

Amani Akkoub (ctf3un), Charli Ahsby (arv2vp), Jonah Cicatko (zvh3zz), Anne Kumashiro (yxt7ue), Jack Nickerson (rze7ud)
DS 3001 - Foundations of Machine Learning
10/04/2024

Data Wrangling/EDA

1. What is in your data?

The source we've elected to use for our project is an online API with data on collegiate level football (<https://collegefootballdata.com/>). The API contains data under a wide variety of different categories, including but not limited to: game and team results/records, historical rankings, various prediction metrics, even betting lines.

2. How will these data be useful for studying the phenomenon you're interested in?

We are interested in predicting the results of the college football 2024-25 season.

3. What are the challenges you've resolved or expect to face in using them?

- Wanted to predict college football playoffs
 - New format
 - Too many teams
 - Resolved by predicting ACC champion
- We will have to collect more data each week as new games are played
- We will need to figure out how to accurately represent out of conference games
- Accounting for injuries
- New teams Cal and Stanford