****

**PES UNIVERSITY**

**(Established under Karnataka Act No. 16 of 2013)**

**100-ft Ring Road, Bengaluru – 560 085, Karnataka, India**

***Project Report (Phase-1)***

***on***

**IPL AUCTION PREDICTION**

*Submitted by*

**Pavan Chakrasali- (PES1PG22CA139)**

**Oct 2023 – Jan 2024**

**under the guidance of**

|  |
| --- |
| **Mr. Santosh S Katti** |
| Assistant Professor |
| Department of Computer Applications,  PESU, Bengaluru – 560085 |

****

**FACULTY OF ENGINEERING**

**DEPARTMENT OF COMPUTER APPLICATIONS**

**PROGRAM – MASTER OF COMPUTER APPLICATIONS**

**CERTIFICATE**

*This is to certify that the project entitled*

**IPL AUCTION PREDICTION**

*is a bonafide work carried out by*

**Pavan Chakrasali- PES1PG22CA139**

in partial fulfillment for the completion of Capstone Project, Phase-1 work in the Program of Study MCA under rules and regulations of PES University, Bengaluru during the period Oct. 2023 – Jan 2024. The project report has been approved as it satisfies the academic requirements of 3rd semester MCA.

|  |  |  |  |
| --- | --- | --- | --- |
| **Internal Guide**  Mr. Santosh S Katti,  Assistant Professor  Department of Computer Applications,  PES University, Bengaluru - 560085 | | **Chairperson**  Dr. Veena S  Department of Computer Applications,  PES University, Bengaluru - 560085 | |
|  | |  | |
|  |  | |  |

**DECLARATION**

I, **Pavan Chakrasali,** bearing **PES1PG22CA139** hereby declare that the Capstone project phase-1 entitled, ***IPL AUCTION PREDICTION,*** is an original work done by me under the guidance of **Mr.Santosh S Katti,** Assistant Professor, PES University, and is being submitted in partial fulfillment of the requirements for completion of 3rd Semester course in the Program of Study **MCA**. All corrections/suggestions indicated for internal assessment have been incorporated in the report.

**PLACE: Bengaluru**

**DATE: 30/01/2024**

**Signature:**

**ABSTRACT**

In this work, we have applied machine learning-based algorithms that predicts the cost at which a player can be sold in the Indian Premier League Auction. we estimated the players' selling price using their past performance parameters like runs, balls, innings, wickets and matches played. Tests were carried out in various machine learning models like Decision Tree Regressor, K-Nearest Neighbors (KNN), Linear Regression, Random Forest Regressor and Catboost.

Among these Catboost and Linear Regression gave best results for predicting batsman and bowlers respectively helping auctioneers make quick decisions. We have also considered inflation factor and mapping of the same to the budget during the training of the model.

**CONTENTS**

**Page No.**

**ABSTRACT**

1. **INTRODUCTION**
   1. PROJECT DESCRIPTION1
2. **LITERATURE SURVEY**
   1. DOMAIN SURVEY 3
   2. RELATED WORK 3
   3. EXISTING SYSTEMS 6
   4. TECHNOLOGY SURVEY 6
3. **HARDWARE AND SOFTWARE REQUIREMENTS** 
   1. HARDWARE REQUIREMENTS 8
   2. SOFTWARE REQUIREMENTS 8
4. **SOFTWARE REQUIREMENTS SPECIFICATION**
   1. USERS 9
   2. FUNCTIONAL REQUIREMENTS 9
5. **SYSTEM DESIGN**
   1. ARCHITECTURE DIAGRAM 10
   2. CONTEXT FLOW DIAGRAM 11
6. **DETAILED DESIGN**
   1. PROCESS FLOW DIAGRAM 12
   2. USE CASE DIAGRAM 14
7. **IMPLEMENTATION**

7.1 SCREEN SHOTS 15

**APPENDIX** 21

1. **INTRODUCTION**
   1. **PROJECT DESCRIPTION**

To predict the price of a batsman there are several important parameters to consider. Runs is an important parameter which tells us the best performance a player will give. This in-turn decides the player’s popularity among fans. Batting average is the total number of runs a batsman has scored divided by the number of times they have been out [3]. This tells us how consistent a batsman is, which is important because the team must score within 10 wickets of which not all are specialized in batting. Batting strike rate is a measure of how frequently a batsman achieves the primary goal of batting, namely scoring runs [4]. This is important because the speed at which the runs are scored are important in high scoring games. Balls faced is the number of balls that the batsman has played in the entire season. The batsman can bat in the first Inning or the second Inning which is decided by the captain of the team which wins the toss. Price is the output of our model.

* + 1. **Problem Scenario:**

Imagine a scenario where an IPL franchise team is preparing for the upcoming auction. They have a limited budget and need to build a competitive team within that constraint. However, they lack insights into player valuations and team compositions, making it challenging to formulate a winning strategy for the auction.

In the time of IPL everyone tries to predict the value of the players viewers cannot find the price for how much players can get with the help of this model viewers can also get to know about predicted price for the players based on their performance in the tournament. It gives more accuracy about the players how much they deserve to get paid.

* + 1. **Proposed Solution:**

The proposed solution for IPL auction prediction involves utilizing machine learning algorithms and data analytics to analyze past auction data and player performances, extract relevant features, train predictive models, and make informed predictions about player prices and team compositions for upcoming auctions.

* + 1. **Purpose:**

The purpose of undertaking IPL auction prediction using machine learning is to revolutionize the decision making process by using of advanced analytics and machine learning algorithms, the aim is to provide stakeholders—franchise owners, coaches, and fans—with accurate ,This initiative seeks to enhance the strategic planning and resource allocation of IPL franchises, optimizing their chances of assembling competitive and balanced teams while navigating the unpredictable dynamics of the auction.

**Team Composition:** Identifying optimal team combinations that complement each other's strengths. This extends to predicting the impact of a new player on the overall team dynamic and performance.

* + 1. **Project Scope:**

**Player Valuation:** The primary focus is on predicting the fair market value of players based on historical performances, statistics, playing conditions, and various contextual factors.

1. **LITERATURE SURVEY**
   1. **DOMAIN SURVEY**

* Existing machine learning models or algorithms used for sports predictions, particularly in cricket or IPL.
* Data sources commonly used for training machine learning models in sports prediction.
* Performance metrics used to evaluate the accuracy and effectiveness of machine learning models in sports prediction.
* Challenges or limitations specific to predicting IPL auction outcomes compared to other sports prediction tasks.
* Any relevant research papers, articles, or case studies related to IPL auction prediction or sports prediction in general.
* Feedback or insights from experts in cricket analytics or sports prediction.
  1. **RELATED WORK**

*Paper 1*

**Prediction of Player Price in IPL Auction Using Machine Learning Regression Algorithms**

* **Authors:** Jhansi Rani, A Kulkarni, A V Kamath, A Menon
* **Published in:** 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)
* **Year:** 02-04 July 2020
* **summary:** This paper explores the use of various machine learning regression algorithms to predict the selling price of players in the IPL auction. Features like runs, balls, innings, wickets, and matches played are used to train the models. Support Vector Regression (SVR) and Linear Regression were found to be the most effective algorithms for predicting batsman and bowler prices respectively.

*Paper 2*

**IPL Players Cost Pay Prediction using Machine Learning Techniques**

* **Author: Na**garaj P, Muneeswaran V, Raja M, M C Prabhu, B Meghana
* **Published in:** 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)
* **Year:** 23 August 2023
* **summery:** The accuracy of an IPL player cost prediction system in anticipating player auction prices is a measure of its success. The system's correctness can be measured using various metrics such as mean absolute error (MAE), mean squared error (MSE), and root mean square error (RMSE). These measures show how well the forecasted prices reflect the actual auction prices.

*Paper 3*

**Cricket Players Performance Prediction and Evaluation Using Machine Learning Algorithms**

* **Authors:**M Sumathi, S Prabu, M Rajkamal
* **Published in:** 2023 International Conference on Networking and Communications (ICNWC)
* **Year:** 05 April 2023
* **summery:** In this paper, different ML techniques are evaluated to predict the performance of a specific player. Linear regression, K-Means, and random forest models are used to predict the performance of a male cricket player. The performance of cricket players are predicted and regressed with linear lines using linear regression to select the relevant attribute for performance analysis.

*Paper 4*

**Predicting Results of Indian Premier League T-20 Matches using Machine Learning**

* **Authors:** Shilpi Agrawal, Suraj Pal Singh, Jayash Kumar Sharma
* **Published in:** 2018 8th International Conference on Communication Systems and Network Technologies (CSNT)
* **Year:** 24 November 2018
* **summery** : In this work, historical data has been collected from real IPL cricket matches and useful features have been extracted after pre-processing of data. Further, suitable data is converted to a numeric form and scale it on three parameters win, loss, and tie. This data is trained and classified with three classifier SVM, CTree and Naïve Bayes using R Tool.

*Paper 5*

**Selection of Players and Team for an Indian Premier League Cricket Match Using Ensembles of Classifiers**

* **Authors:** P Jhansi Rani, Aditya Vidyadhar Kamath, Aadith Menon
* **Published in:** 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)
* **Year:** 02 July 2020
* **summery :** This paper explores the potential of Support Vector Machines and Random Forests for predicting player prices in the IPL auction. It analyzes both traditional performance metrics and advanced features like form index and player impact score. The authors highlight the importance of Realtime data integration and model adaptation to address the dynamic nature of the auction environment.
  1. **EXISTING SYSTEMS**

As of my last update in January 2022, there isn't a single, widely recognized existing system for IPL auction prediction that dominates the market. However, several platforms and companies have developed their own proprietary algorithms and systems for predicting IPL auction outcomes. These systems typically employ various machine learning techniques, statistical analysis, and data mining approaches to make predictions based on historical data, player performances, team requirements, and other relevant factors.

Some well-known analytics companies, sports data firms, and fantasy sports platforms offer services or tools related to IPL auction prediction. However, the specifics of their systems and methodologies are often proprietary and not publicly disclosed.

It's worth noting that while these existing systems can provide valuable insights and guidance for team owners and analysts during the IPL auction process, the unpredictable nature of auctions and the dynamic nature of player performances mean that no system can guarantee perfect predictions.

* 1. **TECHNOLOGY SURVEY**

**Python**

Python is a high level programming language created by Guido Van rossum. It was first released in 1991. Today Python interpreters are available for many operating Systems including Window and Linux. Python is free to use and distribute and is supported by community. Python interpreter is available for every major platform. It is better than traditional and supporting languages because of support for advance reuse mechanism. It is much better than statically typed languages, much smaller language. It is lesser to type, debug and maintain, no lengthy compile and link types. Python programs run unchanged on most platforms, runs on every major platform currently in use. Python has strong support from text pattern matching to networking.

**Pycaret**

PyCaret is an open-source, low-code machine learning library in Python that automates machine learning workflows. It is an end-to-end machine learning and model management tool that exponentially speeds up the experiment cycle and makes you more productive.

Compared with the other open-source machine learning libraries, PyCaret is an alternate low-code library that can be used to replace hundreds of lines of code with a few lines only. This makes experiments exponentially fast and efficient. PyCaret is essentially a Python wrapper around several machine learning libraries and frameworks, such as scikit-learn, XGBoost, LightGBM, CatBoost, spaCy, Optuna, Hyperopt, Ray, and a few more.

The design and simplicity of PyCaret are inspired by the emerging role of citizen data scientists, a term first used by Gartner. Citizen Data Scientists are power users who can perform both simple and moderately sophisticated analytical tasks that would previously have required more technical expertise.

**Seaborn**

Seaborn is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with pandas data structures.

Seaborn helps you explore and understand your data. Its plotting functions operate on data frames and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots. Its dataset-oriented, declarative API lets you focus on what the different elements of your plots mean, rather than on the details of how to draw them.

1. **HARDWARE AND SOFTWARE REQUIREMENTS**
   1. **Hardware Requirements :**

* Computer or Laptop : A computer with sufficient processing power is required for training and running the deep learning model. A machine with a dedicated GPU (Graphics Processing Unit) is recommended to faster model training.
* RAM 8GB : Sufficient RAM is crucial, especially when dealing with large datasets. A minimum of 8 GB of RAM is recommended for efficient model training.
* GPU: I have used NVIDIA GeForce MX250 4gb graphic card, Graphic card in mandated since it takes longer time for larger epochs
* Processor: I have used Intel Core i7-10210U CPU running at 1.60 GHz. These processor helps in faster training, testing and prediction.
  1. **Software Requirements :**
* Operating System: Windows 10
* Python Libraries : Image Processing Libraries like OpenCV. NumPy, Pandas, Seaborn, Pycaret and scikit-learn, for efficient data manipulation and model evaluation.
* Visual Studio Code (VSCode) as the Integrated Development Environment (IDE) (Version: 2023.3.1).
* Development Tools: Python Flask (Version 3.11) for the application, Jupyter Notebooks.

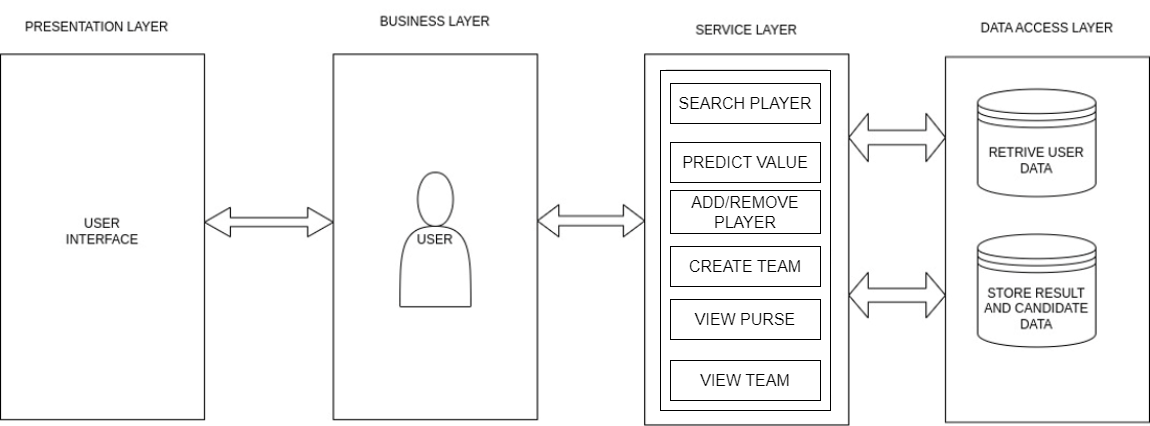
1. **SOFTWARE REQUIREMENTS SPECIFICATION**
   1. **Team Management**

Can search for a player and they can get predicted value and they based on the they can create team

Can view the chart and graph of players performance

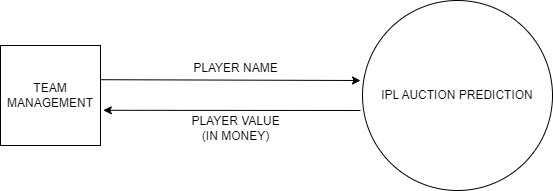
* 1. **Functional Requirements:**
* **Team Management**
  + SEARCH PLAYER
  + PREDICT VALUE
  + ADD/REMOVE PLAYER
  + CREATE TEAM
  + VIEW PURSE
  + VIEW TEAM

1. **SYSTEM DESIGN**
   1. **ARCHITECTURE DIAGRAM**



This Diagram shows how the user interacts with the web interface to search for the player value to know its prediction, when the players name is inserted through presentation layer, the data is passed to service layer where the data is sent to augment, preprocess and apply the trained model to find the predicted value of the player.

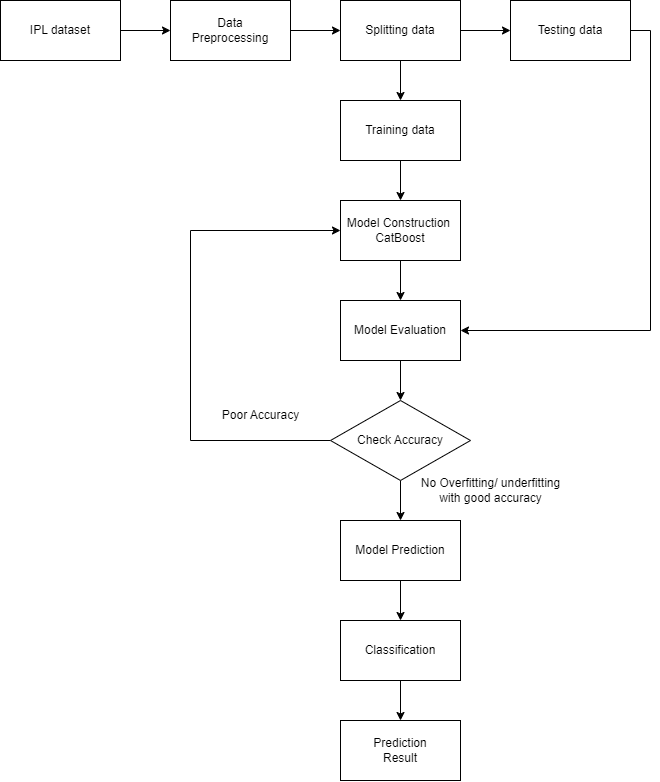
* 1. **CONTEXT DIAGRAM**



A context diagram provides an overview of a system and its interactions with external entities. For an IPL auction prediction project, the context diagram would illustrate the system's components and how it interacts with users and external data sources.

The context diagram provides a high-level view of how the IPL auction prediction system interacts with its environment, including users, external data sources, and feedback mechanisms. It helps stakeholders understand the system's scope, functionality, and dependencies, laying the groundwork for more detailed system design and implementation.

1. **DETAILED DESIGN**
   1. **PROCESS FLOW**

****

**Methodology:**

**IPL Dataset:** These datasets usually include a wide range of information such as match results, player statistics, team performances, venue details, and more, spanning multiple seasons of the Indian Premier League.

**Data Preprocessing:** Clean and preprocess the data to ensure it is ready for analysis. This includes removing duplicates, filling missing data, and standardizing the format of the data. This step is crucial to ensure the accuracy and reliability of the results.

**Splitting Data:** Divide the data into training and testing sets. The training set will be used to train the machine learning models, while the testing set will be used to evaluate the performance of the models. This step is important for ensuring that the models are able to generalize well to new, unseen data.

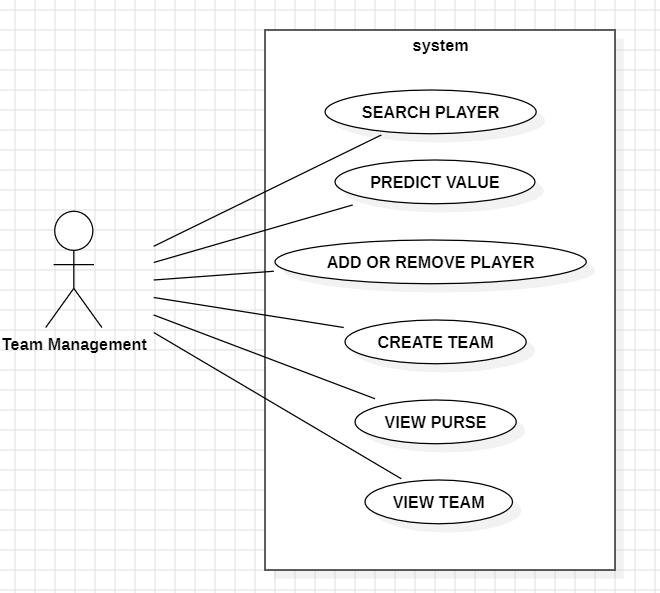
**Model Training:** Train machine learning models, CATBOOST model, on the training data. These models will learn patterns in the data that can be used to make predictions about the value of the player

**Model Evaluation:** Evaluate the performance of the models using the testing set. This will allow you to assess the accuracy of the models and identify areas for improvement.

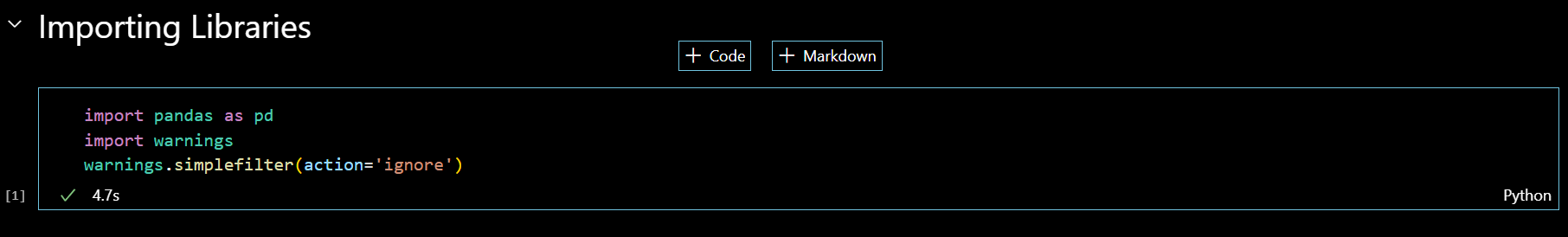
**Generating Models:** Select the best-performing models to make predictions about the value of the player. These models can be integrated into the user-friendly platform to provide accurate and reliable predictions to users.

**Result :** Present the predictions to users through clear and easy-to-understand charts, graphs This will allow users to easily understand the predictions.

* 1. **USE CASE DIAGRAM**

****

1. **IMPLEMENTATION**
   1. **Screenshots**

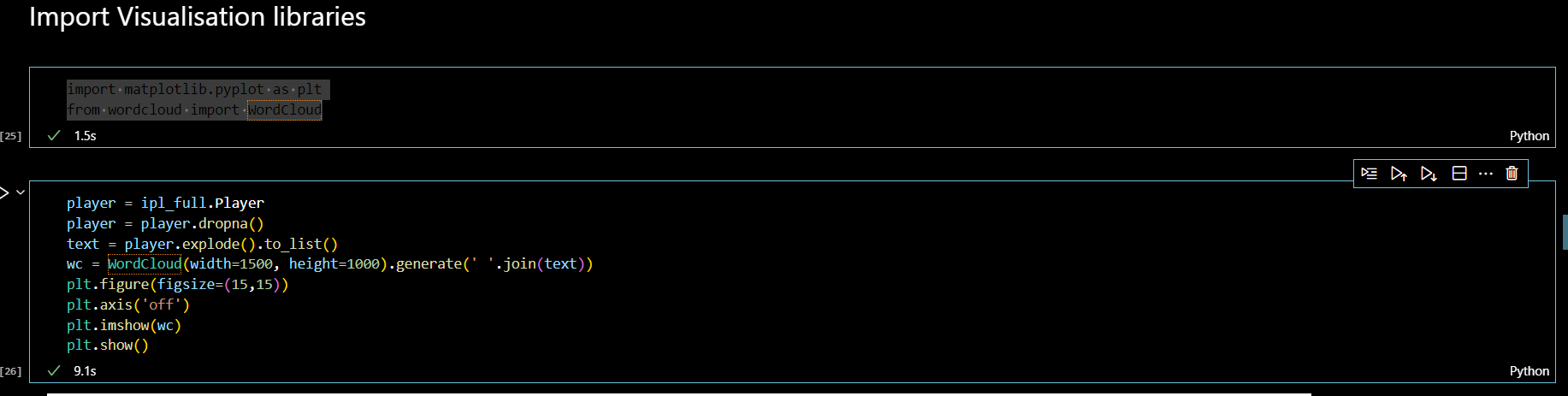
Importing all the necessary libraries to import the dataset 

Reading all the dataset from the year of 2008 to 2023 which includes players performance from past 15 years

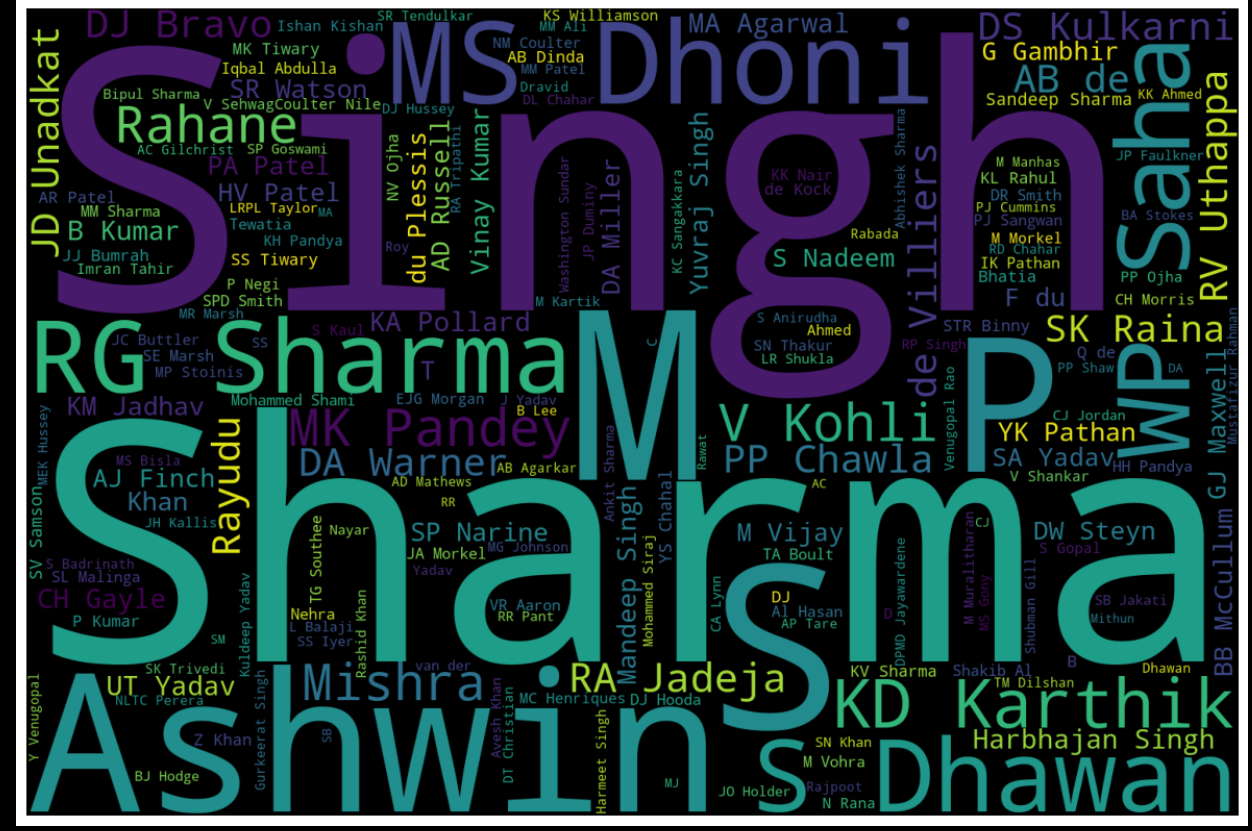


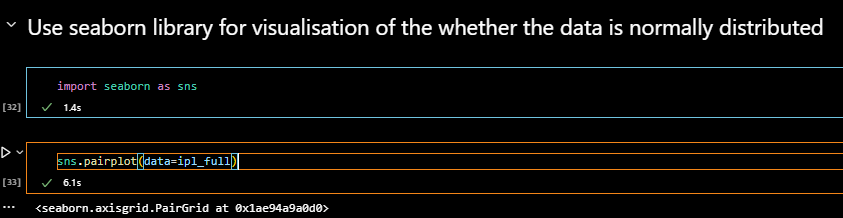
Concatenating all the dataset





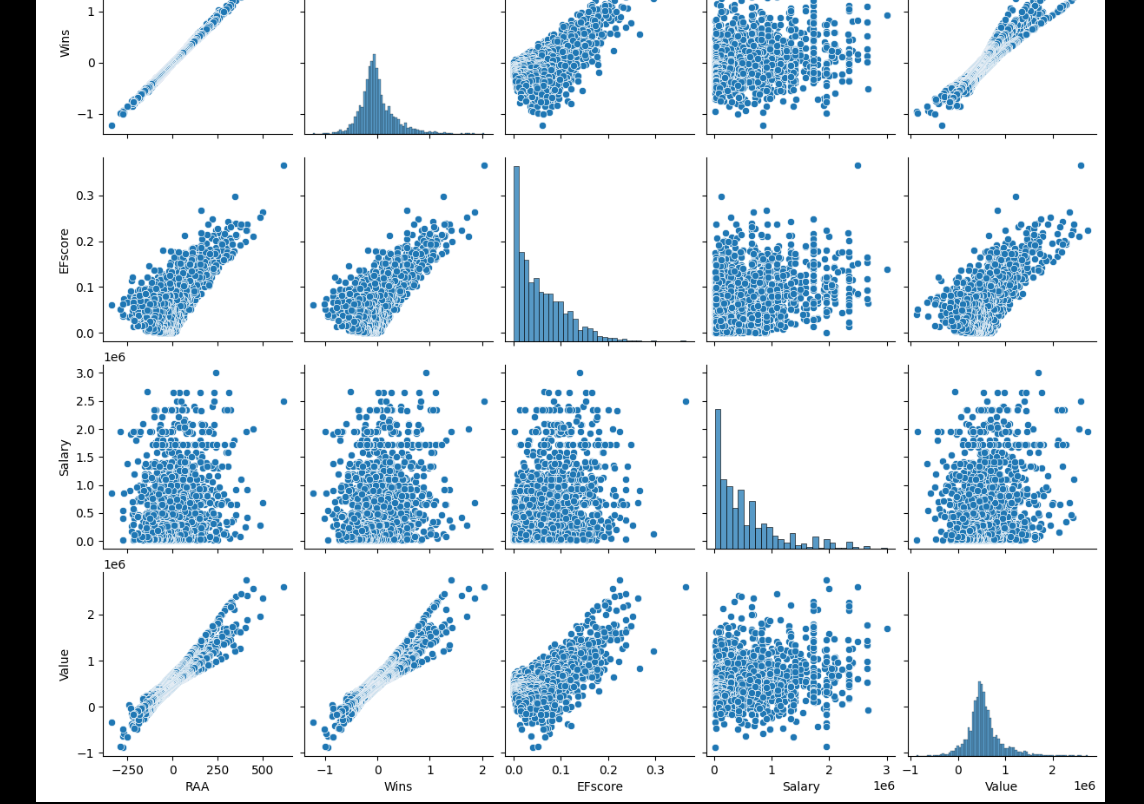
these two lines of code set up the environment for creating word clouds using the WordCloud library and visualizing them using Matplotlib. Once imported, you can use the functionalities provided by these libraries to create and customize word clouds based on your text data and visualize them using Matplotlib's plotting capabilities.

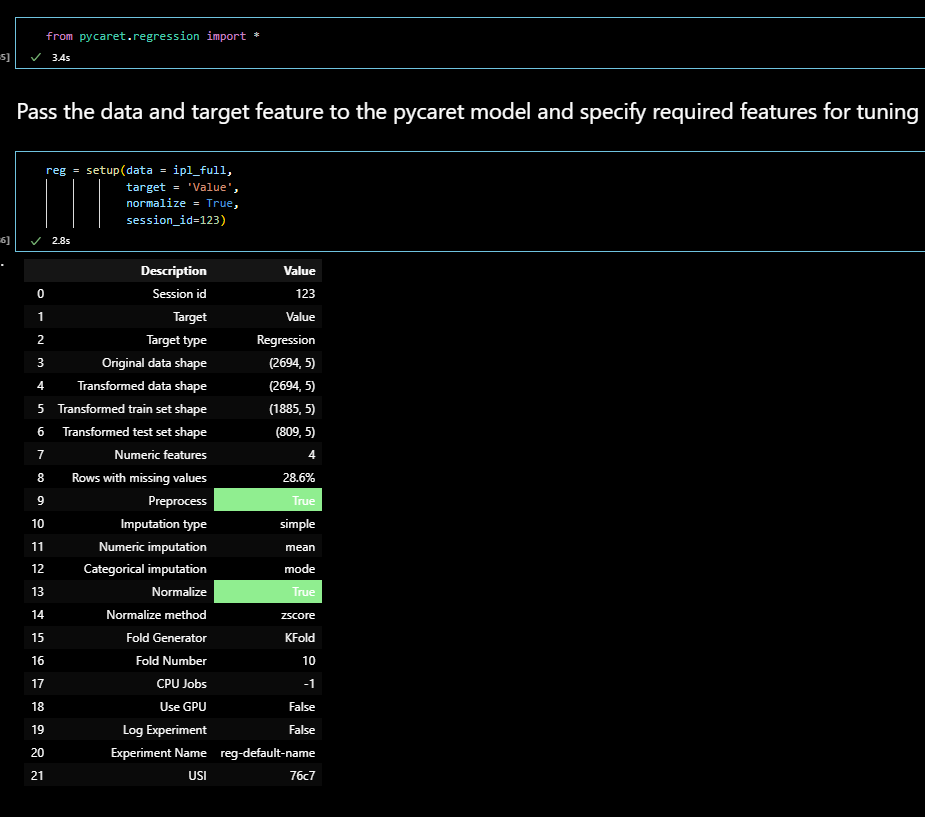




the Seaborn library to create a pair plot visualization based on the ipl\_full Data Frame. The pair plot will contain scatterplots along the diagonal, showing the distribution of each individual variable. The off-diagonal plots will be scatterplots showing the relationship between pairs of variables.

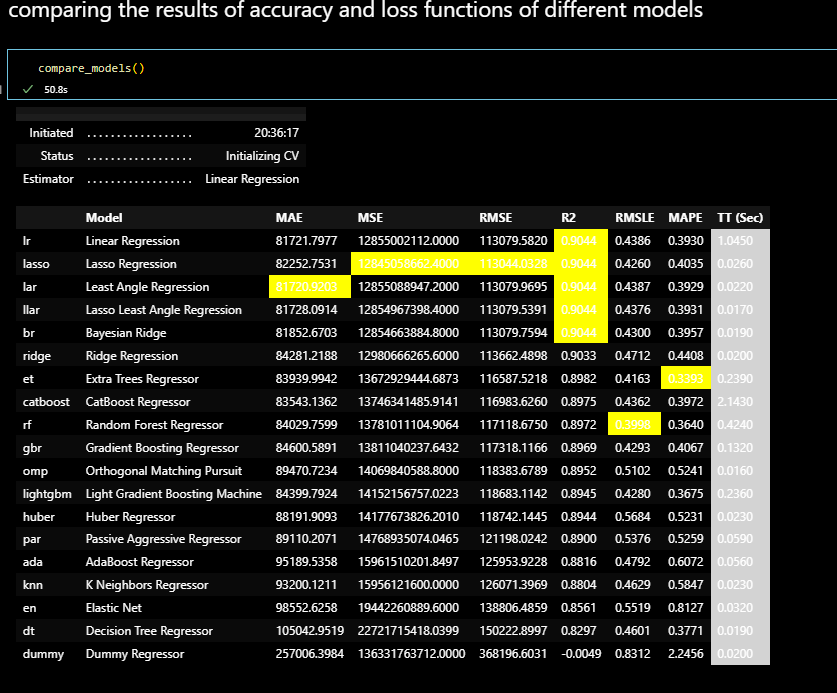
This visualization is particularly useful for exploring relationships between multiple variables in a dataset, identifying patterns, and detecting potential correlations. It's a powerful tool for initial exploratory data analysis (EDA) and gaining insights into the structure of the data.





from pycaret.regression import \*: This line imports the necessary functions and classes from the PyCaret library for regression tasks. PyCaret is a low-code machine learning library in Python that automates various steps in the machine learning workflow, making it easier to perform tasks such as data preprocessing, model training, and evaluation.

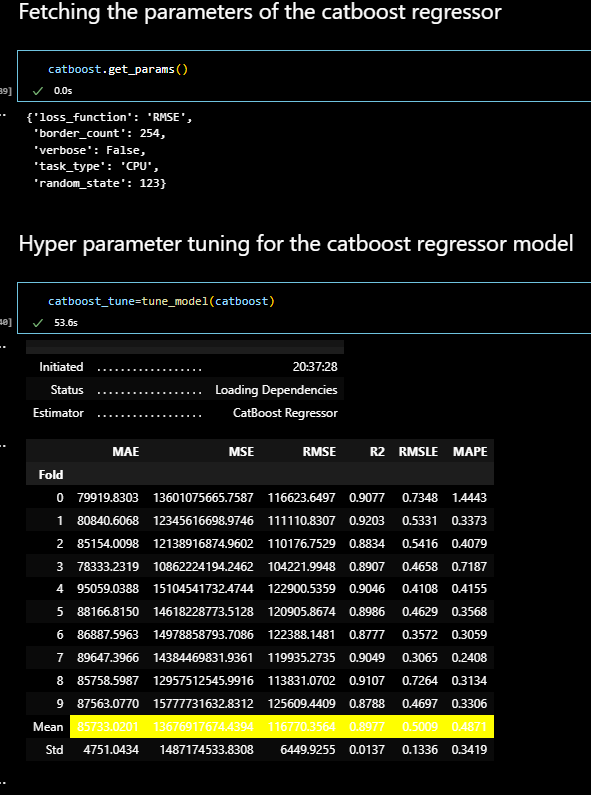
After running this code, the reg object will contain all the preprocessed data, including the transformed features, the target variable, and other information needed for regression modeling. This preprocessed data can then be used to train and evaluate regression models using PyCaret's automated workflow.



The function compares the performance of different regression models using default hyperparameters and evaluates their performance using cross-validation.

After execution, PyCaret generates a table displaying various performance metrics for each model evaluated, allowing you to compare their performance and identify the top-performing ones based on the chosen metric.

Overall, this code snippet simplifies the process of comparing regression models by automating the training and evaluation process, making it easier for users to identify the most suitable model for their regression task.



Top of Form

the process of creating and training a CatBoost regression model by automating the training process and providing a trained model object that can be used for various tasks, such as making predictions or evaluating model performance.

**APPENDIX:**

**BIBIALOGRAPHY**

**Paper 1 :** J. Rani P., A. Kulkarni, A. V. Kamath, A. Menon, P. Dhatwalia and D. Rishabh, "Prediction of Player Price in IPL Auction Using Machine Learning Regression Algorithms," 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2020, pp. 1-6, doi: 10.1109/CONECCT50063.2020.9198668.

**Paper 2 :** N. P, M. V, R. M, M. C. Prabhu, B. Meghana and C. Jahnavi, "IPL Players Cost Pay Prediction using Machine Learning Techniques," 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 2023, pp. 597-602, doi: 10.1109/ICAISS58487.2023.10250755.

**Paper 3 :** M. Sumathi, S. Prabu and M. Rajkamal, "Cricket Players Performance Prediction and Evaluation Using Machine Learning Algorithms," 2023 International Conference on Networking and Communications (ICNWC), Chennai, India, 2023, pp. 1-6, doi: 10.1109/ICNWC57852.2023.10127503.

**Paper 4 :** S. Agrawal, S. P. Singh and J. K. Sharma, "Predicting Results of Indian Premier League T-20 Matches using Machine Learning," 2018 8th International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 2018, pp. 67-71, doi: 10.1109/CSNT.2018.8820235.

**Paper 5:** P. Jhansi Rani, A. Vidyadhar Kamath, A. Menon, P. Dhatwalia, D. Rishabh and A. Kulkarni, "Selection of Players and Team for an Indian Premier League Cricket Match Using Ensembles of Classifiers," 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2020, pp. 1-6, doi: 10.1109/CONECCT50063.2020.9198371.