

Cricket Data Analytics and Score Prediction

A Project Work Synopsis

Submitted in the partial fulfillment for the award of the degree of

**BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE WITH SPECIALIZATION IN
ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING**

Submitted by:

20BCS3848, Puneet Choudhary

Under the Supervision of:

Dr. Sukhmeet Kour



**CHANDIGARH
UNIVERSITY**
Discover. Learn. Empower.

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,

PUNJAB

September, 2023

Abstract

Keywords:

Cricket is one of the most watched sports nowadays. Winning in cricket depends on various factors like the home ground advantage, performances in past matches, the players' experience, performance at the specific venue, performance against the specific team, and the current form of the team and the player. In the recent past, a lot of research has been done which measures the player's performance and predicts the winning percentage. This article briefs about the factors that the cricket game depends on and discusses various research that predicted the winning of a team with the advent of statistical modeling in sports. Cricket is one of the most popular team games in the world. With this article, we embark on predicting the outcome of the Indian Premier League (IPL) cricket match using a supervised learning approach from a team composition perspective. Our work suggests that the relative team strength between the competing teams forms a distinctive feature for predicting the winner. Modeling the team's strength boils down to modeling individual players' batting and bowling performances, forming the basis of our approach. We use statistics and the recent performance of a player to model him. Player-independent factors have also been considered to predict the outcome of a match. Machine learning is used to predict a cricket match's outcome before and during a match. Today's level of sports analytics has evolved both the technology that provides data, and the statistical methodologies that provide the tools for analyzing data improved rapidly. Though sports analytics has been rapidly developing, it has not been the case with cricket. Due to historical reasons where cricket was perceived as a leisurely gentleman's game played without remuneration to players (until recently), cricket was not subject to large financial transactions. It obtained a custom accuracy of 80.86%. In this analysis, Australia emerged as the T20 World Cup 2020 winner. For this purpose, the ESPN Cricinfo dataset has been used.

Table of Contents

Title	Page
Abstract -----	2
1. Introduction-----	4-10
1.1 Problem Definition-----	5
1.2 Project Overview-----	6
1.3 Hardware Specification-----	7-8
1.4 Software Specification-----	9-10
2. Literature Survey-----	11-13
2.1 Existing System-----	11
2.2 Proposed System-----	12
2.3 Literature Review Summary-----	13
3. Problem Formulation-----	14
4. Research Objective-----	15
5. Methodologies-----	16-17
6. Experimental Setup-----	17-18
7. Conclusion-----	18-19
8. Tentative Chapter Plan for the proposed work-----	19-20
9. Reference-----	21

1. INTRODUCTION

Cricket is a well-known sport. The popularity of cricket and its viewership has increased tremendously in the past two decades. To cater to potential future growth, global market research was commissioned by the International Cricket Council (ICC) which revealed that cricket has more than one billion fans worldwide, with the potential for significant growth. Among all formats of cricket, the popularity of T-20 Internationals (T20Is) is the highest. All of these fans of cricket are eager about upcoming cricket events and tournaments. They desire to learn about the prospects of their favorite team.

The 7th edition of the ICC Men's T20 cricket World Cup is going to be held in Australia shortly. The contest will be played between 12 teams. Eight teams have directly qualified for the tournament. These are the top-ranked teams in the T20 form of cricket. Four more teams will be included through a qualifying round. Afterward, all 12 teams (super 12) will be divided into two groups. From each group, two teams will emerge to find their place in the top four teams (semi-finals). Two teams will qualify for the grand finale from semi-finalists. The winner of the final will be the champions of the T20 cricket World Cup for the next two years.

1.1 Problem Definition

In this to design a system that can provide the score and winning prediction in a cricket match, the system can analyze multiple parameters like winning the toss, batting side, DL approach, home ground advantage, player-wise performance, etc. while declaring a time for the particular championship it is very important to select the best team so that the chances of the team winning the championship become easy. This problem had to be solved to generate the best players from both teams for the best battle. To solve this problem we have collected historical data of all teams (India, Australia, New Zealand, South Africa, etc.), and using prediction algorithms like the Naïve Bayes algorithm we are predicting the best-starting players for both the teams can be used in a fantasy league for winning the maximum points.

1.2 Problem Overview

There has been a lot of related study to this problem in various sports. The paper I have used for references is all related work that had been done on this problem. The paper by Trawinski described the prediction of results using a fuzzy classification system. This paper predicts the results for basketball games. I used the attribute selection technique mentioned in this paper for my project. The attribute selection technique proposed in this paper was done using Power BI so it was a good reference point for me too. The wrapper method algorithm and the ranker method algorithm implemented in this paper were also used in my project. But the prediction part was done by using the fuzzy classification system and I did not use that system for my prediction part.

The paper by Haghighat also described the prediction of results in sports using data mining techniques. But this paper was not specific to any particular sport rather it was in for all general sports. The attribute selection algorithm that it used was more of an elimination approach where the attributes were eliminated one by one and the classification accuracy was computed. Once a good subset of attributes is achieved, then the eliminated attributes are again added one by one to see if the accuracy improves, But, in my approach, I did not use this elimination approach for the attribute selection part. Then the paper used various data mining algorithms to perform classification. I used the naïve Bayes and the decision tree algorithms from the paper in my project and compared my accuracy with that of the paper.

1.3 Hardware Specification

a) Server/Desktop Computer:

- (1) Processor: Quad-core processor or higher (Intel Core i5 or equivalent)
- (2) RAM: 8 GB or more
- (3) Storage: 256 GB SSD or higher
- (4) Network Interface Card: Ethernet or Wi-Fi for network connectivity
- (5) Operating System: Windows 10/11, Ubuntu Linux, or macOS

b) Database Server (If Separate):

- (1) Processor: Dual-core processor or higher
- (2) RAM: 4 GB or more
- (3) Storage: 128 GB SSD or higher
- (4) Database Management System: MySQL, SQL, or equivalent

c) Networking Equipment:

- (1) Router: Wi-Fi router for wireless connectivity (if applicable)
- (2) Switch: Ethernet switch for wired connections
- (3) Cabling: Ethernet cables for wired connections

d) Client Devices:

- (1) Desktop/Laptop Computers: To access the system for administrative tasks.
- (2) Smartphones/Tablets: For students and staff to access the system via a web or mobile application.
- (3) Minimum Device Specifications: Modern web browser (Chrome, Firefox, Safari, Edge) with internet connectivity.

e) Printers and Scanners:

- (1) Printer: High-quality printer for generating invoices, reports, and other documents.
- (2) Scanner: For digitizing physical documents, such as student records or maintenance requests.

f) Security Measures:

- (1) Firewall: Hardware or software firewall to protect the server and network from unauthorized access.
- (2) Antivirus Software: To ensure the security of the system against malware and viruses.
- (3) Backup Solution: Regularly backup the database and system files to prevent data loss.

g) Optional Hardware:

- (1) Biometric Devices: For enhanced security, such as fingerprint scanners or facial recognition systems.
- (2) CCTV Cameras: For monitoring common areas and enhancing security within the hostel premises.

h) Development Tools and Accessories:

- (1) Development Environment: Integrated Development Environment (IDE) such as Visual Studio Code, Eclipse, or IntelliJ IDEA for coding.
- (2) Version Control: Git and a platform like GitHub or GitLab for collaborative coding and version control.
- (3) USB Drives or External Hard Drives: For data backup and storage.

i) Internet Connectivity:

- (1) Stable and reliable high-speed internet connection for seamless system access and updates.

j) Power Backup:

- (1) Uninterruptible Power Supply (UPS) or power backup solution to prevent data loss during power outages.

1.4 Software Specification

a) Development Environment:

- i) Integrated Development Environment (IDE): Visual Studio Code, for coding and development.

b) Programming Languages:

- i) Backend: Python, Jupyter Notebook(Pandas)
- ii) Frontend: Power BI, Excel, and a Node.js framework like React or Angular.

c) Database Management System:

- i) Relational database systems like SQL, Power BI, or Pandas.

d) Web Server:

- i) XAMMP or localhost ports / GitHub or Heroku.

e) Version Control:

- i) Git for version control to track changes in your codebase and collaborate effectively with team members.

f) API Development Tools:

- i) API tools for connection of backend and frontend and adding further activities like the weather.

g) User Interface Design Tools:

- i) Graphic design software like Figma, or Sketch for creating mockups and designing user interfaces.

h) Operating System:

- i) Operating system based on your development environment preferences. We will be using Windows 11.

i) Authentication and Security Libraries:

- i) Integrate authentication and security libraries such as Passport.js, Spring Security, or JWT (JSON Web Tokens) for securing user data and access.

- j) **Communication platforms:** like Zoom, Microsoft Teams, or Discord for effective communication within your development team.

k) Deployment Tools:

- i) For deployment, Docker for containerization, along with container orchestration tools like Kubernetes for scaling and management.

2. LITERATURE SURVEY

The paper describes a system that collects raw data for each sport, team, and player, and it is processed into statistical data. These data sets are clustered and stored as the data to be stored is very large. The paper also discusses wearable sensors. These wearable sensors are used in recognizing real-time tasks in sports. The devices are also helping the coaches in the transformation of Decision making. The paper concludes Machine Learning, along with Wearable devices can make a significant impact on the players by creating patterns, strategies, planning, reducing the risk of injury, and improving their performances.

In Cricket AI Tool is discussed it. It can help adjust certain factors to maximize the chances of winning the real game. The paper addressed the problem of predicting the chances of victory in a One Day International cricket match. The paper also shows a comparative evaluation of the classifiers. The systems described have been able to predict the winning criteria formulated using attributes from the dataset.

The Cricket AI tool can be used in real-world applications by teams playing cricket. It can help adjust certain factors to maximize the chances of winning the real game. Weighted Association Rule Mining algorithm for analyzing the Indian cricket team in one-day international cricket matches against Sri Lanka and South Africa is performed and mentioned.

This analysis is used by the team for framing game-winning plans. The author shows that a bigger data set can improve the accuracy of the prediction.

2.1 Existing System

The project is the developed module that has a User login registration and Admin login Registration. The system finds the Generated Winning team by using classification when a user-provided dataset. Admin adds all the information related to team players as well as a Team. The Proposed System is to Find out the Predicated Winning Team and generate the Result. This system works on two-way client-server players extracted from a particular tournament. We will observe that simple features can yield very promising results. Though Sports Analytics has shown a lot of improvement and advancement but still this interesting field has still been lagging in terms of applications in real sports. Many times news channels organize debates on predictions of cricket matches. Sponsors and Businessmen invest a lot of money in teams without knowing whether their team will win or not.

2.2 Proposed System

The proposed system will make use of Machine learning algorithms. ML Algorithms will be trained using data samples. The cricket data will be scraped from the website. The scrapped data will be cleaned and stored. The cleaned data will be processed to use for training purposes. The ML model will be trained using data. The cleaned data will also be used to generate visualization and descriptive statistics. Polyt and matplotlib libraries of Python will be used to draw graphs and charts. The ML model will be implemented using SK learn. The user of the system can set parameters for prediction. Depending on the parameters and model trained the system will make predictions. The predictions will be displayed to the user. Machine Learning algorithms can be broadly classified into three categories algorithms, regression algorithms, and clustering algorithms. One can leverage the advantage of all three categories in cricket analysis and prediction. The system makes use of all three categories for finding out helpful predictions. Random Forest will be used to predict the winner of the match. A lasso regression will be used to predict the score of the match and k means clustering to cluster players depending on their skills. All these machine learning models will be trained using data collected by scraping from the website. The data will be cleaned before saving to CSV and using it for training. Use can check predictions by using Interface. Descriptive statistics will also be provided to the user based on historical data.

2.3 Literature Review Summary

Year and Citation	Article	Author	Tools/Software	Technique	Source	Evaluation Parameter
2022	" r score prediction and players classification "	Sonu Kuma	Power BI, MySQL, HTML,	machine learning	Source	record against the opponent, weather and pitch conditions, and venue
2021	"National cricket team players and predict player's performance "	Aminul Islam Anik	Arduino, Raspberry Pi, IoT sensors	Support Vector Machine (SVM) with the linear, and polynomial kernel	Source	The model predicted runs scored by the batsman
2020	" Prospective analysis of T20 match results "	Gagana S	Biometric scanners	Machine learning	Source	a model for the outcome of ODI matches using decision
2019	"Predictive analysis of the winner using machine learning techniques "	Ch Sai Abhishek	Python, sci-kit learn	Machine learning	Source	Prediction accuracy and occupancy optimization
2018	"T20 International"	Muhammad Yasir	Android Studio, Java	designed	Source	player's history, winning percentage
2017	"Team members based on runs scored "	Kalpdrum Passi	Optimization algorithms	Resource allocation	Source	Resource utilization and cost reduction
2016	" T20 cricket matches "	Stylianios Kampakis	Naïve Bayes	User experience	Source	analyzing of T20 cricket matches
2015	" Cricket Outcome Predictor "	Neeraj Pathak	Naïve Bayes, Support vector machine	Modern classification	Source	ODI cricket matches

3. PROBLEM FORMULATION

Statistical modeling has been used in sports for decades and has contributed significantly to the success of the field. Cricket is one of the most popular sports in the world, second only to soccer. Various natural factors affecting the game, enormous media coverage, and a huge betting market have given strong incentives to model the game from various perspectives. However, the complex rules governing the game, the ability of players and their performances on a given day, and various other natural parameters play an integral role in affecting the outcome of a cricket match. This presents significant challenges in predicting the accurate results of a game.

The game of cricket is played in three formats - Test Matches, ODIs, and T20s. We focus our research on ODIs, the most popular format of the game. To predict the outcome of ODI cricket matches, we will propose an approach where we first estimate the batting and bowling potentials of the 22 players playing the match using their career statistics and active participation in recent games. We will use these player potentials to render the relative dominance one team has over the other.

Taking two other base features into account, namely, the toss decision and the venue of the match, along with the relative team strength, we adopt supervised learning algorithms to predict the winner of the match. The major algorithm used in the project will be SVM is a supervised machine learning algorithm that can be used for both classification and regression challenges. In this algorithm, we plot each data item as a point in n-dimensional space (where n is several features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well.

4. OBJECTIVES

Cricket is being played in many countries all around the world. There are lots of domestic and international tournaments being held in many countries which play cricket. Cricket is a game played between two teams comprising 11 players in each team.

1. Pitch: Unlike other sports, cricket stadiums' shape and size are not fixed except for the dimensions of the inner circle and pitch which are 30 yards and 22 yards respectively. Outfield variations and pitch can have a substantial effect on bowling and batting.

2. TOSS: According to cricket analysts, there is a sure measure of advantage for a team if it wins the toss. This might not be a deciding factor in a match but it would allow the team to choose “what they want”.

3. TEAM STRENGTH: The team strength should be balanced for winning a match. Captainship in a team is also a deciding factor. Records: The past team performances can be considered to predict the outcomes of the match

4. HOME GROUND ADVANTAGE: This is another attribute that determines the winner in the match. If you are playing in the home ground condition everything would be in your hands like climatic factors, pitch nature, and the major role played by the home crowd. The home team gets better motivation.

5. METHODOLOGY

I have followed the following methodology in the course of my project. The methodology consists of 5 different phases as shown in Fig. 1 i.e. Data set generation, Data cleaning, Attribute selection Data mining, and analysis of results. Each of these phases was part of my project milestones submitted and I had created a Gantt chart to keep track of the timeline. I will be discussing each phase in the following section in detail.

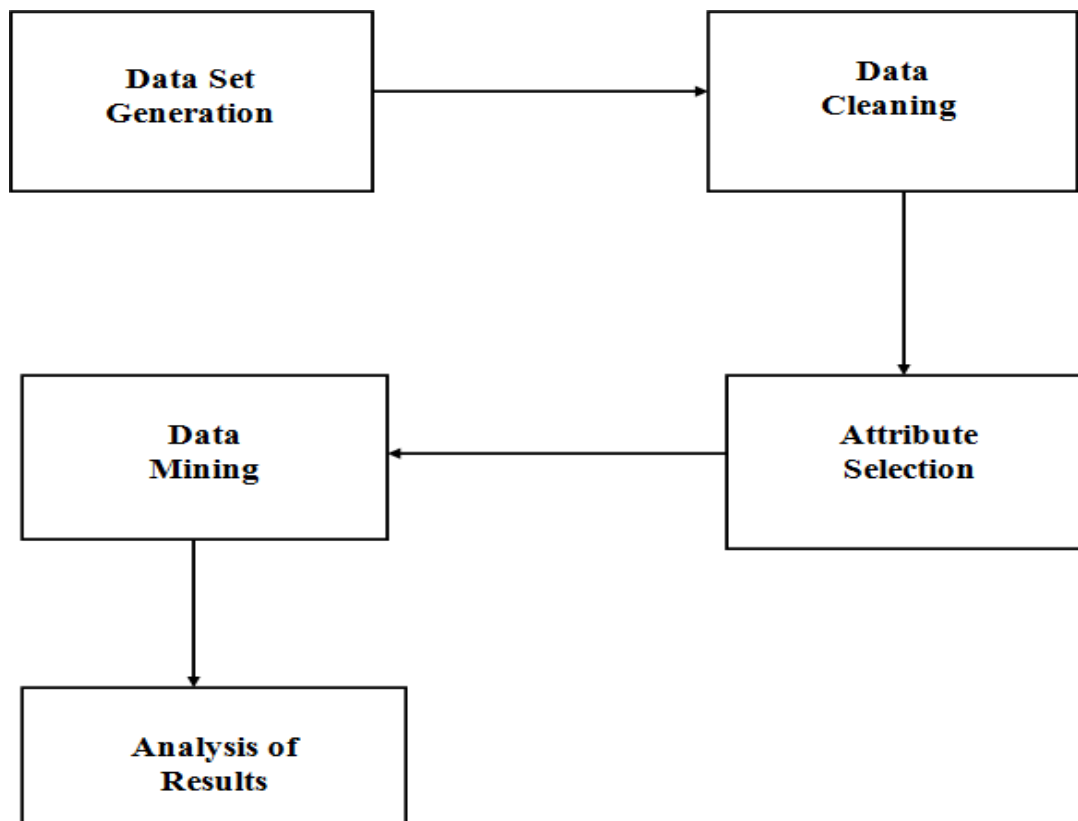


Fig:- Methodology diagram

1. Data Set Generation:

The data was collected from the <http://cricsheet.org> website. The website has data about all matches (from 2008 to 2016) of international tournaments held. It is an ODI format of tournaments. It means that each team bats or bowls for a maximum of overs each. The data set downloaded from this website was in the I used the java classes file and the file writer to read the. CSV file and write the content to a new file. The. CSV file had data about 5219 different entries and the Java code basically copied the important data from all the 5219 files and then combined multiple files into a single file.

2. Data Set Description:

The new combined data set that I generated had the data from the CSV files.

The data set consists of 5 different attributes and 5219 different instances.

The data set spans across all the seasons of the one-day international from 2008 to 2016 season.

3. Data Cleaning:

The data obtained from the <http://cricsheet.org> website was already cleaned. So, I did not have to do any sort of cleaning on the data. I had to tackle the missing values data and the data for the matches which were washed out due to rain. Those match data were present in the .json file. After combining the two data into a single data set, I had to manually fill in the data about the 7 missing teams from the <http://espncricinfo.com/> website. Moreover, 10 matches were washed out so that instances were filled with null values and they were discarded from the data set. So, the final dataset had 5219 instances with 21 attributes.

4. Attribute Selection:

The data set obtained after handling the missing values had to be filtered with the help of the attribute selection algorithm. Since there were 21 attributes it was necessary to identify all important attributes which would be useful for the data mining tasks.

6. EXPERIMENTAL SETUP

In the proposed models, we applied the data analytics technique and then ran the dataset in our model. Thereafter, we assigned values to the x and y variables which we had to predict. We used some libraries such as pandas, NumPy, random forest, linear regression, standard scalar, and Matplotlib libraries. Moreover, we applied some of the results from this model on Google Colab.

6.1. Train and Test Split Data Sets

The dataset is divided into training and test data. The values above the 2016 date year are related to training the dataset and the rest of the values are related to testing the dataset, to check how many predicted values of our model are correct in the test.

Our model's training and testing datasets are divided with a test size of 20% and a random state at zero.

6.2. Converting of String into an Object

This step will convert all the strings present in our dataset into objects to make them understandable for our model. The strings present will be converted into things using the Lambda function and date-time function.

6.3. Building the Model on Scikit Learn and Spark ML

The linear regression model was tested on Spark ML as well as Sklearn. The model is fitted to the encoded train and test samples. Thereafter, we will call a pickle file for the location where the model will work. In this model, building linear regression, we have chosen linear ridge regression, and in model selection, we have chosen the Grid Search CV model selection. In ridge selection, we will define all those parameters on which our prediction is based. Then, we will apply the ridge repressor model to calculate the mean square error of the model and make the prediction.

6.4. Evaluation Result of Linear Regression Using Scikit Learn

Before making the prediction, we successfully did the model creation and implemented the model on the dataset. In this step, we will check how efficient our model is. We will import metrics and NumPy from the Sklearn library to calculate its mean square error, root mean square error, and mean of an absolute error of the model.

6.5. Evaluation Result of Linear Regression Using Spark ML

The primary reason for building this model is to check which framework is better in terms of accuracy. In this model, we have done the same data visualization again and applied the descriptive data analysis on our dataset by checking the covariance of the data attributes that are here considered. Based on these attributes, we have built this model of prediction. In addition, we have used Spark machine learning libraries to apply the model and Spark context libraries to transform our data. The accuracy of the model using this Apache framework is better than the traditional machine learning Scikit. In the Spark machine learning framework, the model's accuracy is increased by 96% as training and 94.5% as testing accuracy. shows the accuracy, mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE).

7. CONCLUSION

The most significant outcomes of this paper can be summarized as follows: (1) The Spark framework is efficient compared to the traditional one, and (2) the linear regression model is the best model used to predict the total score of the match. In this paper, we have built a model that will indicate the winner of a cricket match with your input conditions of the ongoing game. We have seen that the best accuracy we can get is from the linear regression model. It gives us 96% accuracy using the Spark machine learning framework, in terms of the prediction analysis that tells us how efficient our model is in the prediction process. The other models also show excellent results since they do not have good accuracy in the confusion matrix and R mean squared error. Overall, the performance of both models is outstanding and can be used in any match to predict the winner. It is recommended to apply some analytical tools to your dataset to train your model since your accuracy will not be as good if your dataset is not according to the accurate data. In the future, we can build and use these models in many other sports. From the marathon race datasets, we can predict the percentage of winning countries and cities and we can predict which city or country has more chances to win. Furthermore, these kinds of predictions can be performed in the future.

8. TENTATIVE CHAPTER PLAN FOR THE PROPOSED WORK

CHAPTER 1: INTRODUCTION

In this era, while many sports are played in different countries, a few have been liked and encouraged a little more than others. Similarly, cricket is one of the most picked and played sports of this modern era. cricket was introduced in England in the sixteenth century. Initially, cricket was introduced and played in a test format only. After some time, and due to some conditions and policies, the International Cricket Council introduced additional formats such as the T20 and ODI formats. There are three official formats in which cricket is played internationally with varying durations and standards. The one-day international (ODI) cricket is one of the most played and liked structures by everyone. In this format, 100 overs of play are designed. Given that there are 100 overs in a one-day game, each team plays 50 overs, to fight and win. The datasets are available on different electronic databases with the maximum available information.

CHAPTER 2: LITERATURE REVIEW

In this section, we will discuss the previous research conducted on cricket prediction to know the most popular technique used by the researcher for similar works. In this paper, firstly, we have researched the cores of our topic related to predicting the winner in a cricket match. We have seen that many other researchers have done some work and proposed their results using different digital databases and libraries. All of these results are available for access by everyone. For the models that show high accuracy in the prediction model, we have taken these two models and implemented them on the new ODI dataset, as well as checked their performances. Then, the best model with high accuracy is again built on the Spark ML framework to check whether the Spark framework is efficient or not for building the prediction model. To make this model, the datasets were obtained from cricsheet.com, an online data store where you can get any dataset related to any topic. We have built two significant models and compared their performance based on their accuracy and confusion matrixes.

CHAPTER 3: OBJECTIVE

Statistical modeling has been used in sports for decades and has contributed significantly to the success of the field. Cricket is one of the most popular sports in the world, second only to soccer. Various natural factors affecting the game, enormous media coverage, and a huge betting market have given strong incentives to model the game from various perspectives. However, the complex rules governing the game, the ability of players and their performances on a given day, and various other natural parameters play an integral role in affecting the outcome of a cricket match. This presents significant challenges in predicting the accurate results of a game.

CHAPTER 4: METHODOLOGIES

We have followed a general Machine Learning Cycle to build our project which includes the following steps:

1) Data Collection: We collected the dataset from ESPN Cricinfo.

2) Hypothesis Generation: Here entire dataset and its features are analyzed by studying various publications and coming to know about the features that will have an impact in predicting the result of the match.

3) Data-Preprocessing: The dataset we had consists of missing values, outliers, and duplicate records. So, to build a model, we had to first correct these values in the necessary features to improve accuracy and check these features using univariate and bivariate analysis.

4) Data-Encoding: As our project is about a Classification Problem, to analyze some categorical variables we have done encoding of those feature values for better results.

5) Model Selection and Evaluation: We built our model using Random Forest and Support Vector Machine (SVM) and by using the K-Fold cross-validation technique entire data set was used for both training and testing of the model.

CHAPTER 5: EXPERIMENTAL SETUP

As the Random Forest Algorithm has an accuracy that is far greater than that achieved by the Support Vector Machine (SVM) algorithm, we have used the Random Forest in the final model for predicting the winner of the match. Then the inputs from the user are taken after the toss to predict the outcome of the match. So, we can simply say that most of these machine learning problems are optimization problems in which we minimize a target label into different numerical constraints.

CHAPTER 6: CONCLUSION AND FUTURE SCOPE

The research aimed at whether a machine learning model can be used to predict scores in rain-interrupted matches. To address this study, we studied the features that can be useful during our research in section 2.3. Hyperparameter optimization and random search cross-validation methods were used to find the best parameter for all three models. As seen in section 6, all models were trained and tested on the transformed data and the result shows that the margin of error for our model is less compared to the Duckworth-Lewis method. The proposed models produce good results with less RMSE and MAE values as well as good variance scores. The first innings model performance was not however good when compared to the second innings model. The second innings model contains the first innings total similarly future work could involve considering the average score of each team in the IPL. Also features related to team strength based on previous performance or rankings. Batters and bowlers' features can also be considered to check if the model produces better results.

REFERENCES :

1. S. Muthuswamy and S. S. Lam, "Bowler Performance Prediction for One-day International Cricket Using Neural Networks," in Industrial Engineering Research Conference, 2008.
2. G. D. I. Barr and B. S. Kantor, "A Criterion for Comparing and Selecting Batsmen in Limited Overs Cricket," Operational Research Society, vol. 55, no. 12, pp. 1266-1274, December 2004.
3. S.. R. Iyer and R. Sharda, "Prediction of athletes performance using neural networks: An application in cricket team selection," Expert Systems with Applications, vol. 36, pp. 5510-5522, April 2009.
4. I.P. Wickramasinghe, "Predicting the performance of batsmen in test cricket," Journal of Human Sport & Exercise, vol. 9, no. 4, pp. 744-751, May 2014.
5. Lemmer, H. H. (2008). Analysis of players' performances in the first Cricket Twenty-World Cup Series. South African Journal for Research in Sport, 30(2), pp.71-77.
6. Lewis, A. J. (2005). Towards fairer measures of player performance in one-day cricket. Journal of the Operational Research Society, 56, pp.804-815.
7. Saikia, H., Bhattacharjee, D., & Bhattacharjee, A. (2012). Is IPL Responsible for Cricketer's Performance in the Twenty20 World Cup? International Journal of Sports Science and Engineering, 6(2), pp.96-110.
8. Brooks, R. , Bussie`re, L. F., Jennions, M. D., & Hunt, J. (2003). Sinister strategies succeeded at the Cricket World Cup. Proceedings of the Royal Society.
9. "Free Download web scraping tool -web scraper | ParseHub," parsehub,[Online].Available: <https://www.parsehub.com>.
10. "Data extracted from the web," Import.io, [Online].Available: <https://www.import.io>. [11] Tim Kam Ho, "The Random Subspace Method for Constructing Decision Forests," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 8, pp. 832-844, August 1998