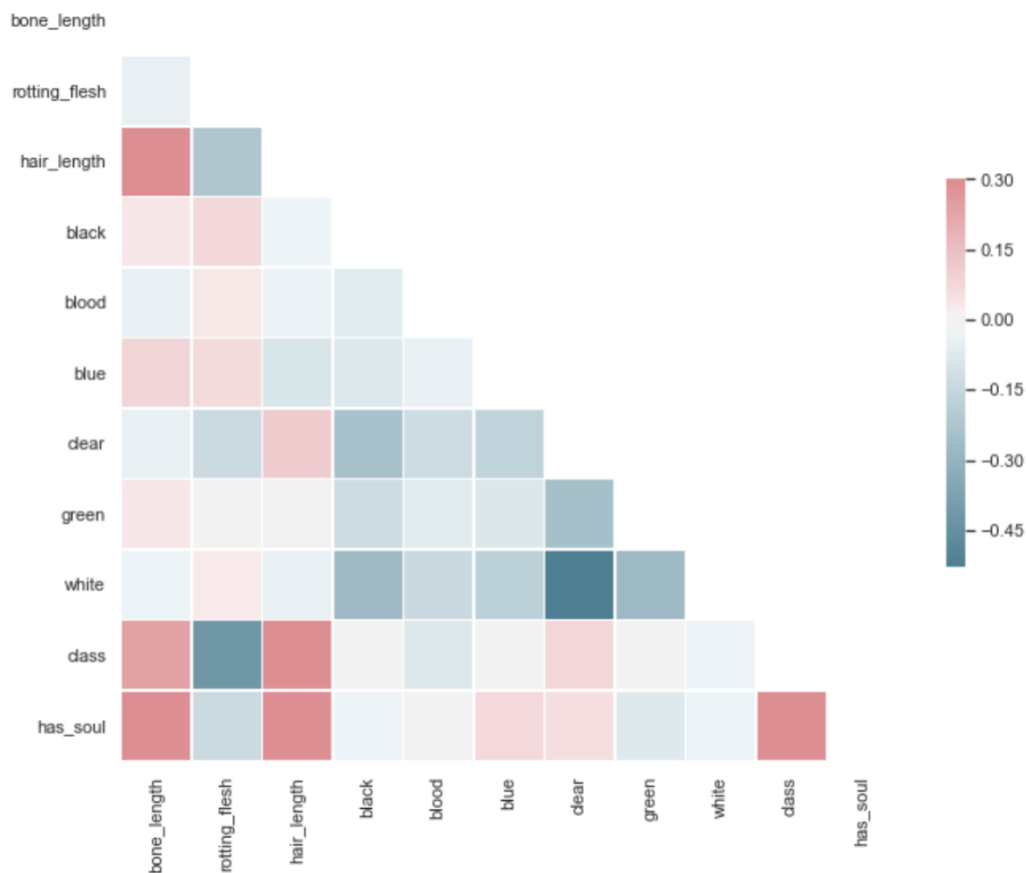


Name : Dhivya Swaminathan  
UID : 2000434729

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

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 screenshot shows the Kaggle interface for the competition "Ghouls, Goblins, and Ghosts... Boo!". The page includes a header with the Kaggle logo and navigation links. A user is signed in as Dhivyaswaminathan. The competition details show it was created 3 years ago by 764 teams. The submission status is "Complete" with a score of 0.71077. The submission file is named "submission.csv". The page also displays instructions for selecting submissions for the final score and a terminal command to submit the file.

Name	Submitted	Wait time	Execution time	Score
submission.csv	2 days ago	0 seconds	0 seconds	0.71077

Complete

[Jump to your position on the leaderboard](#)

You may select up to 2 submissions to be used to count towards your final leaderboard score. If 2 submissions are not selected, they will be automatically chosen based on your best submission scores on the public leaderboard. In the event that automatic selection is not suitable, manual selection instructions will be provided in the competition rules or by official forum announcement.

Your final score may not be based on the same exact subset of data as the public leaderboard, but rather a different private data subset of your full submission — your public score is only a rough indication of what your final score is.

You should thus choose submissions that will most likely be best overall, and not necessarily on the public subset.

```
kaggle competitions submit -c ghouls-goblins-and-ghosts-boo -f submission.csv -m "Message"
```

6 submissions for Dhivyaswaminathan

Sort by: Most recent

Submission and Description	Public Score	Use for Final Score
submission.csv	0.71077	<input type="checkbox"/>

The model accuracy was 0.711 for the test set.

## Reference:

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