Weather and NFL Betting

By: Alex, Addison, Henry, Melanie, Seth

Meet the data!

- Massive 54558 x 86 dataframe
  - One entry for every QB/WR/RB in every game across ~11 yrs
- Many variables of diverse nature
  - Player Stats
  - Weather
  - Betting Info
- Many metrics for weather, gambling odds a POI due to being produced by experts

Observations

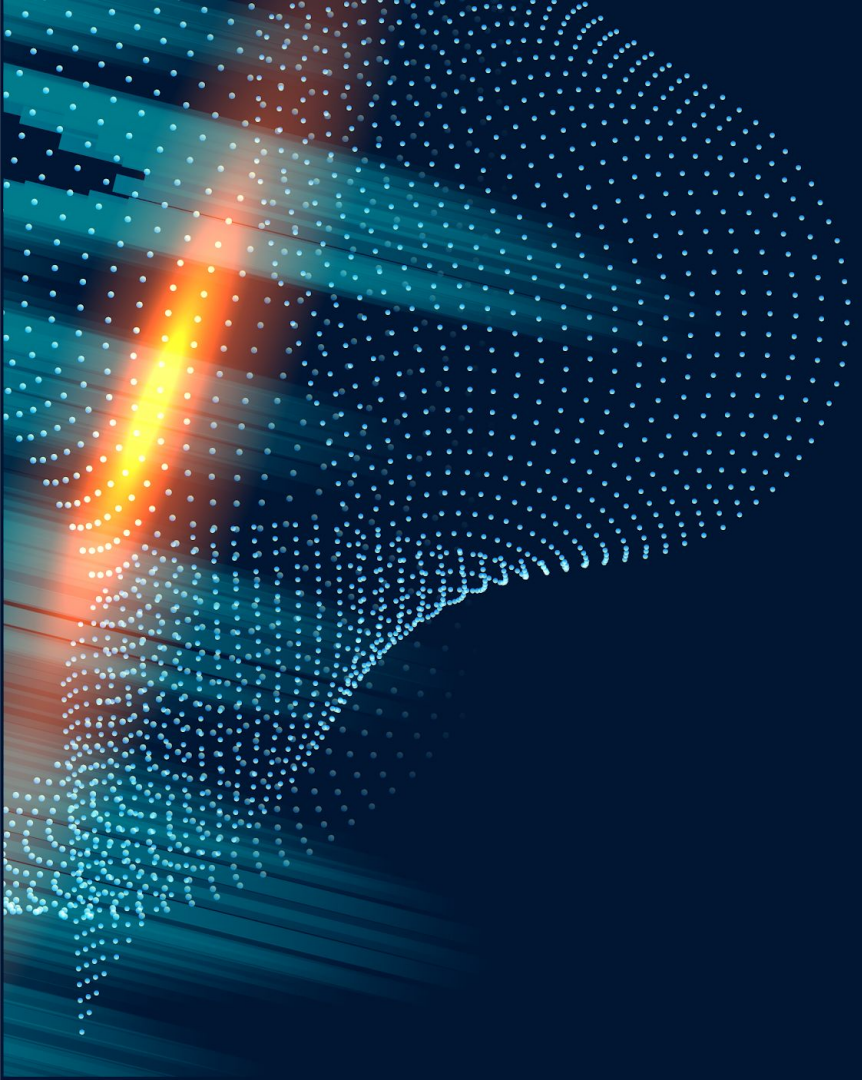
- There is a large amount of data at our disposal
- Gambling is a massive industry
- Games are played in even the harshest conditions
    - Some teams experience this more.
- A successful model could be profitable if accurate.

Questions

- Can weather predict if the spread will be covered?
- Can weather predict if the over/under will be met?
- Can a team's usual weather conditions predict game outcomes?

Feature Engineering

- Missing Data (mostly weather data)
  - Performed imputation for temperature, wind, humidity
  - Dropped NA (~25% of the data)
- Label encoding performed on categorical variables
- Added Variables
  - Winner ID
  - Average Temperature
  - Spread Covered (0, 1)
    - “Spread” = the expected final score difference between the two teams
    - “Covering the spread” = favored team wins by more points than the spread
  - Over/Under Met (0, 1)
    - “Over/under line” = sum of the predicted final point totals of the two teams
    - “Meeting the over/under” = sum of the actual final point total is greater than the over/under line



Question 1: Can weather predict if the spread will be covered?

“Spread” = the expected final score difference between the two teams

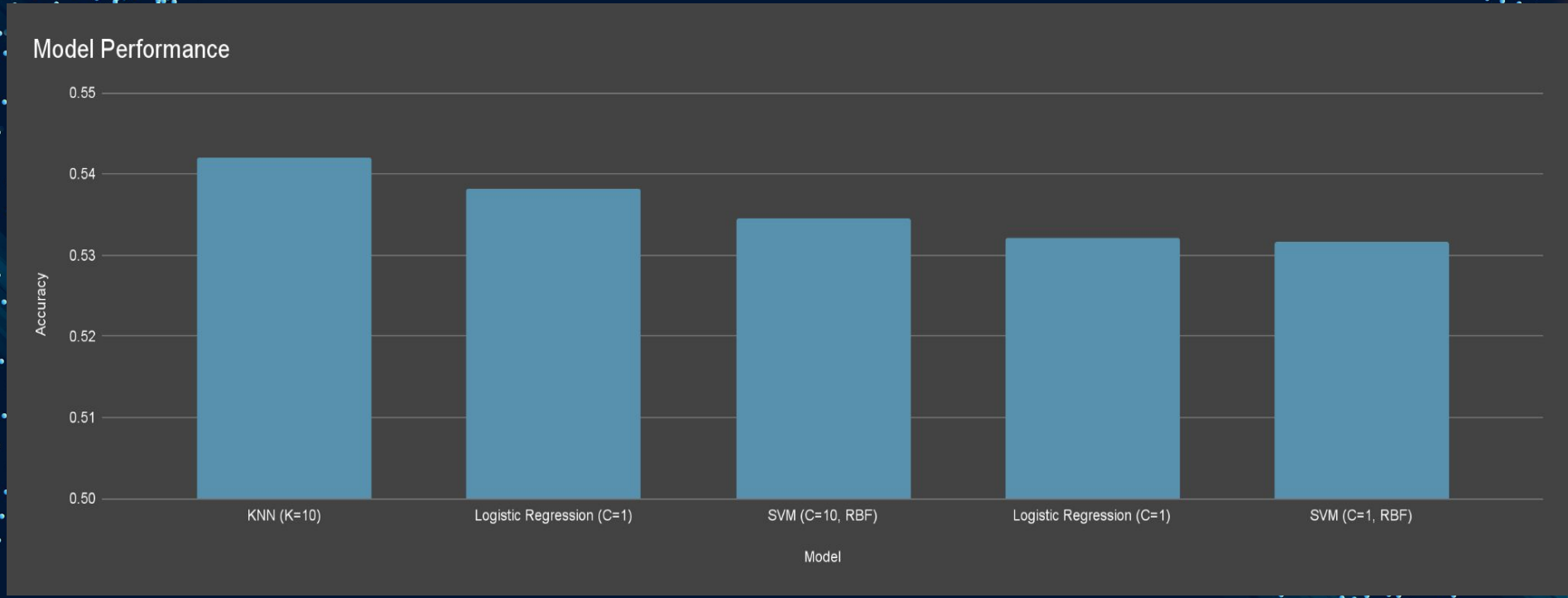
“Covering the spread” = predicted team wins by more points than the spread

Q1 - Can weather predict whether the spread will be covered?

- Aiming for > 52.4% accuracy for profitability
- Trained On
  - Temperature
  - Wind
- Ran Grid Search CV (5-fold)
  - Logistic Regression (c = 0.01, 0.1, 1, 10)
  - KNN (n = 1, 5, 10, 20, 50, 100)
  - SVM (RBF, Linear)

KNN (K=10) Based on Wind, Temperature

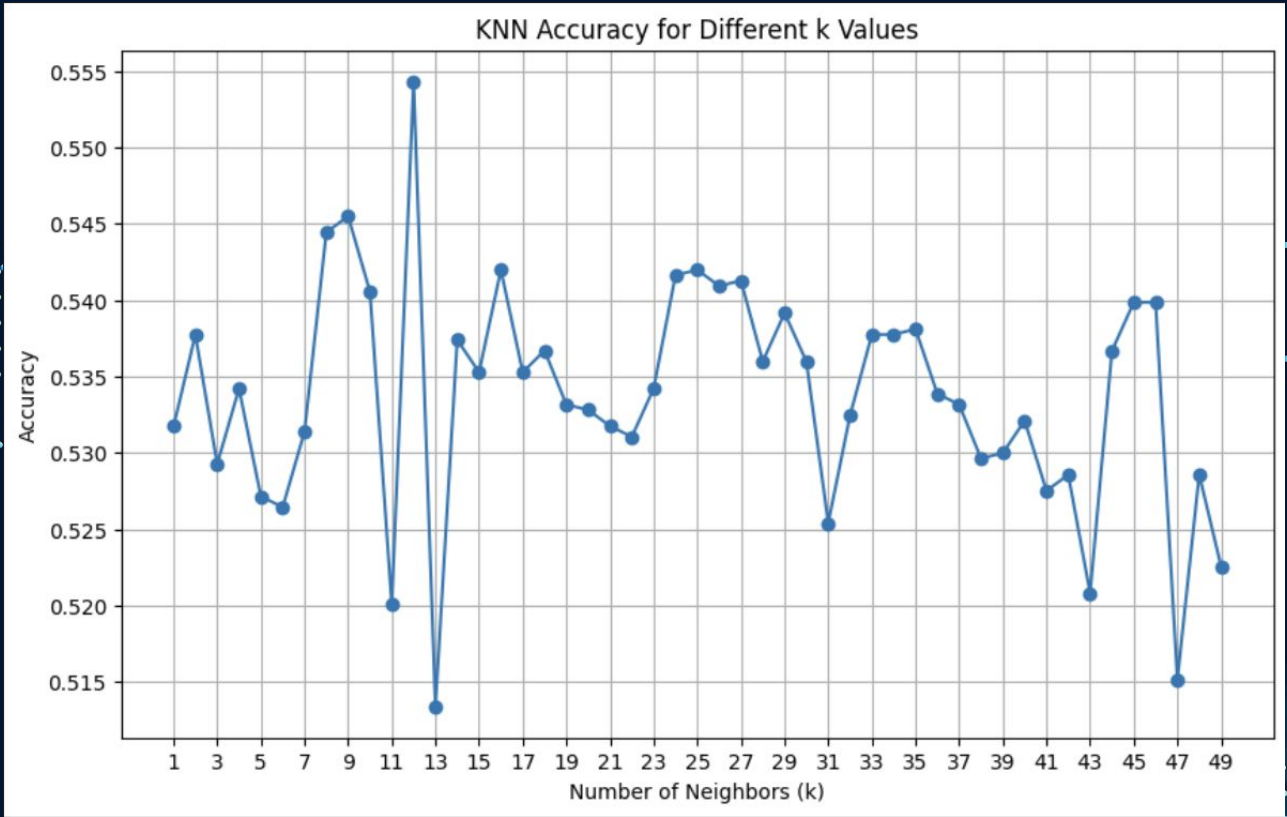
Accuracy = 54.2%

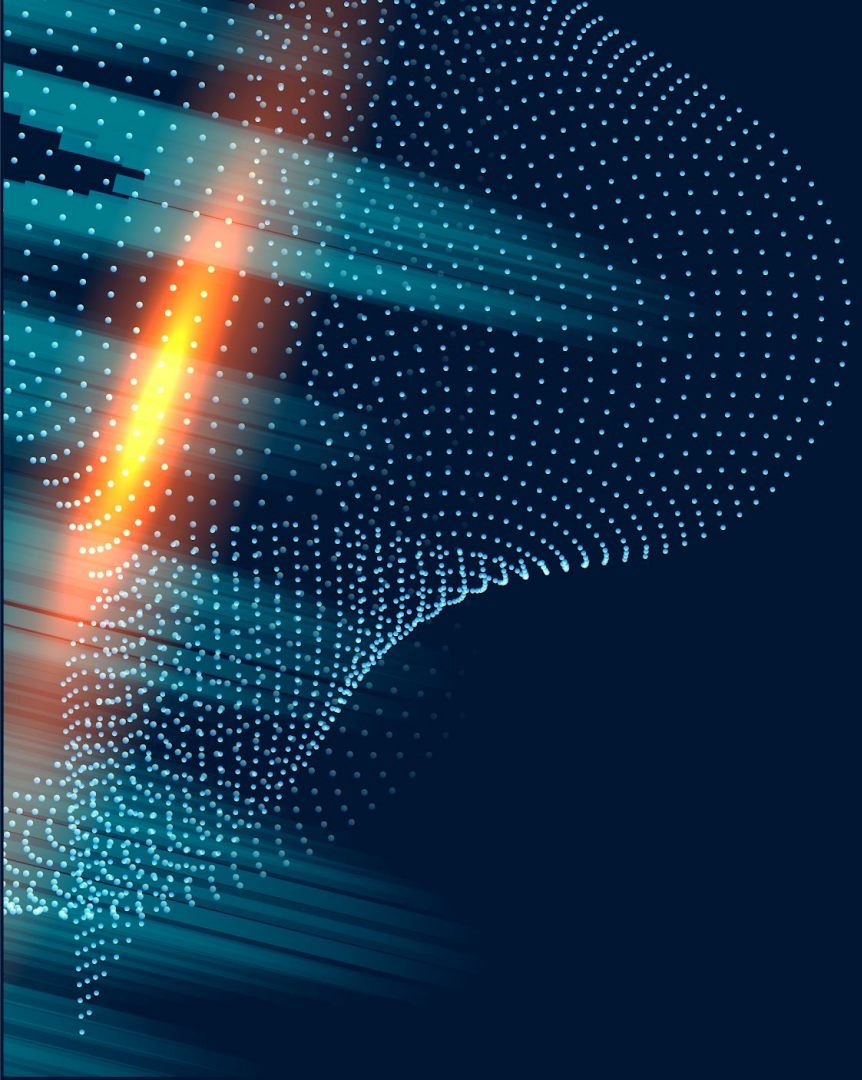


KNN (K=12) Based on Wind, Temperature

Accuracy = 0.554

55.4% > 52.4%, But Still Low





Question 2: Can weather predict if the over/under will be met?

“Over/under line” = sum of the predicted final point totals of the two teams

“Meeting the over/under” = sum of the actual final point total is greater than the over/under line

Q2 - Can weather conditions predict if the over/under will be met?

- Aiming for > 52.4% accuracy for profitability
- Selected weather temperature as sole feature
  - Had greatest correlation
  - Yielded best results
- Ran Grid Search CV (5-fold)
  - Logistic Regression (c = 0.01, 0.1, 1, 10)
  - KNN (n = 1, 5, 10, 20, 50, 100)
  - SVM (RBF, Linear)

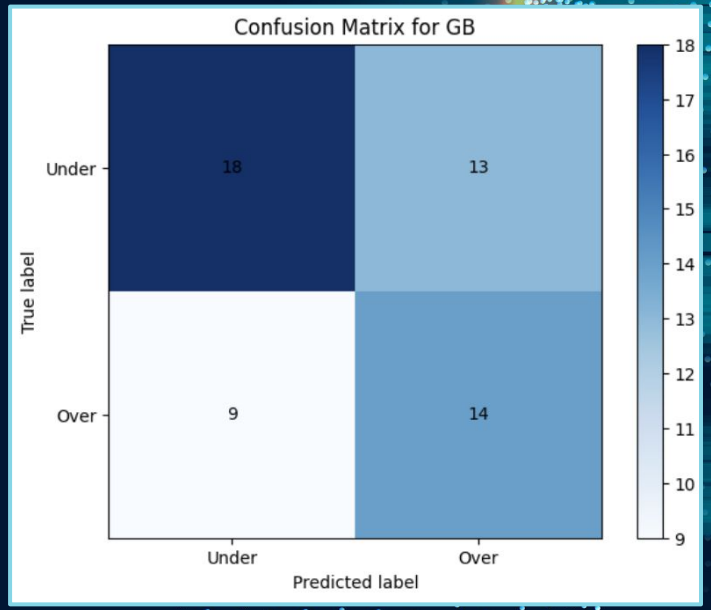
Results

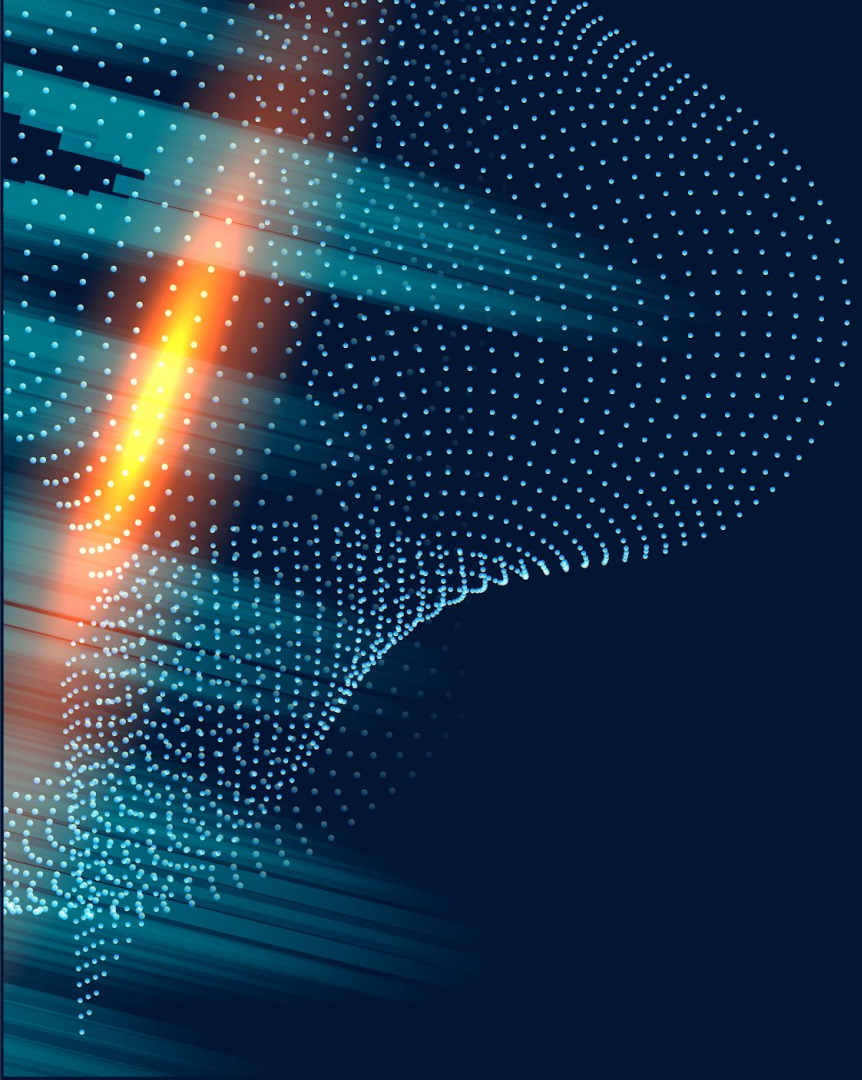
- Most Accurate Model: KNN (k = 5)
  - **53.36% Accuracy**
- Does filtering by team yield different results?

Results - Individual Teams

- Green Bay Packers
  - Most accurate results
  - Using SVM (rbf, C=1)
  - 60.15% accuracy (k-folds CV)
  - Coldest team in NFL

- Results lead us to next question...





Question 3: Can a team's usual temperature predict game outcomes?

Q3 - Team Performance Analysis Based On Temperature

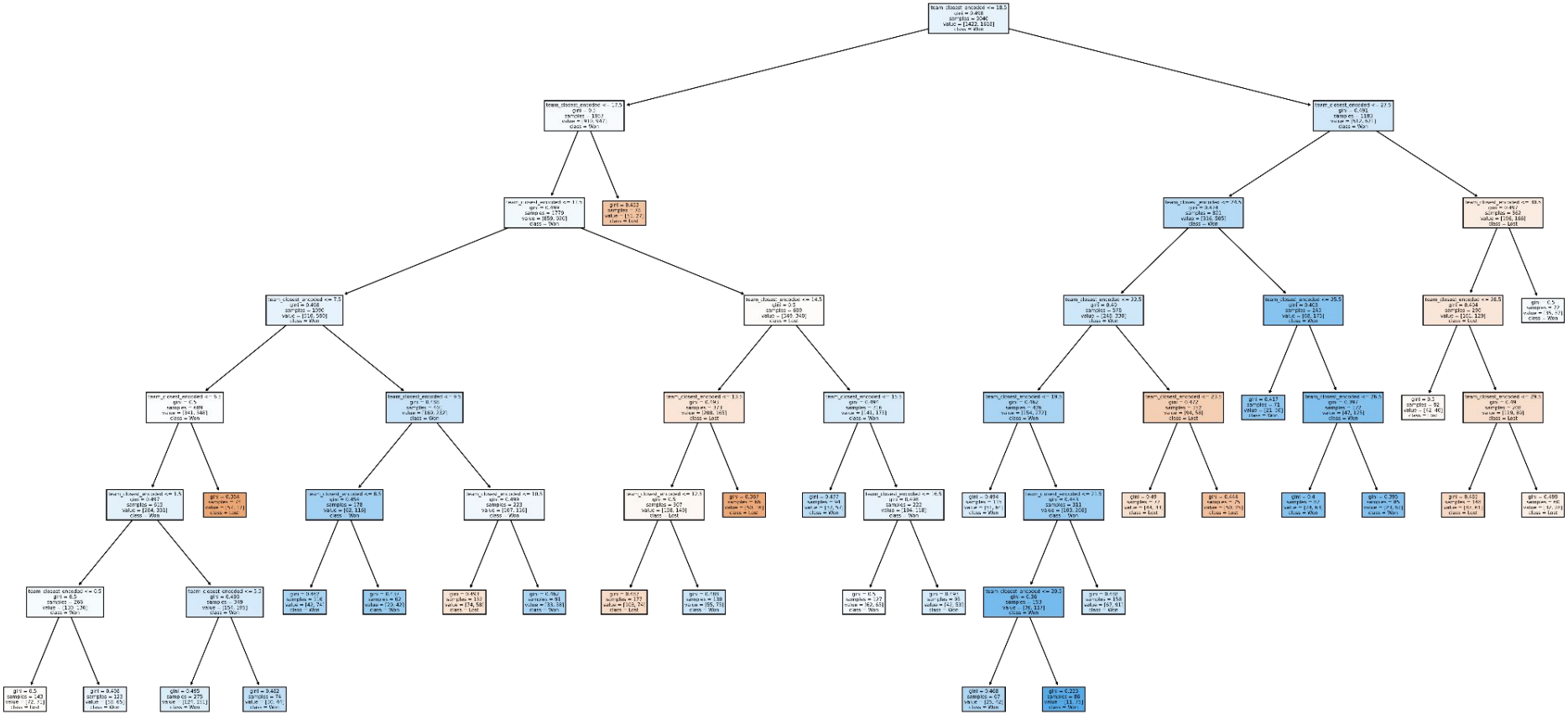
Decision Tree Accuracy: 0.56
- Max depth: 8
- 70/30 train test split

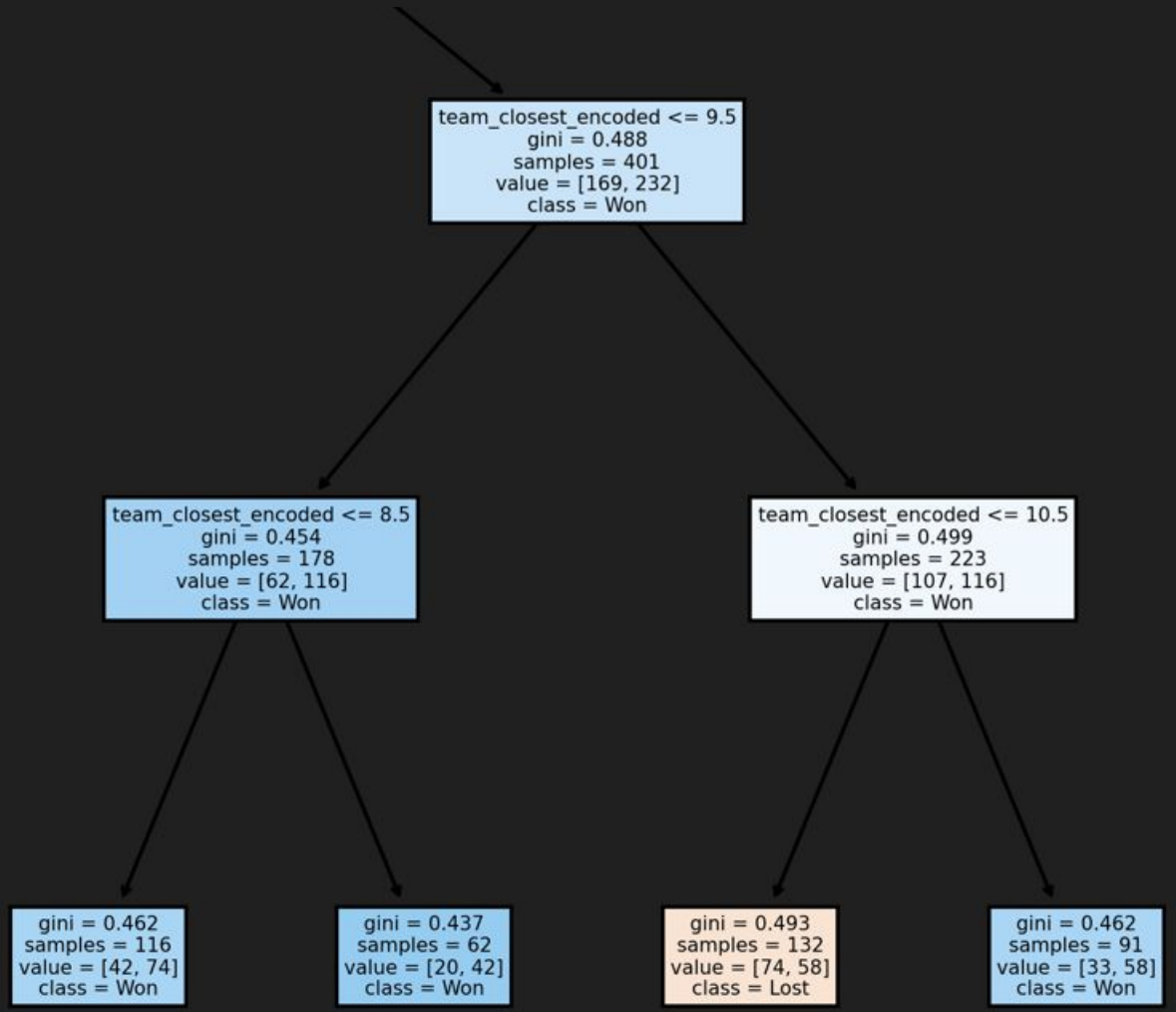
Logistic Regression Accuracy: 0.53
- 70/30 train test split

Sorted by temperature:

team	team_counts	avg_weather
TB	131	71.63
NO	151	69.91
HOU	156	69.53
MIA	123	69.38
ATL	159	68.92
LAC	147	68.53
ARI	154	68.50
IND	157	68.20
DET	156	67.85
LAR	149	67.82
JAX	130	67.81
DAL	156	65.87
LV	149	65.85
SF	128	65.23
CAR	132	64.95
MIN	157	64.46
TEN	127	64.21
DEN	126	61.44
PHI	130	60.81
BAL	121	60.74
SEA	129	60.41
NYJ	122	60.39
WAS	127	60.06
KC	126	58.83
CIN	122	58.62
NYG	125	58.45
CHI	129	58.28
NE	127	57.53
BUF	125	57.51
CLE	123	56.33
PIT	122	55.31
GB	128	54.52

Decision Tree (max_depth = 8)




team_closest_encoded <= 9.5
gini = 0.488
samples = 401
value = [169, 232]
class = Won
team_closest_encoded <= 8.5
gini = 0.454
samples = 178
value = [62, 116]
class = Won
team_closest_encoded <= 10.5
gini = 0.499
samples = 223
value = [107, 116]
class = Won
gini = 0.462
samples = 116
value = [42, 74]
class = Won
gini = 0.437
samples = 62
value = [20, 42]
class = Won
gini = 0.493
samples = 132
value = [74, 58]
class = Lost
gini = 0.462
samples = 91
value = [33, 58]
class = Won

Summary

- Q1: Can weather predict if the spread will be covered?
  - Aimed for >52.4% accuracy
  - Achieved 55.4% accuracy with KNN (k=12)

- Q2: Can weather predict if the over/under will be met?
  - Aimed for >52.4% accuracy
  - Achieved 56.14% accuracy with KNN (k
  - Achieved 60.15% accuracy for Green Bay Packers (linear SVM)

- Q3: Can a team's usual temperature predict game outcomes?
  - Achieved 53 - 56% accuracy

Questions

win 52.4% of bets (Culver 2019). A bettor that has a betting strategy that is successful more than 52.4% of the time will be profitable. The break-even percentage is 52.4%, because of the vig that must be paid on every wager.