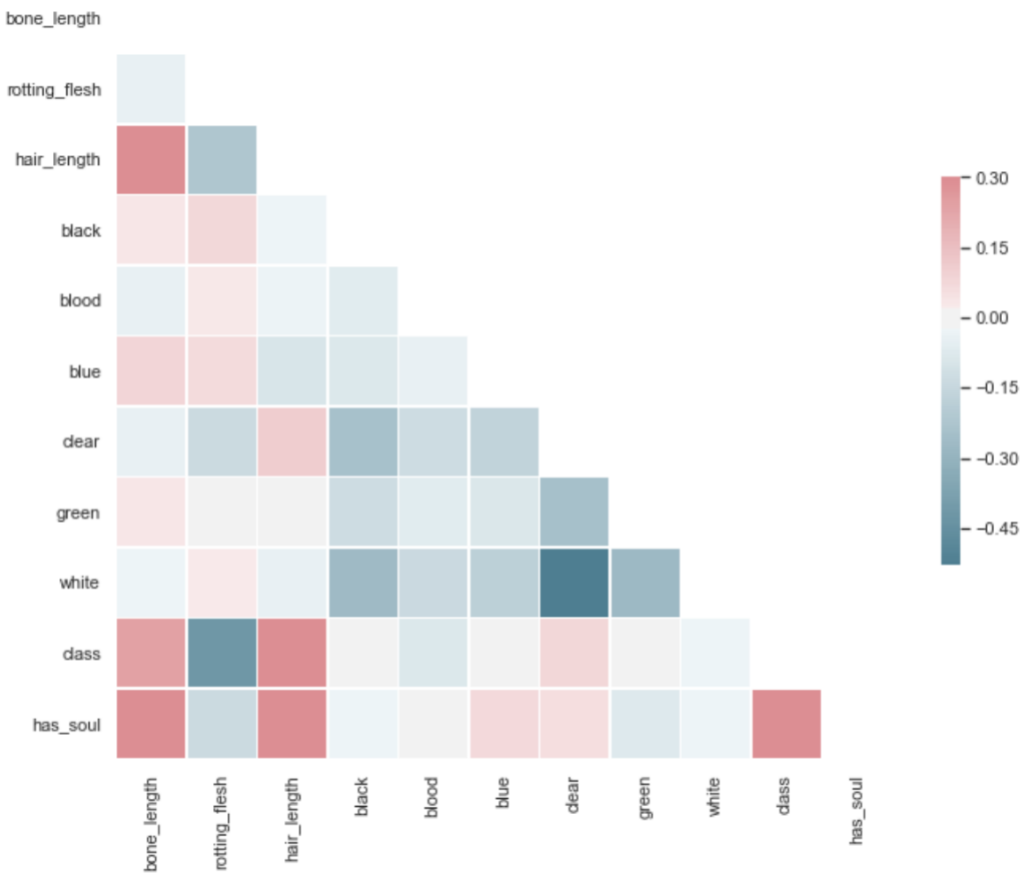
KAGGLE CASES, Summer 2019
Project – 3

Name : Dhivya Swaminathan
UID : 2000434729

There were three steps involved in this project:

1. Data Analysis and Processing
2. Modeling
3. Prediction

Data Analysis and Processing:
The data was label and one hot encoded and processed using sklearn package. Post this, a correlation matrix was plotted. A min max scaler method was employed to scale the data post the processing.



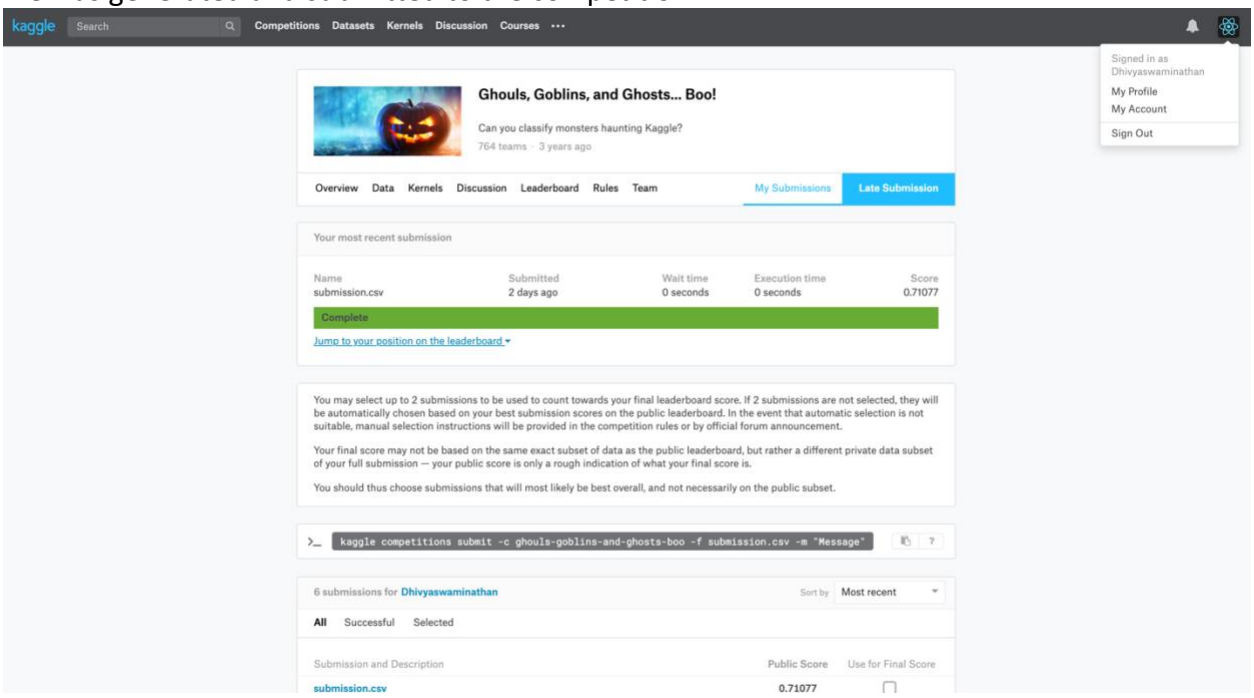
Modeling:

Various baseline models were established, and their accuracies were compared to choose the best model. Adaboost seemed to perform best and its parameters were tuned which resulted in model performance improvement. Finally, a voting classifier was established with all these models and it was used as the final model.

Model	Accuracy
KNN	0.733
Logistic Regression	0.7067
Adaboost	0.7333
Random Forest	0.74
Gradient Boosting	0.6533
Voting Classifier	0.7467

Prediction:

With the voting classifier as the final model, test set predictions were made and the submission file was generated and submitted to the competition.



The model accuracy was 0.711 for the test set.

Reference:

1. https://www.kaggle.com/samratp/machine-learning-with-ghouls-goblins-and-ghosts