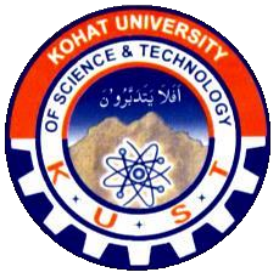
Leveraging Deep Neural Network for Selecting Players of Pakistan Super League on the Root of their Enactment Prediction.

By

Qazi Muzzamil Shah



Institute of Computing

Kohat University of Science & Technology, Kohat-26000

Khyber Pakhtunkhwa, Pakistan

2023

Leveraging Deep Neural Network for Selecting Players of Pakistan Super League on the Root of their Enactment Prediction

Submitted in partial fulfillment of the requirements for the degree of

MS in Computer Science

By

Qazi Muzzamil Shah

(CS320192019)

Institute of Computing

Kohat University of Science and Technology, Kohat-26000

Khyber Pakhtunkhwa, Pakistan 2023

CERTIFICATION FROM THE SUPERVISORS

This thesis entitled **Leveraging Deep Neural Network for Selecting Players of Pakistan Super League on the Root of their Enactment Prediction** submitted by Mr **Qazi Muzzamil Shah** to the Kohat University of Science & Technology for the award of **MS(CS) IN COMPUTER SCIENCE** presents bonafide research work carried out under our supervision. This work (in full or in part) has not been submitted to any other Institution for award of any degree/ diploma/certificate.

Supervisor-I Dr. M. Ufaq Uddin
(Name)

(Signature)

Institute of Computing, KUST

Assistant Professor

Supervisor-II Dr. Muhammad Adnan
(Name)

(Signature)

Institute of Computing, KUST

Lecturer

CERTIFICATION FROM THE EXAMINERS

This thesis entitled **Leveraging Deep Neural Network for Selecting Players of Pakistan Super League on the Root of their Enactment Prediction** submitted presents a bonafide record of original research work carried out by Mr. **Qazi Muzzamil Shah** in partial fulfillment of the degree of **MS(CS) IN COMPUTER SCIENCE**, Kohat University of Science & Technology, Kohat. We find the work satisfactory for the award of the degree if other requirements are met. The

Viva Voce was held on._____

Internal Examiner Dr. Muhammad Roman

(Name)

(Signature)

Lecturer, Institute of Computing, KUST

(Affiliation)

External Examiner: Dr. Fakhre Alam

(Name)

(Signature)

Assistant Professor, Dept. of CS & IT, University of Malakand

(Affiliation)

Chairman/Director: Prof. Dr. Shafiullah Khan

(Name)

(Signature)

KOHAT UNIVERSITY OF SCIENCE & TECHNOLOGY

Kohat 26000, Khyber Pakhtunkhwa, Pakistan Ph # 0922-554563-554565/4786, 4785, Fax # 554556

Directorate of Quality Enhancement

No. 796 /KUST/QEC/PC/22
Date: 21/07/2023

Anti-Plagiarism Test Certificate

(Similarity Index Certificate)

For M.Phil./MS/Ph.D Thesis/Synopsis

Name of Scholar:	QAZI MUZZAMIL SHAH
Discipline (M.Phil./MS/Ph.D.):	MS
Department/Institute:	INSTITUTE OF COMPUTING
Title of Synopsis/Thesis:	LEVERAGING DEEP NEURAL NETWORK FOR SELECTING PLAYERS OF PAKISTAN SUPER LEAGUE ON THE ROOT OF THEIR ENACTMENT PREDICTION
Document Type (Synopsis/Thesis)	Thesis
Words Count	12106
Name and Designation of Supervisor	DR. IRFAN UD DIN, ASSISTANT PROFESSOR , INSTITUTE OF COMPUTING, KUST
Plagiarism software (Similarity Index Checking)	Turnitin.com
Software Generated ID for the document and Date	2134425201 *(Report generated on 21 July., 2023, 10: 56 PKT evidence attached)*
Contents Excluded from the document	Table of Contents, References and other Preliminary Pages
Matching (Similarity) Index found	**13% (Thirteen Percent only)**
Findings/Comments by QEC	The Similarity index of the document lies within the acceptable range (less than or equal to 19%) set by Higher Education Commission (HEC) Pakistan. The document is **Not Plagiarized** on the basis of similarity index.

Prepared by: Naimat Khan, **Office Assistant** 21/9/23

Checked by: Dr. Faisal Mehmood, **Asst. Director**

Dr. Muhammad Zeeshan Bangash
Director Quality Enhancement Cell

Director KUST QEC

Note: *Technical Errors and Omissions are subject to be rectified.*

Turnitin user: datanalyst.qecac@kust.edu.pk

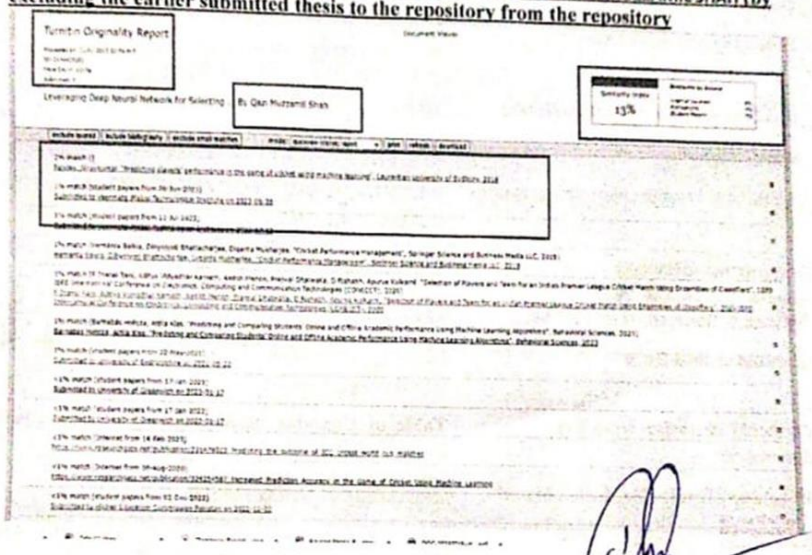
KOHAT UNIVERSITY OF SCIENCE & TECHNOLOGY

Kohat 26000, Khyber Pakhtunkhwa, Pakistan Ph # 0922-554561-554565/4786, 4785, Fax # 554556

Directorate of Quality Enhancement

Earlier Similarity Index Certificate issued vide QEC Letter no: 29 KUST/QEC/ /APC/22/ dated January 06, 2023, (copy attached) with similarity index 6%

Similarity index Certificate after Mphil defense of the scholar QAZI MUZZAMIL SHAH (by excluding the earlier submitted thesis to the repository from the repository



Director Quality Enhancement Cell
KUST
24/7/2023

Director KUST QEC

Note: *Technical Errors and Omissions are subject to be rectified.*

Turnitin user: *dataanalyst.qecoc@kust.edu.pk*

ACKNOWLEGEMENT

All glory be to **Allah**, the Lord of the Worlds, praise be to the numerous amounts of His creatures that supported me in completing this thesis, enormous as the gravity of His crown and enormous as the print required for the writing of His Words. "**O,Allah**! I'm at a loss for words when it comes to expressing my gratitude to you. You are what you think of yourself to be. The completion and eventual result of my MSCS research study necessitated a great deal of help and direction from a large number of people. I consider myself quite fortunate to have received this during the course of my studies. Everything I have accomplished has been made possible by their direction and aid, and I would like to express my gratitude to them.

First and foremost, I would want to convey my heartfelt admiration to Assistant Professor **Dr. Irfan Uddin**, my thesis supervisor. He had been a consistent source of inspiration and support for me to complete my degree, and he had been a leading force for me to complete my research work. Throughout the course of this thesis, he went above and above his duty to give me with direction, support, and critique. I am honoured to have been one of his students. I would like to thank my co- supervisor, Assistant Professor **Dr. Muhammad Adnan**, for his help. He had also provided quick feedback, encouraged me, and directed me during my research. Furthermore, I am grateful to **Dr. Shafiullah**, the Director of my Institute, for his unwavering support and timely direction in completing the graduation requirements throughout my MSCS degree. I would also like to express my gratitude to my entire family, particularly my father, for their persistent blessings and quite well sentiments.

Qazi Muzzamil Shah

DEDICATED

TO

"The Teacher of the Universe (PBUH)" &

TO

MY LOVING PARENTS

The one and only reason of what I became today.

Thanks for your great support and continuous care.

And I am honoured to have you as my parents.

Without you none of my success would be possible.

Having parents like you is a precious gift of Allah.

MY HONORABLE CO-SUPERVISOR

Dr. Muhammad Adnan *for his encouragement, support, invaluable suggestions and guidance.*

Qazi Muzzamil Shah

ABSTRACT

Due to the play's briefness and rapid pace, Twenty20 is the adored format of cricket sport. Pakistan Super League (PSL) is one of the most well-known professional T20 cricket leagues. It was created to improve Pakistan cricket by evaluating emerging talent. However, choosing the greatest players for PSL teams is a crucial step that affects the game's outcomes. Players' performances differ according to a variety of conditions, including the opponent team, the atmosphere, their present form, etc. To prevent selection processes from being biased due to human nature, using a machine learning method, this study seeks to identify and rank the best fifteen players on the PSL squad based on their bowling and batting statistics over the previous five seasons. For this purpose, we used Naïve Bayes, Random Forest, Support Vector Machine, and Decision Tree (classifier) are used to create a prediction model using the individual batting and bowling feature sets. The confusion matrix conclusions have been plotted in terms of precision accuracy, recall, and "f1score" based on a comparison of the algorithms used. To choose both pitchers and batsmen based on runs scored (in terms of wickets taken), Random Forest proved to be the most reliable classifier for both datasets, predicting batsman runs scored by 90.74% and bowler wickets taken by 92.25%, respectively. This study's conclusions also show that bowling performance lags behind batting performance.

Table of Contents

Chapter 1.....**13**
1. Introduction 13
1.1. Formats in cricket ……..……..01
2. 1.2. Importance of domestic league in Pakistan ……..01
3. 1.3. Pakistan Super League……...02
4. 1.4. Motivation…….02
5. 1.5. Problem Statement…….....03
6. 1.6. Aim of Research……03
7. 1.7. Objectives ……....03
8. 1.8. Contributions ……..…..04
9. 1.9. Thesis Organization……..…..04
Chapter 2……..…...**05**
1. 2.1. Back Ground Study ……..……..05
2. 2.2. Literature Review ……..……...06
Chapter 3……..……**13**
3.1. Methodology...…….…13
3.2. Introduction ……..……..….13
3.3. Data and Preprocessing……...….14
3.3.1. Data Collection……....….14
3.3.2. Player Statistics……....15
3.3.3. Batting Measures……..….16
3.3.4. Bowling Measures……..……...…17
3.4. Pre-Processing of Dataset…….…17
3.4.1. Weight Combination…….…18
3.5. Additional Features……..……21
3.5.1. Evaluation of Common Measures……...……...23
3.5.2. Further Input Parameters……...23
3.5.3. Data Cleaning……24
3.5.4. Removing Unwanted Columns...…….....24
3.5.5. Placing Unique Attributes……...……..24
3.5.6. One Hot Encoding……24
3.5.7. Data Transformation……..24
3.6. Feature Selection……..25
Chapter 4.....29
4.1. Result and Discussion…….…29
4.2 Machine Learning Algorithms……29
4.2.1 ML & DL……29
4.2.2 Random Forest……29
4.2.3 Naïve bayes……..30
4.2.4 Decision Tree……30

4.3 EVALUATION PARAMETERS……31
4.3.1. Precision…..……..…….……31
4.3.2. Accuracy…….……...…31
4.3.3. Recall……..…….....…31
4.3.4. F1 Score…….…..….32
4.4. Random Forest……..……..…..32
4.5. Support Vector Machine……..……....33
4.6. Decision Tree…….…..34
4.7. Naïve Bayes…….36
4.8. Summary…….……37
4.9. Batting Analysis……..…….....42
4.10. Batting Index…….....43
4.11. Bowling Analysis……44
4.12 Bowling Index..…….....45
Chapter 5.....**.46**
5.1. Conclusion and Future work……..……..46
5.2. Conclusion…….…47
5.3. Limitation……..……....47
5.4. Future Work……..…….....47
Chapter 6.....**48**
6.1. References……..…….....50

List of Tables

Table 1 Summary of Related Work……12

Table 2 Levels of relative importance for the goals and features…… 19

Table 3 Similarity-based evaluation of the features…….……..…….. 20

Table 4 Batting and Bowling Attributes…….……26

Table 5 Run prediction accuracy with Random Forest…….……... 32

Table 6 Run prediction evaluation of Random Forest with 10% testing set, 90% training set……32

Table 7 Wicket prediction accuracy with Random Forest…….…….. 32

Table 8 Evaluation of Random Forest for predicting Wickets using 10% testing data and 90% training data……..……... 33

Table 9 Run prediction accuracy with SVM…….……..…… 33

Table 10 Performance of SVM for predicting Wickets using 10% testing data and 90% training data…… 33

Table 11 Demonstrate that Support Vector Machine prediction accuracy for Wickets with various training. and test set size……... 34

Table 12 Evaluation of SVM for Predicting Wickets with 10% Test Data and 90% Training Data……... 34

Table 13 Run prediction accuracy using a Decision tree…….…… 34

Table 14 Decision tree performances for predicting runs using 10% testing data and 70% training data…….. 35

Table 15 Indicates that Decision Tree accuracy for predicting Wickets with various training and test set size.. 35

Table 16 Evaluation of Decision Tree for Predicting Wickets with 10% Test Data and 90% Training Data……35

Table 17 Run prediction accuracy with Naive Bayes…….……..…….. 36

Table 18 Performance of Nave Bayes for predicting wickets using 10% testing data and 90% train data…….. 36

Table 19 Naive Bayes wickets prediction accuracy for different training and test set sizes…… 36

Table 20 Nave Bayes' performance in predicting wickets using 10% test data and 90% train data…… 37

Table 21 Summary…….37

Table 22 performance evaluation of run prediction algorithms……

Table 23 Efficiency of such classifiers used to predict wickets…….…….. 41

List of Figures

Figure 1 Graph of the D/L model…… 7

Figure 2 Proposed methodology……..... 13

Figure 3 Batsmen and Bowler's dataset…… 15

Figure 4 Descriptive analysis of top batsmen in PSL…….26

Figure 5 Top Batsman Based on Feature Score……29

Figure 6 A qualitative investigation of the PSL's Top bowlers…….30

Figure 7 Accuracy for top batsmen using Machine learning Algorithm……...38

Figure 8 Accuracy for top bowlers using Machine learning Algorithm……..41

Figure 9 Machine Learning Evaluation of Recall Precision and F1 Score for Top PSL bowlers.....42

Figure 10 Top Batsman Based on Feature Score…… 43

Figure 11 Batting index of top batsmen……..44

Figure 12 Top Bowlers Based on Feature Score……. 44

Figure 13 Bowling index of PSL players…… 45

CHAPTER 1

INTRODUCTION

1.1 Formats in Cricket

Cricket is played in three different formats at the international level, for example, T20 (Twenty20), one-day, and test matches. Each side has one opportunity to bat and one option to bowl during the two innings in these formats. A total of fifty over's can be bowled in one session during one-day international matches. The newest and shortest format of cricket, T20, only allows for 20 overs. The five-day tournament (known as a test) came first in the cricket matches before the one-day game was added. In 2003, the England and Wales Cricket Board (ECB) introduced a format known as Twenty20. (Or T20 cricket). Since then, five world cups and numerous leagues have been played. The Australian BBL Big Bash T20 League and the IPL (Indian premier league) are considered the first successful cricket leagues, respectively. Supporters from one country cheered for a cricket player from another. The potential to compete with other international athletes and evolve from being a national hero to an international celebrity has now been provided. There are numerous well-known and popular T20 leagues, including the (BPL) Bangladesh Premier League, (BBL) Big Bash League, (IPL) Indian Premier League, (PSL) Pakistan Super League, and (CPL) Caribbean Premier League.[1] [2]

1.2 Importance of domestic league in Pakistan

Cricket is the most widely played sport in Pakistan, where the majority of the population is sports-loving for its Field Hockey quash players and cricket greater than 190 million people living there. The Sri Lankan cricket team was the target of extremists on March 3, 2009. The sad incident effectively ended Pakistan's participation in international cricket. As a result of the attack, the International Cricket Council stripped Pakistan of its ability to host the 2011 World Cup. All games were played in the United Arab Emirates until the championship game in 2017, which served as the final test to determine whether top-level cricket could return to Pakistan, due to safety concerns resulting from a terrorist attack in Lahore, Pakistan, in 2009. The Pakistan Super League (PSL) is a significant step in reviving cricket in Pakistan. It is an effective effort to revive international cricket and gives young, skilled players a chance to develop.[3] [5]