

**Project Report  
on  
IPL SCORE PREDICTION**

**Project report submitted in partial fulfillment of the requirement for the award  
of the Degree of  
BACHELOR OF TECHNOLOGY  
IN  
COMPUTER SCIENCE AND ENGINEERING**

**Submitted By**  
SUGALI DIVYA [R180825]  
V LAKSHMI ANJALI [S180937]  
K LAKSHMI SOWJANYA [R181053]

**Under the Esteemed Guidance of**  
**Mr. R.SREENIVASULU,**  
**Assistant Professor**



**RAJIV GANDHI UNIVERSITY OF KNOWLEDGE TECHNOLOGIES(AP IIIT)  
R.K Valley, Vempalli, Kadapa(Dist) – 516330**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**2023-2024**

**RAJIV GANDHI UNIVERSITY OF KNOWLEDGE TECHNOLOGIES(AP IIIT)**  
**R.K Valley, Vempalli(M), Kadapa(Dist) – 516330**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**2023-2024**



## **CERTIFICATE**

This is to certify that the project report entitled “ ***IPL SCORE PREDICTION***” being submitted by V.LAKSHMI ANJALI (S180937),K.LAKSHMI SOWJANYA (R181053) ,SUGALI DIVYA (R180825) under my guidance and supervision and is submitted to **DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING** in partial fulfillment of requirements for the award of Bachelor of Technology in Computer Science and Engineering during the academic year 2023-2024 and it has been found worthy of Acceptance According to the requirements of the University.

**Signature of Internal Guide**

R Sreenivasulu MTech  
Assistant Professor  
Department of CSE

**Signature of HOD**

Dr . P Ravi Kumar, M.E , Ph.D  
Assistant Professor  
Department of CSE

**Signature of External Examiner**

## **ACKNOWLEDGEMENT**

I wish to express our sincere thanks to various personalities who were responsible for the successful completion of the main project.

I am thankful our Director **Dr. A VSS Kumara Swami Gupta**, for providing the necessary infrastructure required for my project.

I am thankful to our Dean **Mrs.CH.RATNA KUMARI**, for permitting and encouraging me in doing this project.

I am grateful to **Dr. P RAVI KUMAR**, Head of the Department, for his motivation and encouragement in completing the project in specified time.

I express my deep felt gratitude to **Mr. R.SREENIVASULU Assistant Professor**, internal guide for her valuable guidance and encouragement which enabled me to successful completion of project in time.

I express my sincere thanks to all other faculty members of CSE Department for extending their helping hands and valuable suggestion when in need.

Finally, my heartfelt thanks to my parents for giving me all I ever needed to be a successful student and individual. Because of their hard work and dedication, I have had opportunities beyond my wildest dreams.

**WITH SINCERE REGARDS**

**V.LAKSHMI ANJALI[S180937]**

**K.LAKSHMI SOWJANYA[R181053]**

**SUGALI DIVYA[R180825]**

## DECLARATION

Hereby declare that this project work entitled “ *IPL SCORE PREDICTION*” submitted to **DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING** is a genuine work carried out by me, for the fulfillment of Bachelor of Technology in the Department of Computer Science & Engineering during the academic year 2023-2024 under the supervision of my project guide **Mr. R.SREENIVASULU Assistant Professor**, Department of **Computer Science & Engineering** in **RAJIV GANDHI UNIVERSITY OF KNOWLEDGE TECHNOLOGIES(AP IIIT), R.K.Valley** and that it has not formed the basis for the award of any degree/diploma or other similar title to any candidate of the university.

WITH SINCERE REGARDS

V.LAKSHMI ANJALI[S180937]

K.LAKSHMI SOWJANYA[R181053]

SUGALI DIVYA[R180825]

## **ABSTRACT**

In today's digital world, the efficient digitization of physical documents plays a crucial role in information management and accessibility. This research introduces an innovative approach to streamline this process through an image-based automatic document scanner and text extractor powered by computer vision techniques. The proposed system aims to revolutionize document handling by seamlessly capturing document images, extracting textual content, and converting them into editable digital formats. The system performs automatic page detection, image enhancement, and perspective correction to ensure accurate and reliable document image acquisition. Subsequently, optical character recognition (OCR) algorithms are employed to decipher text from the acquired images, converting them into searchable and editable text.

Through the fusion of image processing and machine learning techniques, the system offers a robust solution for a wide range of document types, sizes, and orientations. Experimental results demonstrate the system's effectiveness in achieving high accuracy and efficiency in document digitization, paving the way for enhanced document management in various domains, including archives, offices, libraries, and personal use. This image-based automatic document scanner and text extractor harnesses the power of computer vision to bridge the gap between physical and digital document worlds, facilitating seamless access to valuable information and transforming the way we interact with textual content.

**INDEX TERMS** : Computer Vision(CV), Optical Character Recognition(OCR), Image Processing

## CONTENTS

CERTIFICATE	I
ACKNOWLEDGEMENT	II
DECLARATION	III
ABSTRACT	IV
LIST OF FIGURES	V
LIST OF TABLES	VI
	VII

CH.NO .	TITTLE NAME	PAGE NO.
1.	INTRODUCTION	1-2
	1.1 INTRODUCTION	1-2
2.	HISTORY AND BACKGROUND	3-4
	2.1 AIM AND SCOPE	5-5
3.	IPL AND ICL	6-7
	3.1 RULES AND REGULATIONS	8-9
4.	THE TEAMS AT A GLANCE	10-13
5.	SUCCESFUL STAR OF IPL	14-16
6.	TOURNMENT FORMAT	17-18
7.	IPL DATA ANALYTICS	19-21
8.	DATASET DESCRIPTION	22-22
9.	DATA IMPORTING	23-23
10.	DATA CLEANING	24-24

<b>11.</b>	<b>VISUALIZATION</b>	<b>25-25</b>
<b>12.</b>	<b>PRE-PROCESSING</b>	<b>26-26</b>
<b>13.</b>	<b>MODELLING</b>	<b>27-27</b>
	ALGORITHMS	<b>27-33</b>
	14.1 LINEAR REGRESSION	
	14.2 DECISION TREE REGRESSION	
	14.3 RANDOM FOREST REGRESSION	
	14.4 ADABOOST	
	14.5 RIDGE REGRESSION	
<b>14.</b>	<b>OUTPUT</b>	<b>34-36</b>
<b>15.</b>	<b>CONCLUSION</b>	<b>37-38</b>
<b>16.</b>	<b>REFERENCES</b>	<b>39-40</b>

## **CHAPTER - 1**

### **INTRODUCTION**

#### **1.1 INTRODUCTION**

The Indian Premier League (IPL) stands as a groundbreaking phenomenon in the realm of cricket, launched in 2008 under the aegis of the Board of Control for Cricket in India (BCCI). Comprising franchise teams representing cities and regions across India, the IPL has rapidly ascended to become one of the most prestigious and lucrative cricket leagues globally. Its format revolves around the fast-paced and thrilling Twenty20 cricket, captivating audiences with its blend of athleticism, strategy, and entertainment.

What sets the IPL apart is its fusion of cricketing prowess with elements of glamour, star power, and spectacle. Each team is studded with international cricketing icons, creating an unparalleled level of competition and excitement. The league follows a dynamic round-robin format, leading up to playoffs and ultimately the grand finale, where the best vie for the coveted championship title.

Beyond its sporting significance, the IPL serves as a catalyst for economic growth, boasting massive revenues from sponsorships, broadcasting rights, and ticket sales. It has transformed cricket into a year-round spectacle, drawing in millions of viewers both domestically and globally. IPL matches are renowned for their electrifying atmosphere, with packed stadiums resonating with the chants of passionate fans.

Moreover, the league has been instrumental in nurturing young talent, providing a platform for aspiring cricketers to showcase their skills alongside seasoned professionals. IPL franchises invest in grassroots development programs, fostering a sustainable pipeline of talent for Indian cricket.



The IPL's impact extends beyond the boundary ropes, influencing culture, commerce, and the broader sports landscape. It has revolutionized cricket administration, setting new benchmarks in professionalism, transparency, and innovation. As a cultural phenomenon, the IPL transcends cricket, embodying the spirit of unity, diversity, and collective celebration.

In essence, the IPL represents more than just a cricket tournament; it's a spectacle that unites nations, transcends boundaries, and captivates hearts, symbolizing the indomitable spirit of sport and the enduring passion of its followers.

The IPL's impact extends beyond the boundary ropes, influencing culture, commerce, and the broader sports landscape. It has revolutionized cricket administration, setting new benchmarks in professionalism, transparency, and innovation. As a cultural phenomenon, the IPL transcends cricket, embodying the spirit of unity, diversity, and collective celebration.

In essence, the IPL represents more than just a cricket tournament; it's a spectacle that unites nations, transcends boundaries, and captivates hearts, symbolizing the indomitable spirit of sport and the enduring passion of its followers.

The IPL has also become a platform for cricketing innovation, with its strategic timeouts and technological advancements enhancing the game's dynamics. It's not just about cricket; the league's blend of entertainment, celebrity ownership, and off-field drama adds layers of excitement. The IPL's immense popularity has led to the emergence of numerous fan clubs, online communities, and fantasy leagues, further cementing its status as a cultural phenomenon.

## **CHAPTER – 2**

### **HISTORY AND BACKGROUND OF THE IPL**

Kerry Francis Bull more Packer, AC (17 December 1937–26 December 2005), was an Australian media tycoon whose family company owned controlling interests in both the Nine television network and leading Australian publishing company Australian Consolidated Press. Packer was best known for founding World Series Cricket. Top players from several countries rushed to join him at the expense of their international sides. Packer's aim was to secure broadcasting rights for Australian cricket, and he was largely successful. Many of the well-known cricketers of that period left their national team to play in Kerry Packer's World Series cricket. Some of our legendary cricketers also contact to play in that series. But due to some controversies, mainly with Australian board due to television rights, this league could not be successful. On the background of this idea, Zee Entertainment Enterprises.

Organized a league called Indian cricket league. The Indian Cricket League (ICL) was a private cricket league funded by Zee Entertainment Enterprises that operated between 2007 and 2009 in India. In its two seasons included tournaments between four international teams (The World XI, India, Pakistan and Bangladesh) and nine domestic teams notionally located in major Indian cities as well as the champions Lahore Badshah who were based in Lahore, Pakistan. The matches were played in the Twenty20 format. Zee T.V owner Subhash Chandra founded this league in the response of BCCI's some of working style. Many times Subhash Chandra give bid for television rights but every time he was rejected even if his bid was the highest. This results in formation of Indian Cricket League (ICL). But due to some reasons this league could not be successful. Some of reasons are commercial factors, lacking of the support of the BCCI and ICC.

The ICL was set up with a billion dollars Indian Rupee corpus, and was to initially comprise six teams playing Twenty20 cricket, with plans to expand to sixteen teams within three years and to eventually move to 50-over matches. These plans, if they had been realised, would have made the ICL the richest professional league in India. On 24 July 2007, some famous international names were announced to have signed to play in the ICL, including highest innings record-holder Brian Lara. The BCCI refused to recognise the ICL as a cricket league, and criticised Kiran More and Kapil Dev for joining the ICL. On August 21, 2007 Kapil Dev was sacked from his NCA post. The International Cricket Council gave a statement through its chief executive, Malcolm Speed, that the ICC would not recognize the ICL unless the BCCI chooses to recognise it. Faced with the threat of young players joining the ICL, the BCCI jacked up prize money for winners, runners-up and losing semi-finalists across all tournaments. An average domestic cricketer can hope to make around Rs 35,000 per match day from the season of 2007-08: more than double the Rs 16,000 they got in 2005-06. The BCCI has also planned to do away with honorary selectors, who will be paid professionals from September 2008 onwards. Then BCCI started its own international Twenty20 league. The official league, which was launched in April 2008, was called the Indian Premier League Twenty20.

In January 2008, the BCCI announced the formation of the Indian Premier League, unveiling eight franchise teams representing various cities and regions across India. The franchises were sold through a competitive bidding process, with the successful bidders including corporate giants like Reliance Industries, India's biggest film production house, Bollywood stars, and other high-profile investors.

## **2.1 AIM AND SCOPE**

### **AIM OF THE PROJECT**

This project aims at designing an effective result prediction system for a cricket match. The result of a T20 cricket match depends on lots of In-game and pre-game attributes, like venue, Past track-records and toss influence the results of the match predominantly. This project also aims to emphasize on exploratory data analysis, modelling and visualization of data regarding the Indian Premier League. Best possible outcome of a given match will be predicted using different supervised machine learning (Random Forest Classifier) and statistical approaches. For easy access and usage of the outcome, this will be hosted on a user-friendly web application that can run on any browser.

### **OBJECTIVE AND SCOPE**

To predict the outcome of an IPL match. It also aims to analyse and visualize data using various data visualisation techniques for better understanding. The data has to be pre-processed and fed to various supervised machine learning algorithms and analysed in accordance to their accuracies. The best possible outcome will be predicted using a perfect model and will be hosted in a user-friendly web application.

## **CHAPTER – 3**

### **2.1 The Indian Premier League and the Indian Cricket League**

The Indian Premier League (IPL) and the Indian Cricket League (ICL) are two distinct cricket leagues with significant differences in their formation, structure, and impact on Indian cricket. The IPL was established by the Board of Control for Cricket in India (BCCI) in 2008 to capitalize on the growing popularity of Twenty20 cricket. In contrast, the ICL was launched in 2007 by the Zee Entertainment Enterprises as a rival to the BCCI-backed IPL, aiming to provide an alternate platform for cricketers not part of the national team setup.

The IPL operates on a franchise-based model, representing various cities and regions of India, with a round-robin format followed by playoffs to determine the champion. Meanwhile, the ICL also operated on a franchise-based model but featured fewer teams and a more traditional league structure with no playoffs.

The IPL attracted top international and domestic players due to its affiliation with the BCCI and the lure of substantial financial rewards. In contrast, the ICL initially signed several international and domestic players, including some former international stars, who were enticed by lucrative contracts. However, these players were subsequently banned by their respective cricket boards due to their association with the unsanctioned league.

While the IPL was endorsed and organized by the BCCI, the ICL operated independently and faced opposition from the Indian cricket establishment, leading to conflicts and legal battles. The IPL emerged as one of the most successful and lucrative cricket leagues globally, transforming the cricketing

landscape in India and setting new benchmarks in professionalism, entertainment, and revenue generation.

Following the IPL's inception in 2008, it quickly gained traction among cricket fans worldwide, drawing large crowds to stadiums and capturing the attention of television viewers. The IPL's combination of high-quality cricket, star players, and entertainment elements such as cheerleaders and music contributed to its rapid rise in popularity.

The IPL's success also led to the emergence of new revenue streams for cricket, including lucrative sponsorship deals, broadcasting rights, and merchandise sales. Its franchise-based model created opportunities for businesses to invest in cricket and engage with fans on a regional level.

In contrast, the ICL faced challenges from the outset, including legal battles with the BCCI over player contracts and television rights. Despite signing several prominent players and organizing tournaments, the ICL struggled to gain mainstream acceptance and faced financial difficulties.

Ultimately, the BCCI's ban on players participating in unsanctioned leagues, coupled with the lack of support from cricketing authorities, proved to be insurmountable obstacles for the ICL. The league folded in 2009, leaving its players without a platform to showcase their talents.

While the IPL continued to thrive and expand, the ICL's brief existence left a legacy of experimentation and innovation in the cricketing world. It demonstrated the potential for professional Twenty20 leagues outside the traditional cricketing establishment and paved the way for future developments in the sport.

### **3.1 RULES AND REGULATIONS**

The Indian Premier League (IPL) operates under a set of rules and regulations that govern various aspects of the league. These rules cover player eligibility, franchise ownership, player auctions, retention, and trading, as well as match format, scheduling, and venue selection. Additionally, there are codes of conduct in place for players, team officials, and franchise owners, promoting fair play, sportsmanship, and integrity. Anti-corruption measures are stringent, with protocols for reporting suspicious activities and cooperating with anti-corruption authorities. Doping control is conducted in adherence to the World Anti-Doping Agency (WADA) code, ensuring a clean and drug-free competition. Qualified umpires and match officials are appointed to enforce the rules, with criteria for selection, training programs, and performance evaluation. Broadcasting and media rights agreements are crucial, determining the distribution of revenue among franchises and the league. These rules and regulations form the operational framework of the IPL, ensuring transparency, accountability, and the integrity of the competition.

The Indian Premier League (IPL) is governed by a comprehensive set of rules and regulations that dictate every aspect of the tournament. From player eligibility to franchise ownership, player auctions, and match format, these rules ensure the smooth functioning and integrity of the league. One of the key aspects is player eligibility, which specifies criteria such as nationality, age, and cricketing experience for participation in the league. Additionally, franchise ownership is regulated by the IPL, with procedures for bidding, financial obligations, and ownership structures outlined in the rules.

The IPL's player auction is a highly anticipated event, where teams bid for players from around the world to build their squads. Rules govern player

valuation, bidding strategies, and salary caps, ensuring fair and transparent transactions. Moreover, teams have the option to retain a certain number of players from their previous squad and engage in player trading with other franchises. These rules are designed to maintain competitive balance among teams and foster a dynamic player market.

Match format and scheduling are also governed by IPL rules, with matches following a Twenty20 format and a specified number of matches played in a round-robin format followed by playoffs. The rules dictate match scheduling, venues, and logistical arrangements to ensure smooth conduct of matches. Furthermore, the IPL has strict codes of conduct for players, team officials, and franchise owners, promoting fair play, sportsmanship, and integrity. Disciplinary measures are in place for violations of the code, with penalties ranging from fines to suspension.

Anti-corruption measures are paramount in the IPL, with stringent protocols in place to prevent match-fixing, spot-fixing, and other corrupt practices. These measures include reporting suspicious activities and cooperating with anti-corruption authorities to maintain the integrity of the competition. Additionally, doping control is conducted according to the World Anti-Doping Agency (WADA) code, with regular testing to ensure a clean and drug-free competition. The IPL also appoints qualified umpires and match officials to officiate matches, with criteria for selection, training programs, and performance evaluation outlined in the rules. Broadcasting and media rights agreements play a crucial role, determining revenue distribution among franchises and the league, ensuring financial sustainability and growth.



## CHAPTER-4

### THE TEAMS AT A GLANCE



#### 1. Mumbai Indians (MI)

Owner: Reliance Industries

Captain: Rohit Sharma

Overview: Mumbai Indians are one of the most successful teams in IPL history, having won the title multiple times. They are known for their strong batting line up, which includes power hitters like Rohith Sharma, Kieron Pollard, and Hardik Pandya, as well as a formidable bowling attack led by Jasprit Bumrah.

#### 2. Chennai Super Kings (CSK)

Owner: India Cements

Captain: Mahendra Singh Dhoni

Overview: Chennai Super Kings are another powerhouse team with a massive fan following.

Led by the iconic captain MS Dhoni, CSK is known for its consistency and experienced players like Suresh Raina, Ravindra Jadeja, and Dwayne Bravo. They have a reputation for strategic play and strong team spirit.

### **3. Kolkata Knight Riders (KKR)**

Owner: Red Chillies Entertainment, Mehta Group

Captain: Eoin Morgan

Overview: Kolkata Knight Riders, owned by Bollywood superstar Shah Rukh Khan, are known for their aggressive style of play. With players like Andre Russell, Sunil Narine, and Shubman Gill, KKR boasts a formidable lineup capable of turning any match around.

### **4. Royal Challengers Bangalore (RCB)**

Owner: United Spirits

Captain: Virat Kohli

Overview: Royal Challengers Bangalore is one of the most popular teams in the IPL, although they are yet to win the title. Led by Indian cricket captain Virat Kohli, RCB is known for its star-studded lineup, including AB de Villiers, Glenn Maxwell, and Yuzvendra Chahal.

### **5. Sunrisers Hyderabad (SRH)**

Owner: Sun TV Network

Captain: Kane Williamson

Overview: Sunrisers Hyderabad has been a consistent performer in the IPL, known for its balanced team composition. Led by Kane Williamson, SRH has a

strong bowling attack with players like Rashid Khan and Bhuvneshwar Kumar, along with explosive batsmen like David Warner.

#### **6. Rajasthan Royals (RR)**

Owner: Manoj Badale

Captain: Sanju Samson

Overview: Rajasthan Royals, the inaugural IPL champions, are known for nurturing young talent. With a mix of experienced players and emerging stars, RR emphasizes a fair play ethos. Under the leadership of Sanju Samson, they continue to be a competitive force in the IPL.

#### **7. Punjab Kings (PBKS)**

Owner: Mohit Burman, Ness Wadia, Preity Zinta, Karan Paul

Captain: KL Rahul

Overview: Formerly known as Kings XI Punjab, Punjab Kings have had ups and downs in the IPL. Led by KL Rahul, PBKS boasts a strong batting lineup with players like Chris Gayle and Mayank Agarwal. They are known for their aggressive approach to the game.

#### **8. Delhi Capitals (DC)**

Owner: GMR Group, JSW Group

Captain: Rishabh Pant

Overview: Delhi Capitals have emerged as a strong contender in recent years, with a focus on youthful energy and talent. Led by Rishabh Pant, DC has players like Shreyas Iyer, Prithvi Shaw, and Kagiso Rabada, making them a well-balanced team capable of challenging any opponent

Gujarat Titans:

### **9. Gujarat Titans:**

Captain: The captaincy of the team would likely be assigned to a player known for leadership qualities and experience, although the specific captain had not been officially announced at that time.

Owner: The franchise's ownership would typically be attributed to a consortium or an individual entity that won the bid for the team during the IPL auction.

Overview: As a new team in the IPL, the Gujarat Titans would aim to establish themselves in the league, build a strong team roster through the auction, and compete for the IPL title. They would likely focus on recruiting a blend of seasoned veterans and promising young talents to create a competitive squad.

### **10. Lucknow Super Giants:**

Captain: Similar to the Gujarat Titans, the captaincy for the Lucknow Super Giants would be decided based on factors such as experience, leadership, and performance.

Owner: The ownership of the Lucknow Super Giants would also be held by an entity or consortium that successfully bid for the franchise during the IPL auction.

Overview: The Lucknow Super Giants would enter the IPL with the objective of making their mark in the league. They would seek to assemble a formidable team capable of challenging the established franchises and vie for top honors in the tournament. Like other IPL teams, they would focus on strategic player acquisitions, coaching staff appointments, and marketing initiatives to build a strong brand presence and fan base.

## CHAPTER-5

### SUCCESSFUL START OF IPL

The Indian Premier League (IPL) has consistently marked the beginning of the cricketing calendar with resounding success since its inception in 2008. Its innovative format, star-studded line ups, and vibrant fan engagement have catapulted it into one of the most eagerly anticipated sporting events globally.

The success of the IPL can be attributed to several key factors. Firstly, its introduction of the Twenty20 format brought a breath of fresh air to the cricketing world. With matches condensed into approximately three hours, the IPL appealed to a wider audience, including those with limited time for sports consumption.



Secondly, the franchise-based model infused the tournament with a sense of competition and glamour. Each team represents a different city or region of India, allowing fans to rally behind their local heroes. The involvement of prominent business figures, Bollywood celebrities, and international sports stars as team owners further elevated the league's profile and attracted attention from all corners of society.

Moreover, the IPL's emphasis on entertainment and spectacle has been instrumental in its success. Matches are not just about cricket; they are about music, cheerleaders, fireworks, and high-octane performances. This blend of sports and entertainment has transformed cricket stadiums into vibrant arenas of celebration, appealing to fans of all ages and backgrounds.

Another crucial aspect of the IPL's success is its strategic scheduling. Matches are held during the prime-time evening hours, maximizing viewership both in India and around the world. This timing ensures that the IPL captures the attention of a global audience, further enhancing its commercial value and relevance on the international stage.

Furthermore, the IPL has leveraged technological advancements to enhance the fan experience. From live streaming platforms to interactive mobile apps, fans can engage with the IPL in real-time, participate in polls, and access exclusive content, creating a sense of immersion and connectivity with the tournament.

The IPL's success has also been driven by its lucrative commercial partnerships. Sponsors and broadcasters clamor to be associated with the tournament, resulting in lucrative deals that inject substantial revenue into the league. This financial muscle allows the IPL to attract top talent, invest in state-of-the-art facilities, and promote grassroots cricket development initiatives.

Additionally, the IPL's impact extends beyond the cricketing field. It has become a catalyst for economic growth, tourism, and job creation in host cities. Hotels, restaurants, and local businesses thrive during the tournament, while cricketing infrastructure undergoes significant upgrades to meet the demands of hosting IPL matches.

Moreover, the IPL serves as a platform for nurturing young talent and providing opportunities for aspiring cricketers to shine on a global stage. Emerging players rub shoulders with international stars, hone their skills under experienced coaches, and gain invaluable exposure that propels their careers forward.

The IPL's success story is not without its challenges, including controversies surrounding governance, player conduct, and match-fixing allegations. However, the league has demonstrated resilience in addressing these issues, implementing robust governance mechanisms and integrity measures to uphold the integrity of the sport.

In conclusion, the IPL's successful start can be attributed to its innovative format, franchise-based model, emphasis on entertainment, strategic scheduling, technological advancements, commercial partnerships, economic impact, talent development initiatives, and resilience in the face of challenges. As the tournament continues to evolve and grow, it remains a shining example of the transformative power of sports in uniting communities, driving economic growth, and inspiring future generations of cricketers.

## **CHAPTER-6**

### **TOURNAMENT FORMAT**

The tournament format of the Indian Premier League (IPL) has evolved over the years but generally follows a consistent structure. Here's a detailed overview of the tournament format:

#### **League Stage:**

The IPL typically begins with a league stage where all teams compete against each other in a round-robin format.

Each team plays a total of 14 matches, seven at home and seven away.

Matches are scheduled across various venues in India, ensuring each team has a mix of home and away games.

Teams earn points based on the results of their matches: two points for a win, one point for a no result or a tie, and zero points for a loss.

#### **Playoffs:**

After the conclusion of the league stage, the top four teams in the points table advance to the playoffs.

The playoffs consist of three stages: Qualifier 1, Eliminator, and Qualifier 2.

Qualifier 1: The top two teams from the league stage face each other in Qualifier 1. The winner of this match advances directly to the final.

Eliminator: The third and fourth-placed teams from the league stage compete in the Eliminator. The loser of this match is eliminated from the tournament, while the winner proceeds to Qualifier 2.

Qualifier 2: The losing team from Qualifier 1 and the winning team from the Eliminator face each other in Qualifier 2. The winner of this match advances to the final.



**Final:**

The final match of the IPL is contested between the winner of Qualifier 1 and the winner of Qualifier 2.

The victorious team in the final is crowned the IPL champion.

**Prize Money and Awards:**

In addition to the prestige of winning the IPL trophy, teams also compete for significant prize money and individual awards.

Prize money is awarded to the champion team, the runner-up, and other top-performing teams.

Individual awards include Player of the Tournament, Orange Cap (leading run-scorer), Purple Cap (leading wicket-taker), and Emerging Player of the Tournament, among others.

**Player Transfers and Auctions:**

Before each IPL season, teams have the opportunity to retain a certain number of players from their previous squads and release others.

An IPL auction is conducted to fill the vacant slots in each team's roster with new players.

Franchise owners bid for players in a live auction, with each player entering the auction pool at a base price set by the IPL governing council.

The auction process allows teams to strengthen their squads with new talent and strategic acquisitions.

Overall, the IPL tournament format is designed to provide a thrilling and competitive cricketing spectacle, showcasing the talents of top players from around the world while engaging fans through its dynamic and high-stakes matches.

## CHAPTER-7

### IPL DATA ANALYTICS

As the process of analysing raw data to find trends and answer questions, the definition of data analytics captures its broad scope of the field. However, it includes many techniques with many different goals. The data analytics process has some components that can help a variety of initiatives. By combining these components, a successful data analytics initiative will provide a clear picture of where you are, where you have been and where you should go. Statistics have always had a significant role in sports. As I mentioned above, sports analytics is on the rise and will continue to play a significant role in how teams operate, pick their players, how they play the game, etc. Cricket is no different. The runs scored by a batsman, the wickets taken by a bowler, or the matches won by a cricket team – these are all examples of the most important numbers in the game of cricket. Maintaining a record of all such statistics has multiple benefits. The teams and the individual players can dig deep into this data and find areas of improvement. It can also be used to assess an opponent's strengths and weaknesses. Data analytics is a broad field. There are four primary types of data analytics: descriptive, diagnostic, predictive and prescriptive analytics. Each type has a different goal and a different place in the data analysis process. These are also the primary data analytics applications in business. Descriptive analytics helps answer questions about what happened.

Figure 5 : Grayscale image

- These techniques summarize large datasets to describe outcomes to stakeholders. By developing key performance indicators (KPIs,) these strategies can help track successes or failures. Metrics such as return on investment (ROI) are used in many industries. Specialized metrics are developed to track performance in specific industries. This process 7 requires the collection of

relevant data, processing of the data, data analysis and data visualization. This process provides essential insight into past performance. Diagnostic analytics helps answer questions about why things happened.

- These techniques supplement more basic descriptive analytics. They take the findings from descriptive analytics and dig deeper to find the cause. The performance indicators are further investigated to discover why they got better or worse. This generally occurs in three steps:
  - o Identify anomalies in the data. These may be unexpected changes in a metric or a particular market. Data that is related to these anomalies is collected.
  - o Statistical techniques are used to find relationships and trends that explain these anomalies.Predictive analytics helps answer questions about what will happen in the future.

- These techniques use historical data to identify trends and determine if they are likely to recur. Predictive analytical tools provide valuable insight into what may happen in the future and its techniques include a variety of statistical and machine learning techniques, such as: neural networks, decision trees, and regression. Prescriptive analytics helps answer questions about what should be done.

- By Using insights from predictive analytics, data-driven decisions can be made. This allows businesses to make informed decisions in the face of uncertainty. Prescriptive analytics techniques rely on machine learning strategies that can find patterns in large datasets. By analysing past decisions and events, the likelihood of different outcomes can be estimated. These types of data analytics provide the insight that businesses need to make effective and efficient decisions. Used in combination they provide a well-rounded understanding of a company's needs and opportunities. The primary goal of a data analyst is to increase efficiency and improve performance by discovering patterns in data. The work of a data analyst involves working with data

throughout the data analysis pipeline. This means working with data in various ways. The primary steps in the data analytics process are data mining, data management, statistical analysis, and data presentation. The importance and balance of these steps depend on the data being used and the goal of the analysis. Data mining is an essential process for many data analytics tasks. This involves extracting data from unstructured data sources. These may include written text, large complex databases, or raw sensor data. The key steps in this process are to extract, transform, and load data (often called ETL.) These steps convert raw data into a useful and manageable format. This prepares data for storage and analysis. Data mining is generally the most time-intensive step in the data analysis pipeline. Data management or data warehousing is another key aspect of a data analyst's job. Data warehousing involves designing and implementing databases that allow easy access to the results of data mining. This step generally involves creating and managing SQL databases. Non-relational and NoSQL databases are becoming more common as well. Statistical analysis allows analysts to create insights from data. Both statistics and machine learning techniques are used to analyse data. Big data is used to create statistical models that reveal trends in data. These models can then be applied to new data to make predictions and inform decision making. Statistical programming languages such as R or Python (with pandas) are essential to this process. In addition, open-source libraries and packages such as TensorFlow enable advanced analysis. The final step in most data analytics processes is data presentation. This step allows insights to be shared with stakeholders. Data visualization is often the most important tool in data presentation. Compelling visualizations can help tell the story in the data which may help executives and managers understand the importance of these insights.

## CHAPTER-8

### DATASET DESCRIPTION

The dataset is a collection of ipl score from various season and some are inconsistent. This dataset contains the attributes ID, date, venue, batting team, bowling team, batsman, bowler details, runs, wickets, over, runs scored in the last 5 over, wickets win in the last 5 over, striker, non-striker and total runs scored by the team. Language Used: Python.

Date, Venue, Team 1,Team 2,Toss\_winner,Toss\_decision,Winner, Batting team, Bowling team, Runs, Wickets, Over, Run rate, Player, Runs scored, Wickets taken, Strike rate, Weather conditions, Pitch conditions, Player injuries.

Dataset with each line representing a single match entry. Each comma-separated value corresponds to a different attribute of the match, such as date, venue, teams, toss information, innings details, player performances, weather conditions, and player injuries.

You can expand upon this format with additional attributes or more detailed information as needed for your project.

Creating a dataset for IPL (Indian Premier League) score prediction in machine learning involves gathering relevant information that can be used to train a model to predict scores of cricket matches.

Ensure that the dataset is balanced and covers matches from various seasons and venues to make the prediction model robust. Additionally, feature engineering may be necessary to extract relevant information and create new features that can improve the predictive performance of the model.

Including these additional pieces of info can provide more context and depth to your dataset, potentially leading to more accurate predictions when building your machine learning model for IPL score prediction

## CHAPTER-9

### DATA IMPORTING

In the initial phase of the IPL Score Prediction project, importing the dataset emerges as a pivotal step, representing a critical component for training and evaluating the machine learning models. The dataset, typically stored in CSV format, encompasses a comprehensive collection of job postings sourced from online platforms. The utilization of machine learning for ipl score prediction demands a meticulous approach to data importing.

```
df.head()
```

	date	bat_team	bowl_team	runs	wickets	overs	runs_last_5	wickets_last_5	total
0	2008-04-18	Kolkata Knight Riders	Royal Challengers Bangalore	1	0	0.1	1	0	222
1	2008-04-18	Kolkata Knight Riders	Royal Challengers Bangalore	1	0	0.2	1	0	222
2	2008-04-18	Kolkata Knight Riders	Royal Challengers Bangalore	2	0	0.2	2	0	222
3	2008-04-18	Kolkata Knight Riders	Royal Challengers Bangalore	2	0	0.3	2	0	222
4	2008-04-18	Kolkata Knight Riders	Royal Challengers Bangalore	2	0	0.4	2	0	222

Python libraries like Pandas, the CSV file is read, and the dataset is loaded into a structured format suitable for analysis. This process establishes the groundwork for subsequent data cleaning and preprocessing steps, ensuring that the machine learning models are trained on a reliable and representative dataset.

The imported data encompasses various features extracted from job postings, encompassing textual content, formatting details, and any additional relevant information. With the dataset successfully imported, the workflow advances to the crucial steps of data cleaning, exploration, and, eventually, the application of machine learning algorithms for effective Job Scam Detection.

## CHAPTER-10

### DATA CLEANING

After importing the dataset, our IPL Score Prediction project focuses on the critical data cleaning phase. Utilizing Python's Pandas library, we meticulously handle missing values, outliers, and inconsistencies to elevate the overall quality of the dataset. Simultaneously, textual content in job postings undergoes pre-processing for uniformity, enabling more meaningful analysis. Exploratory data analysis plays a pivotal role in identifying patterns, outliers, and potential issues that may impact the model's performance. This process guides the removal of redundant features, streamlining the dataset to optimize machine learning model training. Data cleaning is fundamental for mitigating biases and enhancing the overall effectiveness of our IPL Score Prediction system.

```
# Removing unwanted columns
columns_to_remove = ['mid', 'venue', 'batsman', 'bowler', 'striker', 'non-striker']

print('Before removing unwanted columns: {}'.format(df.shape))
df.drop(labels=columns_to_remove, axis=1, inplace=True)
print('After removing unwanted columns: {}'.format(df.shape))
```

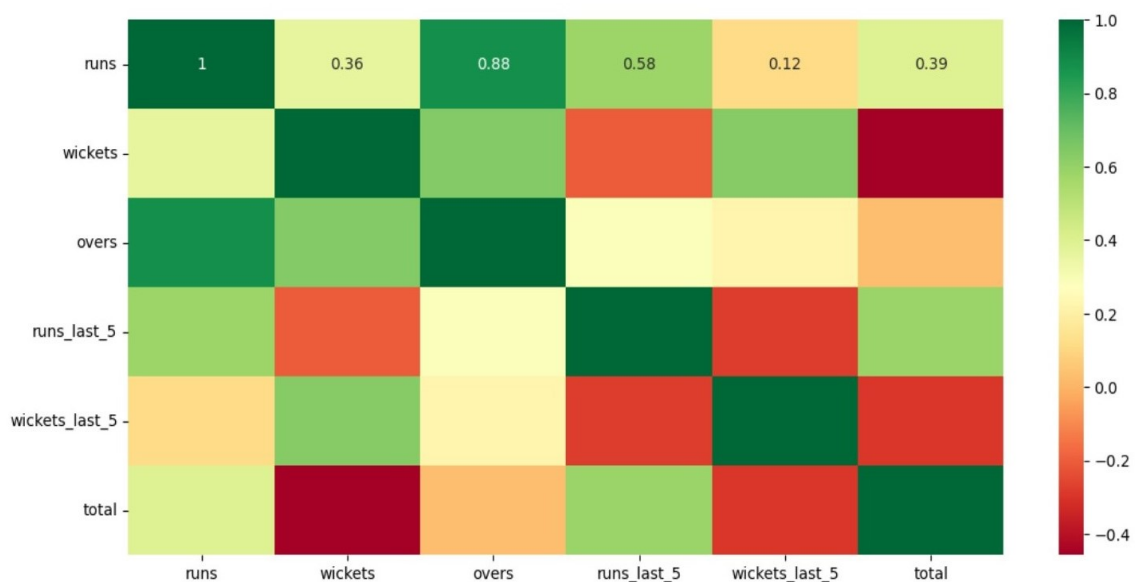
```
Before removing unwanted columns: (76014, 15)
After removing unwanted columns: (76014, 9)
```

## CHAPTER-11

### VISUALIZATION

After meticulous data cleaning, our Ipl Score Prediction project seamlessly moves to data visualization. Using Python's Matplotlib and Seaborn, we create insightful visual representations, including charts and graphs, to understand feature distribution. The objective is a clear depiction of relevant information, aiding in distinguishing features between legitimate job postings and scams.

Heat map for The Dataset





## CHAPTER-12

### PRE-PROCESSING

In Ipl Score Prediction using machine learning, preprocessing is a critical step to refine the dataset and optimize it for effective model training. This involves handling missing values, addressing outliers, handling categorical value and splitting dataset into train and test set in the basis of date. variables to ensure a comprehensive and balanced dataset. Textual data from job postings undergoes normalization and stemming for uniformity, aiding in the extraction of meaningful features. Feature engineering is employed to enhance the model's ability to discern patterns, and data scaling ensures that features are on a consistent scale. The dataset is then split into training and testing sets for model evaluation. Additionally, methods to address class imbalance are applied to prevent bias. This meticulous pre-processing lays the groundwork for robust machine learning models, enhancing their accuracy.

```
#Converting categorical features using OneHotEncoding method  
encoded_df = pd.get_dummies(data=df, columns=['bat_team', 'bowl_team'])  
encoded_df.columns
```

```
Index(['date', 'runs', 'wickets', 'overs', 'runs_last_5', 'wickets_last_5',  
      'total', 'bat_team_Chennai Super Kings', 'bat_team_Delhi Daredevils',  
      'bat_team_Kings XI Punjab', 'bat_team_Kolkata Knight Riders',  
      'bat_team_Mumbai Indians', 'bat_team_Rajasthan Royals',  
      'bat_team_Royal Challengers Bangalore', 'bat_team_Sunrisers Hyderabad',  
      'bowl_team_Chennai Super Kings', 'bowl_team_Delhi Daredevils',  
      'bowl_team_Kings XI Punjab', 'bowl_team_Kolkata Knight Riders',  
      'bowl_team_Mumbai Indians', 'bowl_team_Rajasthan Royals',  
      'bowl_team_Royal Challengers Bangalore',  
      'bowl_team_Sunrisers Hyderabad'],  
      dtype='object')
```

## **CHAPTER-13**

### **MODELLING**

In the modelling phase of IPL Score Prediction, machine learning algorithms are deployed to analyze the preprocessed dataset and identify patterns. Commonly used algorithms such as Random Forest Classifier, linear regression and Decision Tree Regression are applied to learn from the training data and make predictions on new instances. Ensemble methods, like Linear Regression, enhance model robustness by combining multiple classifiers. The choice of algorithms depends on the nature of the dataset and the specific characteristics of score prediction. Model hyperparameters are finetuned to optimize performance, and cross validation is often employed to assess generalization to new data.

The ultimate goal is to develop a predictive model capable of accurately distinguishing Score, contributing to a more secure online recruitment environment.

### **ALGORITHMS**

#### **14.1 Linear Regression:**

Linear regression is a fundamental supervised learning algorithm used in machine learning for modeling the relationship between a dependent variable and one or more independent variables. It assumes a linear relationship between the input variables (features) and the output variable (target).

In linear regression, the algorithm aims to find the best-fitting linear equation that describes the relationship between the independent variables and the dependent variable.

```
▼ LinearRegression
LinearRegression()
```

```
# Predicting results
y_pred_lr = linear_regressor.predict(X_test)
```

```
# Linear Regression - Model Evaluation
from sklearn.metrics import mean_absolute_error as mae, mean_squared_error as mse, accuracy_score
print("---- Linear Regression - Model Evaluation ----")
print("Mean Absolute Error (MAE): {}".format(mae(y_test, y_pred_lr)))
print("Mean Squared Error (MSE): {}".format(mse(y_test, y_pred_lr)))
print("Root Mean Squared Error (RMSE): {}".format(np.sqrt(mse(y_test, y_pred_lr))))

---- Linear Regression - Model Evaluation ----
Mean Absolute Error (MAE): 12.11861754619329
Mean Squared Error (MSE): 251.00792310417415
Root Mean Squared Error (RMSE): 15.843229566732099
```

Linear regression assumes several key properties:

**Linearity:** It assumes a linear relationship between the independent variables and the dependent variable.

**Independence:** The independent variables should be independent of each other.

**Homoscedasticity:** The variance of the residuals (the differences between the predicted and actual values) should be constant across all levels of the independent variables.

**Normality:** The residuals should be normally distributed.

Linear regression is widely used in various fields such as economics, finance, biology, and social sciences for prediction, forecasting, and understanding the relationship between variables. It's a simple yet powerful algorithm, and its interpretability makes it valuable for both exploratory data analysis and predictive modeling tasks.

## **14.2 Decision Tree Regression:**

Decision tree is a versatile supervised learning algorithm used for both classification and regression tasks in machine learning.

It creates a hierarchical structure of decision nodes and leaf nodes, where each decision node represents a feature attribute and each leaf node represents a class label (for classification) or a numerical value (for regression).

The decision tree algorithm recursively splits the data based on the feature attributes to create subsets that are as pure as possible in terms of the target variable.

This splitting process is done by selecting the feature attribute that provides the most information gain (for classification) or the greatest reduction in variance (for regression) at each node.

Decision trees are intuitive and easy to interpret, making them valuable for understanding the underlying patterns in the data. However, they can be prone to overfitting, especially when they grow too deep and capture noise in the training data.

Techniques like pruning, limiting the maximum depth of the tree, and setting minimum sample requirements for splitting nodes can help prevent overfitting and improve generalization performance.

Decision trees are used in various domains, including finance, healthcare, marketing, and natural language processing. They can handle both numerical and categorical data and are capable of capturing complex relationships between features. Additionally, decision trees can be combined into ensemble methods like Random Forest and Gradient Boosting, which further improve predictive performance and robustness.

```
▼ DecisionTreeRegressor  
DecisionTreeRegressor()
```

```
# Predicting results  
y_pred_dt = decision_regressor.predict(X_test)
```

```
# Decision Tree Regression - Model Evaluation  
print("---- Decision Tree Regression - Model Evaluation ----")  
print("Mean Absolute Error (MAE): {}".format(mae(y_test, y_pred_dt)))  
print("Mean Squared Error (MSE): {}".format(mse(y_test, y_pred_dt)))  
print("Root Mean Squared Error (RMSE): {}".format(np.sqrt(mse(y_test, y_pred_dt))))  
  
---- Decision Tree Regression - Model Evaluation ----  
Mean Absolute Error (MAE): 16.965442764578835  
Mean Squared Error (MSE): 525.4521238300936  
Root Mean Squared Error (RMSE): 22.922742502372913
```

Decision trees are versatile supervised learning algorithms used for both classification and regression tasks. They create a hierarchical structure of decision nodes and leaf nodes, where each decision node represents a feature attribute and each leaf node represents a class label (for classification) or a numerical value (for regression).

The decision tree algorithm recursively splits the data based on the feature attributes to create subsets that are as pure as possible in terms of the target variable. Decision trees are intuitive and easy to interpret, making them valuable for understanding the underlying patterns in the data. However, they can be prone to over fitting, especially when they grow too deep and capture noise in the training data.

Additionally, decision trees provide a measure of feature importance, indicating the contribution of each feature to the predictive performance of the model. This information can be useful for feature selection and understanding the most influential factors in the dataset. Despite their advantages, decision trees may not scale well to very large datasets or high-dimensional feature spaces.

### 14.3 Random Forest Regression:

Random Forest is an ensemble learning technique used for both classification and regression tasks in machine learning. It's based on the principle of constructing a multitude of decision trees during training and outputting the average prediction of the individual trees (for regression) or the mode of the classes (for classification).

Random Forest starts by creating multiple random subsets of the training data through a process called bootstrap sampling. Each subset, known as a bootstrap sample, is created by randomly selecting instances from the original dataset with replacement.

```
▼ RandomForestRegressor  
RandomForestRegressor()
```

```
# Predicting results  
y_pred_rf = random_regressor.predict(X_test)  
  
# Random Forest Regression - Model Evaluation  
print("---- Random Forest Regression - Model Evaluation ----")  
print("Mean Absolute Error (MAE): {}".format(mae(y_test, y_pred_rf)))  
print("Mean Squared Error (MSE): {}".format(mse(y_test, y_pred_rf)))  
print("Root Mean Squared Error (RMSE): {}".format(np.sqrt(mse(y_test, y_pred_rf))))  
  
---- Random Forest Regression - Model Evaluation ----  
Mean Absolute Error (MAE): 13.745793676505881  
Mean Squared Error (MSE): 328.97505786478087  
Root Mean Squared Error (RMSE): 18.137669581971682
```

*Note: Since Linear Regression model performs best as compared to other two, we use this model and boost its performance using AdaBoost Algorithm*

This means that some instances may be selected multiple times, while others may not be selected at all. A decision tree is constructed for each bootstrap sample.

However, unlike traditional decision trees, Random Forest introduces randomness during the tree construction process.

At each node of the decision tree, instead of considering all features to determine the best split, a random subset of features is considered. This helps to de-correlate the individual trees and makes them more diverse.

#### 14.4 AdaBoost Algorithm:

AdaBoost, short for Adaptive Boosting, is an ensemble learning algorithm primarily used for classification tasks. It combines multiple weak learners to create a strong classifier. During training, AdaBoost assigns weights to each training instance, focusing more on the difficult instances by increasing their weights. It fits a weak learner (often a decision tree) to the training data in each iteration, aiming to correct the mistakes made by the previous weak learners. The weak learners are combined into a strong classifier by giving more weight to the predictions of the more accurate ones. This iterative process continues until a predefined number of weak learners is reached or until perfect predictions are achieved. The final classifier is a weighted combination of the weak learners' predictions. AdaBoost is powerful, resistant to overfitting, and automatically handles feature selection. However, it may be sensitive to noisy data and outliers, and it can be computationally expensive. Despite its limitations, AdaBoost is widely used for classification tasks where accuracy is crucial.

#### *AdaBoost Algorithm*

```
# AdaBoost Model using Linear Regression as the base learner
from sklearn.ensemble import AdaBoostRegressor
adb_regressor = AdaBoostRegressor(base_estimator=linear_regressor, n_estimators=100)
adb_regressor.fit(X_train, y_train)
```

## 14.5 Ridge Regression:

Ridge regression is a regression technique used in statistical modeling and machine learning to address the problem of multicollinearity and overfitting. It extends ordinary least squares (OLS) regression by adding a penalty term to the loss function, which penalizes large coefficients. This penalty term, known as the L2 regularization term, is proportional to the square of the magnitude of the coefficients.

The objective of ridge regression is to find the set of coefficients that minimizes the sum of squared errors between the predicted values and the actual values, while also penalizing large coefficients to prevent overfitting. The strength of the penalty is controlled by a hyperparameter called the regularization parameter, typically denoted as lambda ( $\lambda$ ) or alpha ( $\alpha$ ). A higher value of the regularization parameter increases the penalty, leading to more regularization and potentially simpler models with smaller coefficients.

Overall, ridge regression helps to stabilize the model by reducing the variance of the parameter estimates, making it particularly useful when dealing with multicollinearity, where predictor variables are correlated with each other, and when working with high-dimensional datasets.

### Ridge Regression

```
## Ridge Regression
from sklearn.linear_model import Ridge
from sklearn.model_selection import GridSearchCV

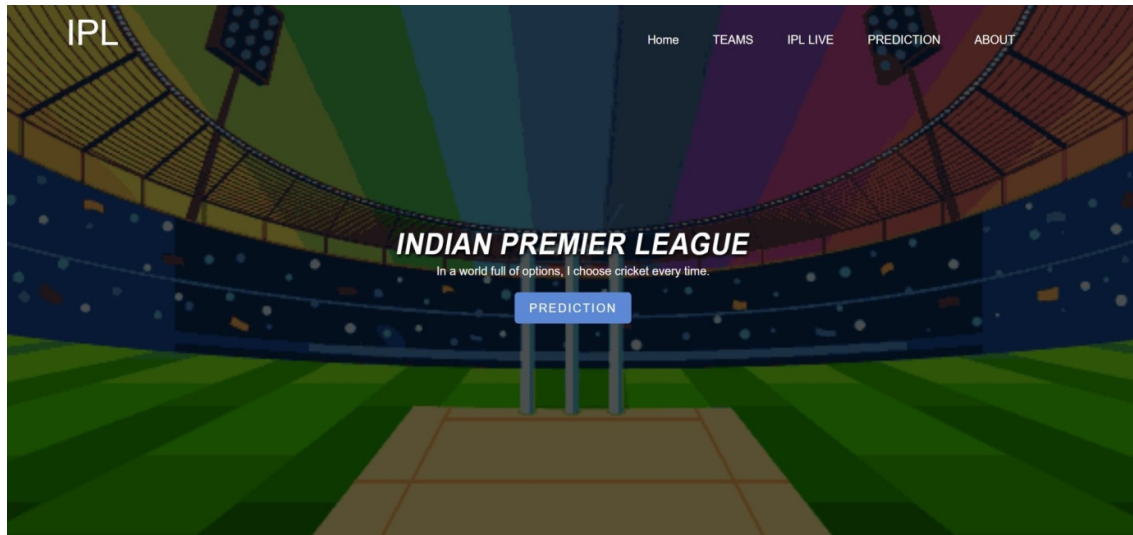
ridge = Ridge()
parameters = {'alpha': [1e-15, 1e-10, 1e-8, 1e-3, 1e-2, 1, 5, 10, 20, 30, 35, 40]}
ridge_regressor = GridSearchCV(ridge, parameters, scoring='neg_mean_squared_error', cv=5)
ridge_regressor.fit(X_train, y_train)
```



## CHAPTER-14

### OUTPUT

#### INTERFACE:



#### ACCURACY COMPARISION:

Comparing the accuracy of different models or methods for predicting first innings scores in the Indian Premier League (IPL) can be quite challenging due to various factors such as the complexity of the game, changing conditions, team compositions, and player performance. However, several approaches can be considered:

**Statistical Models:** These models use historical data and various statistical techniques to predict scores. They might consider factors such as venue, team composition, past performance, player statistics, etc. The accuracy of these models can vary depending on the sophistication of the techniques used and the quality of the data.

**Machine Learning Models:** Machine learning algorithms, such as regression models, decision trees, random forests, or neural networks, can be trained on

historical IPL data to predict first innings scores.

These models can potentially capture complex relationships between various factors and the final score.

The accuracy of these models depends on the quality and quantity of training data, feature selection, and the chosen algorithm.

**Expert Judgement:** Cricket experts, analysts, and commentators often provide predictions based on their knowledge of the game, players, and current form. While expert judgement can be insightful, it's subjective and might not always be accurate.

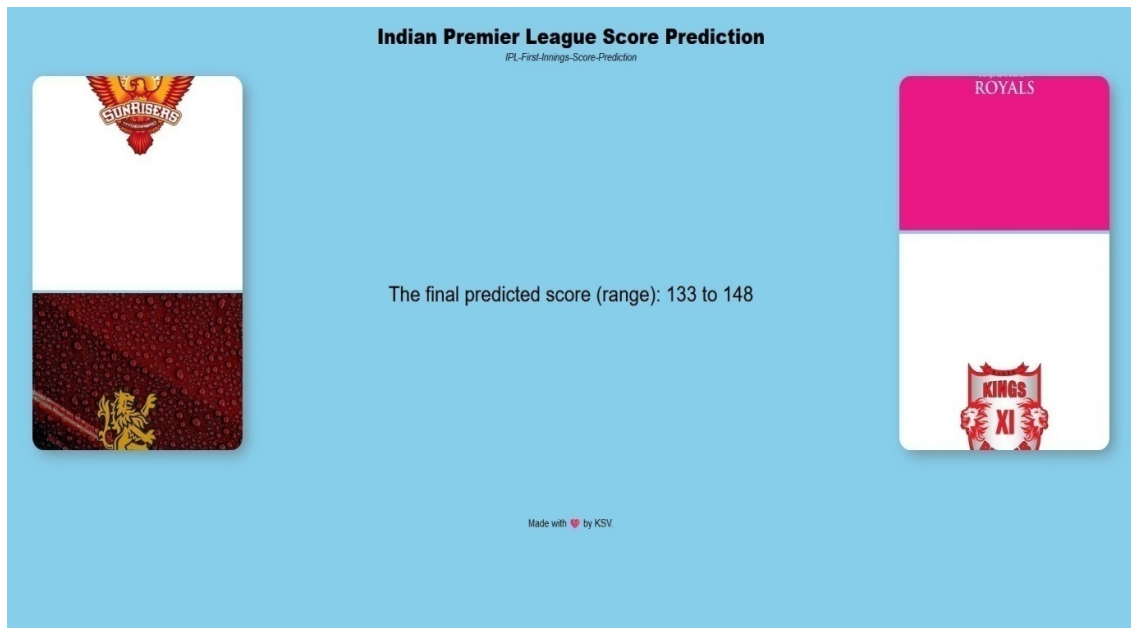
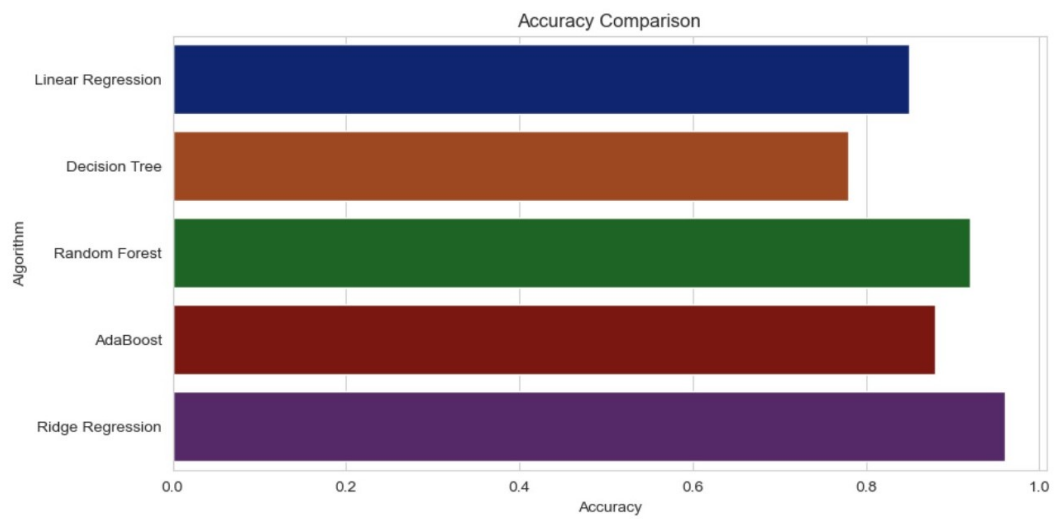
**Ensemble Methods:** Combining predictions from multiple models or experts can sometimes improve accuracy. Ensemble methods like averaging predictions or using techniques such as stacking can be employed for this purpose.

**Real-time Data Analysis:** Some predictions are made during the match based on real-time data such as pitch conditions, weather, team performance in the current match, etc.

These predictions can be more dynamic but are inherently less accurate due to the uncertainty associated with live events.

To compare the accuracy of different methods, one could use metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), or accuracy measures like the percentage of correct predictions within a certain margin of error. Additionally, cross-validation or back testing on historical data can provide insights into the robustness and generalization capability of the models.

## FINAL :



## CONCLUSION

Predicting the first innings score in an IPL match based on various factors like venue, team composition, past performance, etc. Define the scope of the project, including the data sources you'll use, the features you'll consider, and the ML algorithms you'll employ.

Collect historical IPL match data from reliable sources like ESPN info, or official IPL websites. Preprocess the data by cleaning it, handling missing values, encoding categorical variables, and scaling numerical features if necessary. Select relevant features such as venue, team composition (batsmen, bowlers), toss outcome, past performance, etc.

Explore the data to gain insights into the distribution of features, correlations between variables, and any patterns that might exist. Visualize the data using plots like histograms, scatter plots, box plots, etc., to understand the relationships between features and the target variable (first innings score).

Create new features that might capture additional information, such as batting average, strike rate, bowling economy rate, etc. Consider feature transformations or interactions that could enhance the predictive power of the model.

Choose appropriate ML algorithms for regression tasks, such as Linear Regression, Decision Trees, Random Forests, Gradient Boosting, etc. Split the data into training and testing sets to evaluate the performance of the models. Train multiple models using cross-validation and hyperparameter tuning to find the best-performing one.

Evaluate the trained models using appropriate metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), etc. Compare the performance of different models and select the one that provides the best predictive accuracy.

Once you have a trained model with satisfactory performance, deploy it to make real-time predictions. Create a user-friendly interface where users can input relevant information like venue, team composition, etc., and get the predicted first innings score as output.

Document the entire process, including data collection, preprocessing, feature engineering, model selection, and evaluation. Present your findings, insights, and conclusions in a clear and concise manner in your project report. Include visualizations, tables, and graphs to support your analysis and showcase the predictive performance of your model.

Discuss potential areas for improvement, such as incorporating more features, experimenting with different ML algorithms, or refining the existing model. Highlight any limitations or challenges encountered during the project and suggest ways to address them in future research.

By following these steps, you can create a comprehensive and detailed project report on predicting first innings scores in IPL matches using machine learning techniques. Make sure to document your methodology thoroughly and provide evidence to support your findings.

## REFERENCES

- 1) An Intelligent Model for Online Recruitment Fraud Detection,” J. Inf. Secure., vol. 10, no. 03, pp. 155–176, 2019, doi: 10.4236/jis.2019.103009.
- 2) An empirical study of the naive Bayes classifier, || no. January 2001, pp. 41–46, 2014.
- 3) D. E. Walters, —Bayes Theorem and the Analysis of Binomial Random Variables, || Biometrical J., vol. 30, no. 7, pp. 817–825, 1988, doi: 10.1002/bimj.4710300710.
- 4) B. Biggio, I. Corona, G. Fumera, G. Giacinto, and F. Roli, —Bagging classifiers for fighting poisoning attacks in adversarial classification tasks,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 6713 LNCS, pp. 350– 359, 2011, doi: 10.1007/978-3-642-21557-5\_37.
- 5) S. M. Vieira, U. Kaymak, and J. M. C. Sousa, Cohen’s kappa coefficient as a performance measure for feature selection,” 2010 IEEE World Congr. Comput. Intell. WCCI 2010, no. May 2016, 2010, doi: 10.1109/FUZZY.2010.5584447.
- 6) B. Biggio, I. Corona, G. Fumera, G. Giacinto, and F. Roli, —Bagging classifiers for fighting poisoning attacks in adversarial classification tasks,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 6713 LNCS, pp. 350– 359, 2011, doi: 10.1007/978-3-642- 21557-5\_37.
- 7) N. Hussain, H. T. Mirza, G. Rasool, I. Hussain, and M. Kaleem, —Spam review detection techniques: A systematic literature review, || Appl. Sci., vol. 9, no. 5, pp. 1–26, 2019, doi: 10.3390/app9050987, M. Ahmad, A. Doud, L. Wang, H. Hong, October 2016
- 8) “P. Somaskandhan, G. Wijesinghe, L. Bashitha, 2017, Identifying optimal set of attributes that impose high impact on end results of cricket match using machine learning.”

- 9) "M. Rehman, O. Shamim, S. Ismail, October 2018, an analysis of Bangladesh One day international cricket: Machine learning approach".
- 10) "A. Tripathi, J. Vanker, B. Vaje, V. Varekar, Cricket Score Prediction system using clustering algorithm.