Bet Better: Over/Unders

Looking into the performance of machine learning on classifying the outcomes of NBA games

Question:

- Can machine learning models trained on historical game data accurately classify the outcome of a game with regards to the associated total line, set by a bookkeeper

Context

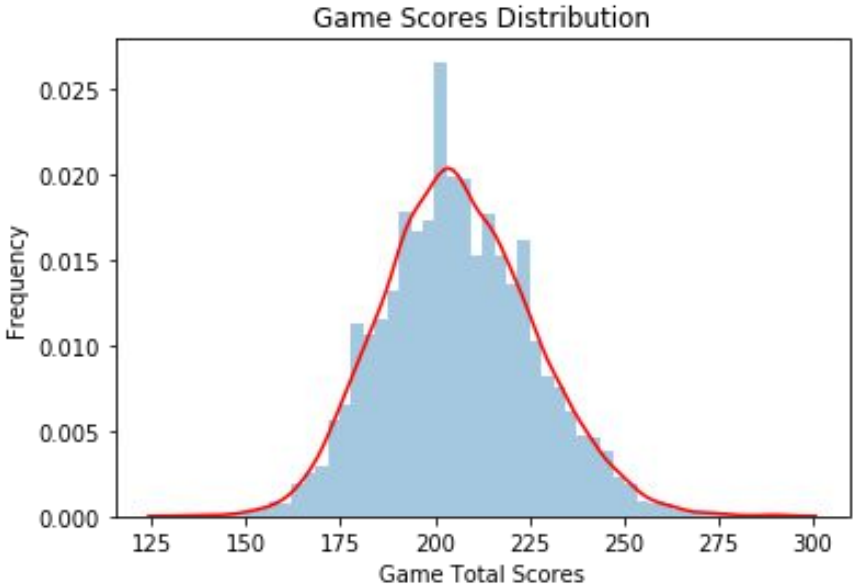
Cleveland Cavaliers: 205 -105

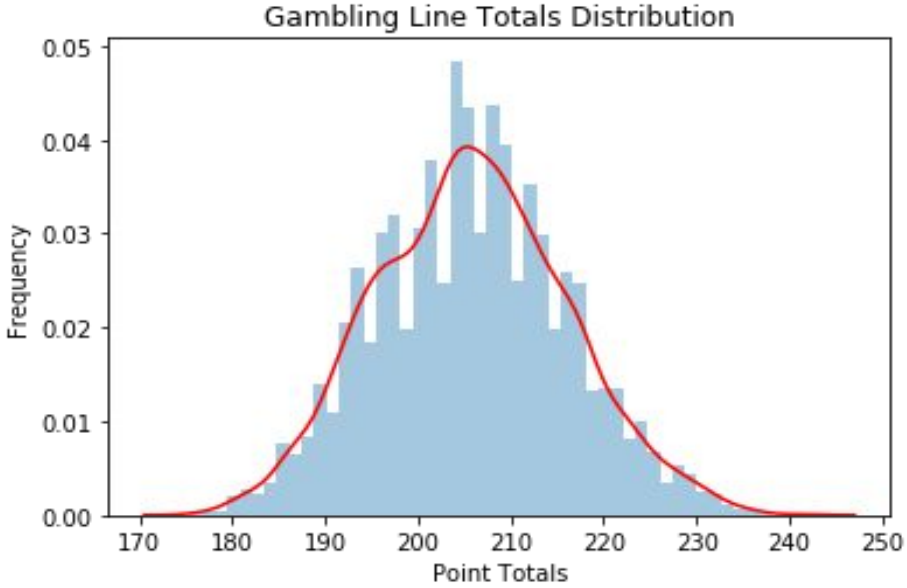
Los Angeles Lakers: -8 -110

Dataset

- Game to game box scores for every NBA game from the past 5 seasons
  - Collected from basketball-refrerence.com

- Totals, set by a bookkeeper, corresponding to every game for the past 5 seasons
  - Bet data collected from sportsbookreview.com

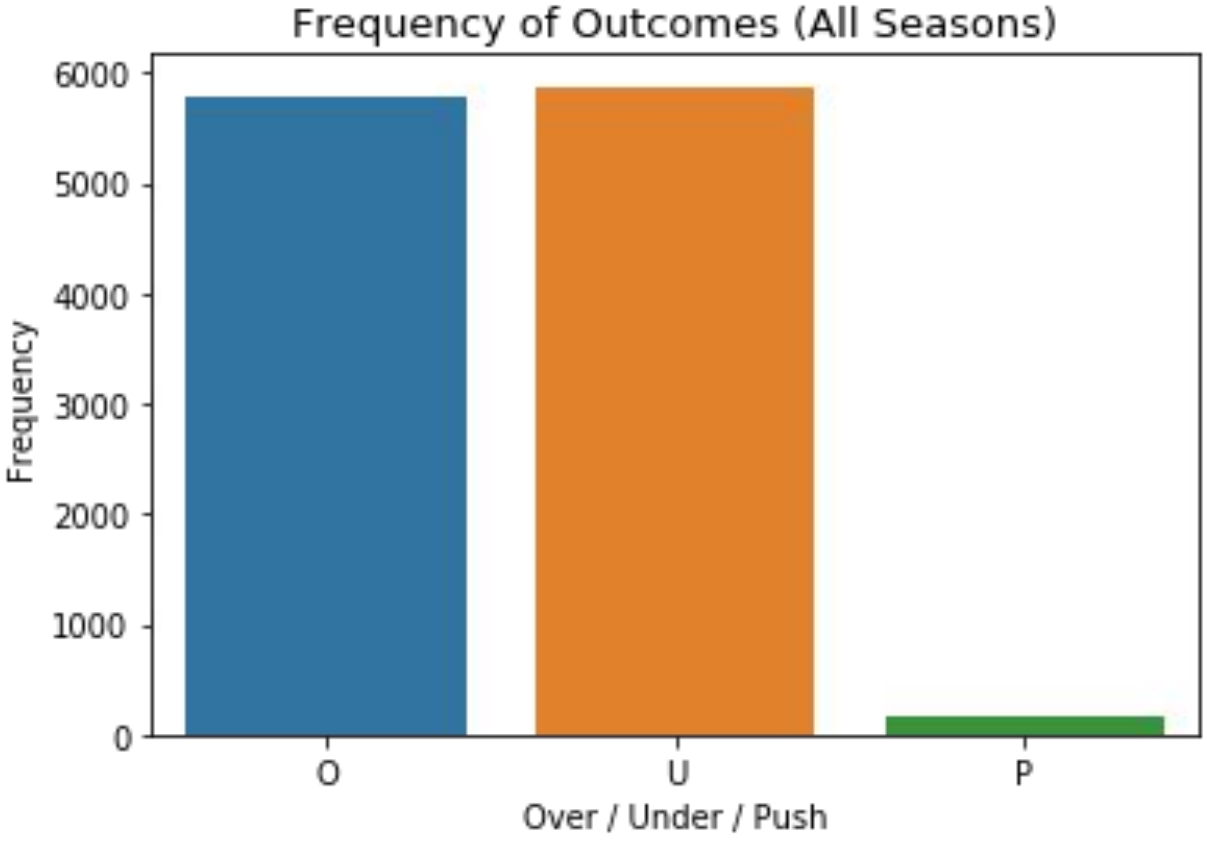

Game Scores Distribution
Frequency
0.025
0.020
0.015
0.010
0.005
0.000
125
150
175
200
225
250
275
300
Game Total Scores


Gambling Line Totals Distribution
Frequency
0.05
0.04
0.03
0.02
0.01
0.00
170
180
190
200
210
220
230
240
250
Point Totals

Features

- Offensive rating

- Imputed whether a game had fallen over the stated total line, under the line, or on the line (Push)


Frequency of Outcomes (All Seasons)
6000
5000
4000
3000
2000
1000
0
Frequency
O
U
P
Over / Under / Push

Time Series Data

- Each game was represented as the box score stats for the previous 3 games, for both the home team and the away team

- Expanded vs. Rolling

- Dropped 1st 3 games of every season

Modeling

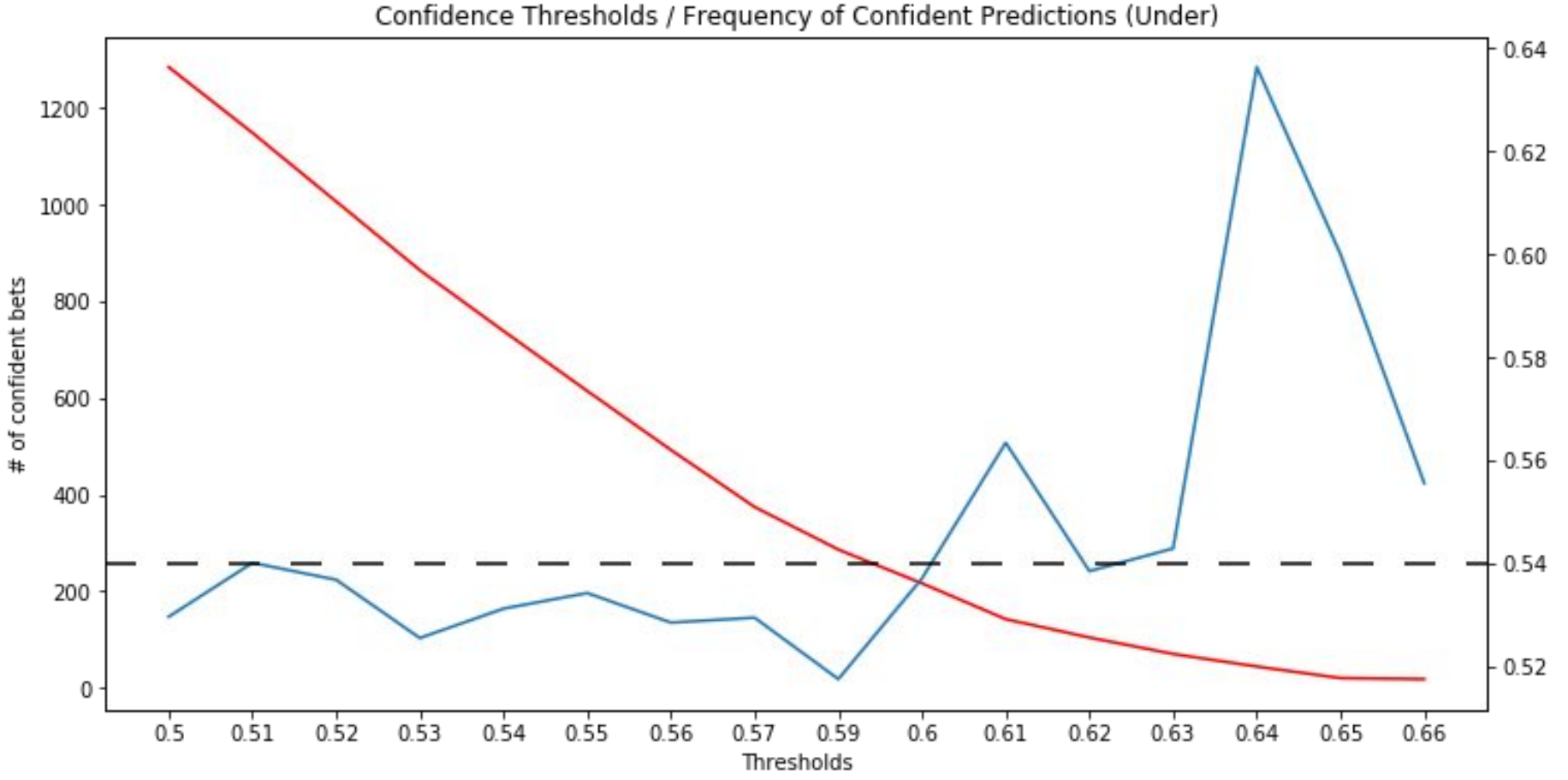
- Best model: Logistic Regression
  - Feature Selection using SelectFromModel transformer (228 final features)

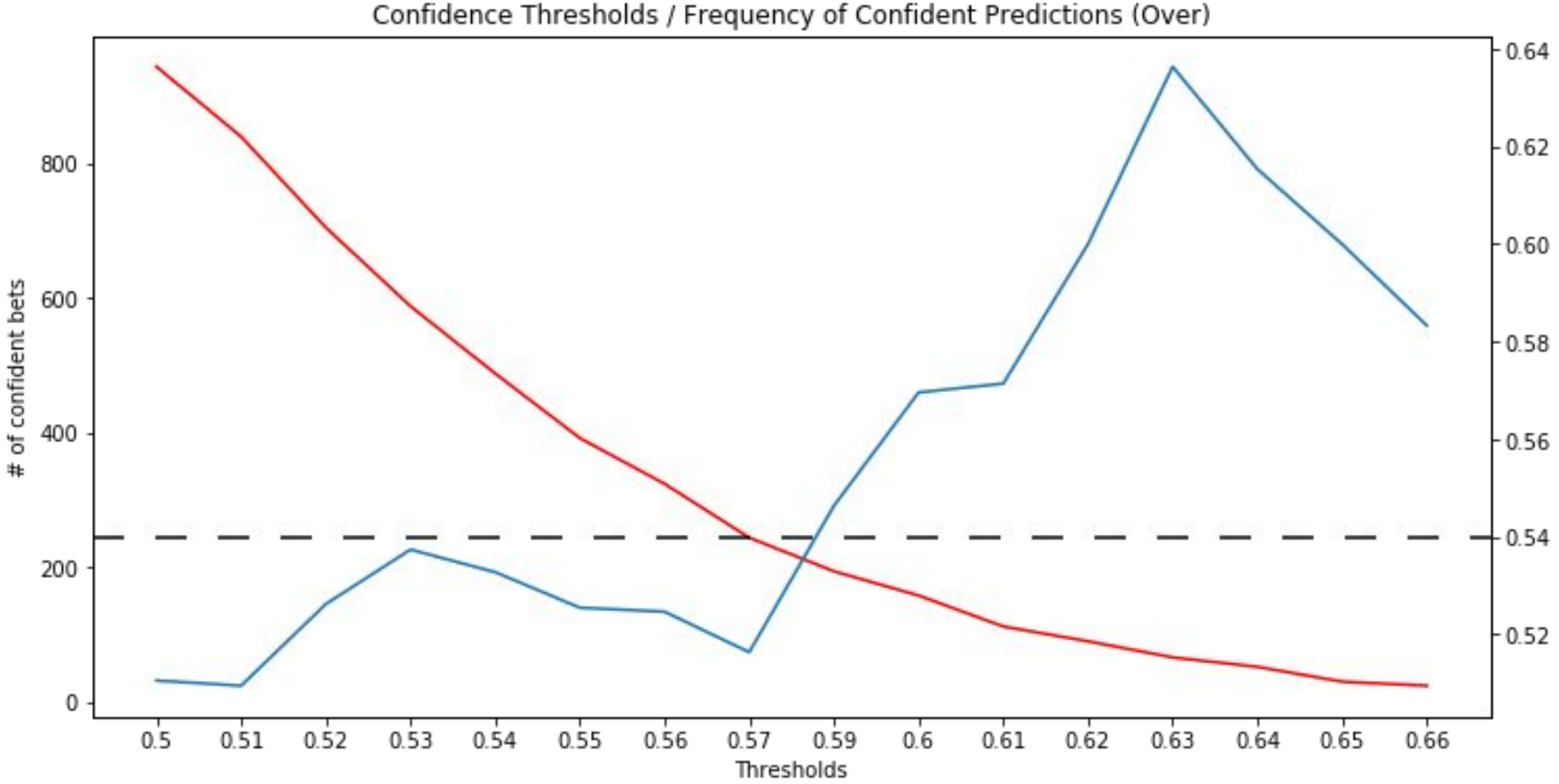
(Scoring: ROC-AUC, Features: SelectFromModel)	Training Data Score	Testing Data Score
Logistic Regression w/ SelectFromModel (Over Bets)	0.586	0.528
Logistic Regression w/ SelectFromModel (Under Bets)	0.584	0.530

“Confident” Predictions

- Creating confidence thresholds

- Tradeoff between # of “confident” predictions and % of correct confident predictions


Confidence Thresholds / Frequency of Confident Predictions (Under)
# of confident bets
Thresholds
0
200
400
600
800
1000
1200
0.52
0.54
0.56
0.58
0.60
0.62
0.64
0.5
0.51
0.52
0.53
0.54
0.55
0.56
0.57
0.59
0.6
0.61
0.62
0.63
0.64
0.65
0.66


Confidence Thresholds / Frequency of Confident Predictions (Over)
# of confident bets
Thresholds
0
200
400
600
800
0.52
0.54
0.56
0.58
0.60
0.62
0.64
0.5
0.51
0.52
0.53
0.54
0.55
0.56
0.57
0.59
0.6
0.61
0.62
0.63
0.64
0.65
0.66

Conclusion

- Model performed well enough to inform a winning betting strategy, given a high enough confidence threshold

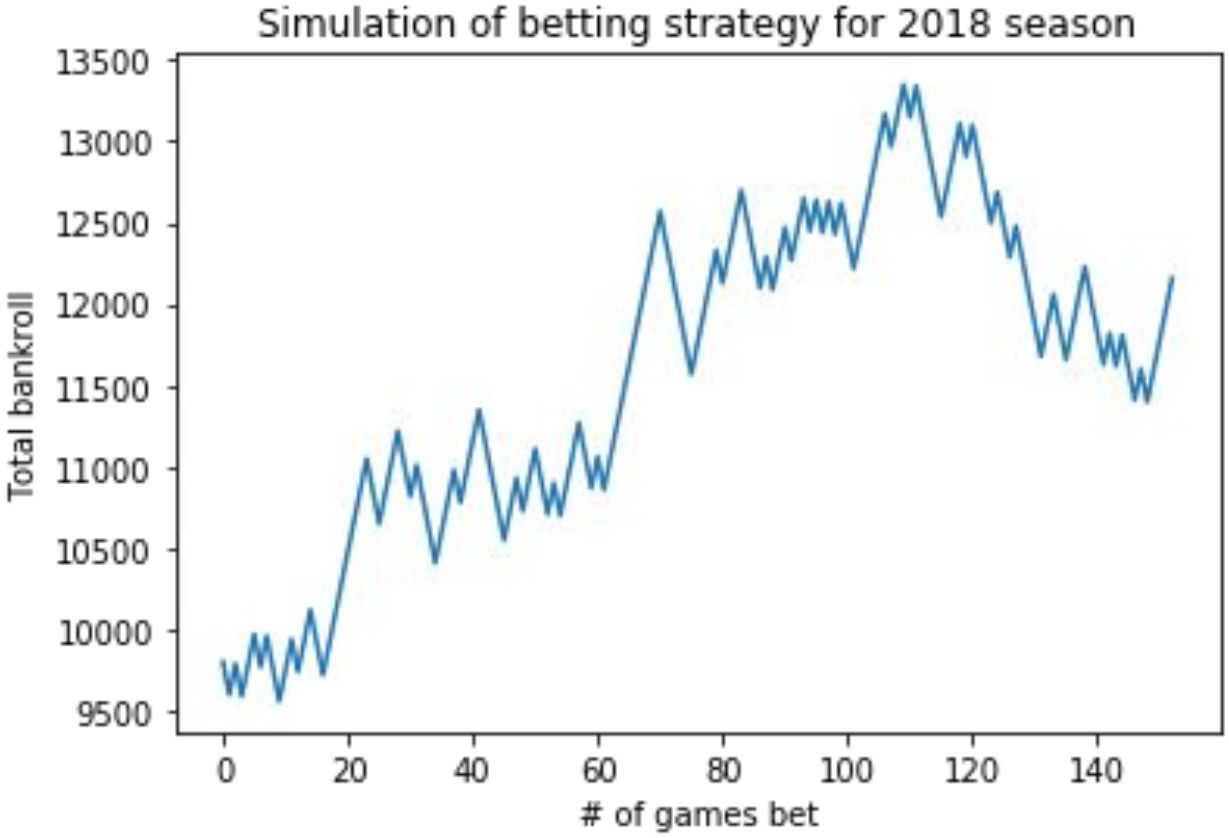
- Recency bias?

Simulation

1. Start with $10,000
2. Bet $200 on every game with either a confident over prediction or a confident under prediction
3. If over prediction or under prediction correct:
    a. Win $190
4. If over prediction or under prediction incorrect:
    a. Lose $200

Predicted 153 games confidently, with a confidence threshold set at 0.60

Result: $12,160 (for 2018 NBA season)


Simulation of betting strategy for 2018 season
Total bankroll
13500
13000
12500
12000
11500
11000
10500
10000
9500
0
20
40
60
80
100
120
140
# of games bet

Takeaways

- Viability of using model to inform betting strategy

- Efficacy of machine learning dealing with this problem

Looking to the Future

- Accounting for variable point totals

- Bayesian Statistics - Live Betting