Analysis of NFL 4th Down Decisions

Frank Tucci

Intro

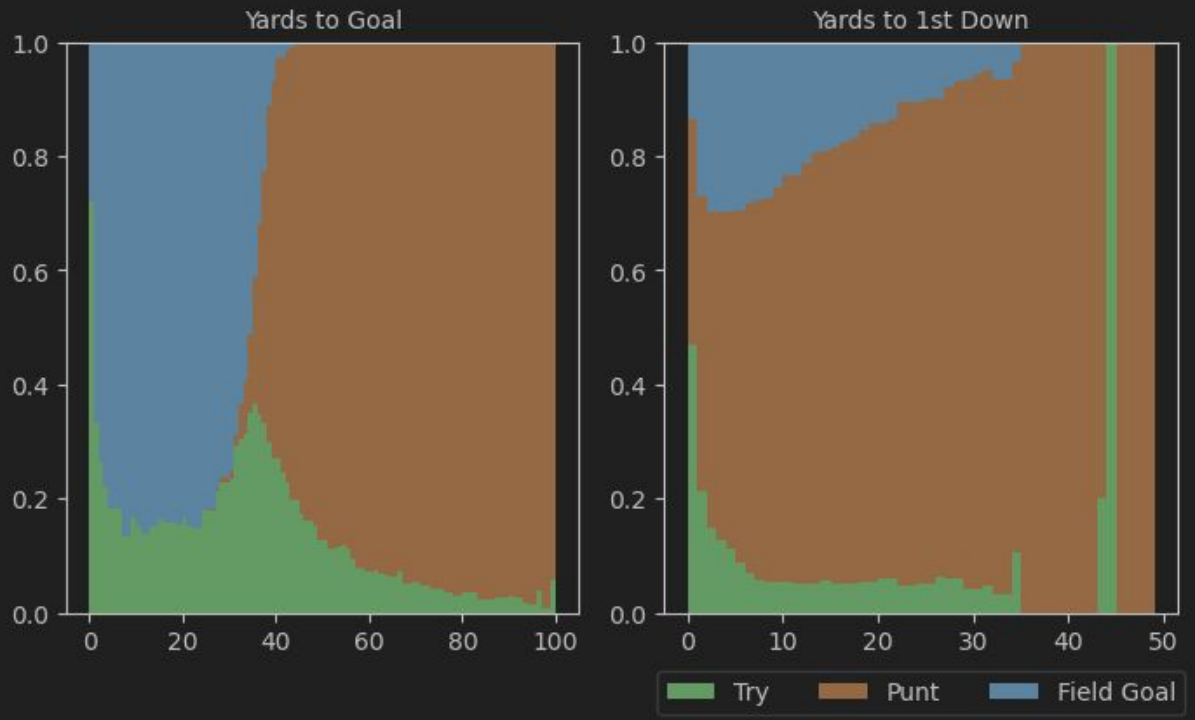
- Possibilities on 4th down:
  - Punt
  - Field Goal
  - Try for a 1st Down
- Purpose:
  - Useful for fans, reporters, sportscasters to know when a play is out of the ordinary.
  - A defensive team can know what to expect.

Data

- Every NFL play from 1999-2023 from nfl_data_py python module
  - 1.2 million play-by-play records
  - 100,000 4th downs
- Features collected:
  - Yards to Endzone
  - Yards to 1st Down
  - Home/Away Score
  - Seconds to the End of Game
  - Weather
- 80/20 Test/Train Validation

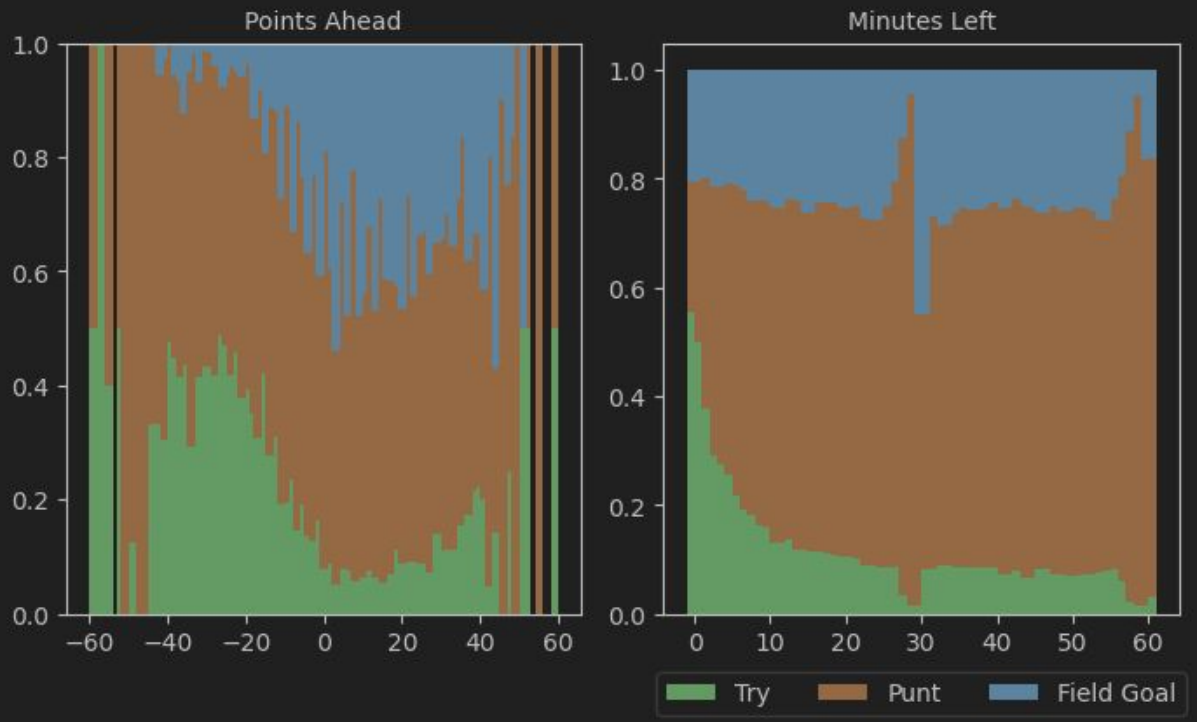
Observations – Yardage

- Field goals are popular inside 36 yards.
- After 36 yards, punting is preferred.
- If the team only has 1 yard until the 1st down, going for it is most common.



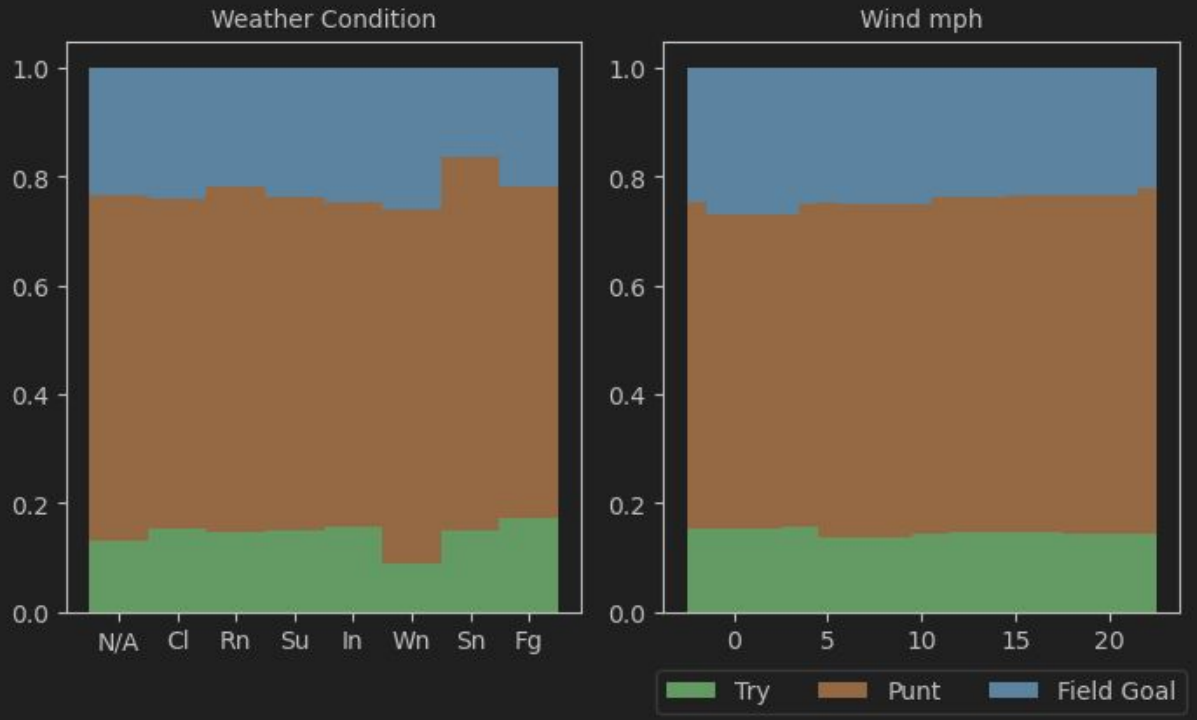
Observations – Point Spread and Time

- A losing team will likely try for 1st down (higher risk/reward).
- A winning team will instead go for a field goal.
- Trying for 1st is most common at the end of the game.



Observations – Weather

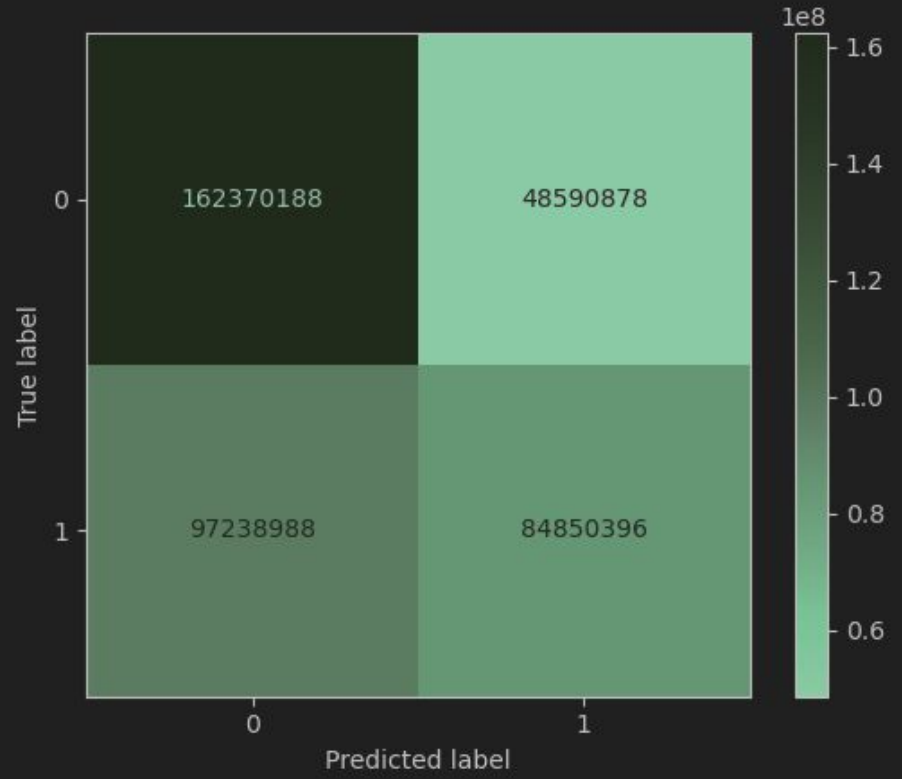
- Weather is not statistically significant and will be ignored.



K Means Clustering

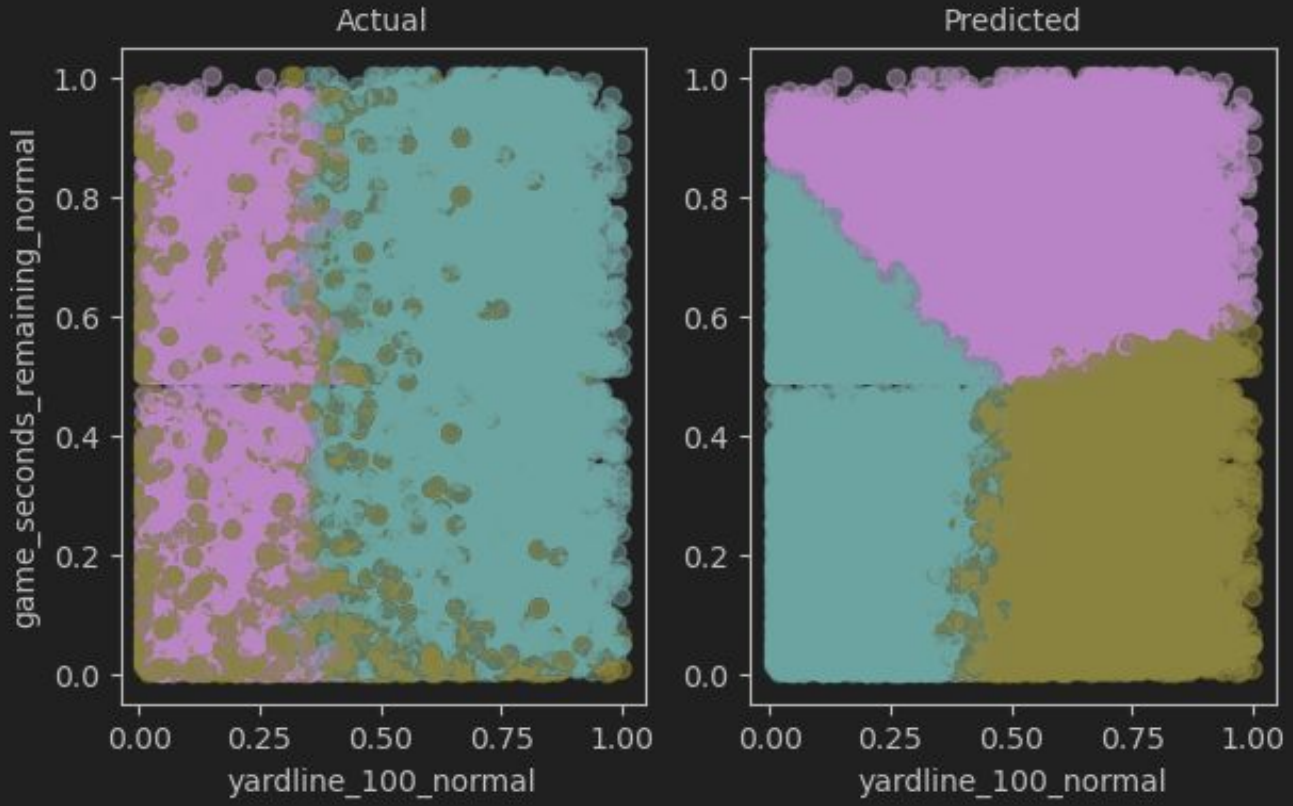
- Didn’t work very well
  - Target categories are different sizes and are too overlapping.

Precision	0.80
Recall	0.65
F1	0.72



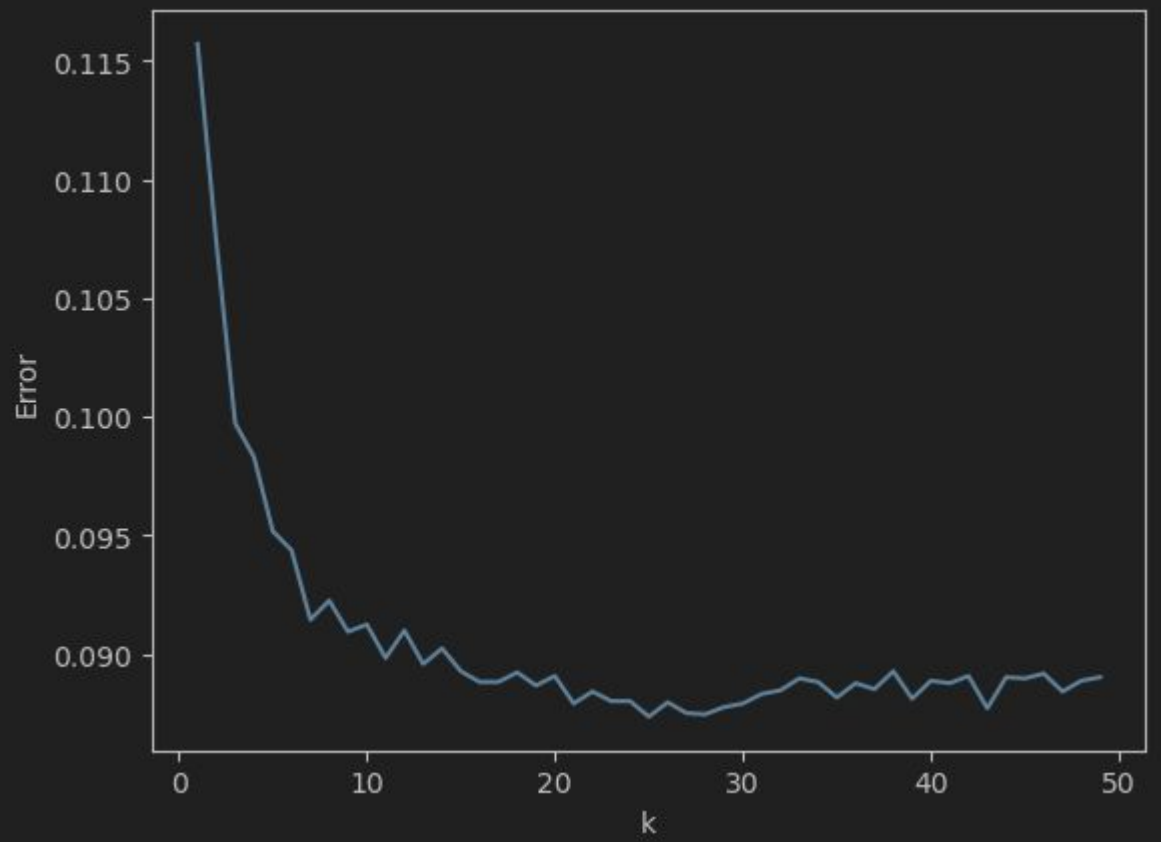
The Problem with K Means Clustering

- Field Goals: 4729
- Punts: 12355
- 1st Down Tries: 2742



K Nearest Neighbor

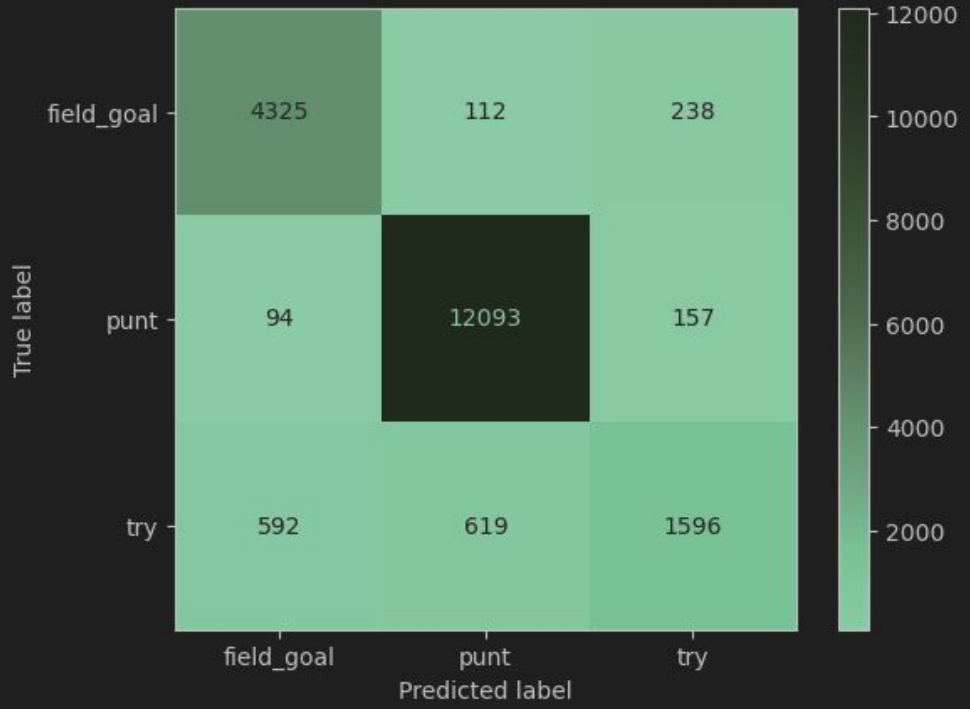
- K selection:
  - Found error term lowest around k = ~20



K Nearest Neighbor

- Worked Significantly Better than K Means

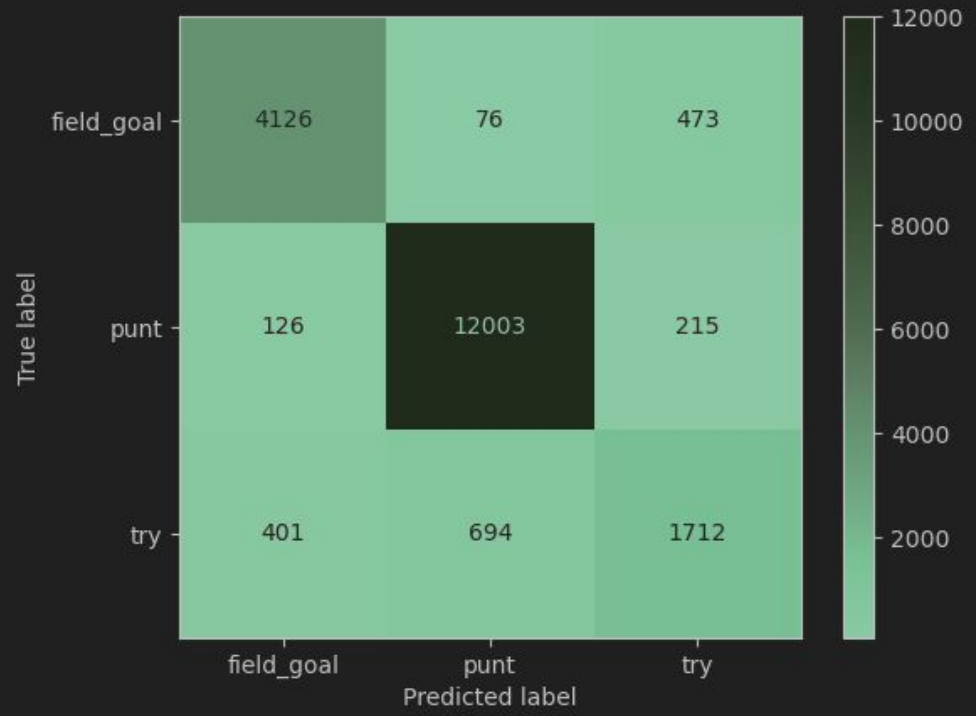
Accuracy	0.91
Precision	0.88
Recall	0.93
F1	0.90



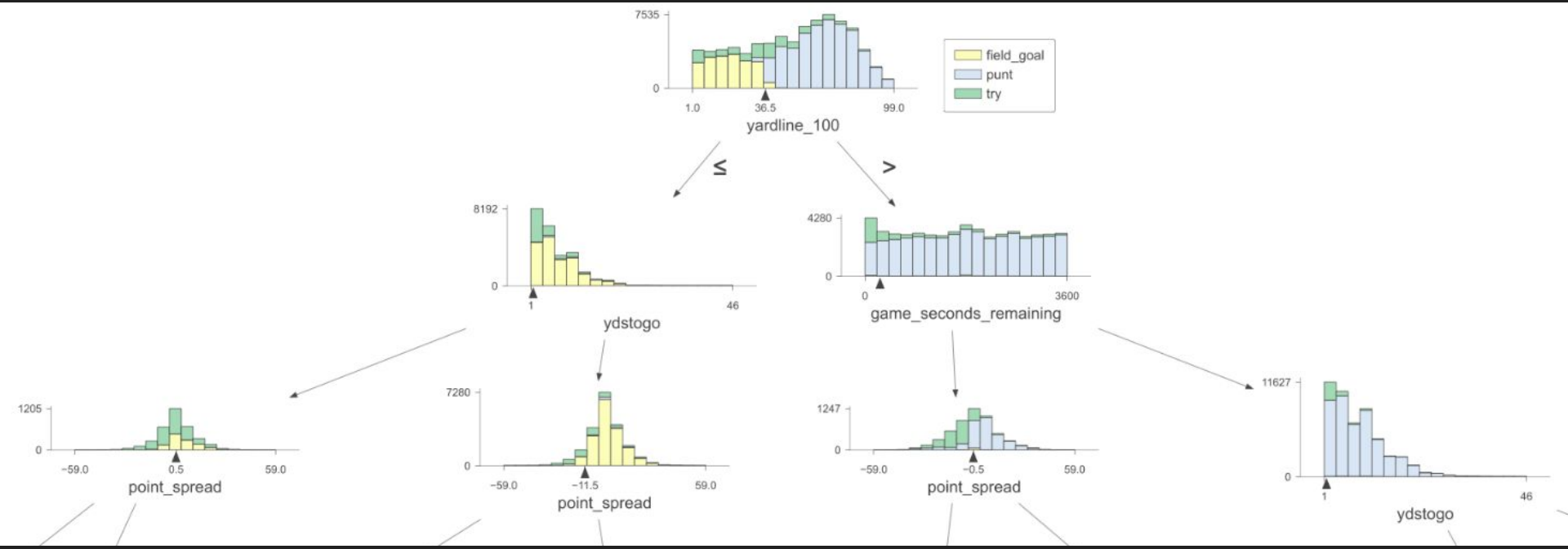
Decision Tree

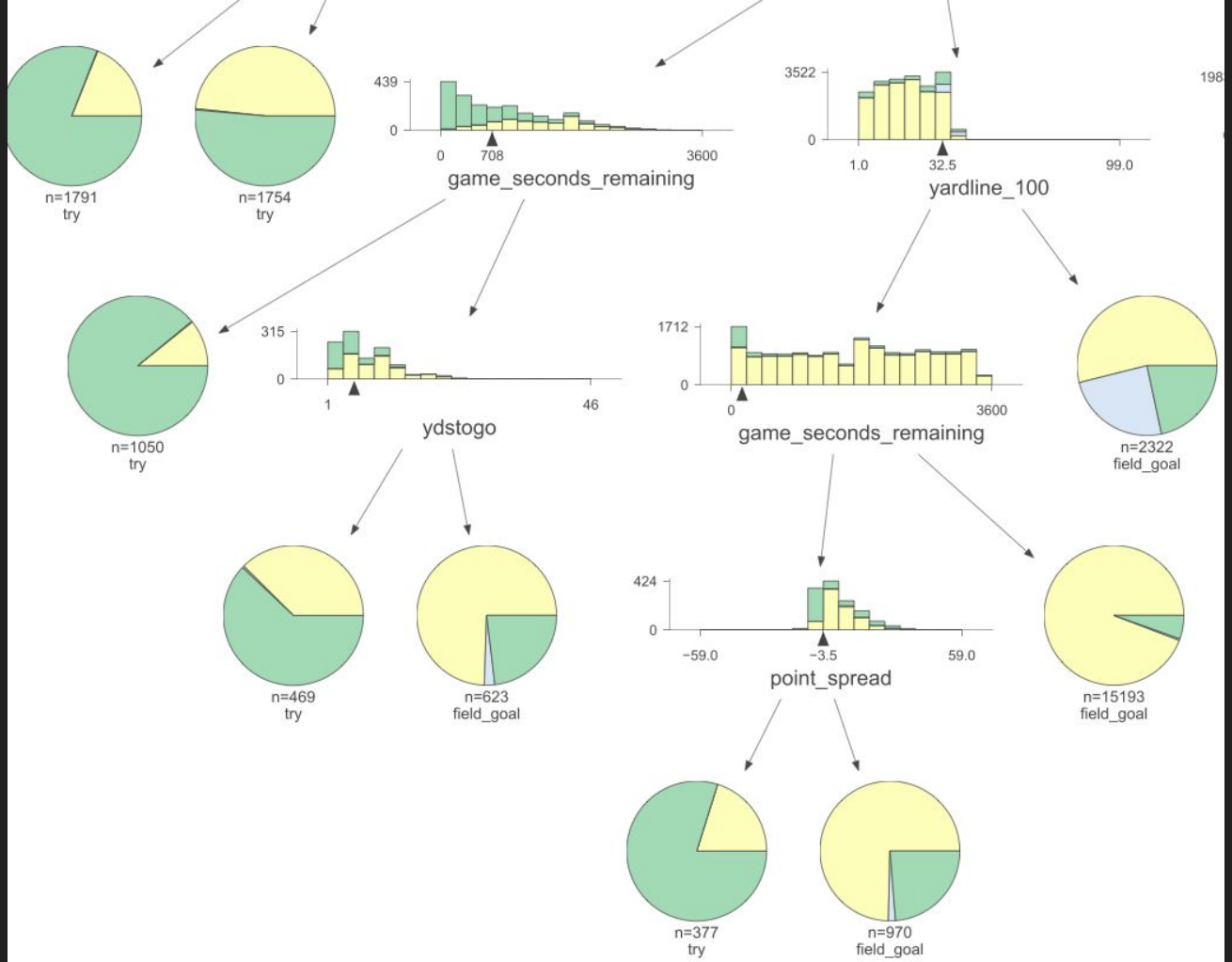
- Similar results as K Means
  - 25 leaf node maximum
  - Better at categorizing tries, the smallest target category

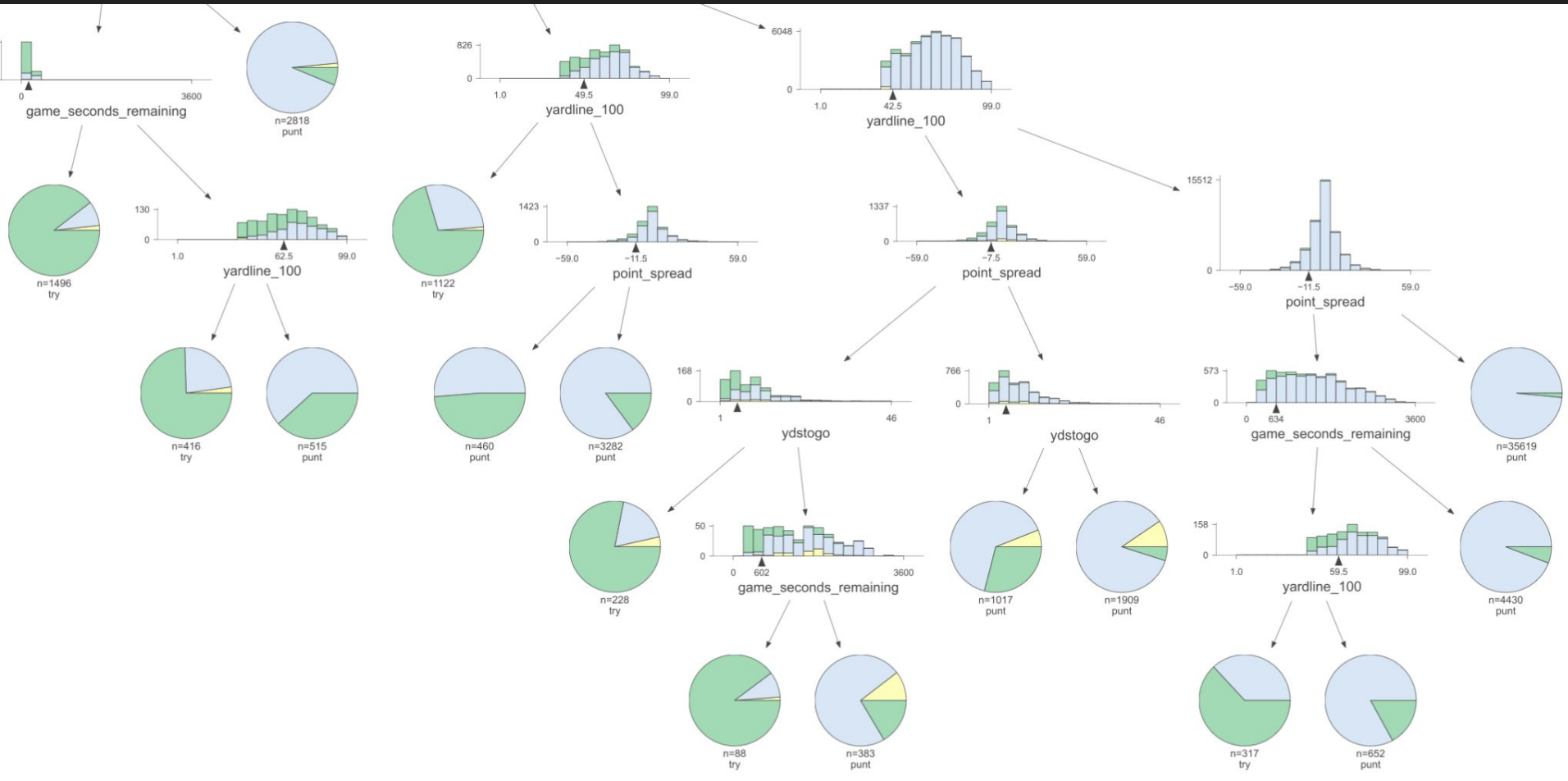
Accuracy	0.90
Precision	0.88
Recall	0.91
F1	0.90



Decision Tree Visualizer




n=1791
try
n=1754
try
439
0
0
708
3600
game_seconds_remaining
3522
0
1.0
32.5
99.0
yardline_100
n=1050
try
315
0
1
46
ydstogo
1712
0
0
3600
game_seconds_remaining
n=2322
field_goal
n=469
try
n=623
field_goal
424
0
−59.0
−3.5
59.0
point_spread
n=15193
field_goal
n=377
try
n=970
field_goal


0
3600
game_seconds_remaining
n=2818
punt
826
0
1.0
49.5
99.0
yardline_100
6048
0
1.0
42.5
99.0
yardline_100
n=1496
try
130
0
1.0
62.5
99.0
yardline_100
n=1122
try
1423
0
−59.0
−11.5
59.0
point_spread
1337
0
−59.0
−7.5
59.0
point_spread
15512
0
−59.0
−11.5
59.0
point_spread
n=416
try
n=515
punt
n=460
punt
n=3282
punt
168
0
1
46
ydstogo
766
0
1
46
ydstogo
573
0
0
634
3600
game_seconds_remaining
n=35619
punt
n=228
try
50
0
0
602
3600
game_seconds_remaining
n=1017
punt
n=1909
punt
158
0
1.0
59.5
99.0
yardline_100
n=4430
punt
n=88
try
n=383
punt
n=317
try
n=652
punt

Conclusions

- Decision tree visualization gave insight *beyond* looking at any feature individually.
- Distance to the end zone is the most influential factor:
  - Having minimal yards to 1st down, or being behind by more than a field goal near the end of the game are two factors that push teams to try for 1st down.
- Hard to get much higher than 90% accuracy:
  - More analysis could be done looking at individual player and/or coach playstyles.
  - There is still an element of randomness involved.

Works Cited

- *The search for the saddest punt in the world* - John Bois
- *NFL fourth-down decisions: The math behind the league's new aggressiveness* - Ben Baldwin
- *Analytics, Have Some Humility: A Statistical View of Fourth-down Decision Making* - Ryan Brill et al