

# Automating NFL Predictions with Ensemble Model

James Fontaine

Q4 2020

# Problem

- Weekly NFL pick league - 5 years running
- I'm tired of thinking (making picks) - and losing



SOLUTION: build a model to make picks for me!



# Training Data

- NFL game results/QB stats since 1970 from Pro Football Reference
- Stadium data from Wikipedia
- Calculate season stats as well as 4,8,16, & 32 game rolling stats for each team and 4,8,16, & all-time stats for each QB
- Use away vs home team differential for input

```
binary = [  
    'ERA_BLOUNT',  
    'ERA_2PT',  
    'ERA_LAW',  
    'ERA_BRADY',  
    'INTERNATIONAL',  
    'FIRST_QB_START_H',  
    'FIRST_QB_START_A']  
  
categorical = (  
    'DAY_ID',  
    '_MONTH',  
    'TIME_OF_DAY')
```

**85 Inputs  
Total**

```
continuous = (  
    'CAPACITY',  
    'DIST_KM_DIFF',  
    'TIMEZONE_DIFF',  
    'ELEVATION_DIFF',  
    'ROOF_DIFF',  
    'FIELD_DIFF',  
    'SEASON_STAT_DIFF_SOS',  
    'SEASON_WIN_PERCENT_DIFF',  
    'SEASON_AVG_PTS_DIFF_DIFF',  
    'SEASON_AVG_YDS_DIFF_DIFF',  
    'SEASON_AVG_TO_DIFF_DIFF',  
    'SEASON_OVERALL_RANK_DIFF',  
    'LAST4_STAT_DIFF_SOS',  
    'LAST4_WIN_PERCENT_DIFF',  
    'LAST4_AVG_PTS_DIFF_DIFF',  
    'LAST4_AVG_YDS_DIFF_DIFF',  
    'LAST4_AVG_TO_DIFF_DIFF',  
    'LAST4_OVERALL_RANK_DIFF',  
    'LAST8_STAT_DIFF_SOS',  
    'LAST8_WIN_PERCENT_DIFF',  
    'LAST8_AVG_PTS_DIFF_DIFF',  
    'LAST8_AVG_YDS_DIFF_DIFF',  
    'LAST8_AVG_TO_DIFF_DIFF',  
    'LAST8_OVERALL_RANK_DIFF',  
    'LAST16_STAT_DIFF_SOS',  
    'LAST16_WIN_PERCENT_DIFF',  
    'LAST16_AVG_PTS_DIFF_DIFF',  
    'LAST16_AVG_YDS_DIFF_DIFF',  
    'LAST16_AVG_TO_DIFF_DIFF',  
    'LAST16_OVERALL_RANK_DIFF',  
    'LAST32_STAT_DIFF_SOS',  
    'LAST32_WIN_PERCENT_DIFF',  
    'LAST32_AVG_PTS_DIFF_DIFF',  
    'LAST32_AVG_YDS_DIFF_DIFF',  
    'LAST32_AVG_TO_DIFF_DIFF',  
    'LAST32_OVERALL_RANK_DIFF',  
    'HTH_SEASON_WIN_PERCENT_DIFF',  
    'HTH_SEASON_AVG_PTS_DIFF_DIFF',  
    'HTH_SEASON_AVG_YDS_DIFF_DIFF',  
    'HTH_SEASON_AVG_TO_DIFF_DIFF',  
    'HTH_LAST4_WIN_PERCENT_DIFF',  
    'HTH_LAST4_AVG_PTS_DIFF_DIFF',  
    'HTH_LAST4_AVG_YDS_DIFF_DIFF',  
    'HTH_LAST4_AVG_TO_DIFF_DIFF',  
    'REST_DAYS_DIFF',  
    'QB_AVG_RATE_ALL_TIME_DIFF',  
    'QB_ALL_TIME_STAT_DIFF',  
    'QB_GP_ALL_TIME_RANK_DIFF',  
    'QB_AVG_RATE_ALL_TIME_RANK_DIFF',  
    'QB_ALL_TIME_OVERALL_RANK_DIFF',  
    'QB_AVG_RATE_SEASON_DIFF',  
    'QB_SEASON_STAT_DIFF',  
    'QB_GP_SEASON_RANK_DIFF',  
    'QB_AVG_RATE_SEASON_RANK_DIFF',  
    'QB_SEASON_OVERALL_RANK_DIFF',  
    'QB_VS_DEF_SEASON_RANK_DIFF',  
    'QB_AVG_RATE_LAST4_RANK_DIFF',  
    'QB_LAST4_OVERALL_RANK_DIFF',  
    'QB_VS_DEF_LAST4_RANK_DIFF',  
    'QB_AVG_RATE_LAST8_RANK_DIFF',  
    'QB_LAST8_OVERALL_RANK_DIFF',  
    'QB_VS_DEF_LAST8_RANK_DIFF',  
    'QB_AVG_RATE_LAST16_RANK_DIFF',  
    'QB_LAST16_OVERALL_RANK_DIFF',  
    'QB_VS_DEF_LAST16_RANK_DIFF',  
    'FIRST_QB_START_DIFF')
```

# Ensemble Model Architecture - Deterministic Stacking

Binary classification - predict win by home team

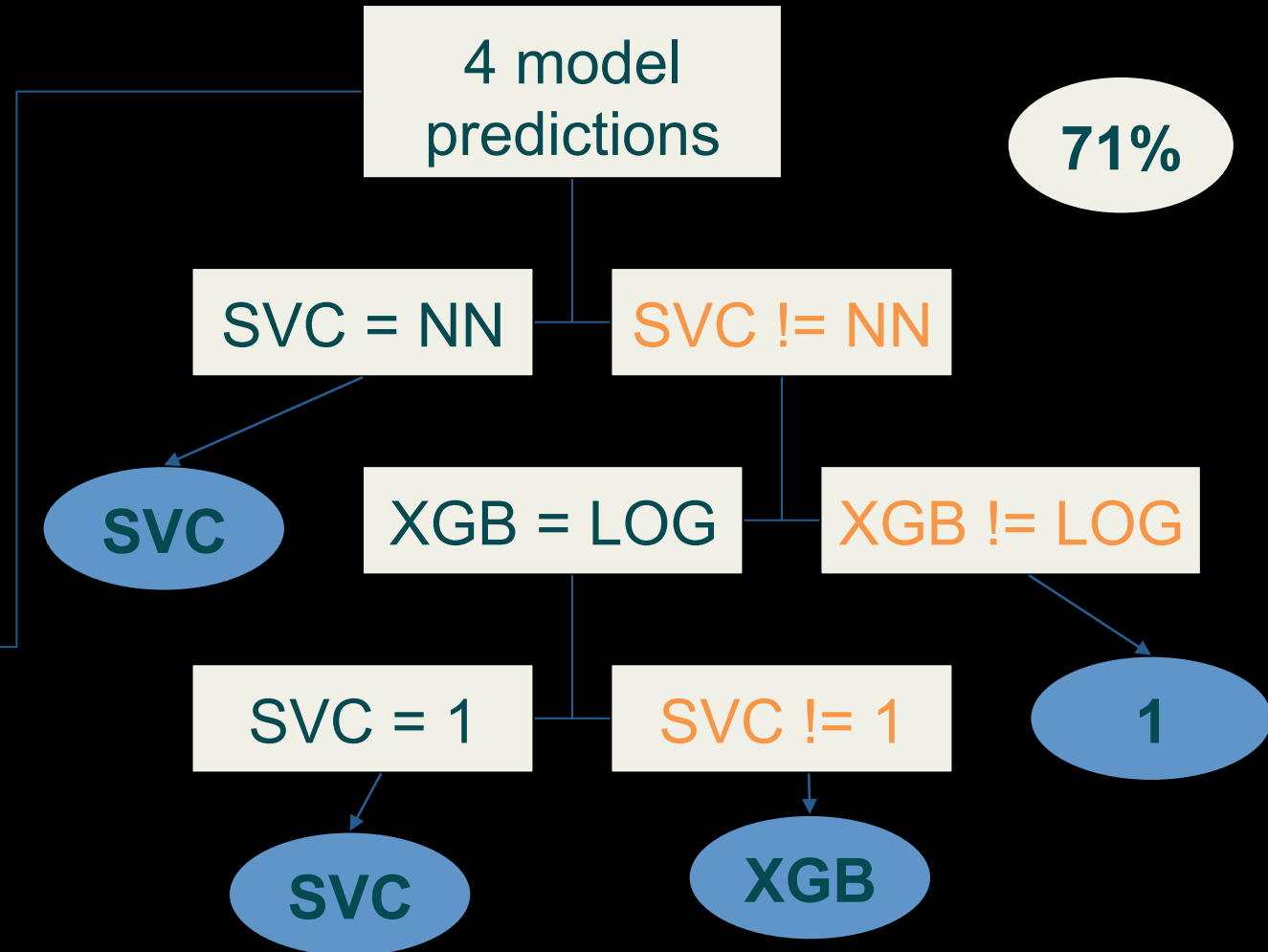
65%

66%

66%

66%

- XGBoost classifier
- Logistic regression (liblinear)
- Support vector classifier (RBF kernel)
- Neural network (3 layers, BCE loss)



# New Problem

- I'm tired running all these scripts... this feels like more work than just making my own picks



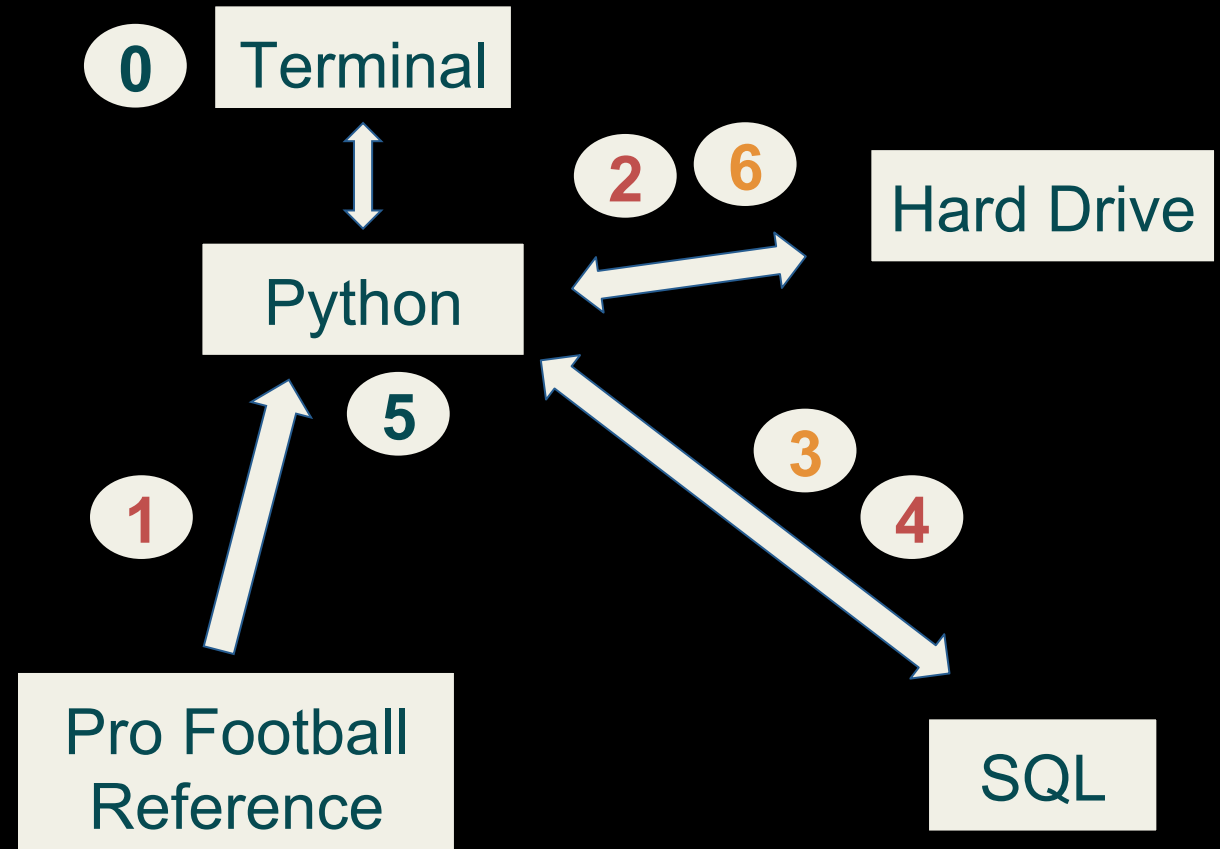
SOLUTION: automate into a single script with command line arguments!



# Automated Script Architecture

Read Write

0. Run script in terminal
1. Scrape last week results from PFF
2. Pull upcoming games data from csv (and stadium data update)
3. Update tables
4. Calculate model inputs
5. Train models, make predictions
6. Store predictions/updated data to hard drive





```
[(base) Jameess-MacBook-Pro:~ james$ cd Documents/NFL_Predictions
[(base) Jameess-MacBook-Pro:NFL_Predictions james$ python nfl_weekly_pred_cli.py
2020 7 --stadium

Connecting...
Done

Updating stadium table...
Done

Pulling game results from week 6 of 2020 season...
Done

Pulling QB results from week 6 of 2020 season...
Done

Saving game & QB results files...
Done

Updating games table...
Done

Updating QB table...
Done

Updating upcoming games table...
Done

Calculating historical team rankings...
Done

Creating model input table...
Done

Pulling model input table into pandas, generating model input file...
Done

Running models...

Model Report

xgb_estimators: 20
xgb Train Accuracy : 0.6794
xgb AUC Score (Train): 0.733709

log Train Accuracy : 0.6654
log AUC Score (Train): 0.711394

svc Train Accuracy : 0.7179
svc AUC Score (Train): 0.814796
```

# Terminal Output

```
Training nn...
Epoch: 7/50.. Training Loss: 0.638.. Test Loss: 0.621.. Test Accuracy: 0.650
Epoch: 14/50.. Training Loss: 0.628.. Test Loss: 0.617.. Test Accuracy: 0.660
Epoch: 21/50.. Training Loss: 0.625.. Test Loss: 0.615.. Test Accuracy: 0.664
Epoch: 28/50.. Training Loss: 0.623.. Test Loss: 0.613.. Test Accuracy: 0.666
Epoch: 35/50.. Training Loss: 0.622.. Test Loss: 0.613.. Test Accuracy: 0.670
Epoch: 42/50.. Training Loss: 0.621.. Test Loss: 0.611.. Test Accuracy: 0.671
Epoch: 49/50.. Training Loss: 0.620.. Test Loss: 0.610.. Test Accuracy: 0.671

Final predictions complete

Saving raw predictions...
Done

Creating final predictions table...
Done

Saving final predictions file...
Done

Printing final results...
49ers @ Patriots: Patriots
Steelers @ Titans: Titans
Seahawks @ Cardinals: Seahawks
Panthers @ Saints: Saints
Packers @ Texans: Packers
Lions @ Falcons: Falcons
Jaguars @ Chargers: Chargers
Giants @ Eagles: Eagles
Cowboys @ Washington Football Team: Cowboys
Chiefs @ Broncos: Chiefs
Buccaneers @ Raiders: Raiders
Browns @ Bengals: Browns
Bills @ Jets: Bills
Bears @ Rams: Rams

Done
Quitting...
```

# Simulated Model Performance

Games Correctly Predicted

YEAR	ME	LEAGUE WINNER	MODEL	DIFF
2015	149	162	180	+18
2016	165	166	190	+24
2017	172	177	171	-6
2018	166	167	177	+10
2019	162	162	175	+13
Total	814	834	893	+59

Model would have won 4 out of last 5 years!



# Next Steps

- Currently 7 games back in 2020 (63%)
- Review features/stacking algorithm to account for reduced home field advantage due to COVID?
- Update model to account for playoff games
- Add features (coach experience, injuries)
- Find someone at STATS LLC (Vista portfolio company) to discuss this with...