***Project Phase II Report***

***On­­­***

**IPL SCORE PREDICTION**

**Submitted for the requirement of**

**Project course**

BACHELOR OF ENGINEERING

**COMPUTER SCIENCE & ENGINEERING**

****

**Submitted to: Monika**

**E-code: E12288**

**Project Teacher (Supervisor)**

**(Name & E-code)**

**Submitted By:**

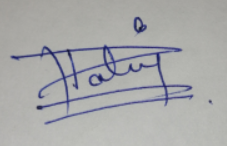
**NAME: JAIMEEN PATEL**

**UID:20BCS1469**

**Student Group:**

|  |  |
| --- | --- |
| **PRIYA CHOUDHARY** | **20BCS1465** |
| **TANISHA** | **20BCS1499** |
| **PANKAJ SHARMA** | **20BCS14** |
| **JAIMEEN PATEL** | **20BCS1469** |

**Co Supervisor Signature**

****

**(Name & E-code)**

**Name: Palvi Sharma**

**E-code: R66**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**CHANDIGARH UNIVERSITY, GHARUAN**

**June 2022**

**ABSTRACT**

Abstract –The main theme of this paper deals with the prediction of IPL Scores as they may vary many times based on different run rate, wickets, average, win ratio of the dataset which affect the winning of the team. There are many ML (machine learning) models which help us to predict mainly the score of the match. The main focus of them is to provide the optimal score. We have considered a dataset which is consisting of last season matches. For this paper, we have plotted dist plots and used some of the algorithms/methods to get the accuracy. This project helps the user to predict the score before ending the match. Out of all the methods used, we found random data gives the most accuracy than the other models.

**TABLE OF CONTENT**

1. Introduction 10
2. Literature Review 12
3. Problem Definition 16
4. Objectives 17

|  |  |  |
| --- | --- | --- |
| References |  |  |
|  | |  |
| **INTRODUCTION**  Overview: Cricket is believed to be a religion in India and people of all faiths, caste creed and language remain glued to the TV or the radio to watch international level matches. It is also the richest sport in India and top players have star status and earn millions of rupees through game appearance fees and endorsements. The Indian Premier League - IPL is a „20-20‟ version of Cricket where professional club teams from different cities plays about 80 games with a final game deciding the champion. The traditional format is the „Test‟ match format with the match of 5 days and two innings. The relatively recent format of one-day ‟50-50‟ series has 50 overs bowled by each team. In the „Twenty20‟ game each team plays 20 overs. Millions of people around the world watch the games and after the Olympics, FIFA World Cup and Euro Cup, cricket is one of the most watched games. Started in 2008, IPL series and brand was valued at 4.13 billion USD in 2009.  About Cricket: The game of Cricket in India is a passion that binds people from different religions, political affiliations, languages and economic background. Introduced by the British when they ruled India, the game is played by across the globe by a handful of common wealth nations such as India, Pakistan, Sri Lanka, Australia, New Zealand, South Africa, West Indies, Zimbabwe, Bangladesh, Britain and other new entrants. The number of people who watch and follow cricket in these cricket-playing nations is collectively more than a billion. Millions of people watch matches between rivals such as India and Pakistan and between England and Australia.  About IPL (Indian Premier League): The Indian Premier League (IPL, officially Vivo Indian Premier League for sponsorship reasons) is a professional Twenty20 cricket league in India contested during April and May of every year by teams representing Indian cities. The league was founded by the Board of Control for Cricket in India (BCCI), however Lalit Modi, the founder and former Commissioner, was the brainchild behind the birth of this league in 2007, which has now become a mammoth, money-spinning cricket venture. Vivo, the Chinese based smart phone multinational company, is serving as the title sponsor since the ninth season of the league.  IPL (Indian Premier League) – the Franchise:  The Indian Cricket League (ICL) was founded in 2007, with funding provided by Zee Entertainment Enterprises. The ICL was not recognized by the Board of Control for Cricket in India (BCCI) or the International Cricket Council (ICC) and the BCCI were not pleased with its committee members joining the ICL executive board. To prevent players from joining the ICL, the BCCI increased the prize money in their own domestic tournaments and also imposed lifetime bans on players joining the ICL, which was considered a rebel league by the board.  On 13 September 2007, the BCCI announced the launch of a franchise based Twenty20 cricket competition called Indian Premier League whose first season was slated to start in April 2008, in a "high-profile ceremony" in New Delhi. BCCI vice-president Lalit Modi, said to be the mastermind behind the idea of IPL, spelled out the details of the tournament including its format, the prize money, franchise revenue system and squad composition rules. It was also revealed that the IPL would be run by a seven-man governing council composed of former India players and BCCI officials, and that the top two teams of the IPL would qualify for that year's Champions League Twenty20. Modi also clarified that they had been working on the idea for two years and that IPL was not started as a "knee-jerk reaction" to the ICL. The league's format was similar to that of the Premier League of England and the NBA in the United States.  In order to decide the owners for the new league, an auction was held on 24 January 2008 with the total base prices of the franchises costing around $400 million. At the end of the auction, the winning bidders were announced, as well as the cities the teams would be based in: Bangalore, Chennai, Delhi, Hyderabad, Jaipur, Kolkata, Mohali, and Mumbai. In the end, the franchises were all sold for a total of $723.59 million. The Indian Cricket League soon folded in 2008. | |  |
|  |  |  |

**LITERATURE REVIEW**

Some researchers discussed about a new dimension for the Cricketing World. They highlighted the initial phase of IPL and the team formation in which players were assigned to various teams by means of auction. Siddhartha K Rastogi asserted about the final bidding Prices, Cricketing attributes of players and to her relevant information. Sanjeet Singh measured the technical efficiency of various teams participating in Indian Premier League. IPL has emerged as a great success and has been called Billion Dollar Baby. However, Coates didn’t agrees with the fact that professional League games help a community or a city to increase revenue. Shashi Kadapa discussed the very high popularity of Indian Premier League with more than 140 million TV audiences and a brand valuation of more than 4 billion USD. He praised the IPL business model for integrating many complex factors like entertainment, glamour, marketing, pricing and the hard-hitting Cricket. Moreover, he conducted a detailed analysis of 10 key issues that threaten the viability of IPL strategy, revenue model and sustainability of IPL. IPL has established a benchmark of Cricket marketing in the whole Cricket World.

IPL is the most fashionable and entertainment Sports League in India. IPL provides a platform where international players from different countries and India’s upcoming talent play together. Many researchers have tried their hand in figuring out the different aspects associated with the Indian Premier League. Some tried to highlight its impact on the national economy, some tried to analyses the strategies used in IPL, some others discussed its popularity by means of the viewership involved and its brand valuation, while some others tried to evaluate its impact on the game of Cricket. Though many researchers have tried to discuss outcome of IPL – its impact on youngsters, opportunities created for uncapped players and its contribution towards national economy.

Since the dawn of the IPL in 2008, it has attracted viewers all around the globe. A high level of uncertainty and last moment nail biters has urged fans to watch the matches. Within a short period, IPL has become the highest revenue-generating league of cricket. In a cricket match, we often see the score line showing the probability of the team winning based on the current match situation. This prediction is usually done with the help of Data Analytics. Before when there were no advancements in machine learning, the prediction was usually based on intuitions or some basic algorithms.

In Machine Learning, the problems are categorized into 2 groups mainly: Regression Problem and Classification problem. The Regression problem deals with the kind of problems having continuous values as output while in the Classification problem the outputs are categorical values. Since the output of winner prediction is a categorical value, the problem which we are trying to solve is a Classification problem.

The below picture clearly tells you how bad is taking run rate as a single factor to predict the final score in a limited-overs cricket match.



**PROBLEM DEFINATION**

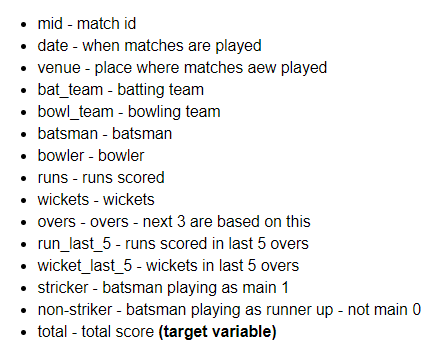
**ABOUT TOOLS AND METHODOLOGY:**

IPL, one of the biggest leagues in T20 cricket with millions of fans all over the world. Around 696 matches have taken place from 2008-2018. There is a huge data which include no. of matches of each innings with match location and all other necessary details. Jupyter notebook is used for integrating this dataset and the python libraries (NumPy, pandas etc.)

Kaggle.com offers various popular scientific packages for deep inspection and exploration of data. Proper Analysis and Visualization performed in jupyter notebook with numerous packages such as NumPy, Pandas, etc. The Pandas library provides a really fast and efficient way to manage and explore data. It does that by providing us with Series and Data Frames, which help us not only to represent data efficiently but also manipulate it in various ways. These features of Pandas is exactly what makes it such an attractive library for data scientists.

Pandas provide a wide array of built-in tools for the purpose of reading and writing data. While analyzing you will obviously need to read and write data into data structures, web service, databases, etc. This has been made extremely simple with the help of Pandas’ inbuilt tools. In other languages, it would probably take a lot of code to generate the same results, which would only slow down the process of analyzing.

For our use case, we are going to use the IPL Scores Dataset (link in reference) which has 76104 observations and 15 features:

****

## EDA – Exploratory Data Analysis for IPL Score Predictor

Having looked at the data quickly, let’s dive deeper into the dataset and explore some of the insights. This procedure is very important and will allow us to understand the data and plan our next steps. Luckily pandas provide easy-to-use functions to perform our analysis.

#### Check Summary Statistics

1. Measurement of Central Tendency – These measures allow us to understand where most of our data lies and mostly include:

* **Mean**– Average of Data
* **Median**– Center for Data
* **Frequency**– No of occurrence of specific data.
* **Mode**– Highest Observation in Data

2. Measures of Dispersion – These are the measures that allow us to understand how widespread data is and mostly includes:

* **Max & Min**– Highest and Lowest Value in the dataset
* **Range**– Highest-Lowest (captures the reach of data)
* **Variance**– Captures the variation of data (how data is varying) – Usually the sum of deviation of actual data from its mean/no of samples – 1
* **Standard Deviation** – Same as standard deviation but on the same scale as data -sqrt (variance)
* **Percentiles**– Capture the spread of data for a specific value i.e how much data is above or below it. 50% – Median

Performing each check will be cumbersome so pandas pack all these in a single function called describe.

**OBJECTIVES**

The main objective of this project is to predict the score before the matches or during the matches using the data set of previous matches and python libraries. The data set contains the runs, wickets, overs, run rate, venue, last wicket, striker, non-striker, total run. On the basis of this data set we will predict the score. We will give the information of both the teams are no. of matches, no. of wickets, run scored in previous 5 overs and wickets taken in previous 5 overs. And then we will be able to predict the score.

**REFRENCES**

* **Goggle for problem solving**
* [**https://www.kaggle.com/**](https://www.kaggle.com/)
* [**https://www.youtube.com/watch?v=4CtyDxfhoN8**](https://www.youtube.com/watch?v=4CtyDxfhoN8)
* [**https://github.com/krishnaik06/IPL**](https://github.com/krishnaik06/IPL)
* [**https://github.com/anujvyas/IPL-First-Innings-Score-Prediction-Deployment**](https://github.com/anujvyas/IPL-First-Innings-Score-Prediction-Deployment)
* [**https://www.analyticsvidhya.com/blog/2021/10/building-an-ipl-score-predictor-end-to-end-ml-project/**](https://www.analyticsvidhya.com/blog/2021/10/building-an-ipl-score-predictor-end-to-end-ml-project/)