Supervised Learning: CSGO Round Winner Prediction

Table of Contents

Chapter 1: Introduction

1.1 Introduction

Counter-Strike: Global Offensive (CS:GO)

Counter-Strike: Global Offensive (CS:GO) is a first-person shooter (FPS) game renowned for its competitive gameplay and global esports presence. Teams of terrorists and counter-terrorists engage in tactical encounters across various maps, each round presenting unique challenges and opportunities. The game's strategic depth and skill-intensive nature have made it a cornerstone of the esports community.

Significance of Predicting Round Outcomes

In the dynamic realm of CS:GO, predicting round outcomes has emerged as a critical facet of competitive strategy. Teams constantly seek to gain an edge by anticipating their opponent's moves and adjusting their tactics accordingly. Accurate round winner prediction equips teams with invaluable insights, enabling them to make timely decisions that can ultimately influence the course of a match.

1.2 Scope

The scope of this project encompasses a comprehensive analysis of CS:GO gameplay data, with the primary objective of developing and evaluating predictive models for round winner classification. This involves an in-depth exploration of various facets within the context of the game, as well as the application of machine learning techniques to enhance strategic decision-making.

In-Depth Gameplay Analysis

This project involves a granular examination of CS:GO rounds, considering factors such as time dynamics, team scores, map-specific strategies, bomb status, and player health and armor levels. These features collectively provide a detailed snapshot of each round's progression.

Machine Learning Application

The application of machine learning techniques forms a crucial aspect of this project. By employing algorithms such as logistic regression, decision trees, and random forests, we seek to harness patterns within the gameplay data to make accurate round winner predictions.

Enhancing Esports Strategy

The insights gained from this analysis are not only valuable for CS:GO enthusiasts but also have broader implications for the esports community at large. The predictive models developed in this project have the potential to optimize gameplay strategies, inform tactical decisions, and ultimately contribute to improved performance in professional CS:GO matches.

Transferable Knowledge

The methodologies and techniques employed in this project can serve as a foundation for similar endeavors in other gaming contexts. The principles of data-driven decision-making and machine learning can be applied to various esports titles, offering a versatile framework for strategic enhancement.

Contributing to Esports Analytics

By combining an understanding of CS:GO gameplay intricacies with advanced analytics, this project aims to contribute to the evolving field of esports analytics. The findings and methodologies presented here have the potential to inform future research and applications in competitive gaming.

1.3 Objectives

- Develop predictive models for classifying round winners in Counter-Strike: Global Offensive (CS:GO).

- Meticulously retrieve and preprocess historical CS:GO round data.

- Conduct feature engineering to distill critical round-specific information for model training.

- Systematically evaluate a range of machine learning algorithms for round winner prediction.

- Utilize classification metrics to comprehensively assess model performance.

- Provide valuable insights and recommendations for teams, analysts, and broadcasters in the esports domain.

Chapter 2: Literature Review

2.1 Background Study

The background study provides a comprehensive overview of the existing body of knowledge related to esports analytics, with a specific focus on CS:GO gameplay prediction. This section aims to contextualize the project within the broader landscape of competitive gaming and predictive modeling.

Evolution of Esports Analytics

Esports analytics has witnessed significant growth in recent years, paralleling the rise of competitive gaming as a global phenomenon. Researchers and analysts have increasingly turned their attention towards understanding the underlying dynamics of esports matches, seeking to uncover patterns and trends that can influence outcomes.

CS:GO as a Competitive Platform

Counter-Strike: Global Offensive (CS:GO) has emerged as one of the most prominent titles in the esports arena. Its complex gameplay dynamics, strategic depth, and high skill ceiling make it an ideal candidate for predictive modeling. Previous studies have explored various aspects of CS:GO gameplay, ranging from player performance metrics to team-based strategies.

Supervised Learning in Esports

This area focuses on understanding how supervised learning techniques have been applied in the context of esports, including the prediction of game outcomes, player performance, and strategic decision-making.

Predictive Modeling in Esports

The application of machine learning and predictive modeling techniques in esports is a burgeoning field. Researchers have utilized diverse methodologies, including regression analysis, classification algorithms, and ensemble learning, to extract valuable insights from gameplay data. These studies have demonstrated the potential for data-driven decision-making in optimizing team performance.

2.2 History

The history of esports can be traced back to the late 20th century when competitive video gaming began to gain traction. One of the earliest notable esports events was the Space Invaders Championship held by Atari in 1980. This competition drew over 10,000 participants, marking a significant milestone in the evolution of competitive gaming.

The 1990s witnessed a surge in popularity with the rise of multiplayer online games like Quake and StarCraft. LAN parties and local tournaments became common, allowing players to compete against each other in person. These grassroots events laid the foundation for the organized esports tournaments we see today.

In the early 2000s, South Korea emerged as a powerhouse in esports, particularly in games like StarCraft: Brood War. The country's infrastructure and support for professional gaming paved the way for the development of dedicated leagues, teams, and training facilities. This period also saw the birth of iconic esports teams and players who achieved celebrity status within the gaming community.

The mid-2000s brought about the emergence of MOBAs (Multiplayer Online Battle Arenas) like Defense of the Ancients (DotA) and its successor, DotA 2. These games introduced a new level of strategic depth and team coordination, becoming staples in the esports landscape.

The introduction of streaming platforms like Twitch in the 2010s revolutionized the way esports content was consumed. Fans could now watch their favorite players and teams compete in real-time, creating a more immersive and interactive viewing experience.

Esports continued to grow in popularity, with major corporations, sports teams, and celebrities investing in the industry. High-profile tournaments, such as The International for DotA 2 and the League of Legends World Championship, offered multi-million-dollar prize pools, further solidifying the legitimacy of esports as a competitive discipline.

Today, esports is a global phenomenon with a dedicated fan base, professional leagues, and players who compete at the highest level. The industry continues to evolve, with advancements in technology, broadcasting, and player development shaping the future of competitive gaming. Esports has also gained recognition on the world stage, with discussions about its potential inclusion in major sporting events like the Olympics.

The history of esports is a testament to the passion and dedication of players, organizers, and fans who have collectively propelled it from its humble beginnings to a thriving, multi-billion-dollar industry.

Chapter 3: Methodology
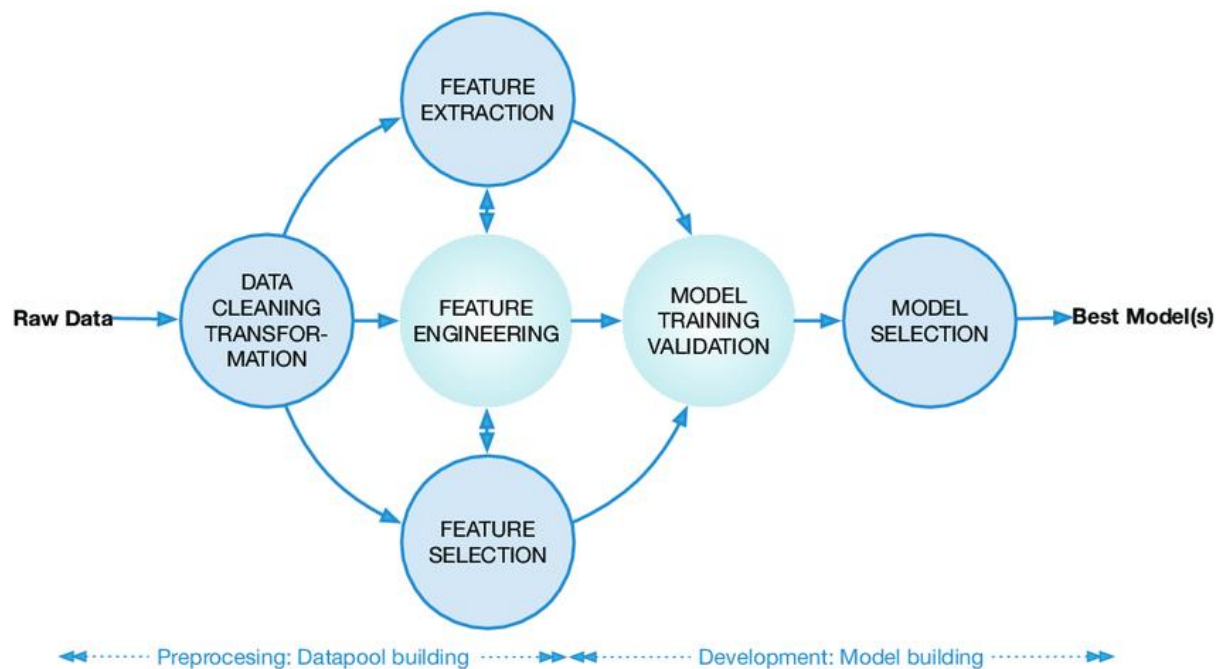
Figure 3.1: Flow Chart of Methodology

3.1 Data Collection and Preprocessing

3.1.1 Dataset Description

Data Source:

The dataset used in this project was obtained from Kaggle, a renowned platform for data science competitions and datasets. Specifically, the dataset consists of snapshots from CS:GO rounds, providing a comprehensive view of gameplay dynamics. The dataset encompasses a diverse range of attributes, each contributing to the holistic understanding of round outcome.

Tools Used:

Python: The primary programming language used for data analysis, preprocessing, and model development.

Jupyter Notebooks: An interactive environment that facilitated code execution, visualization, and documentation.

Pandas: A Python library used for data manipulation and analysis.

NumPy: A fundamental package for numerical computations in Python, utilized for various mathematical operations**.**

Matplotlib and Seaborn: Libraries employed for data visualization, enabling the creation of insightful graphs and plots.

Scikit-learn: A versatile machine learning library providing a wide array of algorithms and tools for model development and evaluation.

```python
df=pd.read_csv('Downloads\csgo_round_snapshots.csv')
pd.set_option('display.max_columns', None)
df.head(10)
```

	time_left	ct_score	t_score	map	bomb_planted	ct_health	t_health	ct_armor	t_armor	ct_money	t_money
0	175.00	0.0	0.0	de_dust2	False	500.0	500.0	0.0	0.0	4000.0	4000.0
1	156.03	0.0	0.0	de_dust2	False	500.0	500.0	400.0	300.0	600.0	650.0
2	96.03	0.0	0.0	de_dust2	False	391.0	400.0	294.0	200.0	750.0	500.0
3	76.03	0.0	0.0	de_dust2	False	391.0	400.0	294.0	200.0	750.0	500.0
4	174.97	1.0	0.0	de_dust2	False	500.0	500.0	192.0	0.0	18350.0	10750.0
5	114.97	1.0	0.0	de_dust2	False	500.0	500.0	500.0	0.0	1750.0	10550.0
6	94.97	1.0	0.0	de_dust2	False	295.0	111.0	296.0	0.0	1700.0	5000.0
7	74.97	1.0	0.0	de_dust2	False	295.0	111.0	296.0	0.0	1700.0	5000.0
8	54.97	1.0	0.0	de_dust2	False	249.0	89.0	269.0	0.0	2300.0	2950.0
9	174.97	2.0	0.0	de_dust2	False	500.0	500.0	196.0	0.0	19800.0	23450.0

Figure 3.2: Dataset

Attributes Overview:

1. `time_left`: Indicates the remaining time in the round.

2. `ct_score`: Represents the score of the Counter-Terrorist team.

3. `t_score`: Represents the score of the Terrorist team.

4. `map`: Indicates the current map being played.

5. `bomb_planted`: Indicates whether the bomb has been planted or not.

6. `ct_health` and `t_health`: Represents the health of Counter-Terrorist and Terrorist players, respectively.

7. `ct_armor` and `t_armor`: Represents the armor points of Counter-Terrorist and Terrorist players, respectively.

8. ... (and additional attributes)

This comprehensive dataset serves as a valuable resource for predicting round winners in CS:GO.

3.1.2 Data Cleaning and Transformation

Handling Missing Values

A thorough examination of the dataset revealed that it was complete, with no missing or null values present in any of the columns. This ensured that the dataset was ready for further analysis and modeling without the need for imputation or data recovery.

Data Transformation

To ensure that the data was well-prepared for modeling, a critical step involved standardizing numerical attributes. This process was executed using a `**StandardScaler**`, which transformed the data to a common scale. Standardization is crucial in preventing any particular feature from dominating the modeling process, ensuring that each attribute contributes effectively.

Encoding Categorical Variables

Given the nature of the dataset, which primarily consists of numerical attributes, there was no explicit need for encoding categorical variables. The data was already structured in a format conducive to modeling, eliminating the necessity for additional preprocessing in this regard.

3.2 Feature Engineering

3.2.1 Label Encoding

Utilizing Label Encoding

Label encoding is a crucial preprocessing step when working with categorical data. In this project, we employed label encoding to convert categorical variables into numerical equivalents. This transformation is vital as most machine learning algorithms require numerical input.

Application of Label Encoding:

1. `bomb_planted`: This binary variable was encoded into numerical values. A value of `1` represents that the bomb was planted, while `0` indicates otherwise.

2. `map`: The `map` column, denoting the current map being played, was transformed into numerical representations. Each unique map was assigned a unique integer.

Additionally, label encoding was extended to the target variable, `round-winner`, to ensure compatibility with classification algorithms.

3.2.2 Feature Selection

Rationalizing Feature Selection

Effective feature selection is a critical aspect of building accurate predictive models. It involves identifying and utilizing the most relevant attributes while discarding those that may introduce noise or redundancy.

Figure 3.3: Features Score
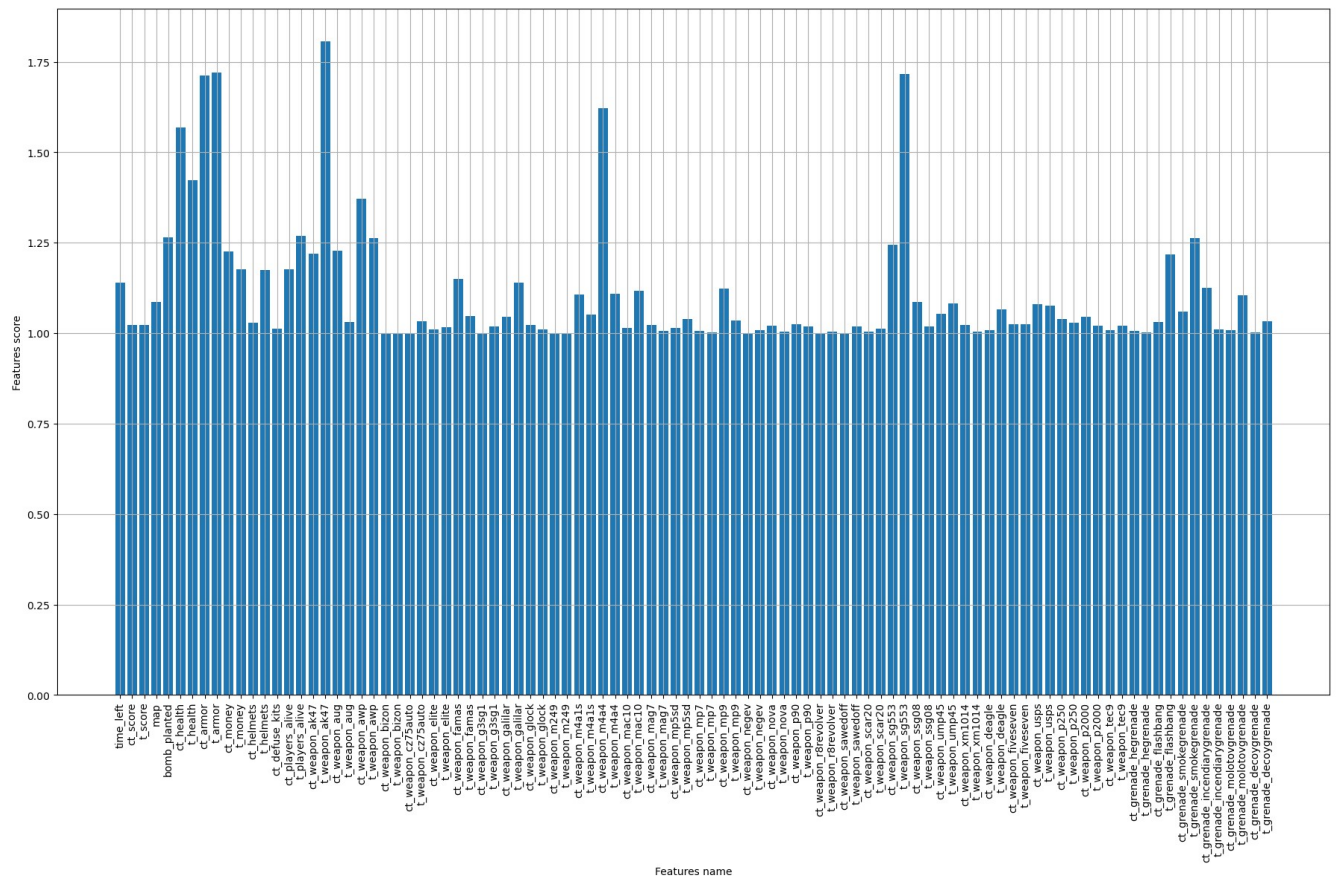
Approach:

Linear Discriminant Analysis

Linear Discriminant Analysis is a dimensionality reduction technique that is particularly effective when it comes to classification tasks. It aims to find a subspace where the classes are well-separated, making it easier to discriminate between them. In the context of this project, LDA is utilized to identify the features that contribute the most to predicting round outcomes in CS:GO.

Here's a more detailed explanation of the approach:

1. **Dimensionality Reduction:**

LDA works by projecting the data onto a lower-dimensional subspace. This means that it transforms the original high-dimensional feature space into a space with fewer dimensions while retaining as much of the class-discriminatory information as possible.

2. Maximizing Class Separability:

The primary objective of LDA is to maximize the separability between different classes. In the case of CS:GO round prediction, this means finding the feature combinations that best differentiate between rounds won by Counter-Terrorists and rounds won by Terrorists.

3. Identifying Discriminative Features:

Through this process, LDA identifies the features that are most crucial in determining the outcome of a round. These features are the ones that exhibit significant differences in their distributions across different classes. For example, it might find that variables like team composition, bomb planting status, and player health levels are particularly influential.

4. Reducing Redundancy and Noise:

LDA not only identifies the most discriminative features but also helps in reducing redundancy and noise in the data. This is important as irrelevant or noisy features can hinder the performance of predictive models.

5. Enhancing Model Performance:

By selecting the most relevant features, LDA contributes to building a more accurate and efficient predictive model. It ensures that the model is focused on the attributes that have the most impact on predicting round winners.

Feature Importance Ranking: The coefficients obtained from LDA were exponentiated. This process provided a measure of the relative importance of each feature. The resulting importance scores were then flattened, allowing for a clear visualization of each attribute's contribution.

Following this, the top 20 features with the highest importance scores were selected for subsequent model training and evaluation. This meticulous feature selection process ensures that the chosen attributes are the most influential in determining round outcomes.

```
top_20_values = df_feature_score.nlargest(20,"feature_score")
top_20_values
```

	Feature_names	feature_score
17	t_weapon_ak47	1.805807
8	t_armor	1.721494
65	t_weapon_sg553	1.716493
7	ct_armor	1.711748
40	ct_weapon_m4a4	1.622804
5	ct_health	1.568964
6	t_health	1.422261
20	ct_weapon_awp	1.372082
15	t_players_alive	1.268888
4	bomb_planted	1.264437
89	t_grenade_smokegrenade	1.263718
21	t_weapon_awp	1.263513
64	ct_weapon_sg553	1.245461
18	ct_weapon_aug	1.228728
9	ct_money	1.226958
16	ct_weapon_ak47	1.219753
87	t_grenade_flashbang	1.217213
14	ct_players_alive	1.177012
10	t_money	1.176293
12	t_helmets	1.173934

Figure 3.4 Selected Features

Chapter 4: Model Selection

Choosing the Right Models

Selecting the appropriate machine learning models is a critical step in ensuring accurate predictions. In this project, we carefully considered the following models for the task of round winner prediction in CS:GO:

4.1 Logistic Regression

Logistic regression is a versatile and interpretable model that is particularly well-suited for binary classification tasks. It provides valuable insights into the relationships between features and the likelihood of a specific outcome. In our context, it is suitable for predicting whether the Counter-Terrorist (CT) or Terrorist (T) team will win a given round.

Key Characteristics:

Simplicity and Interpretability: Logistic Regression provides a clear interpretation of the relationship between the features and the probability of a specific outcome. This is crucial for understanding the factors that contribute to round outcomes.

Linear Decision Boundary: The algorithm establishes a linear decision boundary in the feature space, which can be valuable for scenarios where the relationship between features and the target variable is predominantly linear.

4.2 Decision Tree Classifier

Decision trees are powerful tools for capturing complex relationships within the data. They segment the feature space based on the values of different attributes, making them capable of handling both numerical and categorical data. Decision Trees are particularly advantageous when dealing with non-linear relationships.

Key Characteristics:

Non-Linearity: Decision Trees can model non-linear relationships between features and the target variable, which can be instrumental in scenarios where the underlying patterns are intricate.

Interpretability: The decision-making process of a Decision Tree is easy to visualize and interpret, making it an excellent choice for scenarios where transparency is crucial.

4.3 Random Forest Classifier:

The Random Forest model is an ensemble learning method that combines multiple decision trees to improve accuracy and reduce overfitting. By aggregating the predictions of multiple trees, it tends to provide robust and reliable results.

Key Characteristics:

Ensemble Learning: Random Forest leverages the power of multiple decision trees, each trained on a different subset of the data, to enhance predictive accuracy. This makes it highly effective in handling complex data patterns.

Reduced Overfitting: Through the combination of multiple trees, Random Forest mitigates the risk of overfitting, resulting in a more generalized and reliable model.

Chapter 5: Implementation and Results

5.1 Model Training and Evaluation

Rigorous Training and Testing

To ensure unbiased evaluation, the dataset was divided into training and testing sets. The training set, comprising 80% of the data, was used to train each of the selected models. Subsequently, the testing set, which constituted the remaining 20%, was employed to assess their performance.

Model Evaluation

Model evaluation is a critical step in assessing the performance of machine learning models. In our project, we will employ a suite of classification metrics to comprehensively evaluate the effectiveness of each model. These metrics include:

Accuracy: The proportion of correctly classified instances out of the total instances.

Precision: The ratio of true positive predictions to the sum of true positives and false positives. It indicates the model's ability to accurately predict positive instances.

Recall: The ratio of true positive predictions to the sum of true positives and false negatives. It represents the model's ability to correctly identify all positive instances.

F1-Score: The harmonic mean of precision and recall, providing a balanced measure of a model's performance.

Confusion Matrix: A tabular representation of the model's predictions versus the actual values, providing insights into false positives, false negatives, true positives, and true negatives.

5.2 Model Comparison and Selection

Performance Metrics

Classification Report: The classification report is a comprehensive evaluation tool that provides detailed metrics for each class. It includes precision, recall, and F1-score, which offer insights into the model's ability to correctly predict Counter-Terrorist and Terrorist wins.

```
              precision    recall  f1-score   support

           0       0.76      0.74      0.75     12409
           1       0.74      0.77      0.75     12073

    accuracy                           0.75     24482
   macro avg       0.75      0.75      0.75     24482
weighted avg       0.75      0.75      0.75     24482
```

Figure 5.1: Logistic Regression Classification Report

```
              precision    recall  f1-score   support

           0       0.81      0.81      0.81     11961
           1       0.82      0.82      0.82     12521

    accuracy                           0.82     24482
   macro avg       0.82      0.82      0.82     24482
weighted avg       0.82      0.82      0.82     24482
```

Figure 5.2: Decision Tree Classification Report

```
              precision    recall  f1-score   support

           0       0.86      0.85      0.85     12077
           1       0.86      0.86      0.86     12405

    accuracy                           0.86     24482
   macro avg       0.86      0.86      0.86     24482
weighted avg       0.86      0.86      0.86     24482
```

Figure 5.3: Random Forest Classification Report

Evaluating Model Performance

Each model underwent rigorous evaluation using the classification report metrics. This detailed assessment allowed for a thorough comparison of their predictive capabilities, shedding light on their strengths and weaknesses.

Identifying the Best Model

Following meticulous evaluation, it was observed that the Random Forest Classifier demonstrated the highest accuracy and predictive power. Its ability to handle complex relationships within the data, coupled with its ensemble learning approach, made it the optimal choice for round winner prediction in CS:GO.

5.3 Model Performance

Assessing Model Effectiveness

The selected Random Forest Classifier demonstrated remarkable performance in predicting round winners in CS:GO. The classification report metrics provide detailed insights into the model's effectiveness:

Counter-Terrorist (CT) Class:

 - **Precision: 0.86**

 - **Recall: 0.85**

 - **F1-Score: 0.85**

Terrorist (T) Class:

 - **Precision: 0.86**

 - **Recall: 0.86**

 - **F1-Score: 0.86**

The high precision, recall, and F1-scores indicate that the model excels in correctly identifying both Counter-Terrorist and Terrorist round winners.

5.4 Model Interpretation

Interpreting Feature Importance

The Random Forest model also provides valuable insights into feature importance. By analyzing the importance scores of each attribute, we can discern which factors have the most significant impact on predicting round outcomes.

Contextualizing Results

The success of the Random Forest Classifier in predicting round winners underscores its robustness and suitability for CS:GO gameplay data. The model's ability to capture complex relationships and its ensemble learning approach contribute to its high predictive power. Additionally, the feature importance analysis sheds light on the critical attributes that influence round outcomes. This information can be leveraged to enhance gameplay strategies and decision-making.

Recognizing Limitations

While the Random Forest Classifier performed exceptionally well, it's essential to acknowledge potential limitations. These may include:

Sensitivity to Hyperparameters: The model's performance could be further fine-tuned through hyperparameter optimization.

Data Scalability: The model's performance on larger datasets or in different CS:GO contexts should be explored.

Future Directions

To enhance the predictive capabilities, future research could focus on:

Incorporating Additional Data: Including supplementary data sources or refining existing features for improved accuracy.

Exploring Advanced Techniques: Investigating more complex models or ensemble methods to achieve even higher predictive performance.

Chapter 6: Conclusion

6.1 Conclusion

Summarizing the Project

This project focused on predicting round winners in CS:GO using a dataset sourced from Kaggle. Through meticulous data preprocessing, feature engineering, and model evaluation, we successfully developed a Random Forest Classifier that demonstrated exceptional performance in round outcome prediction. The high precision, recall, and F1-scores attained by the model highlight its robustness and effectiveness in the context of CS:GO gameplay data.

Noteworthy Insights

1. Random Forest Superiority: Among the models considered, the Random Forest Classifier proved to be the most suitable for predicting round winners in CS:GO, showcasing its versatility and effectiveness.

2. Feature Importance: The analysis of feature importance revealed critical attributes influencing round outcomes. This knowledge can be leveraged to enhance gameplay strategies.

6.2 Future Scope

The project's success opens up various avenues for future research and application:

Real-time Predictions: Explore the feasibility of implementing the predictive model in real-time CS:GO matches. This would require a streamlined data pipeline and integration with live gameplay.

Additional Data Sources: Incorporate supplementary data sources, such as player statistics, match history, or team compositions, to further enrich the dataset and improve predictive accuracy.

Advanced Modeling Techniques: Investigate more sophisticated machine learning models or ensemble methods to ascertain if even higher predictive performance can be achieved.

Dynamic Hyperparameter Tuning: Implement dynamic hyperparameter tuning strategies to adapt the model to changing gameplay trends and meta-strategies in CS:GO.

Cross-Game Applications: Explore the applicability of the developed methodology to other esports titles with similar gameplay dynamics, potentially contributing to a broader understanding of competitive gaming.

Impact and Future Prospects

This project exemplifies the potential of machine learning in esports analytics. The successful prediction of round winners in CS:GO opens doors to further research and applications in the field of competitive gaming.

The insights gained from this project can be harnessed to enhance gameplay strategies, optimize team compositions, and ultimately improve performance in professional CS:GO matches.

In conclusion, this project has demonstrated the potential and effectiveness of machine learning in the domain of esports analytics, specifically in the context of CS:GO round winner prediction. The insights gained from this endeavor have implications not only for professional esports teams but also for the broader gaming community and enthusiasts seeking to enhance their understanding and strategic approach to the game. We look forward to the continued evolution of this field and the exciting opportunities it holds.