

TO STUDY THE PERFORMANCE OF THE PLAYERS IN THE IPL SEASONS USING PREDICTIVE ANALYTICS

*Project submitted to the Bharathidasan University, Tiruchirappalli
in partial fulfilment of the requirements for the award of the degree of*

MASTER OF BUSINESS ADMINISTRATION

Submitted by

Chandhu Prakash V

D. No- 21PBA205

Under the Guidance of

Dr. S. SURESH, M.B.A., M.Sc. (CS)., PGDCA., M.Phil., Ph.D.,



ST. JOSEPH'S INSTITUTE OF MANAGEMENT

A Jesuit Business School

St. Joseph's College (Autonomous)

Tiruchirappalli - 620 002, India.

AUGUST 2022

DECLARATION

I, Chandhu Prakash V (21PBA205), hereby declare that the project entitled, “**To study the performance of the players in the IPL seasons using predictive analytics**” is a bonafide record of research work done by me under the supervision of **Dr. S. Suresh**, Assistant Professor, St. Joseph’s Institute of Management, St. Joseph’s College (Autonomous), Tiruchirappalli District-620 002, Tamil Nadu, India, for the PG Programme in MBA at St. Joseph’s Institute of Management, St. Joseph’s College (Autonomous), Tiruchirappalli. I further declare that this work has not been previously formed the basis for the award of any degree, diploma, associate-ship or other similar title.

Place: Tiruchirappalli - 2

Date:

Signature

Chandhu Prakash V

CERTIFICATE

This is to certify that the project entitled **To study the performance of the players in the IPL seasons using predictive analytics** is a record of research work done by Mr. Chandhu Prakash V (D. No.21PBA205) during 2021-2023 at St. Joseph's Institute of Management, St. Joseph's College (Autonomous), Tiruchirappalli, Tamil Nadu, India. This project has not been previously formed the basis for the award of any degree, diploma, associateship, fellowship or any other similar title to anyone.

Dr. S. Suresh
Project Guide

Dr. P. Jega Patrick
Dean Academics

Rev. Dr. P. Paulraj, SJ
Director

Place: Tiruchirappalli - 2

Date:

VIVA VOCE Examination has been conducted on:

Internal Examiner

External Examiner

ACKNOWLEDGEMENT

I commence with the holy name of Almighty God whose benevolence and beneficence enabled me to complete this project.

I express my heartfelt thanks to our **Director Rev. Dr. P. Paulraj**, SJ, St. Joseph's Institute of Management, St. Joseph's College (Autonomous), Tiruchirappalli, District - 620 002, Tamil Nadu, India, for his constant encouragement in completing this project.

I would like to thank to our **Administrator Rev. Fr. I. Antony Inico**, SJ, St. Joseph's Institute of Management, St. Joseph's College (Autonomous), Tiruchirappalli, District - 620 002, Tamil Nadu, India, for his continual encouragement in completing this project.

I would like to express my sincere thanks to our **Dean (Academics) Dr P Jega Patrick**, St. Joseph's Institute of Management, St. Joseph's College (Autonomous), Tiruchirappalli District - 620 002, Tamil Nadu, India for his motivation to do better every day.

I earnestly express my sincere gratitude to my beloved guide **Dr S Suresh, Assistant Professor**, St. Joseph's Institute of Management, St. Joseph's College (Autonomous), Tiruchirappalli District - 620 002, Tamil Nadu, India, for his able guidance and persistent help throughout this project.

Abstract

IPL is the most famed T20 cricket celebration in the world in which 8 teams give their Best Performance to reach the top squad in the tournament. In such a contest there are many players from different nations playing for different teams. As we know that only a certain number of players can play one match, so there's a problem for squad selection to choose the decent combination of players for the match. In this paper, we're calculating the Most Valuable Player (MVP) by using the Player Batting Points (PBT) and Experience, IPL Stats is used to classify the Matches Played into victories and lost. Further, formerly again MVP is calculated by the Player Bowling Points (PWT), and experience is used for the selection of bowlers, Co-variance analysis is used to find the similarity among players. Eventually, multiple regression is used to analyze the statistics of players by their strike rate, runs, and total score. After opting the best players, we're forecasting individual strike rates with total team scores. This paper is going to give them a result to shutting out non-performing players using the customized techniques of their performance analysis in earlier matches, assembling a decent player for any match using revolutionary techniques, and deciding batting order in an effective manner.

Keywords:

MVP (Most Valuable Player), IPL stats, Co-variance Analysis, Multiple Regression, Score Prediction.

CONTENTS

Chapter	Title	Page
	Title page	
	Declaration	ii
	Certificate	iii
	Acknowledgement	iv
	Abstract	v
	Contents	vi
	List of tables	vii
	List of Charts	viii
I	Introduction to the Study and Profile of the Industry & the Company	1
II	Review of Literature	11
III	Research Methodology	16
IV	Data Analysis and Interpretation	20
V	Findings, Recommendations and Conclusion	45
	References	i

List of Tables

No.	Title	Page
4.1	IPL Stats	20
4.2	MVP Calculation for Batsman	23
4.3	MVP Calculation for Bowlers	25
4.4	Co-Variance Analysis of KL Rahul and Kane Williamson	27
4.5	Co-Variance Analysis of Andre Russell and Rishabh Pant	29
4.6	Last 10 IPL Matches of Andre Russell	31
4.7	Last 10 IPL Matches of Rishabh Pant	33
4.8	Last 10 IPL Matches of KL Rahul	35
4.9	Last 10 IPL Matches of David Warner	37
4.10	Last 10 IPL Matches of Ishan Kishan	39
4.11	Last 10 IPL Matches of Shreyas Gopal	41
4.12	Top 5 Players Predicted Score for IPL 2022	43

List of Charts

No.	Title	Page
4.1	Clustered column chart showing IPL Stats	22
4.2	Clustered column chart showing MVP of Top 10 Batsman	24
4.3	Clustered column chart showing MVP of Top 10 Bowlers	26
4.4	Clustered column chart showing Predicted Score of Top 5 Players	44

CHAPTER I

1 INTRODUCTION

Indian Premier League (IPL) is a Twenty20 (T20) cricket game started by BCCI (Board of Control for Cricket in India) in 2008, which is held annually in the month of April-June. The first season of IPL was financed by DLF, which is a leading real estate company in India. The original season of the event took place from 18 April-1 June 2008. The challenge is that during an IPL purchase only named players can play the justice match; hence the team holder must opt the optimal combination of players. As of now, there is no fool- substantiation result to this challenge nor is there any result ranging from the selection of bowlers to the selection of batsmen maintaining experience within the team. A possible methodology consists of trying to choose the swish payer to buy among all the actors in an IPL sale for the team using a measure called MVP. MVP is dynamic in nature, which implies that comparison criteria change over the proceedings of the sale. We also classify players according to their complete performance measured points, called TCP. Also, we essay to choose the swish set of bowlers and batsmen using the MVP generality. We essay to decide the swish fur order using friction analysis and using some interesting measures, we predict the results for the game.

In a new departure for the cricketing world, players were assigned to armies primarily through a sale, which makes it possible directly to observe the valuations placed on individual players. Using this data, information on the former performance, experience, and other characteristics of individual players, we are suitable to explore the determinants of valuations and inquiry a number of suppositions related to the design of the sale. Presently, with eight armies, each team plays each other twice in a home-and-down round-robin format in the league phase. At the conclusion of the league stage, the top four armies will qualify for the playoffs. The top two armies from the league phase will play against each other in the first Qualifying match, with the winner going straight to the IPL final and the clunker getting another chance to qualify for the IPL final by playing the alternate Qualifying match. Meanwhile, the third and fourth place armies from the league phase play against each other in an eliminator match, and the winner

from that match will play the clunker from the first Qualifying match. The winner of the alternate Qualifying match will move onto the final to play the winner of the first Qualifying match in the IPL Final match, where the winner will be crowned the Indian Premier League titleholders.

Currently, with eight teams, each team plays each other twice in a home-and-away round-robin format in the league phase. At the conclusion of the league stage, the top four teams will qualify for the playoffs. The top two teams from the league phase will play against each other in the first Qualifying match, with the winner going straight to the IPL final and the loser getting another chance to qualify for the IPL final by playing the second Qualifying match. Meanwhile, the third and fourth place teams from league phase play against each other in an eliminator match and the winner from that match will play the loser from the first Qualifying match. The winner of the second Qualifying match will move onto the final to play the winner of the first Qualifying match in the IPL Final match, where the winner will be crowned the Indian Premier League champions.

A team can acquire players through any of the three ways: the annual player auction, trading players with other teams during the trading windows, and signing replacements for unavailable players. Players sign up for the auction and also set their base price, and are bought by the franchise that bids the highest for them. Unsold players at the auction are eligible to be signed up as replacement signings. In the trading windows, a player can only be traded with his consent, with the franchise paying the difference if any between the old and new contracts. If the new contract is worth more than the older one, the difference is shared between the player and the franchise selling the player. There are generally three trading windows—two before the auction and one after the auction but before the start of the tournament. Players cannot be traded outside the trading windows or during the tournament, whereas replacements can be signed before or during the tournament.

IPL games utilize television timeouts and hence there is no time limit in which teams must complete their innings. However, a penalty may be imposed if the umpires find teams misusing this privilege. Each team is given a two-and-a-half-minute "strategic timeout" during each

innings; one must be taken by the bowling team between the ends of the 6th and 9th overs, and one by the batting team between the ends of the 13th and 16th overs.

Since the 2018 season, the Umpire Decision Review System is being used in all IPL matches, allowing each team one chance to review an on-field umpire's decision per innings.

The 2019 season of the IPL offered a total prize money of ₹500 million (equivalent to ₹530 million or US\$7.0 million in 2020), with the winning team netting ₹200 million (equivalent to ₹210 million or US\$2.8 million in 2020). The first and second runners up received ₹125 million (US\$1.7 million) and ₹87.5 million (US\$1.2 million), respectively, with the fourth placed team also winning ₹87.5 million (US\$1.2 million). The other teams are not awarded any prize money. The IPL rules mandate that half of the prize money must be distributed among the players.

Out of the thirteen teams that have played in the Indian Premier League since its inception, one team has won the competition five times, one team has won the competition four times, one team has won the competition twice and three other teams have won it once. Mumbai Indians is the most successful team in league's history in terms of the number of titles won. The Chennai Super Kings have won 4 titles, the Kolkata Knight Riders have won two titles, and the other three teams who have won the tournament are the Deccan Chargers, Rajasthan Royals and Sunrisers Hyderabad. The current champions are the Chennai Super Kings who defeated the Kolkata Knight Riders by 27 runs in the final of the 2021 season securing their fourth title.

1.1 Introduction

In this section, there is an understanding of the techniques that have used in this study. The discussed PBT (Player Batting Points), PBW (Player Bowling Points), MVP (Most Valuable Player), Co-Variance Analysis, Multiple Regression. Along with each of them provided are the implementation details of the models, such as the cross validation, tables, charts, etc. used for the prediction of results. Throughout this section we try and maintain a balance between the intuitive understanding and the mathematical formulation, though the former overshadows the other in certain cases for better expression of ideas.

1.2 Concepts Used and their Definitions

1.2.1 Data Collection

Gathering data can be fulfilled through a primary source (the experimenter is the first person to gain the data) or a secondary source (the experimenter obtains the data that has formerly been collected by other sources, similar as data circulated in a scientific journal). Data analysis methodologies vary and include data triangulation and data percolation. The ultimate offers an eloquent system of collecting, classifying, and assaying data using five possible angles of analysis (at least three) to maximize the exploration's neutrality and permit an understanding of the marvels under disquisition as complete as possible qualitative and quantitative styles, literature reviews (including scholarly papers), interviews with experts, and computer simulation. The data are later "oozed" using a series of pre-determined way to prize the most applicable information. Regardless of the field of study or preference for defining data (quantitative or qualitative), accurate data collection is essential to maintain research integrity. The selection of appropriate data collection instruments (existing, modified, or newly developed) and delineated instructions for their correct use reduce the likelihood of errors. A formal data collection process

is necessary as it ensures that the data gathered are both defined and accurate. This way, subsequent decisions based on arguments embodied in the findings are made using valid data. The process provides both a baseline from which to measure and in certain cases an indication of what to improve.

1.2.2 Data Cleaning

Data cleansing or data cleaning is the process of detecting and correcting (or removing) loose or inaccurate records from a record set, table, or database and refers to relating deficient, incorrect, inaccurate, or inapplicable corridor of the data and also replacing, modifying, or deleting the dirty or coarse data. Data cleansing is also appertained to as data drawing or data recalling. After cleansing, a data set should be consistent with other similar data sets in the system. The inconsistencies detected or removed may have been originally caused by user entry errors, by corruption in transmission or storage, or by different data dictionary definitions of similar entities in different stores. Data cleaning differs from data validation in that validation almost invariably means data is rejected from the system at entry and is performed at the time of entry, rather than on batches of data.

1.2.3 Data preprocessing

Data preprocessing can relate to the manipulation or dropping of data before it's used in order to insure or enhance performance, and is an important step in the data mining process. The expression "scrap in, scrap out" is particularly applicable to data mining and machine literacy systems. Data-gathering styles are frequently approximately controlled, performing in out-of-range values (e.g., Income – 100), insolvable data combinations (e.g., Sex Male, Pregnant Yea), missing values, etc. Assaying data that has

not been precisely screened for similar problems can produce deceiving results. Therefore, the representation and quality of data are first and foremost before running any analysis. Frequently, data preprocessing is the most important phase of a machine literacy design

1.2.4 Data visualization

Data visualization refers to ways wanted to communicate perceptivity from data through visual representation. Data visualization is the process of rephrasing large data sets and criteria into maps, graphs, and other illustrations. The performing visual representation of data makes it easier to identify and partake real- time trends, outliers, and new perceptivity about the information represented in the data. After data is collected, reused, also the connections need to be anatomized. From an academic point of view, this representation can be considered as a mapping between the original data (usually numerical) and graphic elements (for example, lines or points in a chart). The mapping determines how the attributes of these elements vary according to the data. In this light, a bar chart is a mapping of the length of a bar to a magnitude of a variable. Since the graphic design of the mapping can adversely affect the readability of a chart, mapping is a core competency of Data visualization. Data visualization has its roots in the field of Statistics and is therefore generally considered a branch of Descriptive Statistics. However, because both design skills and statistical and computing skills are required to visualize effectively, it is argued by some authors that it is both an Art and a Science. Research into how people read and misread various types of visualizations is helping to determine what types and features of visualizations are most understandable and effective in conveying information.

1.2.5 MVP (Most Valuable Player)

MVP is comparing all types of players on the base of a single reckoned parameter called MVP value. The parameter is well calculated taking account of numerous sets of attributes that define a player's performance. In addition to player performance, the MVP value also takes account of the current demand of the platoon in the form of Batting Demand points and Experience Demand points. These Demand points are decided by the possessors who need to buy players from the transaction after retaining players of the platoon from the former event. In other words, these points are depicting the anticipation of possessors from the transaction.

1.2.6 Co-Variance Analysis

Analysis of covariance (ANCOVA) is used in examining the differences in the mean values of the dependent variables that are related to the effect of the controlled independent variables while taking into account the influence of the unbridled independent variables. An analysis of covariance, ANCOVA, is a normal direct model that contains at least one factor and one nonstop variable as explicatory variables. The nonstop variable is also called a covariate, hence the name analysis of covariance. To fit the model, it's important to first check whether the factor is indeed defined as a factor and the nonstop variable contains figures (i.e., numeric or integer values) in the data frame.

1.2.7 Multiple Regression

Multiple regression is a statistical approach that can be used to analyze the relationship between a single dependent variable and several independent variables. The ideal of multiple regression analysis is to use the independent variables whose values are known to foretell the value of the single dependent value. Each predictor value is counted, the weights denoting their relative donation to the overall forecasting.

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

Here Y is the dependent variable, and X_1, \dots, X_n are the n independent variables. In calculating the weights, a, b_1, \dots, b_n , regression analysis ensures maximal prediction of the dependent variable from the set of independent variables.

1.2.8 Predictive Analysis

Predictive analytics is a branch of advanced analytics that makes predictions about unborn issues using factual data combined with statistical modeling, data mining approaches, and machine learning. Companies employ predictive analytics to find patterns in this data to identify pitfalls and openings. Predictive analytics is frequently associated with big data and data science. Companies' moment is swimming in data that resides across transactional databases, outfit log lines, images, videotape, detectors, or other data sources. To gain insights from this data, data scientists use deep learning and machine learning algorithms to find patterns and make forecasts about upcoming events.

1.2.9 Predictive modeling

Predictive modeling, a tool used in predictive analytics, refers to the process of using fine and computational approaches to develop predictive models that examine current and literal datasets for underpinning patterns and calculate the probability of an outgrowth. The prophetic modeling process starts with data collection, also a statistical model is formulated, forecasts are made, and the model is revised as new data becomes available.

1.2.10 Exploratory Data Analysis

In statistics, exploratory data analysis is an approach to analyzing data sets to abstract their main characteristics, frequently using statistical visuals and other data visualization approaches. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task. Exploratory data analysis has been promoted by John Tukey since 1970 to encourage statisticians to explore the data, and conceivably formulate suppositions that could lead to new data collection and trials. EDA is different from original data analysis (IDA), which focuses further hardly on checking hypotheticals needed for model fitting and thesis testing, handling missing values, and making conversions of variables as required.

1.3 Importance and Scope of Study

- There can be entries in dataset which also will fit a pattern of legitimate behaviour. Also, the problem has many constraints.

- The improvement of methods is more difficult by the fact that the security concern imposes a limitation to exchange of ideas and methods. Lastly, the data sets are continuously evolving and changing making the profiles of normal.
- In this study we have evaluated five advanced approaches. MVP (Most Valuable Points), IPL stats, Co-variance Analysis, Multiple Regression, Score Prediction.
- The Scope of this study is to evaluate the player by taking their performance in to the account we can predict and bid the maximum for the best player for the upcoming matches.

1.4 Limitations of the Study

There are some limitations for this study that will be helpful for the people

- The need for future researchers is to revise the specific system, for gathering data. It's also a limitation.
- Self- reported data can contain several implicit sources of bias, one should warn to note as limitations. These impulses came apparent if they're incongruent with data from other sources.

The chapter dealt with major sources that are much needed for research study. Data collection and data analysis methods were discussed. The proposed framework used information gain measure to select important data features and utilized various approaches for the prediction task.

CHAPTER II

2 REVIEWS OF LITERATURE

This chapter explains the article used for this research and will be helpful for the future research to be done in this topic. It has the information connecting to the current topic that has used to develop this project.

2.1 Review of Literature

Review of literature is a written work that compiles significant exploration published on a content by accredited scholars and experimenters, checks scholarly papers, books, compositions, conference proceedings, and other sources, examines differing perspectives, theoretical approaches, methodologies, findings, results, conclusions. Reviews critically, analyses, and synthesizes being exploration on a content and performs a thorough-real view, - overview, or- look again of history and current workshop on a subject, issue, or proposition. Literature reviews are important because they're generally a needed step in a thesis offer (Master's or Ph.D.). The offer won't be well supported without a literature review. Also, literature reviews are important because they help to learn important authors and ideas in this field. This is useful for the coursework and the jotting. Knowing crucial authors also helps to come acquainted with other experimenters in this field. In this chapter, the experimenter has explained the need and significance of prophetic analytics. This chapter also explains the papers which helped the experimenter to dissect colorful motifs in prophetic analytics like machine literacy, retrogression model, and backward accretive elimination system.

Barr and Kantor. (2019) states that "How to Compare and Select Cricketers in Overs Cricket,". In this paper, the author outlines the main criteria for comparing the selection of strikers in limited sodalities. This paper shows a clear 2D representation of the strike rate on one axis and the probability of exit, i.e., $P(\text{exit})$ on another axis. We're erecting a strike selection strategy grounded on this 2D frame that combines scale and strike rate as an illustration of this operation we use this principle in beating the 2003 World Cup performance to show the strong and harmonious performance of batsmen playing in Indian and Australia brigades.

Prakash. (2016) proposed the ranking methodology for ranking the batsmen and bowlers in the fray in the IPL IX is described that imposes an ordering on the players in terms of their fur and bowling capabilities. This ranking scheme is grounded on a machine approach using Recursive Point elimination to produce a comprehensive performance indicator for fur and another for bowling.

Jhanwar and Paudi. (2016) forecast the outgrowth of a justice match by comparing the strengths of the players. For this, they measured the performances of individual players. They developed an analysis to model the performances of batsmen and bowlers where they determine the eventuality of a player by examining his career performance and also his recent performances.

Kumar and Sindhu. (2014) discussed some prevailing criteria for assessing players performance in justice is reviewed along with their limitations and strengths. The Random Timbers approach and the new criteria evolved using this machine literacy approach are presented.

Sankaranarayanan, Sattar, and Lakshmanan. (2014) discusses modeling home-runs and non-home runs vaticination algorithms and considers taking runs, lattices, frequency of being each-out as literal features into their vaticination model. But they haven't abused bowler's features and have given further emphasis to batsmen.

Kalgotha, Sharda, and Chakraborty. (2013) found that predictive models aid directors to elect players for a talented platoon at the least possible price. This is calculated on the base of the player's history performance. The optimal model is named on the base of the rate of confirmation data misclassification. This model helps in the selection of players by abetting in the author's bidding equation. This exploration also facilitates the directors to set the hires for players.

Kumar. (2013) attempts to identify the most important attributes of a player's performance that determine expert conditions. In this work, a series of groups with three different pruning strategies and an array of machine literacy- grounded algorithms are executed. Also, a list of the most important performance criteria for each of the four playing positions which approximates the attributes considered by the experts is attained while assigning conditions.

Lakkaraju and Sethi. (2012) described a Sabermetrics style principle to dissect fur performance in justice. Also considered performance analysis using pars and strike rates for bowling and fur and estimated the performance on all-rounders in IPL.

Sonali and Shubhasheesh. (2012) hypothecated how brigades strategically decide on the final shot quantum grounded on once player & platoon performance in IPL and formats analogous to IPL. The authors also exfoliate light on how the personalities of players can affect platoon performance. They dissect the possible factors grounded on which stab decide and make a

prophetic model for pricing in the transaction. The analysis is done collectively for all the brigades.

Singh and Gupta. (2011) proposed an integer programming real- time model for an optimal strategy for binding processes. Spreadsheets were used to validate & calculate the results since it was the optimal choice considering the inflexibility of objectification for further weight- age grounded on the recent performance of a player to estimate the final outgrowth.

Singh. (2011) anatomized using Data Envelopment Analysis to measure how effective brigades are in IPL. The author calculates awarded points, total run rate, profit, and returns by determining total charges including the paycheck- price of players and staff as well as other charges. Effectiveness score is generally directly related to the performance of the player in the league. On putrefying the inefficiencies into specialized and scale inefficiency, it's realized that the inefficiency is primarily due to the unoptimized scale of product & unoptimized metamorphosis of the results and the considered data.

Ahmed, Jindal, and Deb. (2011) discussed the NSGA-II algorithm to propose a new representation scheme & a multi-objective approach for opting players on a limited budget considering the fur & bowling strengths along with the platoon conformation. Factors similar as contending farther optimize the results. The dataset to define performance is taken from IPL-4th Edition. The author shows analysis in real- time transaction events, opting players one-by-one. The author argues that the methodology can be enforced across other fields of sports similar as soccer etc.

Saikia and Bhattacharjee. (2011) determined performances of all-rounders into 'Performer', 'Batting-All Rounder', 'Bowling All Rounder' and 'Under- pantomime'. Further, they suggest and consider independent variables that impact on each- rounder's performance by using Step-wise Multinomial Logistic Regression (SMLR). The independent variables are used to prognosticate the class of an each-rounder player using the Naive Bayes Bracket conception.

Norusis. (2011) discussed initially, the Kohonen networks have a competitive literacy algorithm where each cluster competes for compliances and on landing a record; it's awarded by conforming its parameters so that it'll have an easier time landing analogous records in the future.

Sharpet.al. (2011) has formulated an Integer programming system for T20 platoon selection as a maximization problem, wherein the decision variable indicates whether a certain player has a specific skill is named or not. The portions for different chops similar as fur, bowling, and fielding were regularized using a system that employed certain indicators established indicators similar as pars and strike rates for batsmen and bowlers using a combination of different performance statistics.

Ahmedet.al. (2011) have represented player selection as an evolutionary multi-objective optimization algorithm (NSGA-II). In the after corridor of their paper, after taking into account the trade-offs between batsmen and bowlers using Dynamic optimization, they've proposed a set of high- performance brigades that can give a base for platoon selection during IPL Player Deals.

Budayan. (2008) proposed the family of hierarchical styles is less suitable for segmentation because apriorism there's no reason to anticipate these parts to have a hierarchical structure. In non-hierarchical styles, k- means, an iterative portioning algorithm is popular and it doesn't put a hierarchical structure.

CHAPTER III

3 RESEARCH METHODOLOGY

Research methodology is a way to totally break the exploration problem. It may be understood as a wisdom of studying how exploration is done scientifically. In it, we study the colorful way that are generally espoused by an experimenter in studying his exploration problem along with the sense behind them. Research may be veritably astronomically defined as a methodical gathering of data and information and its analysis for the advancement of knowledge in any subject. The exploration attempts to find answers to intellectual and practical questions through the operation of methodical styles. In this chapter, we banded study design, Ideal, Variable, etc.

3.1 Study Design

An Exploratory Data Analysis (EDA) technique is being implemented. The comparison is made for data sets such as PBT (Player Batting Points), PBW (Player Bowling Points), MVP (Most Valuable Player), Co-Variance Analysis, Multiple Regression, to determine the prediction of best players.

3.2 Objectives of the Study

- To Analyze the Most Valuable Player (MVP) by using a Set of Rules.
- To Analyze the Dynamic Nature of MVP Calculation After Each Player's Selection Process.
- To Analyze and find similarity & dis-similarity among players using co-variance analysis.

- To Analyze and Predict the team scores with individual score & strike rate of a player using multiple linear regression analysis.

3.3 Data Source

The type of data which is used in the study is Secondary data. The data is collected from multiple websites. Websites are major source of data collection place where large amount of data are collected and converted into useful information. In this study variety of data sets were taken for analysis. Mainly the datasets from cricket websites gave a good representation and wide knowledge about the widespread. The secondary source of data collected were need to be cleansed, looked up for missing variables and then need to be analyzed.

Websites:

<https://www.iplt20.com/>

<https://www.kaggle.com/vamsikrishna/exploratory-data-analysis-of-ipl-data>

<https://www.mykhel.com/cricket/ipl-2021-s4/>

<http://www.howstat.com/cricket/Statistics/IPL/MatchList.asp>

<https://www.sportskeeda.com/cricket/most-ipl-wins>

3.4 Variables considered

A variable in exploration simply refers to a person, place, thing, or miracle that you're trying to measure in some way. The stylish way to understand the difference between a dependent and independent variable is that the meaning of each is inferred by what the words tell us about the variable you're using. Once the content has been perfected, the experimenter should identify the crucial variables of interest. These variables give a good sapience into the data. Choosing the right variable for the modeling will help us to prognosticate the needful outgrowth.

IPL Stats – Matches Played, Matches Won, Matches Lost.

Most Valuable Player (MVP) – Player Batting Points, Player Bowling Points, Experience.

Co-Variance Analysis – Runs.

Multiple Regression – Runs, Total Score, Strike Rate.

Score Prediction – Runs, Average, Strike Rate.

3.5 Sampling Design

There is no specific sample design, because it is a cricket data analysis, from that we are choosing a certain data to predict, analyze and experimenting the chosen data, to give approximate solution. This can be helping the future researchers while doing this topic.

3.6 Methods and Tools of Data Collection

A set of data seems to be from the Kaggle website, as well as the explanation could be discovered on both the website of iplt20 and mykhel. The whole data sets seem to be strongly unbalanced so this set of data may have features applicable mostly to analysis, etc. This data has been manually entered by surfing through multiple cricket websites.

3.7 Methods and Tools of Data Analysis

The data from secondary sources has been arranged and the missing data were cleansed. The cleaned data has been analyzed with various cricket websites and entered manually these data compared with accuracy using various analysis in Microsoft Excel.

- Data Cleaning – MS Excel
- Data Visualization – MS Excel
- Data Analysis – MS Excel
- Predictive Modelling Techniques – IPL Stats, MVP (Most Valuable Player) calculation, Co-Variance Analysis, Multiple Regression, and Predicted Score.

In the above chapter it is deeply discussed about the research methodology, objectives, study design, data source, variables, methods and tools of data collection, methods and tools of data analysis.

CHAPTER IV

4 DATA ANALYSIS AND INTERPRETATION

This is the chapter of the research study that implies the data analysis and interpretation done and interpreted in the previous chapter. It also gives the recommendation and suggestion based on the hypotheses of the study.

4.1 Descriptive Presentation and Analysis of the Concept

4.1.1 IPL Stats

Table 4.1

IPL STATS			
Team	Played	Won	Lost
Mumbai Indians	217	125	88
Chennai Super Kings	195	117	76

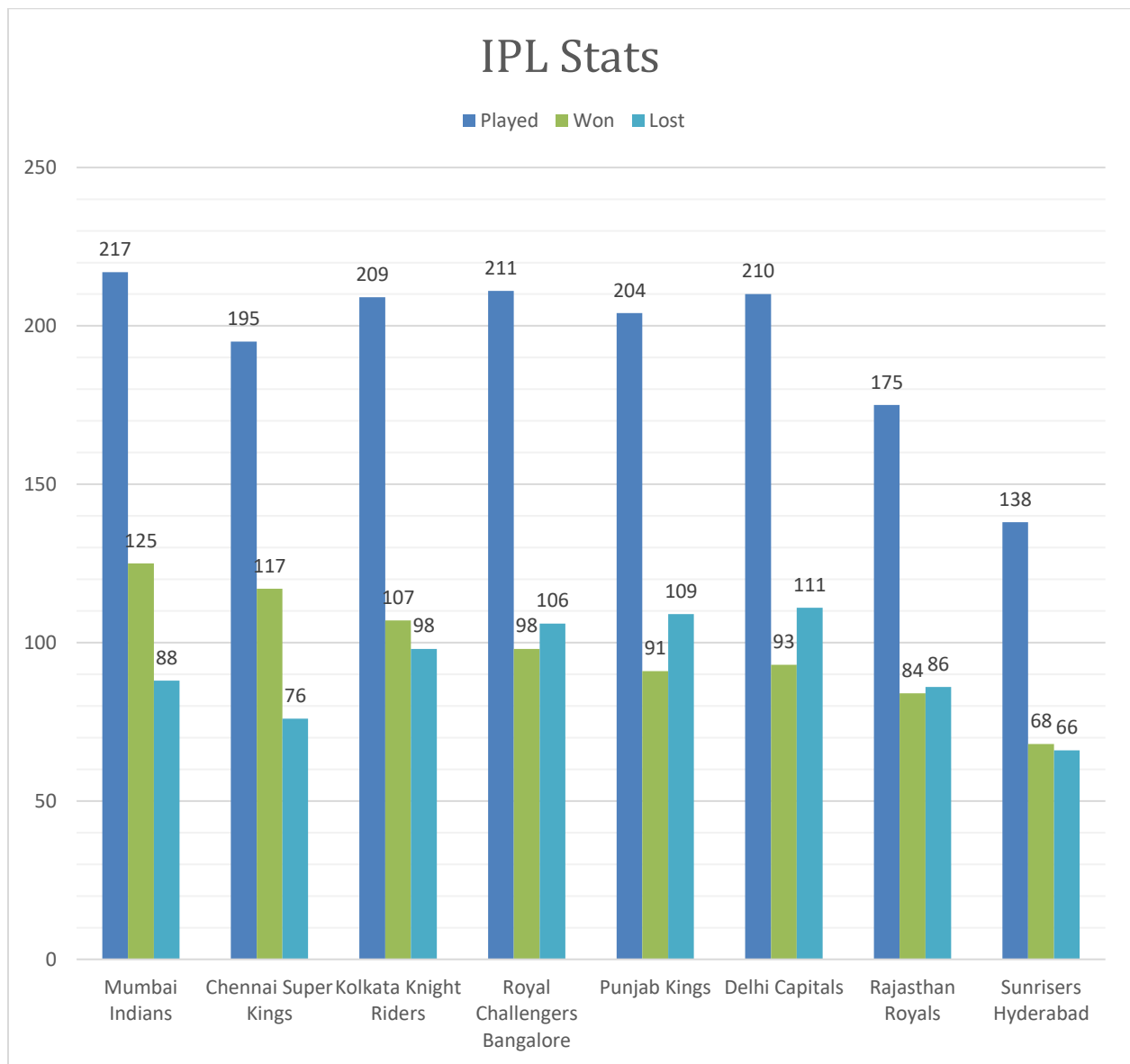
Kolkata Knight Riders	209	107	98
Royal Challengers Bangalore	211	98	106
Punjab Kings	204	91	109
Delhi Capitals	210	93	111
Rajasthan Royals	175	84	86
Sunrisers Hyderabad	138	68	66

Source: Secondary Data.

From Table 4.1.1, It is understood that team Mumbai Indians have played the greatest number of matches and they also won the greatest number of matches. And the team Sunrisers Hyderabad has played the least number of matches in the IPL and they have won the least number of matches.

Chart 4.1

IPL Stats



4.1.2 MVP Calculation for Batsman

Table 4.2

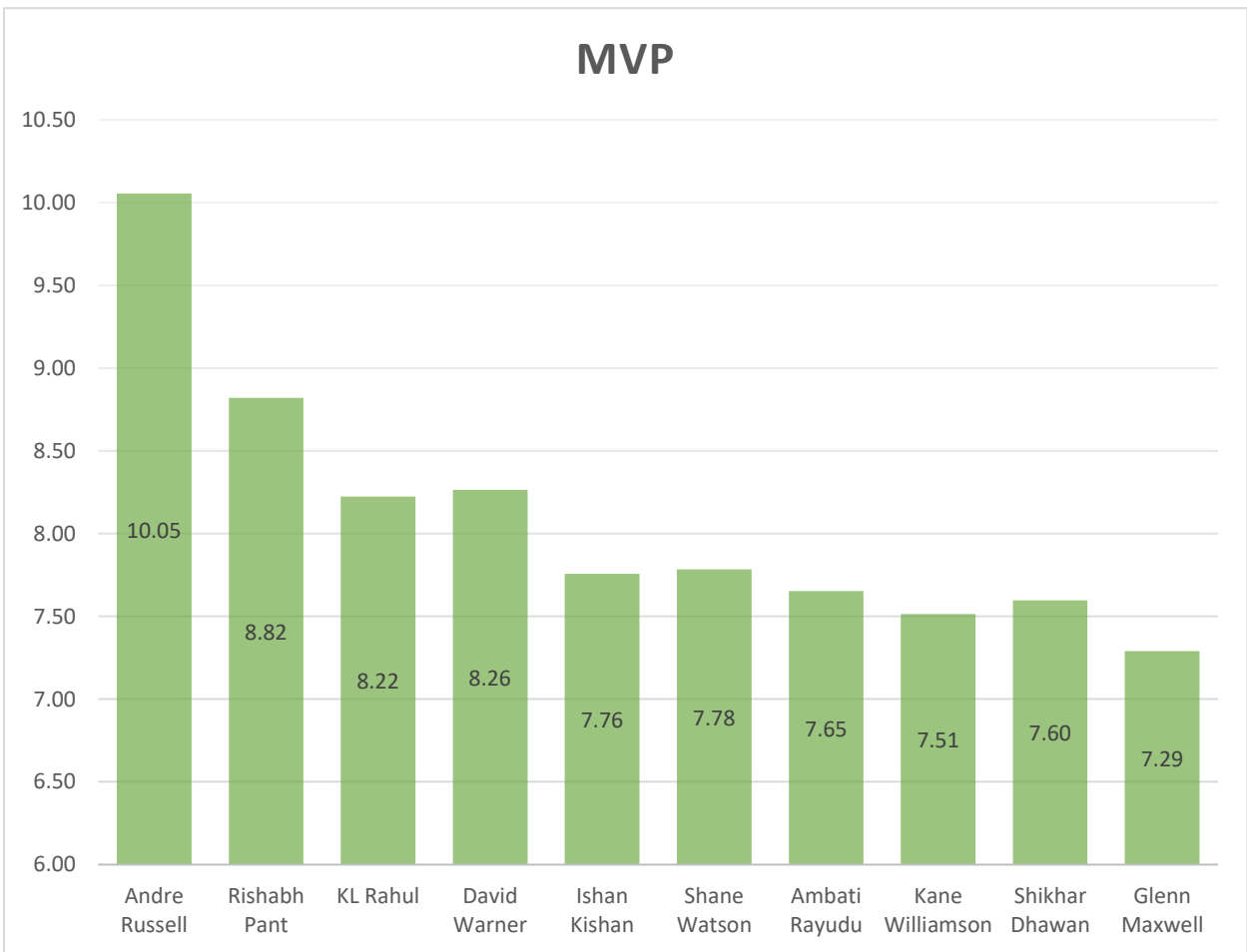
Top 10 Batsman from IPL Season 2017-2021											
Top 10 Batsman	Team	Most Runs	Average	Strike Rate	50s	100s	PBT (Player Batting Points)	No.of Matches Played	Total No.of Matches in IPL	EXP	MVP
Andre Russell	KKR	510	56.66	204.81	4	0	9.97	64	780	0.082	10.05
Rishabh Pant	DC	684	52.61	173.60	5	1	8.72	76		0.097	8.82
KL Rahul	PK	659	54.91	158.41	6	0	8.10	94		0.121	8.22
David Warner	SRH	692	69.20	143.86	8	1	8.09	135		0.173	8.26
Ishan Kishan	MI	516	57.33	145.76	4	0	7.63	99		0.127	7.76
Shane Watson	CSK	555	39.64	154.59	2	2	7.61	134		0.172	7.78
Ambati Rayudu	CSK	602	43.00	149.75	3	1	7.44	166		0.213	7.65
Kane Williamson	SRH	735	52.50	142.44	8	0	7.43	63		0.081	7.51
Shikhar Dhawan	DC	618	44.14	144.73	4	2	7.39	159		0.204	7.60
Glenn Maxwell	RCB	513	42.75	144.10	6	0	7.17	97		0.124	7.29

Source: Secondary Data

From Table 4.1.2, It is understood that from IPL season 2017-2021 Andre Russell stands as the top player of all other players. He scored 510 as his most runs, holds 56.66 as his average and strike rate of 204.81. Here he is selected as the Most Valuable Player among the other batsman.

Chart 4.2

MVP of Top 10 Batsman



4.1.3 MVP Calculation for Bowlers

Table 4.3

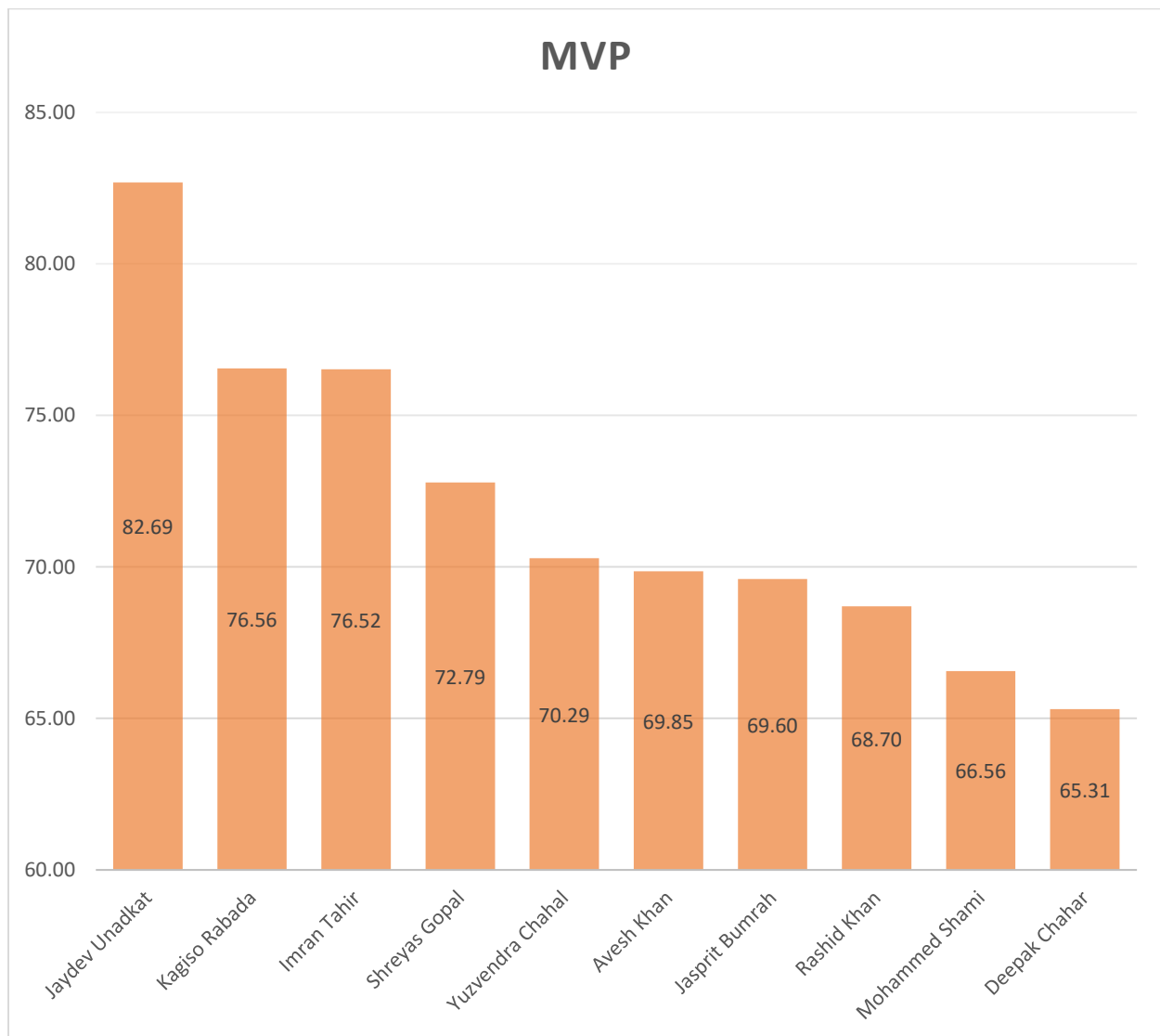
Top 10 Bowlers from IPL Season 2017-2021											
Top 10 Bowlers	Team	Most Wickets	Average	Strike Rate	Economy	4W	5W	PWT	No.of Matches Played	EXP	MVP
Jaydev Unadkat	RPS	24	13.41	11.45	7.02	0	1	82.58	86	0.110	82.69
Kagiso Rabada	DC	25	14.72	11.28	7.82	2	0	76.49	50	0.064	76.56
Imran Tahir	CSK	26	16.57	14.84	6.69	2	0	76.45	59	0.076	76.52
Shreyas Gopal	RR	20	17.35	14.4	7.22	0	0	72.73	48	0.062	72.79
Yuzvendra Chahal	RCB	21	19.28	16.33	7.08	0	0	70.18	84	0.108	70.29
Avesh Khan	DC	24	18.75	15.25	7.37	0	0	69.82	25	0.032	69.85
Jasprit Bumrah	MI	19	21.52	19.47	6.63	0	0	69.46	106	0.136	69.60
Rashid Khan	SRH	21	21.8	19.42	6.73	0	0	68.64	46	0.059	68.70
Mohammed Shami	PK	19	20.78	16.63	7.5	0	0	66.46	77	0.099	66.56
Deepak Chahar	CSK	22	21.9	17.59	7.47	0	0	65.23	63	0.081	65.31

Source: Secondary Data

From Table 4.1.3, It is understood that from IPL season 2017-2021 Jaydev Unadkat stands as the top player of all other players. He took 24 as his most wickets, holds 13.41 as his average and economy of 7.02. Here he is selected as the Most Valuable Player among the other bowlers.

Chart 4.3

MVP of Top 10 Bowlers



4.1.4 CO-Variance Analysis

Table 4.4

Co-Variance Analysis of KL Rahul and Kane Williamson

Matches	A (KL Rahul)	B (Kane Williamson)
1	4	16
2	60	66
3	19	26
4	91	20
5	49	18
6	21	1
7	21	51
8	67	11
9	39	26
10	98	31

Co-Variance Analysis of 2 Players		
	A (KL Rahul)	B (Kane Williamson)
A (KL Rahul)	915.89	
B (Kane Williamson)	62.56	329.64

Manual Calculation	
Average (A)	46.9
Average (B)	26.6
Sum of A*B	13101
Avg of A* Avg of B	1247.54
Sum/10	1310.1
(Sum/10) - (Average of A * Average of B)	62.56

Source: Secondary Data

From Table 4.1.4, It is Understood that there is a strong bond between the two players KL Rahul and Kane Williamson. The Co-Variance between the two players is a positive value of 62.56. By this we can say that they are the consistent players in IPL.

Table 4.5

Co-Variance Analysis of Andre Russell and Rishabh Pant

Matches	Andre Russell	Rishabh Pant
1	20	6
2	0	51
3	0	10
4	45	15
5	10	26
6	9	39
7	54	24
8	31	35
9	9	14
10	5	16

Co-Variance Analysis of 2 Players		
	Andre Russell	Rishabh Pant
A (Andre Russell)	324.01	
B (Rishabh Pant)	-32.58	184.24

Manual Calculation	
Average (A)	18.3
Average (B)	23.6
Sum of A*B	3993
Avg of A * Avg of B	431.88
Sum/10	399.30
(Sum/10) - (Average of A * Average of B)	-32.58

Source: Secondary Data

From Table 4.1.4.1, It is Understood that there is no strong bond between two players Andre Russell and Rishabh Pant. The Co-Variance between the two players is a negative value of - 32.58. By this we can say that they are non-consistent players in IPL.

4.1.5 Multiple Regression

Table 4.6

Last 10 IPL Matches of Andre Russell

Match	(X) Strike Rate	(Y) Runs	(Z) Total Score
1	133.33	20	171
2	0	0	0
3	0	0	0
4	166.67	45	154
5	111.11	10	126
6	128.57	9	133
7	254.45	54	202
8	155	31	166
9	60	9	142
10	100	5	187

Regression Statistics of Andre Russell

Regression Statistics					
Multiple R	0.960853195				
R Square	0.923238861				
Adjusted R Square	0.901307108				
Standard Error	24.2856388				
Observations	10				
ANOVA					
	df	SS	MS	F	Significance F
Regression	2	49655.77585	24827.88792	42.09598863	0.000125313
	Coefficients	Standard Error	t Stat	P-value	
Intercept	1.438160005	16.50522047	0.087133644	0.93300545	
(Y) Runs	2.419284297	0.546924607	4.423432893	0.003068515	
(Z) Total Score	0.508992485	0.145281707	3.503486416	0.009946221	

Source: Secondary Data

From Table 4.1.5, We can see that the significance value is 0.0001, which is below 0.05. So, there is a statistically significant difference between the runs and the total score. Therefore, it affects the player's strike rate.

Table 4.7**Last 10 IPL Matches of Rishabh Pant**

Match	(X) Strike Rate	(Y) Runs	(Z) Total Score
1	100	6	135
2	145.71	51	172
3	125	10	164
4	125	15	139
5	118.18	26	132
6	108.33	39	127
7	100	24	154
8	166.67	35	139
9	127.27	14	167
10	200	16	156

Regression Statistics of Rishabh Pant

Regression Statistics					
Multiple R	0.332911326				
R Square	0.110829951				
Adjusted R Square	-0.143218634				
Standard Error	33.68379181				
Observations	10				
ANOVA					
	df	SS	MS	F	Significance F
Regression	2	989.9478231	494.9739115	0.436254943	0.662898275
	Coefficients	Standard Error	t Stat	P-value	
Intercept	37.46297319	105.7171695	0.354369809	0.733496744	
(Y) Runs	0.316805873	0.784900277	0.403625635	0.698533357	
(Z) Total Score	0.583679517	0.699685534	0.834202636	0.431709916	

Source: Secondary Data

From Table 4.1.5.1, We can see that the significance value is 0.66, which is above 0.05. So, there is no significant difference between the runs and the total score. Therefore, it does not affect the player's strike rate.

Table 4.8

Last 10 IPL Matches of KL Rahul			
Match	(X) Strike Rate	(Y) Runs	(Z) Total Score
1	233.33	98	139
2	111.43	39	158
3	121.82	67	168
4	95.45	21	135
5	100	21	125
6	148.48	49	183
7	159.65	91	179
8	95	19	123
9	115.38	60	132
10	66.67	4	120

Regression Statistics of KL Rahul

Regression Statistics					
Multiple R	0.88591683				
R Square	0.78484864				
Adjusted R Square	0.72337682				
Standard Error	24.5279989				
Observations	10				
ANOVA					
	df	SS	MS	F	Significance F
Regression	2	15362.57738	7681.28869	12.767617	0.0046196
	Coefficients	Standard Error	t Stat	P-value	
Intercept	95.2753232	54.82394461	1.73784145	0.1258041	
(Y) Runs	1.39143371	0.313114513	4.4438493	0.0029938	
(Z) Total Score	-0.244956	0.419702909	-0.5836414	0.5777747	

Source: Secondary Data

From Table 4.1.5.2, We can see that the significance value is 0.004, which is below 0.05. So, there is a statistically significant difference between the runs and the total score. Therefore, it affects the player's strike rate.

Table 4.9

Last 10 IPL Matches of David Warner

Match	(X) Strike Rate	(Y) Runs	(Z) Total Score
1	66.67	2	120
2	0	0	134
3	103.64	57	171
4	75	6	159
5	100	37	121
6	105.88	36	137
7	145.95	54	143
8	75	3	177
9	66.67	2	172
10	100	17	132

Regression Statistics of David Warner

Regression Statistics					
Multiple R	0.775797365				
R Square	0.601861551				
Adjusted R Square	0.488107709				
Standard Error	27.22617408				
Observations	10				
ANOVA					
	df	SS	MS	F	Significance F
Regression	2	7843.930806	3921.965403	5.290911831	0.039821666
	Coefficients	Standard Error	t Stat	P-value	
Intercept	44.11148176	63.48280678	0.694857143	0.509544353	
(Y) Runs	1.307853008	0.402106645	3.252502847	0.014008738	
(Z) Total Score	0.080364692	0.422569331	0.190181081	0.854564901	

Source: Secondary Data

From Table 4.1.5.3, We can see that the significance value is 0.039, which is below 0.05. So, there is a statistically significant difference between the runs and the total score. Therefore, it affects the player's strike rate.

Table 4.10**Last 10 IPL Matches of Ishan Kishan**

Match	(X) Strike Rate	(Y) Runs	(Z) Total Score
1	138.46	84	235
2	169.23	50	94
3	75	9	111
4	107.69	14	155
5	110	11	136
6	35.29	6	131
7	92.86	26	137
8	57.14	12	150
9	33.33	1	152
10	147.37	28	159

Regression Statistics of Ishan Kishan

Regression Statistics					
Multiple R	0.793309853				
R Square	0.629340522				
Adjusted R Square	0.523437814				
Standard Error	32.30356609				
Observations	10				
ANOVA					
	df	SS	MS	F	Significance F
Regression	2	12402.50933	6201.254667	5.942629175	0.031003654
	Coefficients	Standard Error	t Stat	P-value	
Intercept	128.6737402	45.94273425	2.800741886	0.026495987	
(Y) Runs	1.732843515	0.507632765	3.413576966	0.01123305	
(Z) Total Score	-0.505467595	0.344597717	-1.466833849	0.185855887	

Source: Secondary Data

From Table 4.1.5.4, We can see that the significance value is 0.031, which is below 0.05. So, there is a statistically significant difference between the runs and the total score. Therefore, it affects the player's strike rate.

Table 4.11**Last 10 IPL Matches of Shreyas Gopal**

Last 10 IPL Matches of Shreyas Gopal			
Match	(X) Economy	(Y) Runs	(Z) Wicket
1	9	9	0
2	11.7	35	0
3	13.3	40	0
4	14.7	44	1
5	10	10	0
6	7.5	30	2
7	8	32	0
8	3.5	14	1
9	8	32	1
10	7.8	31	1

Regression Statistics of Shreyas Gopal

Regression Statistics					
Multiple R	0.74211721				
R Square	0.55073795				
Adjusted R Square	0.42237736				
Standard Error	2.45600995				
Observations	10				
ANOVA					
	Df	SS	MS	F	Significance F
Regression	2	51.76110589	25.8805529	4.2905534	0.060778375
	Coefficients	Standard Error	t Stat	P-value	
Intercept	5.83790115	2.025057891	2.88283173	0.0235561	
(Y) Runs	0.17445341	0.067489149	2.58491046	0.0362123	
(Z) Wicket	-2.2004342	1.192660353	-1.8449797	0.1075564	

Source: Secondary Data

From Table 4.1.5.5, We can see that the significance value is 0.060, which is above 0.05. So, there is no significant difference between the runs and the total score. Therefore, it does not affect the player's strike rate.

4.1.6 Score Prediction

Table 4.12

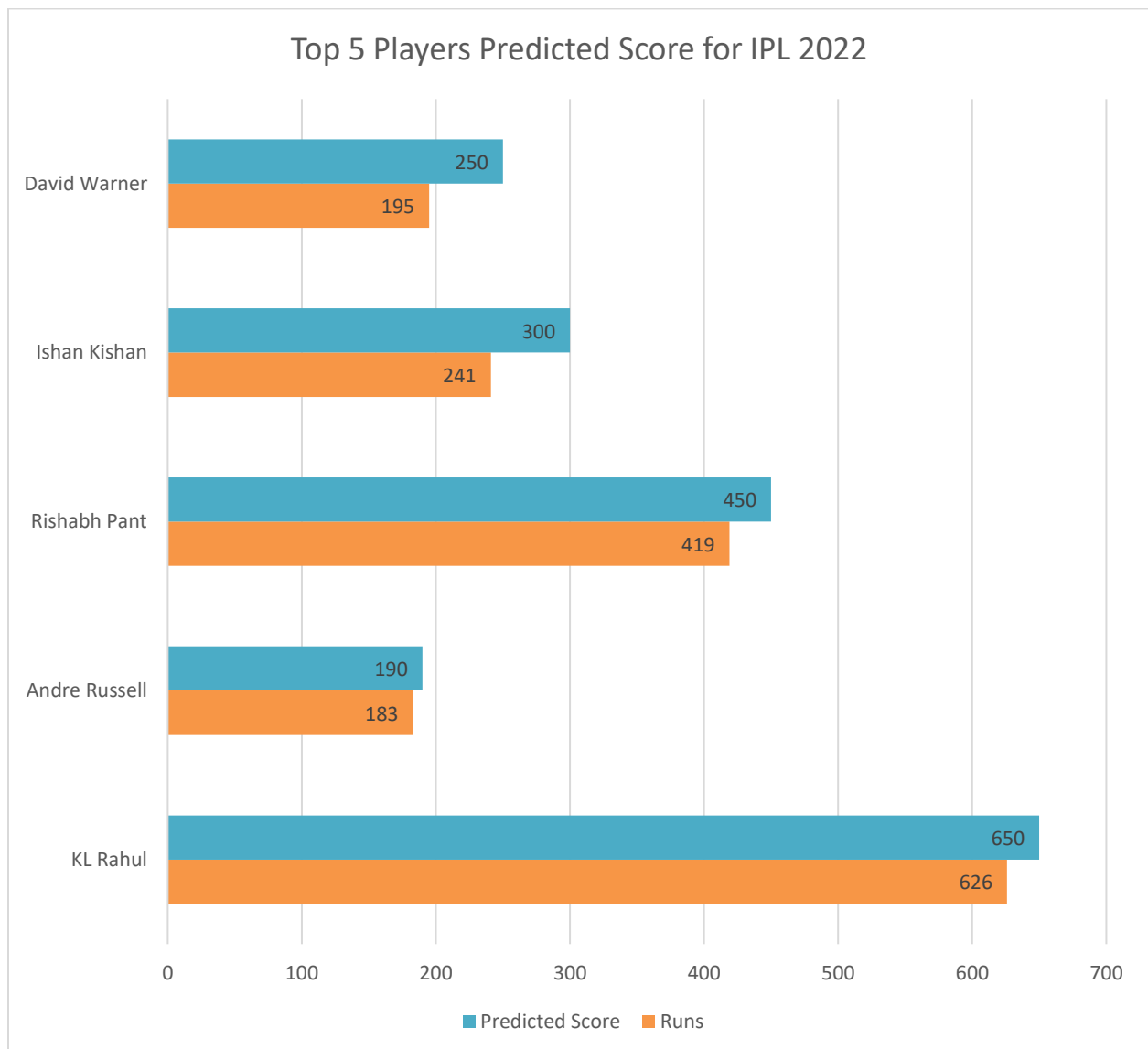
Top 5 Players Predicted Score for IPL 2022				
Players of IPL 2021	Runs	Average	Strike Rate	Predicted Score
KL Rahul	626	62.6	138.8	650
Andre Russell	183	26.4	152.5	190
Rishabh Pant	419	34.91	128.52	450
Ishan Kishan	241	26.77	133.88	300
David Warner	195	24.37	107.73	250

Source: Secondary Data

From Table 4.1.6, It is understood that the individual scores of the top 5 players of IPL 2021. By taking into the consideration, we can predict the scores for upcoming matches of these players that is IPL 2022.

Chart 4.4

Predicted Score of Top 5 Players



CHAPTER V

5 FINDINGS, RECOMMENDATIONS AND CONCLUSION

In this chapter they discussed about the findings, recommendations and conclusion of the study, which will be helpful for the future study based on this topic.

5.1 Findings of the Study

- ✓ IPL Stats of matches played, wins, and lost.
- ✓ MVP calculation for player points.
- ✓ Co-Variance analysis for two different players.
- ✓ Multiple Regression for each player.
- ✓ Predicting scores for each player for the upcoming matches.

In this study I used predictive analytics and found the results to choose the best and continuous performing players in the IPL seasons using various analysis. This creates a best team when every individual player is analyzed by their performance in the previous matches played by them.

5.2 Recommendation on the Findings

This study recommends to use various types of analysis used for the data analytics. It results to selection of best-balanced team for the team management. The random selection of batsmen and batting order may give a better result in the ever-unpredictable game.

5.3 Suggestion for the further study

Researcher took limited amount of data and this gave a result which is somewhat acceptable but having more dataset will help the researcher for perfect forecast of future outbreak. We discussed IPL Stats, MVP (Most Valuable Player), Co-Variance Analysis, Multiple Regression and Score Prediction. Further research proceeding must be concerned with taking more data into account.

5.4 Conclusion

This paper focuses on the colorful bracket of forecasting modeling approaches. we demonstrate the dynamic changing demand of a player in the duration of a purchase, it results in the selection of the best-balanced squad for the squad handling. When a team has a full set of players, the system classifies them grounded on their part and performance using MVP computation. This will give further options for the captain to use his bowling strength during the match. Also, the system utilizes the differences and parallels in performance of batsmen using friction to constrain the possible accept cases for better affair. The arbitrary selection of batsmen and fur order may give a better result in the ever-changeable game. The total score forecasting's delicacy using regression shows the reliance of a platoon's fur on an individual player.

Reference

Khandelwal, M., Prakash, J., & Pradhan, T. (2016). An Analysis of Best Player Selection Key Performance Indicator: The Case of Indian Premier League (IPL). In *Intelligent Systems Technologies and Applications* (pp. 173-190). Springer, Cham.

Singh, S; Gupta, S; & Gupta, V. (2011). 'Dynamic bidding strategy for players auction in IPL,' *International Journal of Sports Science and Engineering*, 05(01), 03-16.

Singh, S. (2011), 'Measuring the performance of teams in the Indian Premier League,' *American Journal of Operations Research*, 01, 180-184.

Pankush Kalgotra, Ramesh Sharda, Goutam Chakraborty, 'Predictive Modelling in sports leagues: an application in Indian Premier League,' in *SAS Global Forum 2013*, pp. 019-2013.

F. Ahmed, A. Jindal & K. Deb, 'Cricket Team Selection Using Evolutionary Multi-Objective Optimization', 71-78

Sonali B. and Shubhasheesh B, 'Auction of players in Indian premier league: the strategic perspective' *International Journal of Multidisciplinary Research*, ISSN 2231 5780.

H. Saikia and D. Bhattacharjee, 'On classification of all-rounders of the Indian premier league (IPL): A Bayesian approach', *Vikalpa* 36 (4), 25-40.

Santra, A., Mitra, A., Sinha, A., & Das, A. K. (2021). Prediction of Most Valuable Bowlers of Indian Premier League (IPL). In *Data Management, Analytics and Innovation* (pp. 211-223). Springer, Singapore.

Ghosh, A., Sinha, A., Mondal, P., Roy, A., & Saha, P. (2021, January). Indian Premier League Player Selection Model Based on Indian Domestic League Performance. In *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 0277-0282). IEEE.

Sankaran, S. (2014). Comparing pay versus performance of IPL Bowlers: an application of cluster analysis. *International Journal of Performance Analysis in Sport*, 14(1), 174-187.

Chakraborty, S., Sen, A. K., & Bagchi, A. (2015). Combinatorial auctions for player selection in the Indian Premier League (IPL). *Journal of Sports Economics*, 16(1), 86-107.

Kansal, P., Kumar, P., Arya, H., & Methaila, A. (2014, November). Player valuation in Indian premier league auction using data mining technique. In *2014 international conference on contemporary computing and informatics (IC3I)* (pp. 197-203). IEEE.

Ghosh, A., Dey, M., Guha, B., Jana, S., & Sarkar, A. (2021). Performance evaluation & rankings of players in IPL 2019 by DEA & SEM. *Malaya Journal of Matematik*, Vol. S, (1), 46-56.

Prakash, J., Khandelwal, M., & Pradhan, T. (2015, September). Evaluation of IPL teams and players using association, correlation and classification rules. In *2015 International Conference on Computer, Communication and Control (IC4)* (pp. 1-6). IEEE.

Amin, G. R., & Sharma, S. K. (2014). Cricket team selection using data envelopment analysis. *European journal of sport science*, 14(sup1), S369-S376.

Dey, P. K., Banerjee, A., Ghosh, D. N., & Mondal, A. C. (2014). AHP-neural network-based player price estimation in IPL. *International Journal of Hybrid Information Technology*, 7(3), 15-24.

Patel, N., & Pandya, M. (2019). IPL Player's Performance Prediction. *International Journal of Computer Sciences and Engineering*, 7, 478-481.

Singla, S., & Shukla, S. S. (2020). Integer Optimization for Dream 11 Cricket Team Selection. *International Journal of Computer Sciences and Engineering*.

Prakash, C. D. (2016). A new team selection methodology using machine learning and memetic genetic algorithm for ipl-9. *Int. Jl. of Electronics, Electrical and Computational System IJECS* ISSN.

Lamsal, R., & Choudhary, A. (2018). Predicting Outcome of Indian Premier League (IPL) Matches Using Classification Based Machine Learning Algorithm. *CoRR*.

Kashid, D., Mehta, S. N., & Basotia, V. (2019, October). A Hybrid AHP-TOPSIS Mathematical Multi Criteria Decision-Making Model for Players Performance Evaluation and Selection in IPL a Case Study. In *National Conference on " New Frontiers of Innovation in Management, Social Science, and Technology and their Impact on Societal Development"* AU-FAIT (Vol. 5).

Singh, S., & Kaur, P. (2017). IPL visualization and prediction using HBase. *Procedia computer science*, 122, 910-915.

Dey, P. K., Ghosh, D. N., & Mondal, A. C. (2011). A MCDM approach for evaluating bowlers' performance in IPL. *Journal of emerging trends in Computing and Information Sciences*, 2(11), 563-573.

Prakash, C. D., Patvardhan, C., & Lakshmi, C. V. (2016). Team selection strategy in ipl 9 using random forests algorithm. *International Journal of Computer Applications*, 975, 8887.