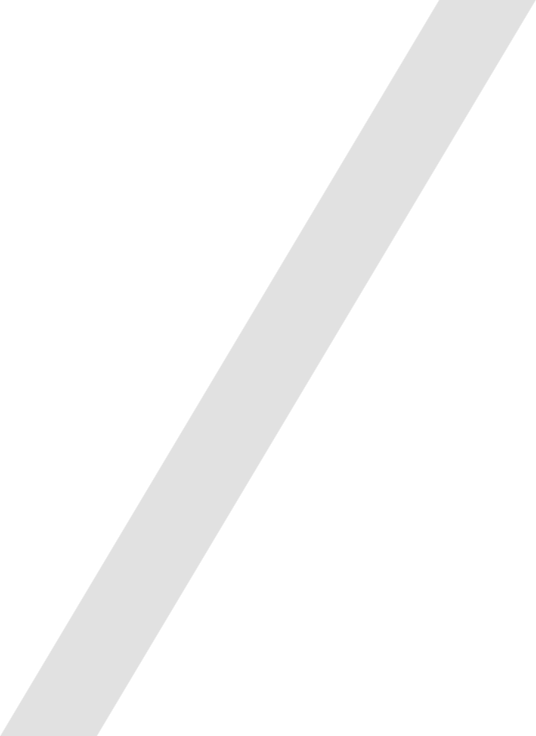
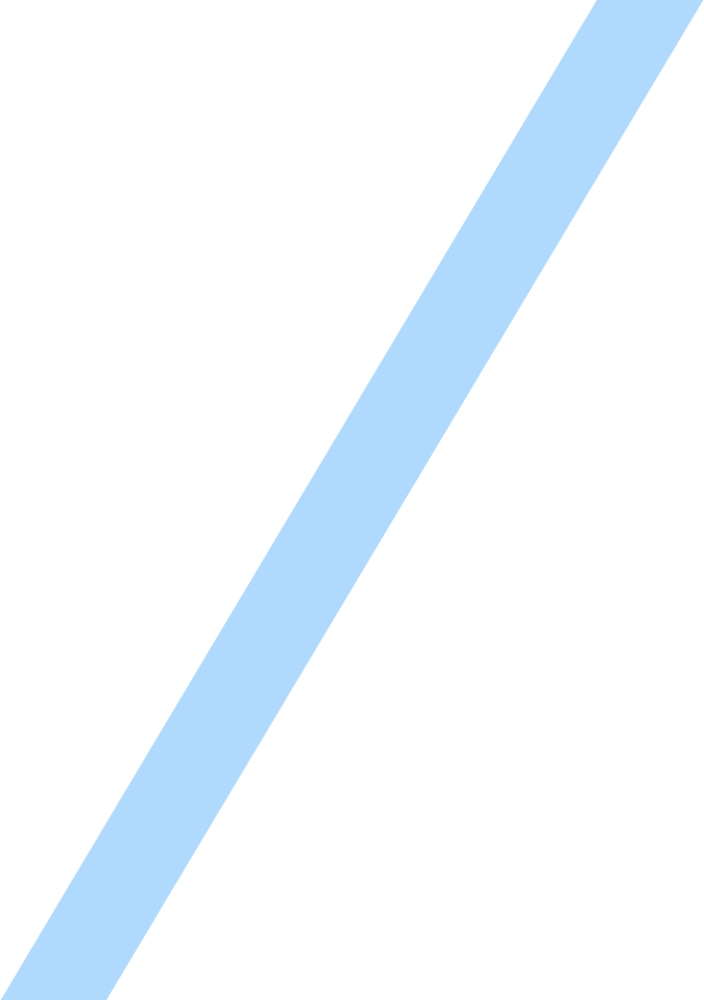
|  |
| --- |
| TECHNICAL REPORT TEMPLATE |

|  |
| --- |
| Electrical & Computer Engineering & Computer Science (ECECS) |

|  |  |
| --- | --- |
| SPRING 2024 |  |



Contents

[Project Name 2](#_Toc96341549)

[Executive Summary 2](#_Toc96341550)

[Technical Report 3](#_Toc96341551)

[Highlights of Project 3](file:///C:\Users\ardia\Desktop\TEACHING\SPRING%2022\DSCI6002%20Data%20Science\Projects\Technical%20report%20Template.docx#_Toc96341552)

[Submitted on: 3](file:///C:\Users\ardia\Desktop\TEACHING\SPRING%2022\DSCI6002%20Data%20Science\Projects\Technical%20report%20Template.docx#_Toc96341553)

[Abstract 4](#_Toc96341554)

[Methodology 5](#_Toc96341555)

[Results Section 5](#_Toc96341556)

[Discussion 6](#_Toc96341557)

[Conclusion 6](#_Toc96341558)

[Contributions/References 7](#_Toc96341559)

|  |
| --- |
| DATA ANALYSIS FOR NEWCRICKET TEAM FRANCHISE |

|  |  |  |
| --- | --- | --- |
| Executive Summary Cricket is one of the most popular sports in the world. The IPL allowed many talented players to showcase their skills on a platform where many international cricketers compete, and as a result, cricket standards in India reached their pinnacle. Therefore, we made the decision to use a machine learning algorithm to forecast player performances, team performance, and the likelihood of winning the cup. This will enable a new IPL franchise to make decisions about their initial course of action with ease. | | |
| person at a table writing in a notebook with people around | | |
| **Team Members:**  **Hemanth Pathipati**  **Srikanth Dumpala**  **Ramana Bellamkonda**  **Naga Venkata Bharath Lanka** | **Questions?**  Contact : |  |

|  |
| --- |
| Technical Report |

|  |  |
| --- | --- |
| **DATA ANALYSIS FOR NEW CRICKET TEAM FRANCHISE** |  |
| Highlights of Project • Projecting player outcomes to help a new IPL franchise choose which players to select and which ones to pass on.   • By comparing their results with those of other franchises, teams can use this information to analyze their own performance.    • We used data engineering techniques to tune our data to meet our needs, and we also added numerous visualizations that used exploratory data analysis to describe the different features of the dataset.   • To improve usage and interaction, we deployed the model using the Flask app. Submitted on: |

The core of data science and data engineering is to extracting meaningful insights from the given dataset. Alternatively looking for answers that could help a business. We had examined how to get started with Python data analysis in this article. In this notebook, we use the Python packages NumPy, Pandas, Matplotlib, and Seaborn.

## Abstract

The Indian Premier League (IPL) Dataset that we used was obtained from cricsheet and was made available on Kaggle Datasets. Fans of cricket should be able to relate to the questions we pose and the issues we try to address because the Indian Premier League is one of the most well-known cricket competitions in the world.

[Here you can find a few useful tips on coming up with a great pitch](https://www.youtube.com/watch?v=bZTWx2bftaw)

Executive Summary

1.Initially, the data was cleaned up by adding missing values and removing superfluous Nan values as part of the preprocessing step.   
2. Later, in order to better understand the dataset, we plotted multiple times.  
3. Training and testing models is the next step.   
  
4. To predict, we employed 8 machine learning algorithms.   
5. Five-fold cross-validation methods were employed.   
6. Combined the best model into a single layer using super-learning.   
7. Stored the output locally using the Pickle Library.   
8. Using the Flask application, the model was finally deployed.

Introductory Section

The Indian Premier League (IPL) is the world's most popular cricket league, and each franchise in the IPL requires assistance in locating the best players. We can use the IPL dataset to evaluate a player's statistical worth before purchasing them.

Review of available research

Many features needed for IPL evaluation are visualized, and the entire IPL dataset is carefully examined. Many machine learning algorithms have been used to compare the winners of any two teams. There aren't many models that try to rank players based on a limited number of mathematical models or simple formulas. Efficiency was very low when significant quantities of valuable data sets were missing. This is due to the fact that there was insufficient data available when these models were suggested. The current models are unable to satisfy the unique requirements of the T20 game.

## Methodology

Using CRISP-DM methodology,Logistic Regression, Random Forest, KNeighbors, Decision Tree, Ada Boost, Bagging, Gradient Boosting, and XGBoost were the eight machine learning algorithms we employed. There is a 5-fold cross-validation. Although it doesn't really help, PCA makes the dataset's condition worse. merged our best model into a single layer using Super Leaner. With a Super Leaner, the maximum accuracy we can attain is 94%, which is quite good!

## Results Section

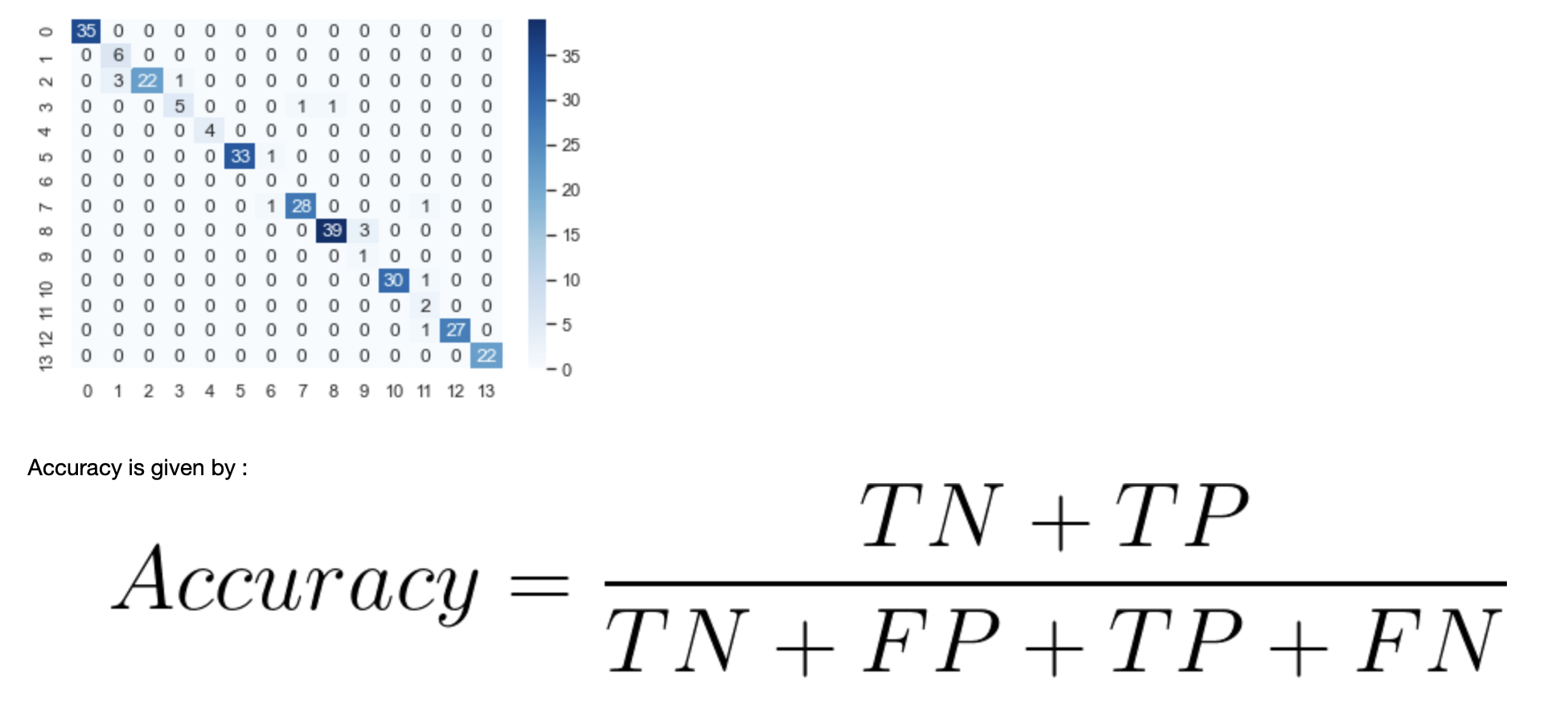
94% of the time, the classification is accurate; roughly 3.2% of losing teams are deemed winners. We are trying to show with this example that accuracy is a poor metric when the data set is not uniform. In these situations, relying solely on accuracy could result in an incorrect reading of the data.   
R^2 is the determination coefficient.   
RSS is the sum of the residual squares.

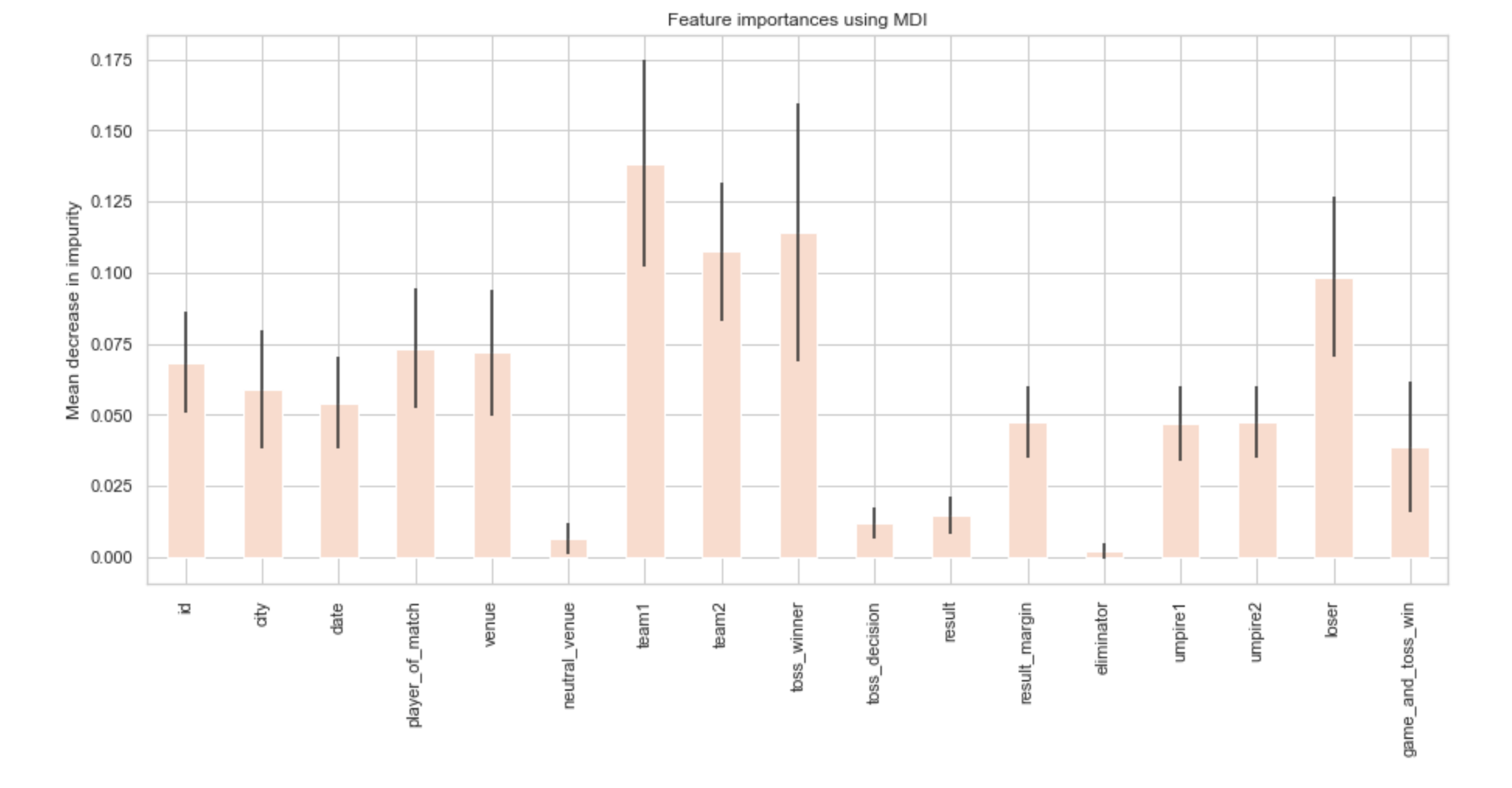
TSS = total sum of squares

R^2 = 1 - (RSS/TSS)



An excellent classifier ought to have a high r2 score of 1. Sum of squares of residuals (RSS) in the numerator tends to one as precision decreases.





A screenshot of a computer

Description automatically generated

It is evident from observation that certain traits are more important than others, such as team1, team2, loser, and toss winner.

## Discussion

In this report, we use a comprehensive dataset of all the IPL matches ever played to present the results of a detailed analysis of several metrics used in the game along with useful visualizations. We had trouble ranking the players, but we overcame it by simulating their advantages and disadvantages against particular opponents, how they performed on a particular surface, etc. These are specific details that can be very helpful and give the team a big competitive edge. The Player Ranking Index has also been used to rank the players using machine learning techniques. When the recommended method was applied, forecast accuracy increased to 94%.

## Conclusion

The objective is to develop a model that predicts the game's winner using the most effective machine learning technique. Our goal is also to deliver dependable prediction with minimal features and testing. In this study, only 17 crucial traits are taken into account. Various classification techniques were used. Techniques like gradient boosting, Ada boost, boost, and bagging are employed. Before the data was used in the model, it underwent pre-processing. The Super Learner method yielded the best outcomes in this paradigm. To expand on this, additional strategies like association rules, clustering, and evolutionary algorithms could be applied. Owing to the limitations of this research, more intricate models combining a variety of data types must be developed in order to improve early match prediction accuracy.

## Contributions/References

<https://www.kaggle.com/datasets/patrickb1912/ipl-complete-dataset-20082020>

<https://ieee-dataport.org/documents/ipl-dataset>