College Football FBS Classifier

Saehej Kang (ssk2353)
Tyler Stubbs (tjs986)

Introduction:

College football is broken down into a few divisions with FBS being the most popular and most competitive division (this is the division that UT participates in). College Football's FBS comprises 134 teams spread out among 10 conferences.

Of the 134 teams, approximately 84 of those teams are eligible to play in a bowl game at the conclusion of the season (typically all teams with 6 wins or more will play in a bowl game plus some 5 win teams). From the teams that do qualify to play in a bowl game, the top 12 are selected to play in a playoff tournament where the winner is crowned the National Champion.

We want to take metrics from the college football landscape and see if we can use a k-NN algorithm based on those metrics to classify teams into one of three categories:

1. Playoff eligible
2. Bowl eligible but not playoff eligible
3. Not eligible for bowl game nor playoffs

Data Acquisition: We found some older season's data on Kaggle relating to college football but nothing for this season. We would like to try and use this past season's data so part of the project would be writing a script to scrape the metrics we care about from ncaa.com. In the event we could not perform this task, we could resort to using some of the available data sets from previous seasons.This table below shows the available team stats we want to consider:

Total Offense	1	543.5
Rushing Offense	10	204.5
Passing Offense	3	338.9
Team Passing Efficiency	1	192.53
Scoring Offense	1	45.5
Total Defense	105	416.6
Rushing Defense	85	161.0
Passing Yards Allowed	115	255.6
Team Passing Efficiency Defense	101	143.58
Scoring Defense	T-78	28.0
Turnover Margin	T-23	0.46
3rd Down Conversion Pct	2	0.550
4th Down Conversion Pct	106	0.400
3rd Down Conversion Pct Defense	114	0.448
4th Down Conversion Pct Defense	T-31	0.433
Red Zone Offense	20	0.908
Red Zone Defense	109	0.891
Net Punting	37	40.19
Punt Returns	33	10.57
Kickoff Returns	83	19.00
First Downs Offense	5	336
First Downs Defense	T-103	274
Fewest Penalties Per Game	T-71	6.08
Fewest Penalty Yards Per Game	67	51.69
Time of Possession	94	28:51

Data Processing: Once we have data from the web, we would want to look into isolating the 5-7 key metrics that we think could be indicators on if a team had a good season or not. In this case, we are calling qualifying for a bowl game as a good season indicator. A great season is one where a team qualifies for the playoff.

We want to use something like the Lasso Regression model or Recursive Feature Elimination (RFE) to isolate the metrics that are the best indicators for a good/great team.

We will have data metrics for 134 teams. We would like to use 100 teams for the training set and then see how accurately we can classify the remaining 34. Ideally, we would prefer to have a balanced sampling of training data that encompasses very bad, bad, good, and very good teams.

Model Training/Evaluation: We will implement a k-NN algorithm to train the model on the training data on the key metrics we derived from above. We then would apply the model to the test data and compare the labels.

Special Note:
Last season, the NCAA College Football Playoff was only 4 teams, not 12. We are going to use ESPN's final playoff rankings and take the top 12 teams from those rankings and consider them the teams who would have been eligible for a playoff spot had the playoffs been 12 teams.