Economy	Low	Z-shaped	[0.279829 0.87033]
	Medium	Gaussian	[0.905331 0.021240]
	High	S-shaped	[0.828802 0.399067]
Bowling Strike Rate	Low	Z-shaped	[0.513679 0.442904]
	Medium	Gaussian	[0.208522 0.913153]
	High	S-shaped	[0.878228 0.90905]
Output Name	**Class**	**Membership Function Type**	**Parameters**
Wickets Conceded	Low	Triangular	[-0.325502 0.288239 0.733567]
	Medium	Triangular	[0.18727 0.306396 0.537321]
	High	Triangular	[0.5 1 1.5]

Figure 17. Input membership distribution after Bowling Performance FIS tuned

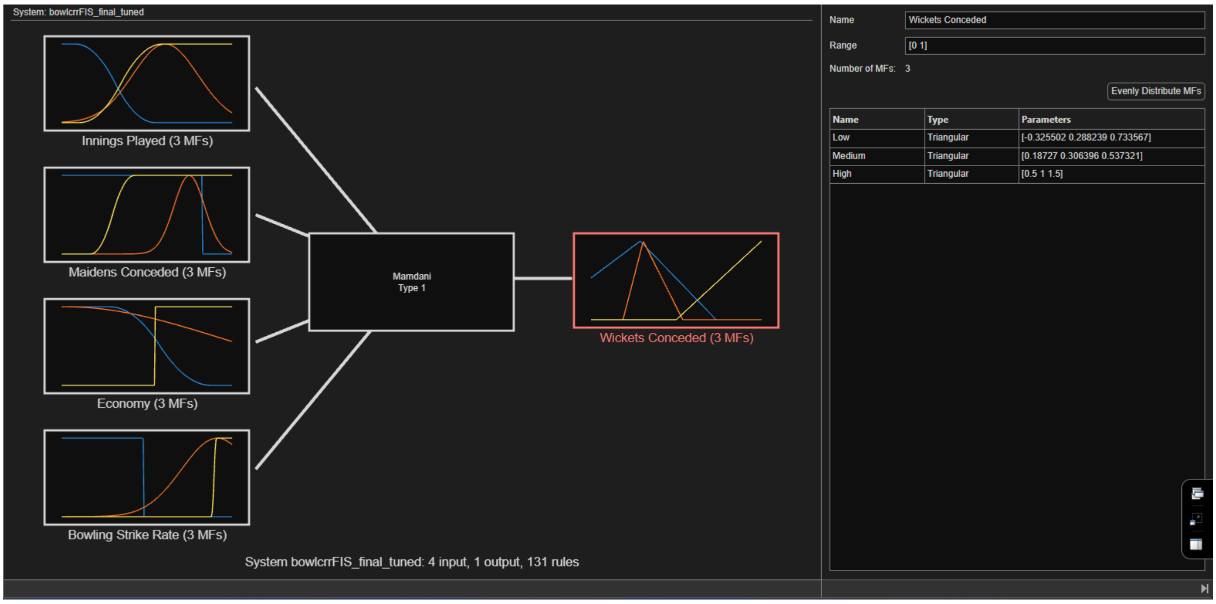


Figure 18. Output membership distribution after Bowling Performance FIS tuned

The validation accuracy fluctuations of tuned Bowling Performance FIS are shown in the Figure 19 below.

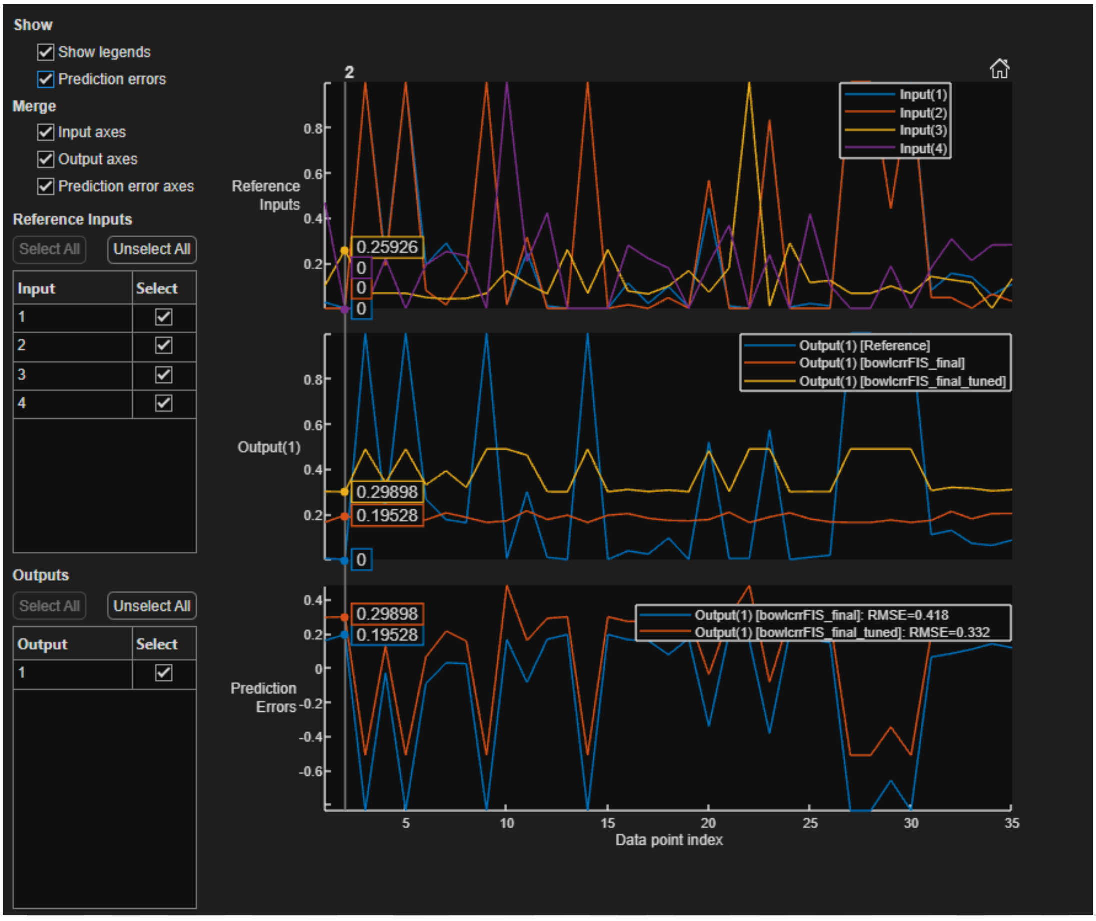


Figure 19. Validation plots of Tuned Bowling Performance FIS, RMSE = 0.33

4.2.3. Allrounder Performance FIS

The FIS validating performance of the initial Allrounder Performance FIS is depicted in the below Figure 20.

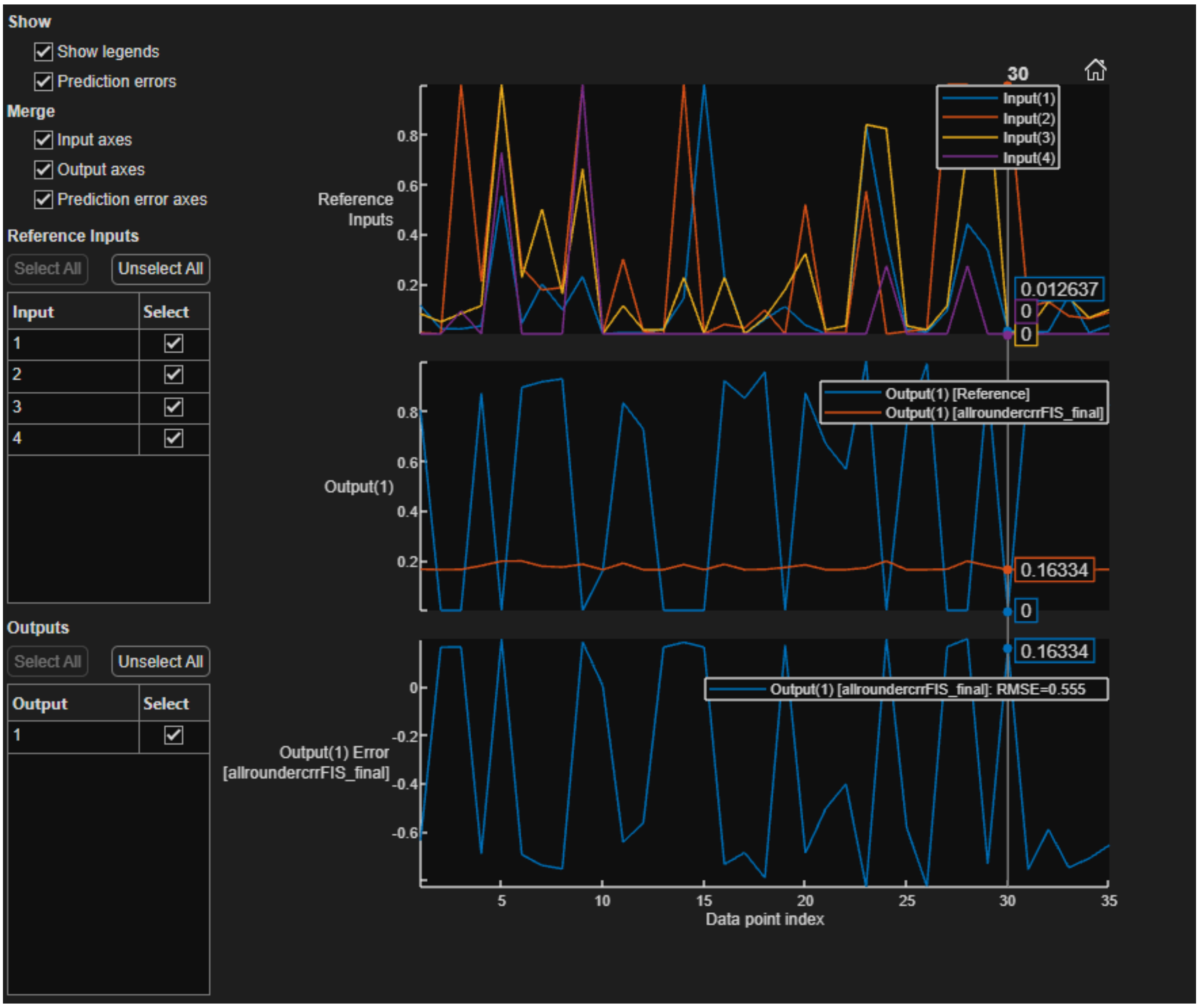
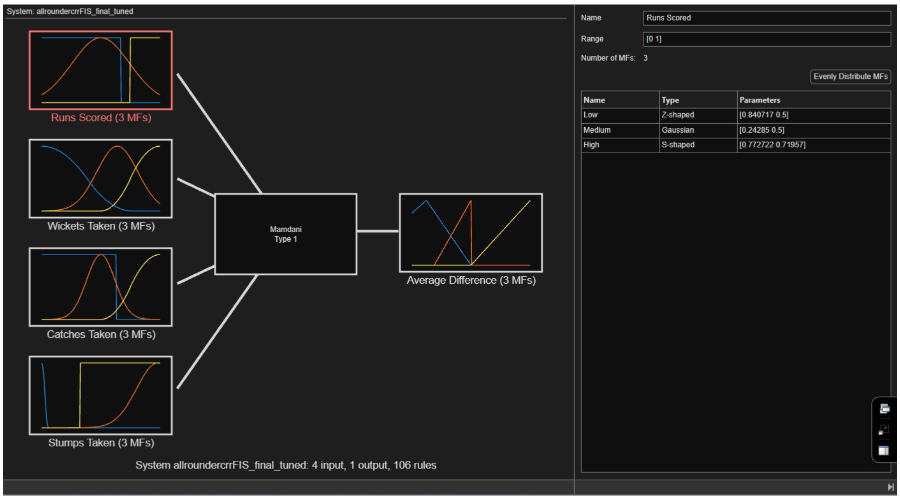


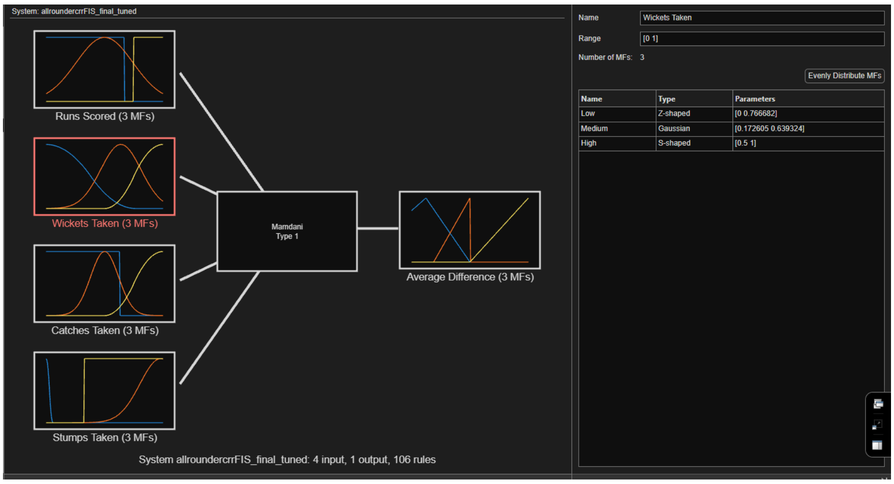
Figure 20. Initial Validation Performance of Allrounder Performance FIS

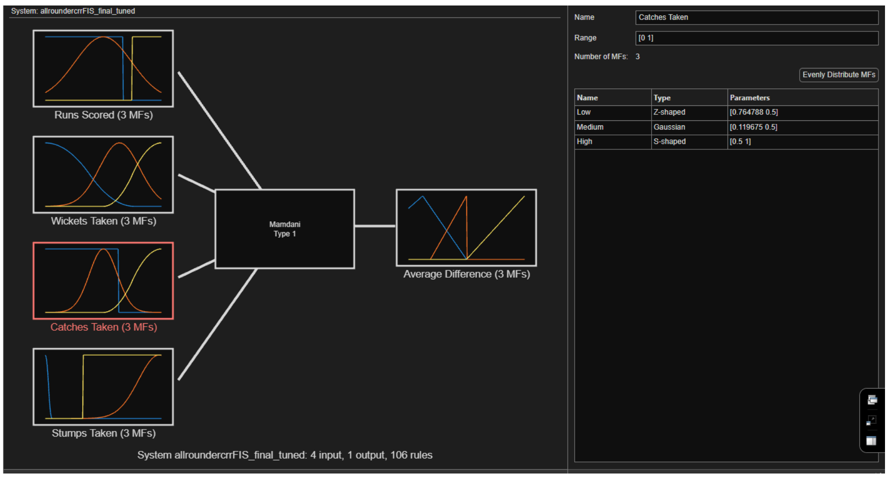
According to the Figure 20 above, the Root Mean Square Error (RMSE) is nearly 55%, which have yielded 45% accuracy when ranking allrounders. Yet, in order to tune the rules, Genetic Algorithmic approach was considered (All FISs tuned using Genetic Algorithm). Table 13 tabulates specifications of the tuned allrounder FIS as follows.

Table 13. FIS specifications of Tuned Allrounder FIS.

Input Name	**Class**	**Membership Function Type**	**Parameters**
Runs Scored	Low	Z-shaped	[0.840717 0.5]
	Medium	Gaussian	[0.24285 0.5]
	High	S-shaped	[0.772722 0.71967]
Wickets Taken	Low	Z-shaped	[0 0.766682]
	Medium	Gaussian	[0.172605 0.639324]
	High	S-shaped	[0.5 1]
Economy	Low	Z-shaped	[0.764788 0.5]
	Medium	Gaussian	[0.119675 0.5]
	High	S-shaped	[0.5 1]
Stumps Taken	Low	Z-shaped	[0 0.766682]
	Medium	Gaussian	[0.172605 0.639324]
	High	S-shaped	[0.5 1]
Output Name	**Class**	**Membership Function Type**	**Parameters**
Average Difference	Low	Triangular	[-0.5 0.123236 0.5]
	Medium	Triangular	[0.189413 0.5 0.504716]
	High	Triangular	[0.5 1 1.5]


System: allroundercrrFIS_final_tuned
Runs Scored (3 MFs)
Wickets Taken (3 MFs)
Catches Taken (3 MFs)
Stumps Taken (3 MFs)
Mamdani
Type 1
Average Difference (3 MFs)
System allroundercrrFIS_final_tuned: 4 input, 1 output, 106 rules
Name
Runs Scored
Range
[0 1]
Number of MFs: 3
Evenly Distribute MFs
Name Type Parameters
Low Z-shaped [0.840717 0.5]
Medium Gaussian [0.24285 0.5]
High S-shaped [0.772722 0.71957]


System: allroundercrrFIS_final_tuned
Runs Scored (3 MFs)
Wickets Taken (3 MFs)
Catches Taken (3 MFs)
Stumps Taken (3 MFs)
Mamdani
Type 1
Average Difference (3 MFs)
System allroundercrrFIS_final_tuned: 4 input, 1 output, 106 rules
Name
Wickets Taken
Range
[0 1]
Number of MFs: 3
Evenly Distribute MFs
Name Type Parameters
Low Z-shaped [0 0.766682]
Medium Gaussian [0.172605 0.639324]
High S-shaped [0.5 1]


System: allroundercrrFIS_final_tuned
Runs Scored (3 MFs)
Wickets Taken (3 MFs)
Catches Taken (3 MFs)
Stumps Taken (3 MFs)
Mamdani
Type 1
Average Difference (3 MFs)
System allroundercrrFIS_final_tuned: 4 input, 1 output, 106 rules
Name
Catches Taken
Range
[0 1]
Number of MFs: 3
Evenly Distribute MFs
Name Type Parameters
Low Z-shaped [0.764788 0.5]
Medium Gaussian [0.119675 0.5]
High S-shaped [0.5 1]

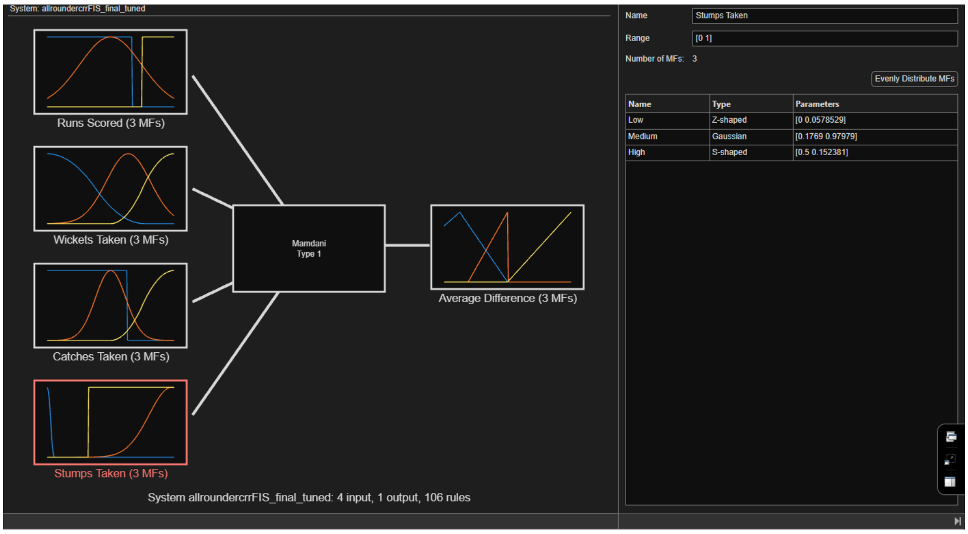


Figure 21. Input membership distribution after Allrounder Performance FIS tuned

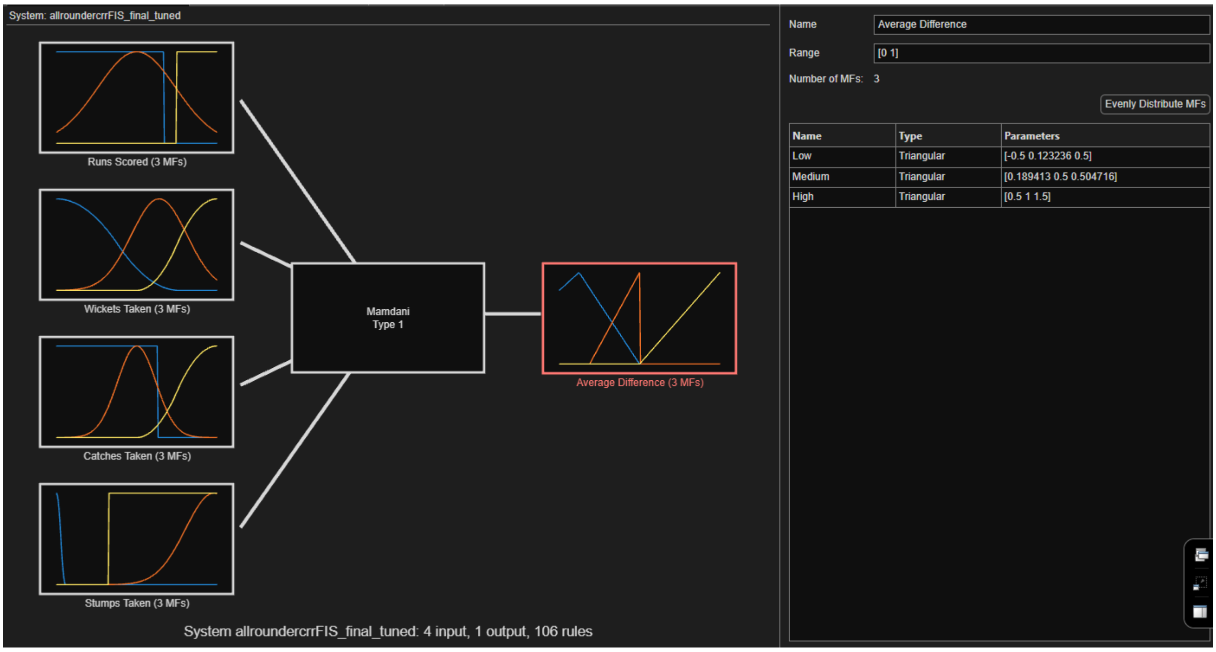


Figure 22. Output membership distribution after Allrounder Performance FIS tuned

The validation accuracy fluctuations of tuned Allrounder Performance FIS are shown in the Figure 23 below.

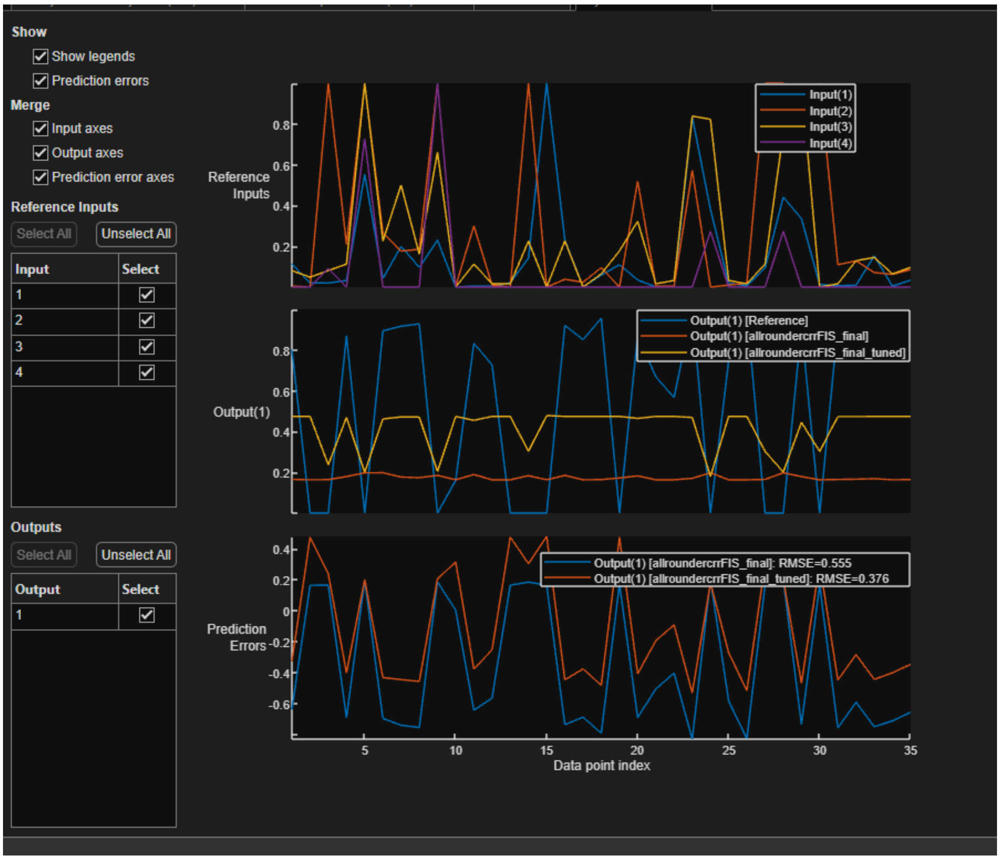


Figure 23. Input membership distribution after Batting Performance FIS tuned

4.3. Summary

This chapter discusses the research findings that were obtained after conducting the systematic research methodology. The results tabulated in this chapter showcase how the research objectives have been achieved in terms of developing models for Team Composition Selection and Player Rating and Ranking Systems. The final TCS model has yielded a 76% of overall accuracy in predicting the number of batsmen, bowlers and allrounders for a match. 75%, 67% and 62% accuracies have been yielded for Batting Performance FIS, Bowling Performance FIS and Allrounder Performance FIS respectively.

CHAPTER 5

CONCLUSION AND FUTURE WORK

This chapter summarizes the work, discusses its findings and contributions, points out limitations of the current work, and also outlines directions for future research.

The research, “Cricket Team Selection Based on Complex Dynamics using Machine Learning” has adopted a data driven approach to predict cricket team composition for a match and rate and rank players in the squad based on their personal historic performances.

5.1. Predicting Team Composition

As discussed in the literature review and as well as considering expert opinions, team composition selection is an intuitive effort that selection committees need to carry out. All most every team intends to win matches, and decisions taken about the team orients to winning. Hence, Team Selection is an important and crucial factor. According to the experts’ experiences there are numerous factors that affecting the outcome of a match and as well as the team selection. The external factors which are not in control of players, such as the weather, pitch condition, crowd support and opponent performance level etc. play a major role in deciding the team selection in cricket.

The research employs a neural network that operates as a multi-output, multi-label system, which can be utilized to develop a decision support system for team selection. The developed model exhibits a commendable accuracy of 76%. Yet, it should be noted that there is a high imbalance in target variables of the prediction task data. This imbalance, inherent to the problem at hand, as the ration among batsmen, bowlers and allrounders tends to fluctuate in the range of four to three to four as shown in the Figure 24.

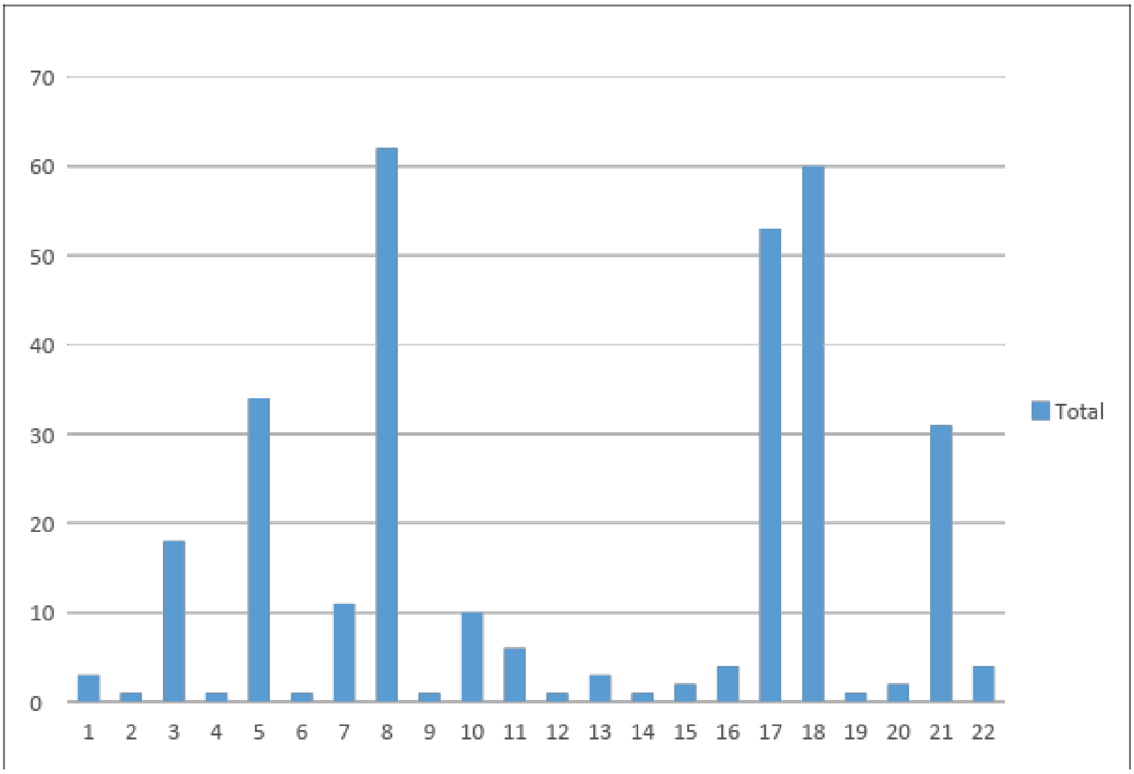


Figure 24. Distribution of Team Combinations

This poses a unique set of challenges, making it difficult to rectify through conventional methods. Despite the inherent data imbalance, the neural network model can be considered to be a robust tool to predict team composition in cricket. Further, the consideration of the performance level of the opposing team has not considered for the model, which is very integral factor that should be considered, as acclaimed by the experts. Yet, experts state that the formulation of the team composition considering these mentioned factors may not be enough as there are unquantifiable factors, such as players' physiological and psychological capacities. Further, the game as a whole is very complex and unpredictive with a high uncertainty. Hence, it can be concluded that the performance of the model built with this research is fairly effective for the prediction of team composition in cricket.

5.2. Player Rating and Ranking Model

An important tool for evaluating the performance of individual cricket players is the player rating and ranking model, which offers a numerical way to calculate each player's contribution to the team. It is imperative to recognize the inherent difficulties that come with player ratings, though, as these evaluations could be prone to prejudices and subjective interpretations. Conditions of play, opponents, and subtleties of the surrounding environment can all affect how a player is regarded to have performed.

The research aims to reduce biases by developing Fuzzy Logic algorithms to tackle the complexity of player rating. A sophisticated method for assessing player performance is provided by fuzzy logic, which is renowned for its capacity to manage ambiguity and imprecision. The study particularly focuses on three important areas: Allrounder Performance, Bowling Performance FIS, and Batting Performance FIS.

5.3. Future Prospects

Looking ahead, there are several future enhancements that can be implemented to this cricket team selection research. Firstly, the integration of advanced machine learning techniques, such as deep learning models, could offer increased accuracy and predictive capabilities. Exploring neural networks with more complex architectures may provide deeper patterns in player performance data, providing an understanding of the factors influencing team success.

Further, addressing the challenge of imbalanced data is crucial for refining the accuracy of predictive models. Future research efforts could explore approaches to handle imbalanced datasets, potentially incorporating advanced resampling techniques or ensemble learning methods to mitigate the impact of data imbalances.

The inclusion of real-time and dynamic factors in the model, such as player injuries, recent form fluctuations, and emerging talent, could enhance the accuracy of team selection process. Integrating these dynamic elements could be achieved through continuous data updates and real-time analytics. Furthermore, collaborations with cricket experts, coaches, and players could provide valuable qualitative insights to complement quantitative data. To sum up, future improvements ought to aim for a more complex and flexible model of squad selection, utilizing state-of-the-art technologies, resolving data imbalances, embracing real-time dynamics, and capitalizing on the abundance of expertise from cricket experts. These developments will further enable selection committees to make strategic and knowledgeable selections for the sport's future while also advancing cricket analytics.

5.4. Summary

This chapter provides concluding remarks and future prospects of the study. This research has achieved four models for decision support in cricket team selection with a robust approach and it should be noted that the domain inherits data imbalance. In future it is considered to explore approaches to limit the data imbalance problem. Further real-time and dynamic factors, such as player injuries and emerging talent, are deemed essential additions continuous data updates and collaboration with cricket experts to provide qualitative insights. The ultimate goal is to create a more intricate and adaptable model for squad selection, thereby empowering selection committees to make strategic and informed decisions in the dynamic landscape of cricket.

APPENDICES - I

Data Scrapping Scripts

````
```python
import requests
from bs4 import BeautifulSoup
import pandas as pd
import numpy as np
```
```python
response =
requests.get("https://stats.espncricinfo.com/sl/engine/records
/averages/batting.html?class=2;current=2;id=8;type=team")
response.ok
```
```python
response.headers['Content-Type']
```
```python
content = response.content
```
```python
read = BeautifulSoup(content, "html.parser")
```
```python
response
```
```python
read.div
```
```python
div = read.find_all("div", class_= "pnl650M")
```
```python
len(div)
```
```python
test_tables = read.table
test_tables
```
```python
all_tables = read.find_all("table" , class_="engineTable")
```
```python
len(all_tables)
```
```python
test_batting = all_tables[0]
````
````

````
```
```python
len(test_batting)
```
```python
col = []
for i in test_batting.find_all("th"):
col.append(i.text)
```
```python
col
```
```python
table_data = test_batting.find_all("td")
table_data
```
```python
table_list = []
start = 0
end = 15
for i in range(37):
Player, Span, Mat, Inns, NO, Runs, HS, Ave, BF, SR, Hundreds,
Fifties, Zeros, Fours, Sixes = table_data[start:end]
table_list.append([Player.text, Span.text, Mat.text,
Inns.text, NO.text, Runs.text, HS.text, Ave.text,
BF.text, SR.text, Hundreds.text, Fifties.text, Zeros.text,
Fours.text, Sixes.text])
start += 15
end += 15
```
```python
table_list
```
```python
odi_battingavg = pd.DataFrame(table_list , columns=col)
```
```python
odi_battingavg
```
```python
odi_battingavg.to_csv("ODI_battingavg.csv",index=False)
```
````
````

```python
import requests
from bs4 import BeautifulSoup
import pandas as pd
import numpy as np
```
```python
response =
requests.get("https://stats.espncricinfo.com/sl/engine/records
/averages/bowling.html?class=2;current=2;id=8;type=team")
response.ok
```
```python
response.headers['Content-Type']
```
```python
content = response.content
```
```python
read = BeautifulSoup(content, "html.parser")
```
```python
response
```
```python
read.div
```
```python
div = read.find_all("div", class_= "pnl650M")
```
```python
len(div)
```
```python
test_tables = read.table
test_tables
```
```python
all_tables = read.find_all("table" , class_="engineTable")
```
```python
len(all_tables)
```
```python
test_batting = all_tables[0]
```
```python
len(test_batting)
```

```python
col = []
for i in test_batting.find_all("th"):
col.append(i.text)
```
```python
col
```
```python
table_data = test_batting.find_all("td")
table_data
```
```python
table_list = []
start = 0
end = 16
for i in range(35):
Player, Span, Mat, Inns, Overs, Mdns, Runs, Wkts, BBI, Ave,
Econ, SR, Fours, Fives, Ct, St = table_data[start:end]
table_list.append([Player.text, Span.text, Mat.text,
Inns.text, Overs.text, Mdns.text, Runs.text, Wkts.text,
BBI.text, Ave.text, Econ.text, SR.text, Fours.text,
Fives.text,
Ct.text, St.text])
start += 16
end += 16
```
```python
table_list
```
```python
odi_bowlingavg = pd.DataFrame(table_list , columns=col)
```
```python
odi_bowlingavg
```
```python
odi_bowlingavg.to_csv("ODI_bowlingavg.csv",index=False)

```

APPENDICES - II

Neural Network Model

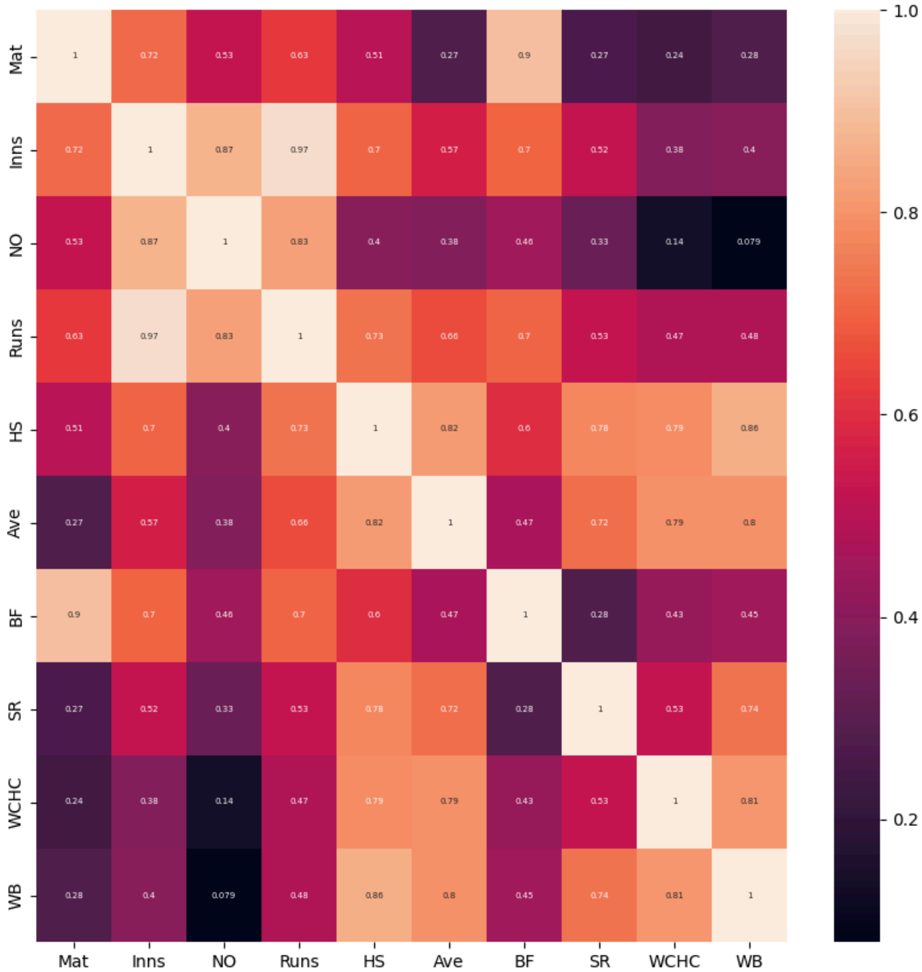
```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
from tensorflow.keras.layers import Dense
from tensorflow.keras.optimizers import SGD
from tensorflow.keras.models import Sequential

import random
random.seed(2)
import tensorflow as tf
tf.random.set_seed(2)
import numpy as np
np.random.seed(2)

from tensorflow.keras.optimizers import Adam
from tensorflow.keras import regularizers
from tensorflow.keras.layers import Dropout

cricdata =
pd.read_csv('/content/drive/MyDrive/Research/CD_Feature_Engine
ered_3.0.csv')

num_columns = cricdata.select_dtypes(include=np.nu
mber).columns
cat_columns = []
for col in cricdata.columns:
     if col not in num_columns:
     cat_columns.append(col)
cat_columns

for cc in cat_columns:
     cricdata[cc] = pd.Categorical(cricdata[cc])
     cricdata[cc] = cricdata[cc].cat.codes

training_dataset = cricdata.sample(frac=0.8)
testing_dataset =
cricdata[~cricdata.index.isin(training_dataset.index)]

cricdata_model = Sequential()
cricdata_model.add(Dense(128, input_dim=10, ctivation='tanh'))
cricdata_model.add(Dense(128, activation='tanh'))
cricdata_model.add(Dense(64, activation='tanh'))
cricdata_model.add(Dense(64, activation='tanh'))
cricdata_model.add(Dense(32,
activation='tanh',kernel_regularizer=regularizers.l2(0.01)))
cricdata_model.add(Dense(16, activation='softmax'))
cricdata_model.add(Dense(3, activation='tanh'))
adam = Adam()
```

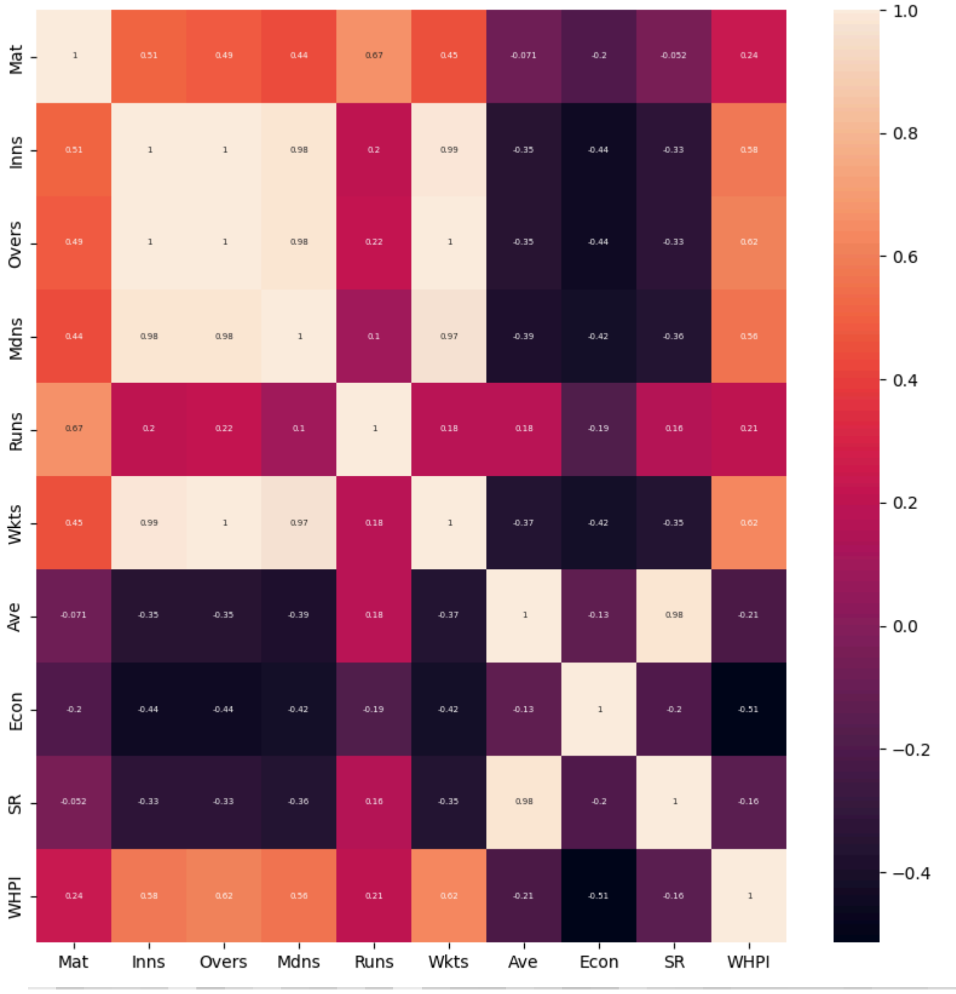
```
batch_size = 50
cricdata_model.compile(loss='categorical_crossentropy',
optimizer=adam, metrics=['accuracy'])
cricdata_model.summary()

history_adam_composition_model =
cricdata_model.fit(training_dataset[['Stadium', 'Stadium
Label', 'Weather', 'Minimum Temperature Label', 'Maximum
Temperature Label',
'Average Wind Speed Label', 'Team 1', 'Team 2', 'Margin',
'Winner',]],
    training_dataset[['Batsman', 'Bowler', 'Allrounder']],
    batch_size=batch_size,
    epochs=75,
    validation_split=0.2,
)

test_loss,                  test_acc                  =
cricdata_model.evaluate(testing_dataset[['Stadium',   'Stadium
Label',  'Weather',  'Minimum  Temperature  Label',  'Maximum
Temperature                                          Label',
'Average  Wind  Speed  Label',  'Team  1',  'Team  2',  'Margin',
'Winner',]],           testing_dataset[['Batsman',   'Bowler',
'Allrounder']])

print(f"Evaluation  result  on  Test  Data  :  Loss  =  {test_loss},
accuracy = {test_acc}")

cricdata_model.save('/content/drive/MyDrive/Research/models/TC
S/cricdata_model_0.75_tanh.keras')
```

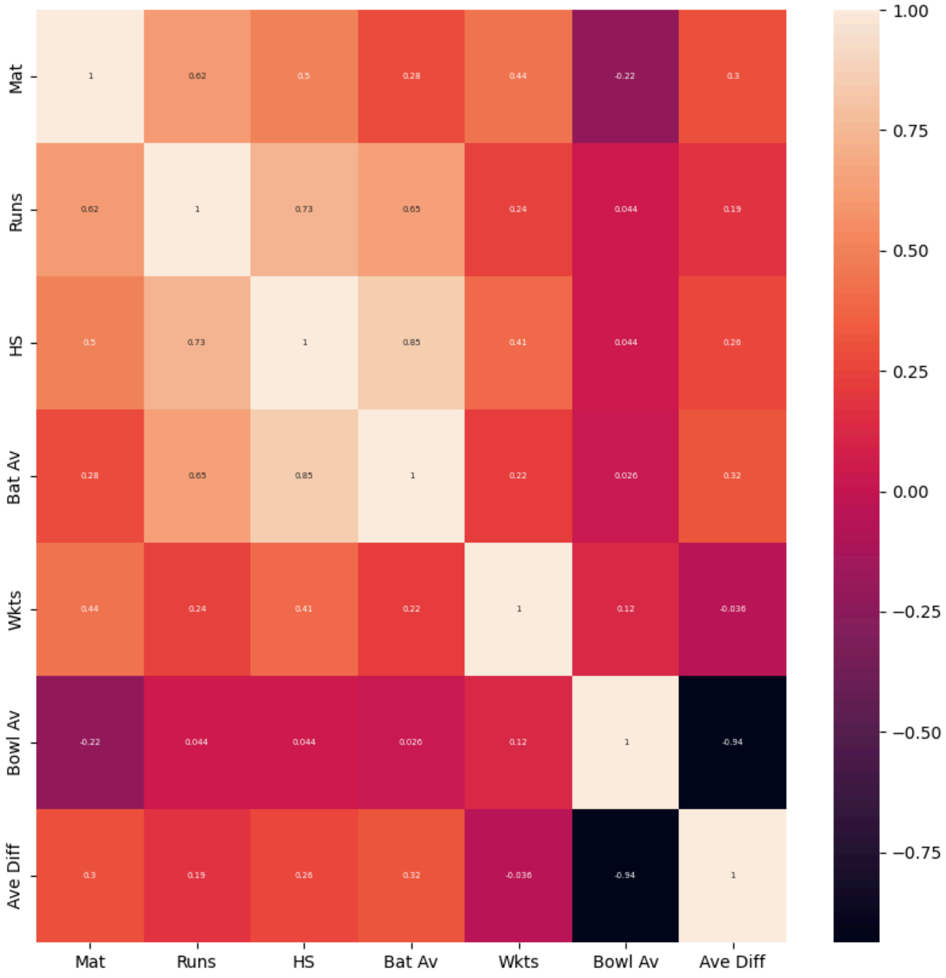
APPENDICES - III

Correlation Analysis on Player Performance Data

Following heatmaps depicts correlation results conducted for player performance data in batting performance, bowling performance and allrounder performance.




Mat
Inns
Overs
Mdns
Runs
Wkts
Ave
Econ
SR
WHPI
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4


Mat
Runs
HS
Bat Av
Wkts
Bowl Av
Ave Diff
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
1 0.62 0.5 0.28 0.44 -0.22 0.3
0.62 1 0.73 0.65 0.24 0.044 0.19
0.5 0.73 1 0.85 0.41 0.044 0.26
0.28 0.65 0.85 1 0.22 0.026 0.32
0.44 0.24 0.41 0.22 1 0.12 -0.036
-0.22 0.044 0.044 0.026 0.12 1 -0.94
0.3 0.19 0.26 0.32 -0.036 -0.94 1

REFERENCES

[1] I. C. C. ICC, “About ICC Cricket | International Cricket Council.” [Online]. Available: https://www.icc-cricket.com/about/cricket/history-of-cricket/early-cricket

[2] I. C. C. ICC, “The three formats of cricket,” ICC. [Online]. Available: https://www.icc-cricket.com/about/cricket/game-formats/the-three-formats

[3] R. Palace, “Powerplay In Cricket Everything You Need To Know,” *R. Palace*, 2022, [Online]. Available: https://rp777news.com/powerplay-in-cricket/

[4] R. M. Silva, A. B. W. Manage, and T. B. Swartz, “A study of the powerplay in one-day cricket,” *Eur. J. Oper. Res.*, vol. 244, no. 3, pp. 931–938, 2015, doi: 10.1016/j.ejor.2015.02.004.

[5] M. Tavana, F. Azizi, F. Azizi, and M. Behzadian, “A fuzzy inference system with application to player selection and team formation in multi-player sports,” *Sport Manag. Rev.*, vol. 16, no. 1, pp. 97–110, 2013, doi: 10.1016/j.smr.2012.06.002.

[6] F. Ahmed, K. Deb, and A. Jindal, “Evolutionary Multi-Objective Optimization and Decision Making Approaches to Cricket Team Selection,” *Proc. Second Int. Conf. Swarm Evol. Memetic Comput.*, pp. 1–23, 2011.

[7] I. Preston and J. Thomas, “Batting strategy in limited overs cricket,” *J. R. Stat. Soc. Ser. Stat.*, vol. 49, no. 1, pp. 95–106, 2000, doi: 10.1111/1467-9884.00223.

[8] G. R. Amin and S. K. Sharma, “Cricket team selection using data envelopment analysis,” *Eur. J. Sport Sci.*, vol. 14, no. sup1, pp. S369–S376, 2014, doi: 10.1080/17461391.2012.705333.

[9] M. F. Fiander, J. Stebbings, M. C. Coulson, and S. Phelan, “The information coaches use to make team selection decisions: a scoping review and future recommendations,” *Sports Coach. Rev.*, vol. 00, no. 00, 2021, doi: 10.1080/21640629.2021.1952812.

[10] S. Trninic, V. Papic, V. Trninic, and D. Vukicevic, “Player selection procedures in team sports games,” *Acta Kinesiol.*, vol. 2, no. February, pp. 24–28, 2008.

[11] ngel Pé rez-Toledano, F. J. Rodriguez ID, J. García-Rubio, and S. José Ibañez, “Players’ selection for basketball teams, through Performance Index Rating, using multiobjective evolutionary algorithms,” 2019, doi: 10.1371/journal.pone.0221258.

[12] F. Ahmed, A. Jindal, and K. Deb, “Cricket Team Selection Using Evolutionary Multi-objective Optimization,” pp. 71–78, 2011.

[13] V. Sarda, P. Sakaria, and P. K. Deulkar, “Football Team Selection Using Genetic Algorithm,” *Int. J. Eng. Tech. Res. IJETR*, vol. 3, no. 2, pp. 153–156, 2015.

[14] S. S. Sathya and M. S. Jamal, *Applying Genetic Algorithm to Select an Optimal Cricket Team*. 2009.

[15] I. H. Kusumsiri and S. Perera, “Optimal One day International Cricket Team Selection by Genetic Algorithm,” *Conf. Pap.*, vol. 4531, no. March, pp. 0–1, 2017.

[16] S. M.AqilBurney, N. Mahmood, K. Rizwan, and U. Amjad, “A Generic Approach for Team Selection in Multiplayer Games using Genetic Algorithm,” *Int. J. Comput. Appl.*, vol. 40, no. 17, pp. 11–17, 2012, doi: 10.5120/5071-7440.

[17] S. Gursgaran, B. Nitin, and S. Sawtantar, “Fuzzy Logic based Cricket Player Performance Evaluator | Semantic Scholar,” *Int. J. Comput. Appl.*, pp. 11–16, 2011.

[18] P. Premkumar, J. B. Chakrabarty, and S. Chowdhury, “Key performance indicators for factor score based ranking in One Day International cricket,” *IIMB Manag. Rev.*, vol. 32, no. 1, pp. 85–95, 2020, doi: 10.1016/j.iimb.2019.07.008.

[19] W. Sałabun *et al.*, “A Fuzzy Inference System for Players Evaluation in Multi-Player Sports: The Football Study Case,” *Symmetry*, vol. 12, no. 12, pp. 2029–2029, 2020, doi: 10.3390/sym12122029.

[20] S. R. Iyer and R. Sharda, “Prediction of athletes performance using neural networks: An application in cricket team selection,” *Expert Syst. Appl.*, vol. 36, no. 3, Part 1, pp. 5510–5522, Apr. 2009, doi: 10.1016/j.eswa.2008.06.088.

[21] M. Shetty, S. Rane, C. Pandita, and S. Salvi, “Machine learning-based Selection of Optimal sports Team based on the Players Performance,” presented at the 2020 5th International Conference on Communication and Electronics Systems (ICCES), IEEE, 2020, pp. 1267–1272. doi: 10.1109/ICCES48766.2020.9137891.

[22] V. Punjabi, R. Chaudhari, D. Pal, K. Nhavi, N. Shimpi, and H. Joshi, “a Survey on Team Selection in Game of Cricket Using Machine Learning,” *Int. Res. J. Eng. Technol.*, pp. 2442–2446, 2019.

[23] R. Al-Shboul, T. Syed, J. Memon, and F. Khan, “Automated Player Selection for Sports Team using Competitive Neural Networks,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 8, pp. 457–460, 2017, doi: 10.14569/IJACSA.2017.080859.