**Final Project Report:**

In this project, the primary goal was to analyze and predict the performance of soccer teams in the top 5 global leagues. The analysis involved two key steps: categorizing teams based on their previous season's performance and predicting their positions at the end of the current season using a Random Forest Classifier.

1. Introduction

The project commenced with a meticulous classification process, assigning teams to three distinct categories: the prestigious Top 5, the stable Mid-table, and those on the brink of Relegation. Leveraging logistic regression, we carefully selected features known to influence a team's success. The predictive prowess of machine learning was then harnessed through a Random Forest Classifier, utilizing 100 estimators to discern intricate patterns within the chosen features.

Model evaluation extended beyond accuracy metrics, encompassing a detailed classification report, a confusion matrix, and a scrutiny of feature importance. Simultaneously, a critical step of data cleaning fortified the dataset, ensuring its reliability for subsequent analyses.

This comprehensive approach continued into the ongoing season. Merging datasets from the past and present, we created a dynamic representation of team dynamics, capturing the essence of real-time soccer.

Our project transcends conventional predictive modeling. It endeavors not just to foresee outcomes but to unravel the intricate tapestry of factors shaping a team's destiny. Beyond the algorithms and classifications, this project embodies an immersive exploration into the heartbeat of soccer, offering insights that resonate with enthusiasts, analysts, and the strategic minds steering the fate of soccer clubs.

2. Data Cleaning and Preparation

Data cleaning was vital to ensure the accuracy of predictions. Columns such as 'PlAge,' 'Possession,' 'GlsAsistPer90,' 'SoTPercentage,' 'GlsperSoT,' 'DistofShot,' and 'PassCmpPer' were converted to numeric format, removing potential inconsistencies.

2.1 Classification of Previous Season

Post data cleaning, the classification process involved selecting pertinent features for logistic regression. The model was trained using a Random Forest Classifier, emphasizing accuracy metrics, the classification report, and the visualization of feature importance. The correlation between various features and the classification was also visualized.

2.2 Data Update for Current Season

Merging datasets from the previous and current seasons, we focused on selected features to create training and testing sets for predicting the current season's outcomes.

3. Model Training and Prediction

3.1 Training the Model

Adapting the Random Forest Classifier to predict the final position of teams for the current season involved converting selected features to float and training the model on the updated dataset.

3.2 Evaluation and Predictions

The model's performance was assessed, showcasing an impressive accuracy of 85%, and predictions were made for the teams' positions at the end of the current season.

4. Results and Insights

The project yielded substantial results, showcasing an accuracy of 85% in predicting team positions for the current season. The classification report further detailed the precision, recall, and F1-score for each category, providing a comprehensive understanding of the model's performance. The confusion matrix illustrated the model's ability to correctly classify teams into their respective categories.

Feature importance analysis uncovered critical determinants of team performance. 'GoalDiff' emerged as the most influential factor, followed by 'GlsAsistPer90,' 'TotalGlsAsist,' and 'Possession.' This insight provides a strategic understanding of the features that significantly impact a team's classification.

The predictions for the current season unveiled a nuanced perspective on each team's expected performance, laying the groundwork for informed decision-making by soccer enthusiasts, analysts, and team management.

5. Future Considerations

5.1 Model Improvement

Future enhancements may include exploring alternative algorithms, conducting hyperparameter tuning, and incorporating additional relevant features.

5.2 Real-time Updates

Considering real-time data updates could enhance the accuracy of predictions as the current season progresses.

6. Conclusion

In conclusion, the project successfully classified teams based on the previous season and predicted their positions for the current season, achieving an accuracy level of 85%. The insights gained have valuable applications for soccer enthusiasts, analysts, and team management.