INDIAN INSTITUTE OF INFORMATION TECHNOLOGY

Dharwad, Karnataka- 580009

A Project Report On

"IPL Match Score Prediction"

Submitted by

M Joshasree (19BCS069)

P Lalithaanjale (19BCS087)

S Pranay Sai Teja (19BCS102)

Mohammed Abdul Sohail (19BCS122)

M Anupama (19BCS123)

Under the Guidance of

Dr. Deepak K T

Asst. Professor

Electronics and Communication Engineering

Indian Institute of Information Technology Dharwad

CERTIFICATE

This is to certify that the Project Work entitled — **"IPL Match Score Prediction"** is a bonafide work carried out by M. Joshasree(19BCS069), P. Lalithaanjale(19BCS087), S. Pranay Sai Teja(19BCS102), Mohammed Abdul Sohail(19BCS122), M. Anupama(19BCS123) in fulfillment for the Machine learning Project of Bachelor of Technology in Computer Science & Engineering of the Indian Institute of Information Technology Dharwad during the year 2021-2022. The Project Report has been approved as it satisfies the academics prescribed for the Bachelor of Technology degree.

Signature of Supervisor(s)

Name(s)

Department(s)

(Month, Year)

DECLARATION

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

M. Joshasree

19BCS069

P. Lalithaanjale

19BCS087

S. Pranay Sai Teja

19BCS102

Mohammed Abdul Sohail

19BCS122

M. Anupama

19BCS123

ACKNOWLEDGEMENT

It is a great pleasure for us to acknowledge the assistance and support of a large number of individuals who have been responsible for the successful completion of this project.

It is our privilege to express our sincerest regards to our project coordinator, **Dr. DEEPAK K T**, Asst. Professor, Department of Electronics and Communication Engineering, Indian Institute of Information Technology Dharwad, for his valuable inputs, able guidance, encouragement, whole-hearted cooperation and constructive criticism throughout the duration of our project.

We take this opportunity to thank all our lecturers who have directly or indirectly helped our project. We pay our respects and love to our parents and all other family members and friends for their love and encouragement throughout our career. Last but not the least we express our thanks to our friends for their cooperation and support.

ABSTRACT

We worked on a model that can predict the score of an IPL Team from a point in time. The constraints though being that a minimum of 5 overs need to be taken as the input for prediction of score.

We've tried out 2 Models/Methods for the Prediction Model :
  a) Random Forest Classifier
  b) Lasso Regression

For a smooth experience for User, a Web Application is provided where the user enters the Input Parameters.

The Input Parameters include :
  a) Batting Team
  b) Bowling Team
  c) Venue
  d) Overs Bowled (Eg : 6.4)
  e) Runs Scored (Eg : 72)
  f) Wickets Fallen (Eg : 2)
  g) Runs scored in previous 5 Overs (Eg : 52)
  h) Wickets taken in previous 5 Overs (Eg : 1)

Here, the Venue at which a match is being played is also important since the Score varies. For Example, in a smaller venue such as M Chinnaswamy (Bengaluru), the score will be quite higher compared to larger venues such as Eden Garderns (Kolkata).

As a result, we get the predicted score after providing the required inputs. The predicted score is generated using Lasso Regression since it was more effective than Random Forest Classifier

CONTENTS

CHAPTER 1

INTRODUCTION

Cricket is ingrained in Indian culture, and elite players such as Sachin Tendulkar, Virat Kohli, MS Dhoni, and Rohit Sharma frequently achieve celebrity status and are among the country's most powerful individuals. In this, IPL is the feast which everyone awaits for. The best method for IPL prediction is machine learning. Based on prior datasets available on Kaggle from 2008 to 2017, a model utilizing machine learning methods is suggested in this project to predict the score of each match. Reinforced, unsupervised, and supervised learning techniques can all be classified. These methods are utilized depending on the application and the outcome.

The Indian Premier League (IPL) is a Twenty20 professional cricket league controlled by the Board of Control for Cricket in India. Every year, the league takes place, with competing clubs representing various Indian cities. Twenty20 cricket leagues are organized in a number of countries. Data from Indian Premier League matches has also been used in several intriguing machine learning experiments.Here we trained a model from scratch and integrated it into a web app using basic and powerful frameworks like sklearn, pandas, and flask. There is also some web development being done.

1.1 Problem Statement

Our main objective is to design an optimized machine learning model to predict the score of a batting team in the first innings of each IPL match using machine learning techniques or algorithms (basically comparing different algorithms in terms of different comparison metrics of machine learning models and selecting the optimized one) with available datasets on Kaggle from 2008 to 2017 for training the model.

In addition, we also aim to develop a website and host it to make it look good while predicting the runs scored by the batting team in the first innings.

CHAPTER 2

LITERATURE SURVEY

2.1 Introduction

In this section, we are interested to elaborate the main articles that we have followed to implement the project. This section contains two subsections having explanations and observations from the articles that used required algorithms.

2.2 Review 1

From the article **Prediction of IPL Match Score and Winner Using Machine Learning Algorithms**,we learned about :
Algorithms like :
● Lasso Regression.
● Random Forest.
These algorithms are proposed for Score prediction. SVM(Linear, RBF), Logistic Regression classifier is for the match-winning prediction. The algorithm is selected based on its Precision, Recall and F-Measure.

For reference : [Prediction of IPL Match Score and Winner Using Machine Learning Algorithms](#)

2.3 Review 2

From the article,**"Predictive Analysis of an IPL Match Winner using Machine Learning Techniques"**,We learned about various factors that influence the outcome of an Indian Premier League match.
In this article,
● Logistic Regression
● Decision Trees
● Random Forest
● K-Nearest Neighbors
These Algorithms are used and observed that Random Forest Classifier and Decision Tree provided the highest accuracy.

For reference : [Predictive Analysis of IPL Match Winner using Machine Learning Techniques](#)
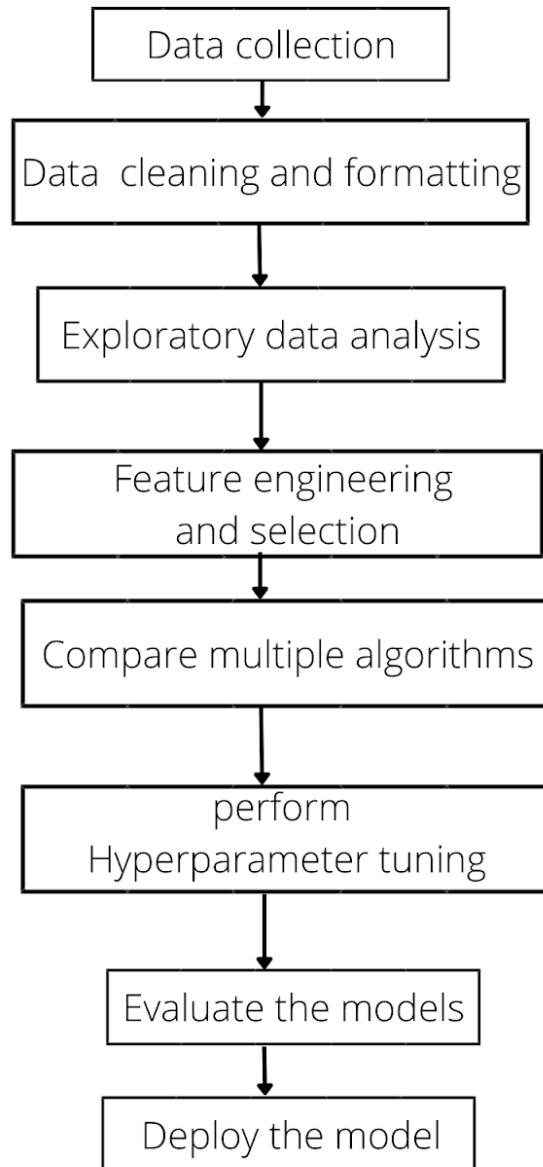
CHAPTER 3
WORKFLOW

```
┌─────────────────────────────┐
│       Data collection        │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│  Data  cleaning and formatting │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│   Exploratory data analysis  │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│      Feature engineering     │
│        and selection         │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│  Compare multiple algorithms │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│           perform            │
│    Hyperparameter tuning     │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│      Evaluate the models     │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│       Deploy the model       │
└─────────────────────────────┘
```

Figure 3 - Flow Chart

CHAPTER 4

PROCEDURE

4.1 Data Description

Data description involves both data collection and data preprocessing.

4.1.1 Data Collection

The first step in the machine learning workflow is to collect data for training the ML model. The accuracy of ML systems' predictions is only as good as the data used to train them. Some of the data sources on which we can rely on for data collection are :

- Dataset search by Google
- Open ML
- UCI: Machine Learning Repository
- Public datasets on GitHub
- Visual Data Discovery
- Kaggle
- Amazon Datasets

We have collected our data from Kaggle and data world(website).

For reference :

https://www.kaggle.com/ramjidoolla/ipl-data-set?select=deliveries.csv

https://data.world/raghu543/ipl-data-till-2017/workspace/file?filename=Match.csv

4.1.2 Data Preprocessing

Raw data and images from the real world are frequently incomplete, unreliable, and lacking in specific behaviors or trends. They're also likely to be riddled with errors. As a result, when they've been gathered, they're pre-processed into a format that the machine learning algorithm can utilize to build the model.Pre-processing includes a number of techniques and actions:

- Data cleaning
- Data imputations
- Oversampling
- Data integration
- Data normalization

We have used *OneHotEncoding* and also converted the string date into a datetime object and rearranged the columns.

4.2 Random Forest Regression

Random Forest Regression is a supervised learning approach for regression that employs ensemble learning. The ensemble learning method combines predictions from numerous machine learning algorithms to produce a more precise forecast than a single model.
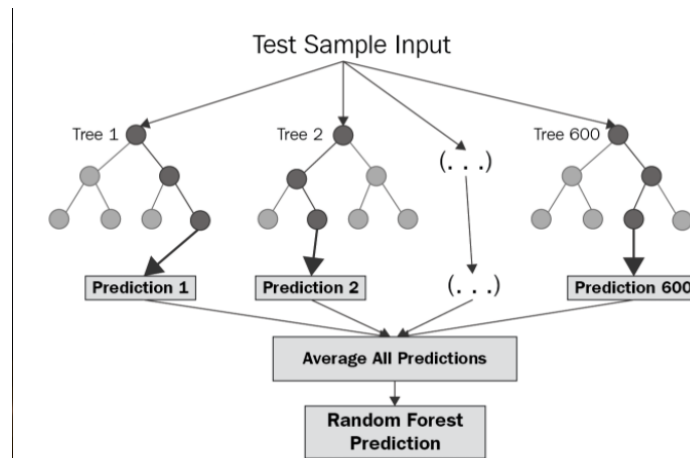

Figure 4.1- Visualization of Random forest regression workflow

Random Forest Regression is a powerful and precise model. It usually works well on a wide range of issues, including those with non-linear relationships. However, there are some drawbacks: there is no interpretability, overfitting is a possibility, and we must choose the amount of trees to include in the model.For Random Forest, we have used *RandomizedSearchCV* for Hyper parameter tuning.

4.3 Lasso Regression

The word LASSO stands for "*Least Absolute Shrinkage and Selection Operator*".

Statistical formula for

i) Regularization of Data Models

ii) Feature Selection

Lasso regression is a type of linear regression that uses shrinkage. Shrinkage is where data values are shrunk towards a central point, like the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). For Lasso regression, we have used *GridSearchCV* for Hyperparameter tuning.

4.4 Regularization

Regularization is an important concept that is used to avoid overfitting of the data, especially when the trained and test data are much varying. In regularization, what we do is normally keep the same number of features but reduce the magnitude of the coefficients. We can reduce the magnitude of the coefficients by using different types of regression techniques that use regularization to overcome this problem.
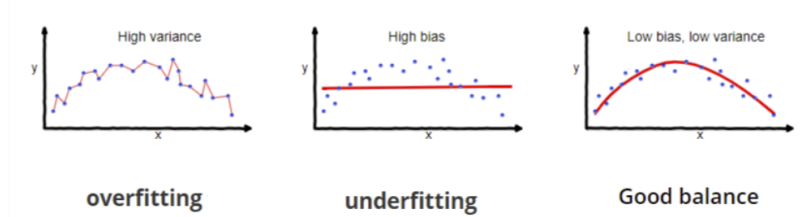

Figure 4.2

4.5 Comparing the Algorithms

For evaluating the Lasso Regression model and Random Forest Regression model we used **Distplot** and **Sklearn Metrics**.It can be observed that most of our values are 0 or close to 0 in Lasso Regression. In Random Forest we observed that not many of our values are 0 or close to 0 when compared to Lasso Regression. Also the error values are higher than Lasso Regression. Therefore we can state that the Lasso regression model has a better accuracy than Random Forest.

CHAPTER 5

RESULTS

Figure 5.1 - Web Application Provided to User

IPL Batting Score Predictor 🏆

Batting Team:
Delhi Daredevils

Bowling Team:
Chennai Super Kings

Venue:
Feroz Shah Kotla

Overs Bowled:
9

Runs Scored:
81

Wickets Fallen:
2

Runs Scored in previous 5 Overs:
45

Wickets taken in previous 5 Overs:
1

Predict Score

Figure 5.2 - Providing Input Parameters

IPL Batting Score Predictor 🏆

The final predicted score (range):
165 to 180

Figure 5.3 - Predicted Score

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

The numerous factors that impact the result of an Indian Premier League match were found in this research. The seven factors which significantly influence the result of an IPL match include the batting team, bowling team, venue, overs bowled, runs scored, wickets fallen, runs scored in previous 5 overs, wickets taken in previous 5 overs. All match data from 2008 to 2017 was included in the dataset. Lasso regression has been used for the prediction as we are getting less mean square error, less mean absolute error when compared with other algorithms. A lasso regression based ML model is used to predict the runs scored by the batting team in their first innings based on above mentioned features. A constraint of considering runs scored in previous 5 overs and wickets taken in previous 5 overs has been followed in order to use Lasso regression.

In Future, we can consider adding a "Power Factor" for each player which we can take into consideration while predicting the score. The reason being some players have a greater ability to hit the ball for sixes such as Andre Russell & Keiron Pollard when compared to inexperienced players such as Riyan Parag & Ripal Patel. Also in Future, the Web Application will be converted to a Mobile based Application for both Android & ioS Users.

CHAPTER 7

SOURCE CODE

Source code uploaded in GitHub contains Dataset(CSV file), Machine learning source code, pickle file(containing Lasso regression model), static files, templates, requirements text file and web application code done using python.

Github Link (for reference) : https://github.com/ANUPAMA0221/ML_Project

CHAPTER 8

REFERENCES

- Prediction of IPL Match Outcome Using Machine Learning Techniques by Srikantaiah K C, Aryan Khetan, Baibhav Kumar, Divy Tolani, Harshal Patel,Department of CSE, SJB Institute of Technology, Affiliated to Visveswaraya Technological University BGS Health & Education City, Bengaluru-560060, Karnataka, India.
- Live Cricket Score Prediction Web Application using Machine Learning by Eeshan Mundhe, Ishan Jain and Sanskar Shah
- Daniel Mago Vistro, Faizan Rasheed and Leo Gertrude David, "The Cricket Winner Prediction With Application Of Machine Learning And Data Analytics", International Journal of Scientific & Technology Research, vol. 8, no. 09, 2019.
- Predicting Outcome of Indian Premier League (IPL) Matches Using Machine Learning by Rabindra Lamsal and Ayesha Choudhary, School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi 11006.
- LASSO: A feature selection technique in predictive modeling for machine learning by R Muthukrishnan, R Rohini, Coimbatore, India, 2017.
- IPL Winner prediction using ML algorithms by Ruchitha M and Dr. Preeti Savant, Department of Computer Application Jain Knowledge Campus, Bangalore Jain University, India.
- [IPL Score Prediction Using Machine Learning | by Chitwan Manchanda | MLearning.ai | Medium](#)
- https://www.statisticshowto.com/lasso-regression/
- [Prediction of IPL Match Score and Winner Using Machine Learning Algorithms](#)
- [IPL CRICKET SCORE AND WINNING PREDICTION USING MACHINE LEARNING TECHNIQUES](#)