League of Legends SoloQ matches at 15 minutes 2024

- Group_A2_20
- João Coelho, up202004846
- João Mota, up202108677
- Pedro Landolt, up202103337

Specification

- Dataset contains data about the first 15 minutes of gameplay for over 24 thousand solo queue matches;
- There are 14 features for red, and 14 features for the blue team;
- Main purpose of the dataset is to help train models for predicting the winner based on how the first 15 minutes of the match played out;

Related Work

- There are a few notebooks that can serve as reference:
  - https://www.kaggle.com/code/kirklin/game-winner-prediction-best-76-9-w-eda-finetune
  - https://www.kaggle.com/code/fariborz2023/svm-with-linear-kernel-achieving-75-accuracy
  - https://www.kaggle.com/code/jordandanh/lol-game-predictive-model

Description of Tools and Algorithms

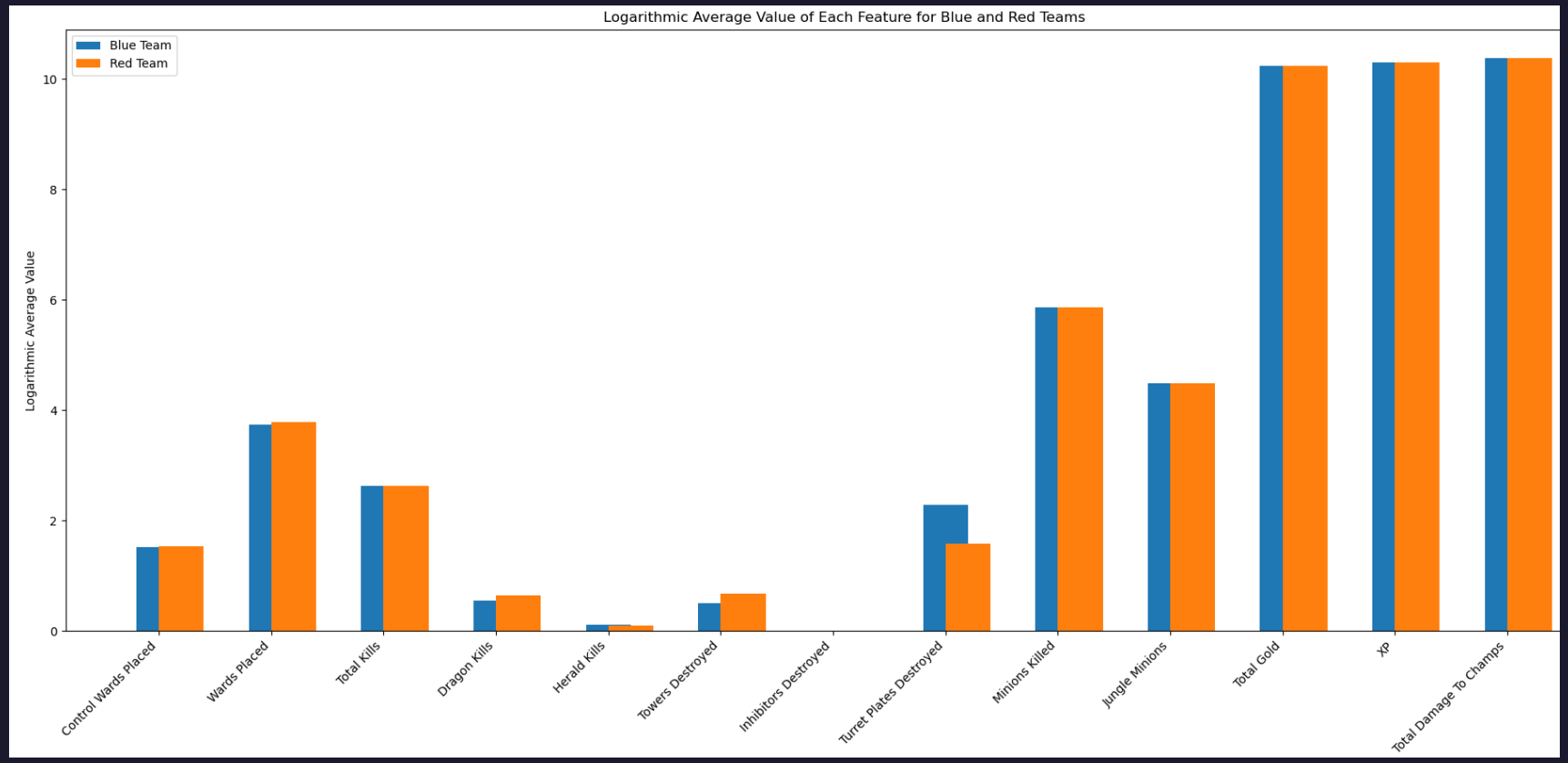
- SciKit Learn – Machine Learning library that will allow us to train the model;
- Pandas – read csv;
- Matplotlib – plot stats;
- NumPy – useful for operations;
- Algorithms:
  - 7 different types;

Implementation progress

- Can read files and extract data;
- Can receive user input to predict outcome;

Data Preprocessing

- Implemented functions to check for missing values, average values and outliers: no missing values and outliers;
- Scaled all features using a StandardScaler;
- Created models using PCA, reducing features to 17 (90% of variance);

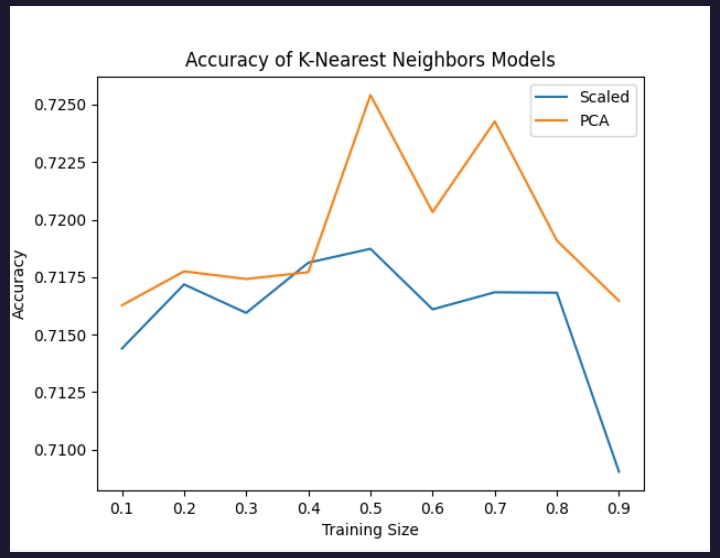


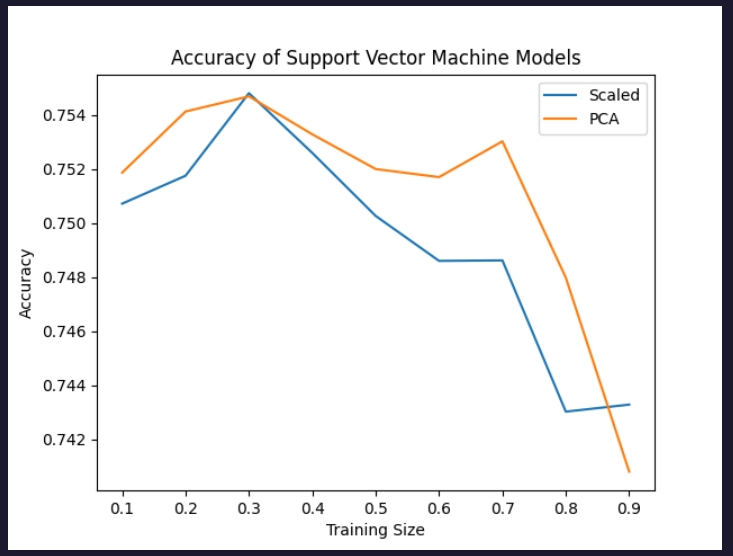
Created models

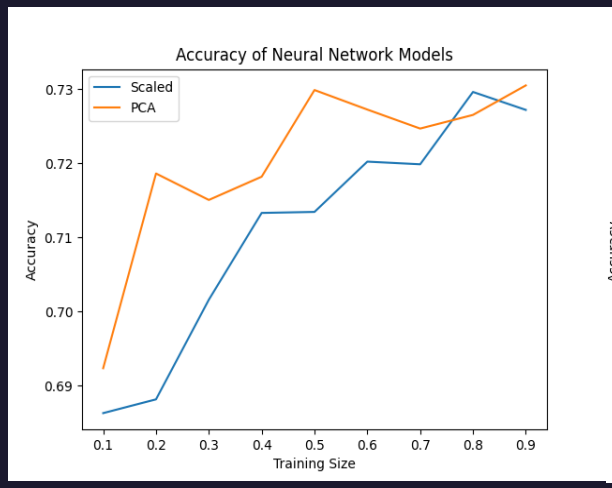
- 7 different models, one of each of the different models:
  - Decision Tree;
  - K-NN;
  - SVM;
  - Naïve Bayes;
  - Random Forest;
  - Logistic Regression;
  - Neural Network;

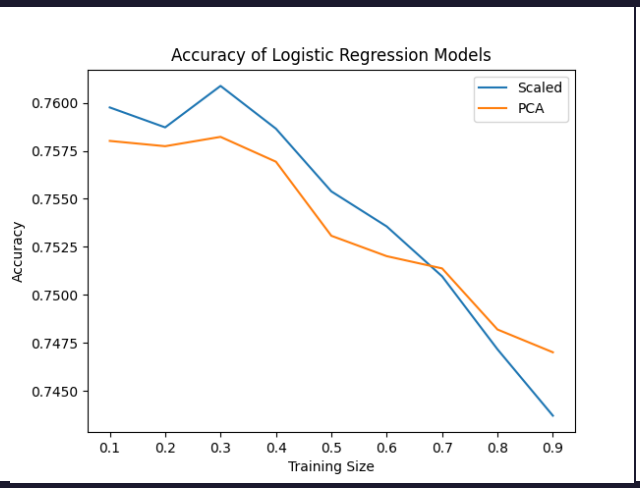
Stats


Accuracy of Decision Tree Models
Scaled
PCA
Accuracy
Training Size

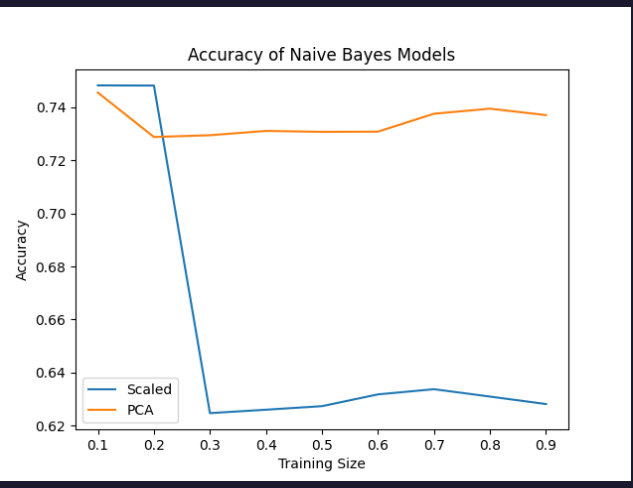

Accuracy of K-Nearest Neighbors Models
Scaled
PCA
Accuracy
Training Size


Accuracy of Support Vector Machine Models
Scaled
PCA
Accuracy
Training Size

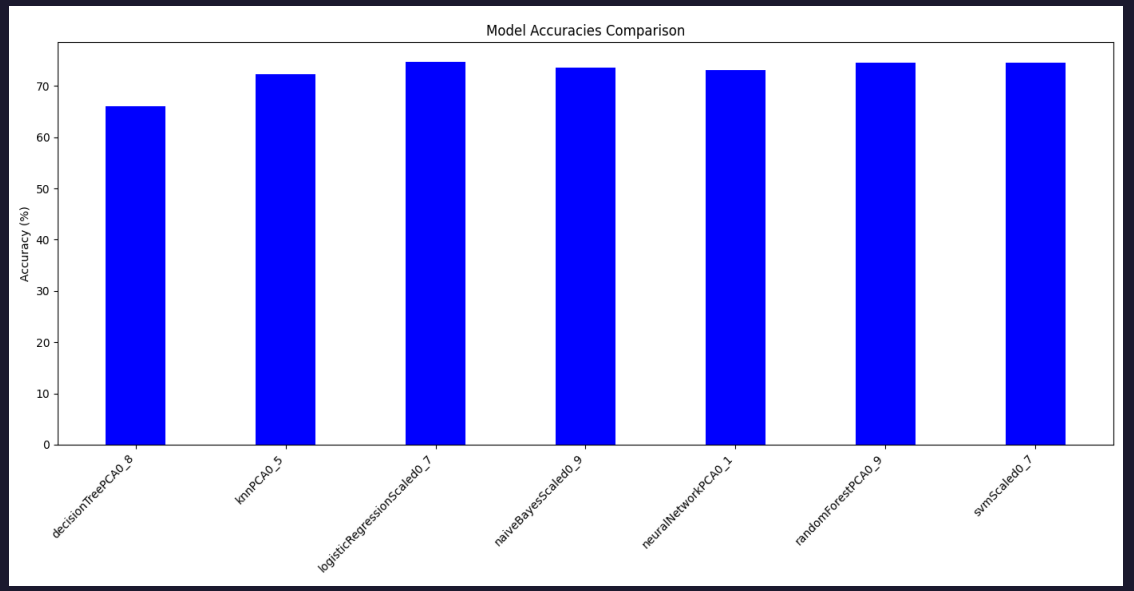

Accuracy of Neural Network Models
Scaled
PCA
Accuracy
Training Size

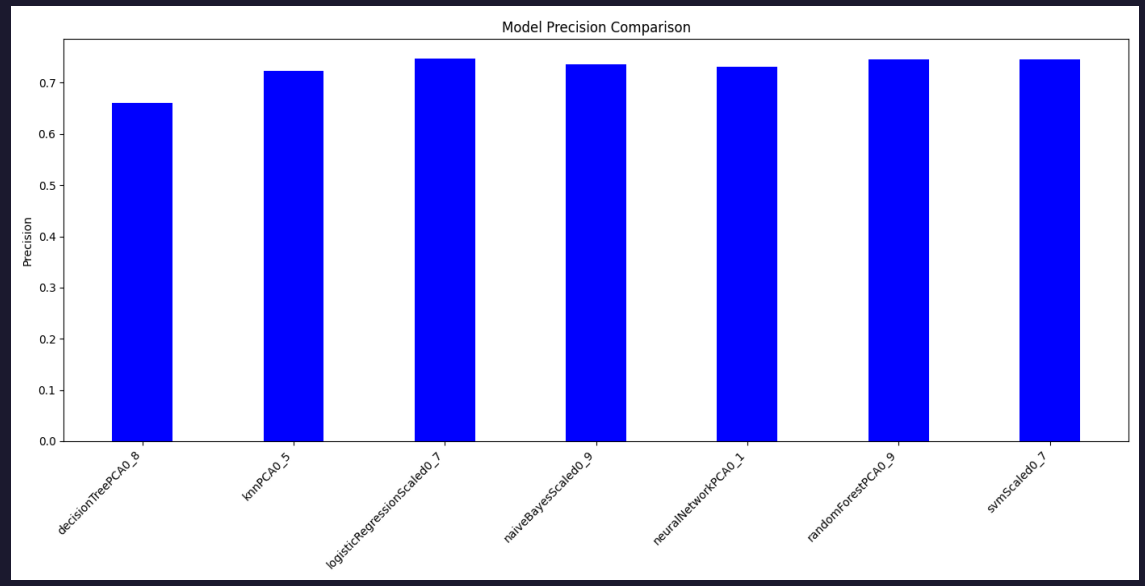

Accuracy of Logistic Regression Models
Scaled
PCA
Accuracy
Training Size

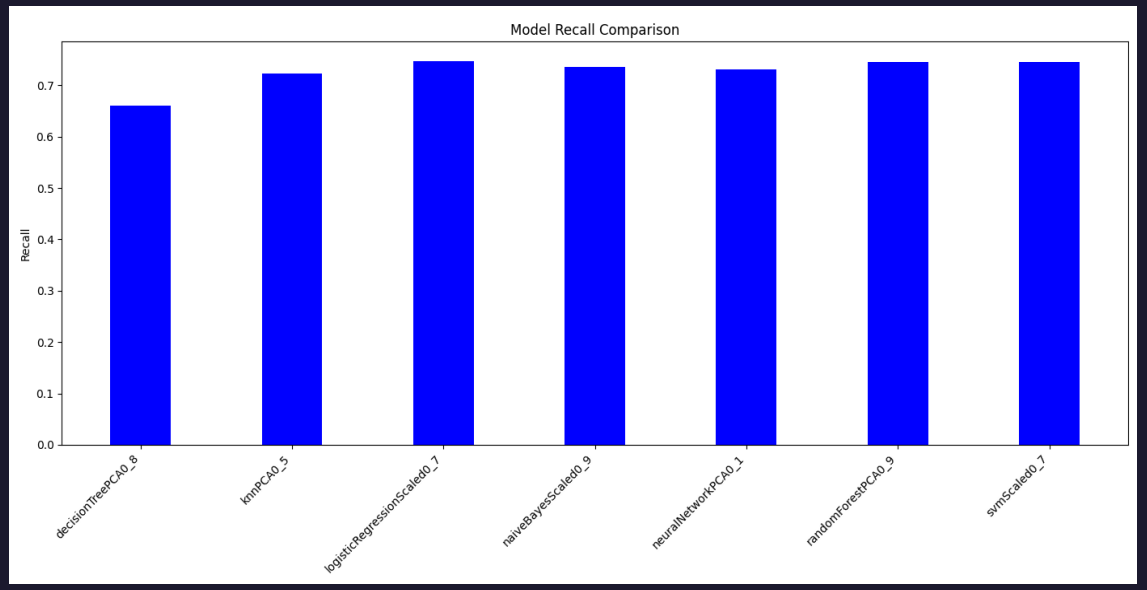

Accuracy of Random Forest Models
Scaled
PCA
Accuracy
Training Size

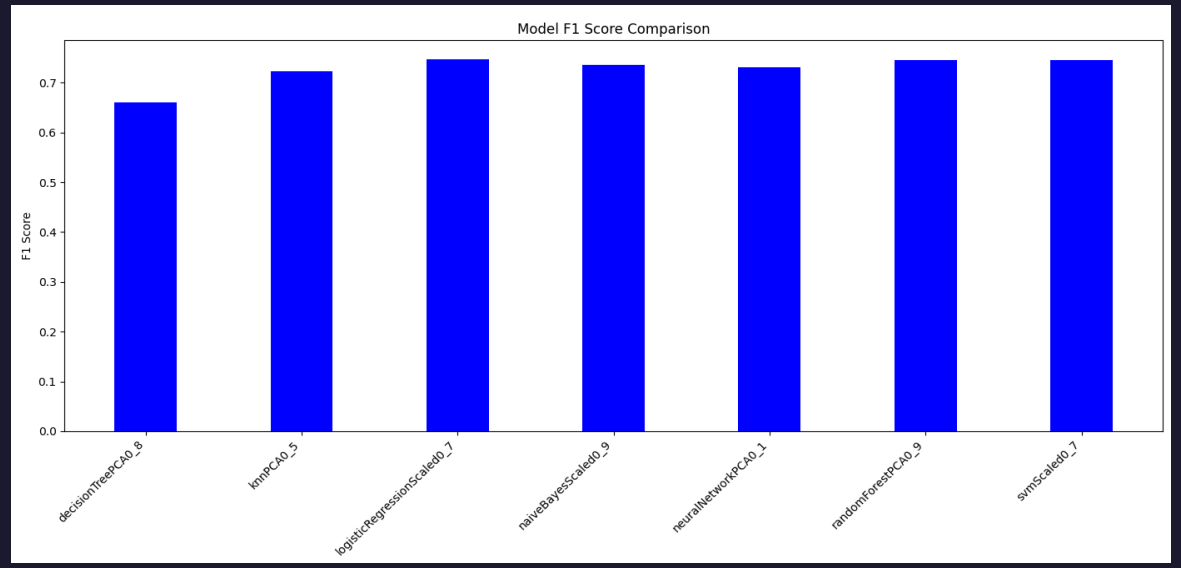

Accuracy of Naive Bayes Models
Scaled
PCA
Accuracy
Training Size

Stats – best Models comparison









Confusion Matrices of best Models

```
Model Name: decisionTreePCA0_8.sav
TP: 812, FP: 413
FN: 411, TN: 787

Model Name: knnPCA0_5.sav
TP: 881, FP: 344
FN: 326, TN: 872

Model Name: logisticRegressionScaled0_7.sav
TP: 916, FP: 309
FN: 302, TN: 896

Model Name: naiveBayesScaled0_9.sav
TP: 909, FP: 316
FN: 324, TN: 874

Model Name: neuralNetworkPCA0_1.sav
TP: 907, FP: 318
FN: 335, TN: 863

Model Name: randomForestPCA0_9.sav
TP: 919, FP: 306
FN: 309, TN: 889

Model Name: svmScaled0_7.sav
TP: 909, FP: 316
FN: 302, TN: 896
```