Classification Analysis of NFL Spreads

Gavin Jones, Metis Data Science 4Q20

You're doing what now?

- The best team in the NFL is playing the worst team. Who would you bet on?

- A **Spread** allows Vegas bookmakers (or people who create betting odds) to handicap a good team in an effort to even out a bettor's odds of winning or losing
    - The spread is a point total that is subtracted from the favorite, or added to an underdog, to give the underdog a head start

...Huh?

MATCHUP	SPREAD	MONEYLINE	TOTAL
New York Giants	+5.5 (-110)	+150	49.5 (-110) O
Dallas Cowboys	-5.5 (-110)	-200	49.5 (-110) U

Would you win if you bet on the Giants?

- Giants 3 to Cowboys 13 = No, you lose
- Giants 7 to Cowboys 10 = Yes, you win

More to know

- In this style of betting: $1 down = $2 back if you win, **minus a commission paid to your bookmaker** (this is called a ‘Vig’)

- Because of the commission, you need to win 52.5% of the time to be profitable

More to know

- In this style of betting: $1 down = $2 back if you win, **minus a commission paid to your bookmaker** (this is called a ‘Vig’)

- Because of the commission, you need to win 52.5% of the time to be profitable

QUESTION: CAN WE USE A CLASSIFIER TO ‘BEAT THE SPREAD’ AND STAY ABOVE A 52.5% SUCCESS RATE?

Feature Engineering

- We only care if a team is better than expectations
  - Each feature must be related in some way to the spread
  - The spread represents expectations

- Target ÷ Opponent = ‘how much better’
- ‘how much better’ ÷ spread = ‘how much better after a handicap’

- Spreads are given in terms of - / +, so we engineer the spread to adjust for sign change

Final Features Reflect Expectations

- Points per game
- Points let up by defense per game (‘Points Against’)
- “Spread Record” up through prior game, as a percentage
- Are they the favorite?

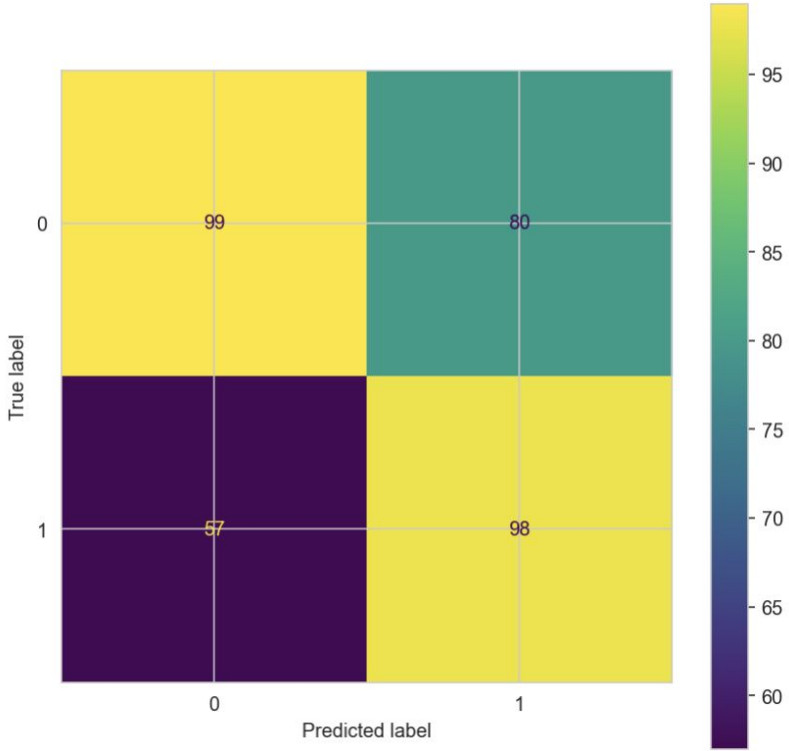
Models

Tried all, settled on XGBoost:

Model Breakdown

Model	Accuracy
XGBoost	**0.58982038**
SVC	0.53892215
KNN	0.52395209
Linear SVC	0.52095808
Logistic Regression	0.51197604
Naive Bayes (Gaus)	0.50898203

True Results

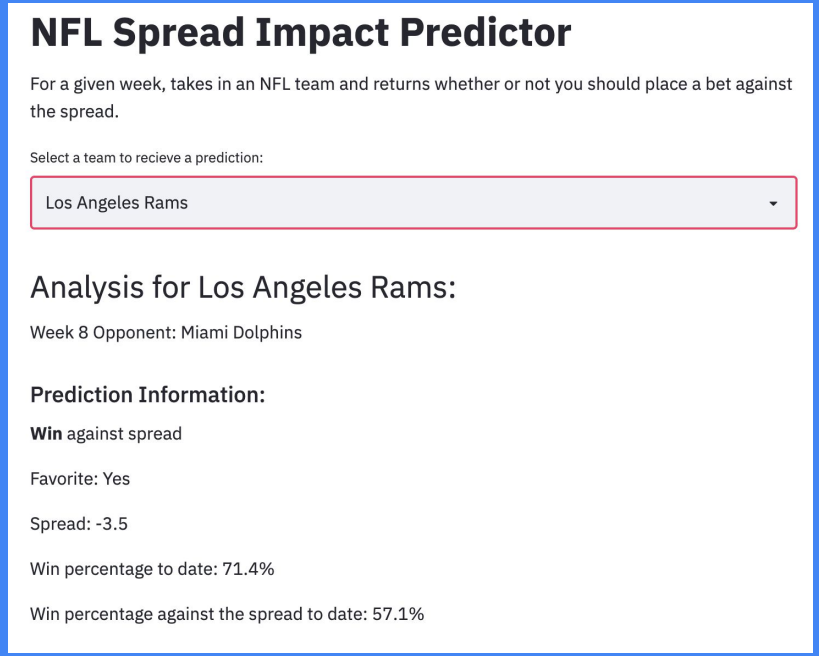


We care more about **Precision** than accuracy

Precision about 55% - above our 52.5% goal!

We can disregard any negatives as they do not drive any actions on the bettor's part

Continued concerns around variance, data size



Streamlit Demo

Continued exploration

- **Matchup Data:** what does the model think will happen to the opponent?

- **Previous Game Influence, etc:** Factors such as how the team performed in the previous game can over-influence spreads

- **Google Search Results:** potential as a basis for chatter/popularity of team, which could indicate required adjustments to force 50/50 betting split

Questions?