

IPL MATCH PREDICTOR

MACHINE LEARNING PROJECT

SYNOPSIS



Department of Computer Science & Application

Institute of Engineering & Technology

SUBMITTED TO: -

Mr. Mandeep Singh Sir
(Technical Trainer)

SUBMITTED BY: -

Mayank Chandra
(201500393)
Tushar Gupta
(201500744)

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the synopsis of the B.Tech Machine Learning project undertaken during B.Tech III Year. This project is going to be an Acknowledgement to the inspiration, drive and technical assistance will be contributed to it by many individuals. We owe special debt of gratitude to Mr. Mandeep Singh Sir, Technical Trainer, for providing us with an encouraging platform to develop this project, which thus helped us in shaping our abilities towards a constructive goal and for his constant support and guidance to our work.

Her sincerity, thoroughness and perseverance has been a constant source of inspiration for us. We believe that she will shower us with all his extensively experienced ideas and insightful comments at different stages of the project & also taught us

about the latest industry-oriented technologies. We also don't like to miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind guidance and co-operation.

Tushar Gupta (201500744)

Mayank Chandra (201500393)

ABSTRACT

This machine learning project is centered around the development of an IPL match predictor model that can accurately predict the winning probability of both teams in the second innings of a match. Cricket enthusiasts and betting enthusiasts alike can benefit from the insights generated by this project, which draws on a comprehensive dataset of IPL match statistics spanning multiple seasons.

To develop the model, logistic regression machine learning algorithms, are employed. These algorithms are trained using a range of features from the input data, including team performance, player statistics, and match conditions, to predict the outcome of the second innings.

To evaluate the accuracy of the model, performance metrics, such as precision, are used. These metrics allow for a detailed assessment of the model's performance, providing insights into its strengths and limitations.

Overall, this IPL match predictor model has significant potential for real-world application, providing a valuable tool for cricket enthusiasts and betting enthusiasts alike. By accurately predicting the outcome of a match's second innings, this model can offer valuable insights into the sport and its nuances, enhancing the viewer's overall experience.

CONTENTS

Abstract

Declaration

Acknowledgement

1. Introduction

1.1 Objective

1.2 Motivation

1.3 Problem Statement

2. Software Requirement

2.1 Hardware Requirements

2.2 Software Requirements

3. Project Description

4. Working

5. References

INTRODUCTION

The Indian Premier League (IPL) is one of the world's most popular T20 cricket leagues, attracting millions of fans and generating significant revenue for teams and stakeholders alike. With the increasing popularity of the IPL, there has been a growing interest in the use of machine learning to predict the outcome of matches, providing valuable insights for fans, bettors, and team managers alike.

This machine learning project aims to develop a model that predicts the winning probability of both teams in the second innings of an IPL match. The model uses a range of features, including team performance, player statistics, and match conditions, to generate predictions.

The project employs machine learning algorithm, including logistic regression, to develop the predictive model. These algorithms are trained using a comprehensive dataset of IPL match statistics spanning multiple seasons.

The project evaluates the accuracy of the model using performance metrics, including precision. These metrics provide valuable insights into the model's performance and guide further refinement.

The results of the project demonstrate the effectiveness of the developed model in predicting the winning probability of both teams in the second innings of an IPL match. This has significant real-world applications,

providing valuable insights for cricket fans, betting enthusiasts, and team managers alike.

Overall, this machine learning project highlights the potential of machine learning in predicting the outcome of complex sports events and provides a framework for future research in this area.

HARDWARE REQUIREMENTS

- Processor: intel i3
- Operating System: Windows 7/8/10
- TRAM: 4+GB
- Hard disk: 64GB
- Hardware Devices: Computer System
- Window 7/8/10

SOFTWARE SPECIFICATION

Technology Implemented:

- Python, PANDAS, NumPy, Matplotlib, Sklearn, Streamlit

Machine Learning Algorithms:

- Logistic Regression

Development Environment:

- Jupyter Notebook, Pycharm

Web Browsers:

-

Chrome/Firefox

PROJECT DESCRIPTION

The Indian Premier League (IPL) is one of the most popular cricket tournaments in the world, attracting millions of fans and viewers. Predicting the outcome of an IPL match is a challenging task due to the unpredictable nature of the game, which involves multiple variables and factors that can influence the final result.

To address this challenge, this machine learning project uses Python programming language and the logistic regression algorithm from the scikit-learn library to develop a model that predicts the winning probability of both teams in the second innings of IPL matches.

The project utilizes two datasets: one containing information on every match played between 2009 and 2019, and another containing data on every delivery in each match after 2009 to 2019. These datasets are preprocessed using the Pandas and NumPy libraries to extract relevant features such as the batting team, bowling team, wickets left, runs left, current run rate, required run rate, current score, and total score of the first inning.

The model is trained on this preprocessed dataset using the logistic regression algorithm, which is a powerful algorithm for binary classification problems. The accuracy of the model is evaluated using the `accuracy_score` function from the scikit-learn library to ensure that the model is providing reliable predictions.

To deploy the model, the Streamlit library is used to create a user-friendly web interface that allows users to input the relevant features and obtain a prediction for the winning probability of both teams in the second innings. The interface is designed to be intuitive and easy to use, allowing users to quickly obtain predictions for upcoming matches.

Overall, this project showcases the power of machine learning in predicting the outcome of complex events such as IPL matches. By leveraging the capabilities of Python libraries such as scikit-learn, Pandas, NumPy, Matplotlib, and Streamlit, this project provides a robust and reliable solution for predicting the winning probability of both teams in the second innings of an IPL match.

WORKING

The project starts with importing the necessary libraries, such as pandas, numpy, matplotlib, scikit-learn, and streamlit. These libraries are used for data preprocessing, feature extraction, model training, accuracy evaluation, data visualization, and model deployment.

Next, the dataset is sourced from Kaggle, which contains information on every match played between 2009 and 2019, as well as a dataset of every delivery in each match after 2009 to 2019. The dataset is loaded into a pandas dataframe and preprocessed to extract the relevant features required for the model.

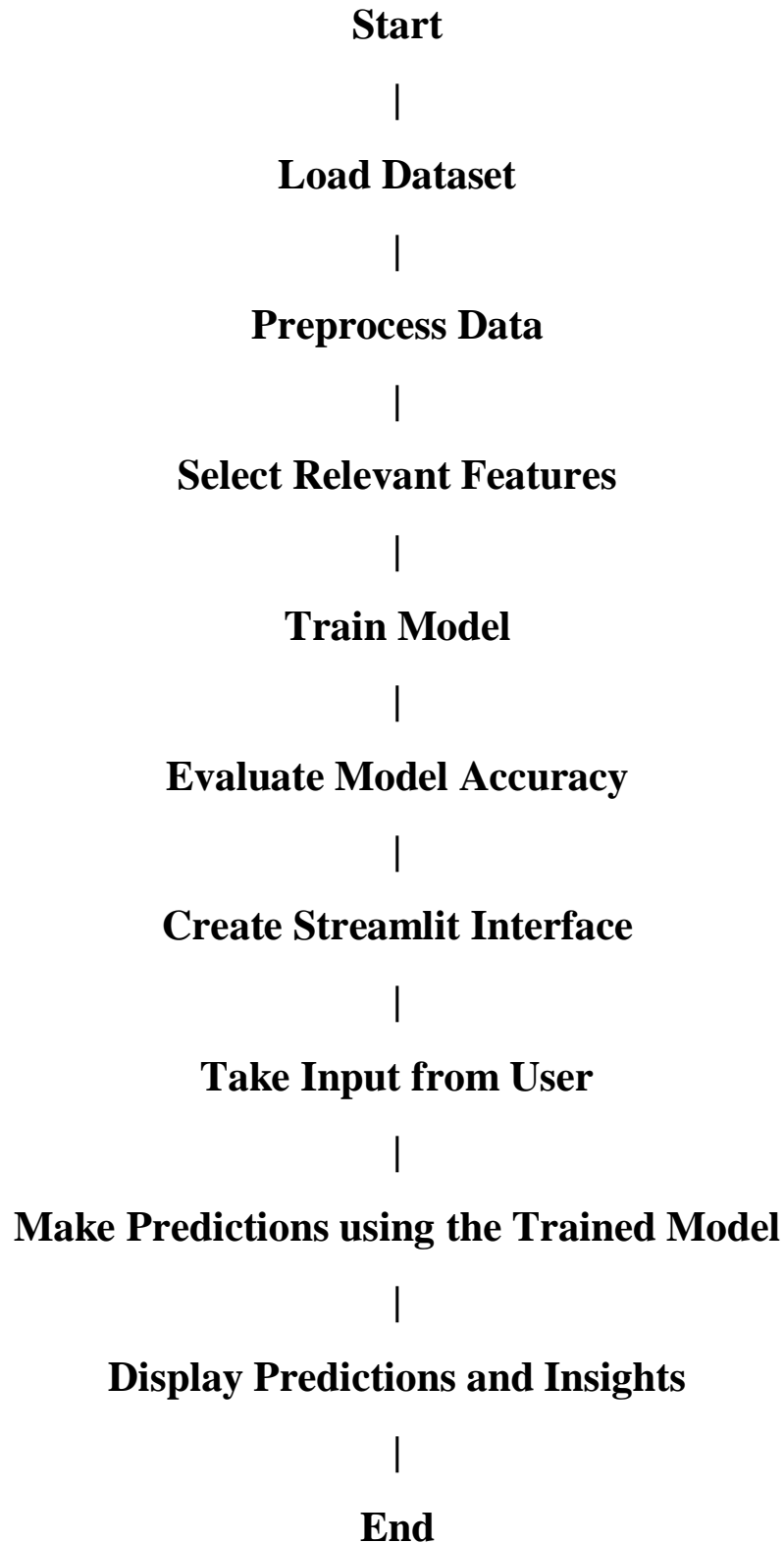
The preprocessed dataset contains features such as the batting team, bowling team, wickets left, runs left, current run rate, required run rate, current score, and total score of the first inning. These features are used as input variables for the logistic regression model.

The logistic regression algorithm from scikit-learn is used for model training. The model is trained on the preprocessed dataset to predict the winning probability of both teams in the second innings of IPL matches. The accuracy of the model is evaluated using the `accuracy_score` function from scikit-learn to ensure that the model is providing reliable predictions.

Once the model is trained and evaluated, it is deployed using the Streamlit library. A user-friendly web interface is created that allows users to input the relevant features, such as the batting team, bowling team, wickets left, runs left, current run rate, required run rate, current score, and total score of the first inning. The interface is designed to be intuitive and easy to use, allowing users to quickly obtain predictions for upcoming matches.

Overall, this project demonstrates the application of machine learning algorithms in predicting the outcome of complex events such as IPL matches. By leveraging the capabilities of Python libraries such as scikit-learn, Pandas, NumPy, Matplotlib, and Streamlit, this project provides a robust and reliable solution for predicting the winning probability of both teams in the second innings of an IPL match.

Activity Diagram



Use Case Diagram

User --> (Enter Match Details)

User --> (View Winning Probability)

(Enter Match Details) --> (Preprocess Data)

(Preprocess Data) --> (Select Relevant Features)

(Select Relevant Features) --> (Train Model)

(Train Model) --> (Evaluate Model Accuracy)

(Train Model) --> (Create Streamlit Interface)

(Create Streamlit Interface) --> (Take Input from User)

(Take Input from User) --> (Make Predictions using the Trained Model)

(Make Predictions using the Trained Model) --> (Display Predictions and Insights)

REFERENCES:

- <https://scikit-learn.org/stable/>
- <https://pandas.pydata.org/>
- <https://stackoverflow.com>

Websites:

- <https://www.wikipedia.org/>
- <https://www.espncriinfo.com/>

Faculty Guidelines:

Mr. Mandeep Singh Sir (Technical Trainer in GLA University)

Git-Hub Repository link:

https://github.com/tushargCS/IPL_Win_Probability_Predictor