Predicting the winner in CS:GO matches

(STATS/CSE 780 course project)

Yan Min
Department of Electrical and Computer Engineering
McMaster University

2023-12-05

Outline

- Motivation
- EDA
- Methods
  - Random Forest
  - DNN
- Results
- Discussion

Motivation

- Try to predict the winner in a match
- Growing interest in eSports.
- Develop an advanced predictive Model basd on past data

Data

- The dataset("CS:GO Round Winner Classification" 2020) was originally published by Skybox as part of their CS:GO AI Challenge, running from Spring to Fall 2020.
- 122410 entries and 97 colums
  - dtypes: bool(1), float64(94), object(2)
  - $n > p$
- Results of EDA
  - Winnner: CT(60004), T(62406)
  - No null values
  - Scaling and PCA
- Feature selection
- Spliting
  - Train/test datasets(test_size=0.3)

Data

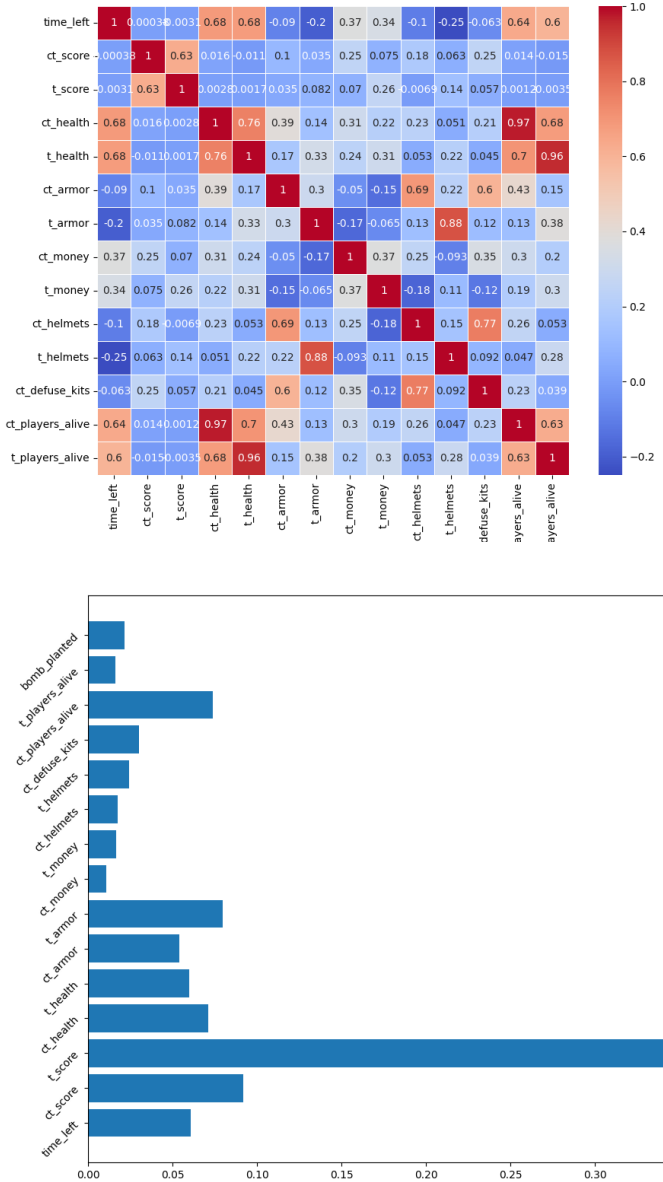

Figure 1: Correlation Matrix and Feature Importance

Methods

- Random forest
  - Classification(CT/T)
  - PCA
- DNN
  - Suitable for large datasets
  - `n_layers = 4`
  - `n_nodes = 300`
  - `epochs = 28/50`

Results

- Random forest
  - Accuracy: 0.753
- DNN
  - Accuracy: 0.743

Discussion

- The two accuracies are close.
- Features on health are more important.
- We can use this to predict the winner in today's matches.
- Game updates can significantly impact performance.
- Random forest:
  - Overfitting? `0.781`
  - GridSearch? It takes resources.
  - more interpretable
- DNN
  - Compared to RM, more suitable for large datasets
  - less interpretable

Thank You!

References

"CS:GO Round Winner Classification." 2020. https://www.kaggle.com/datasets/christianlillelund/csgo-round-winner-classification.