SUPER BOWL LIII

Team NB

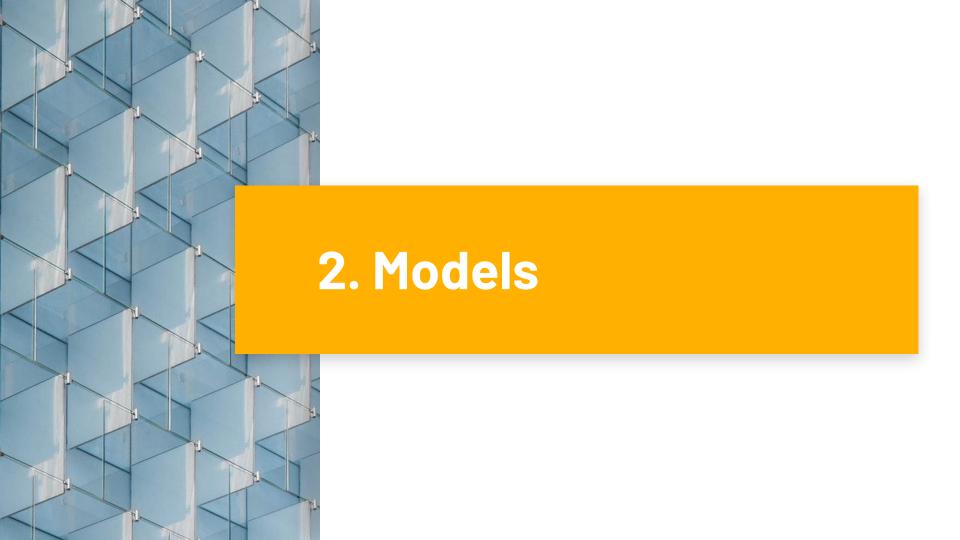—— **Shen Yang, Guanhang Chen, Xinran Su, Qinyu Xia, Xinqi Lin**

What We've Learned

- Under-confidence VS over-confidence

- Share information and perspectives helps make dragonfly eye work

- Keep bias away

1. Dataset

Super Bowl LIII Prediction Dataset

Time	2013-2018
Independent variables	Win or Lose
	Points scored
22 Dependent variables	Yards gained by passing/rushing(offense & defense)
	Team turnovers lost(offense) / turnovers gained(defense)
	Expected points
	Weather condition, stadium type
	Las Vegas betting odds
	……

2. Models

To Predict Who Will Win:

- Logistic Regression

- Decision Tree

To Predict the Exact Score:

- Linear Regression

- Poisson Regression
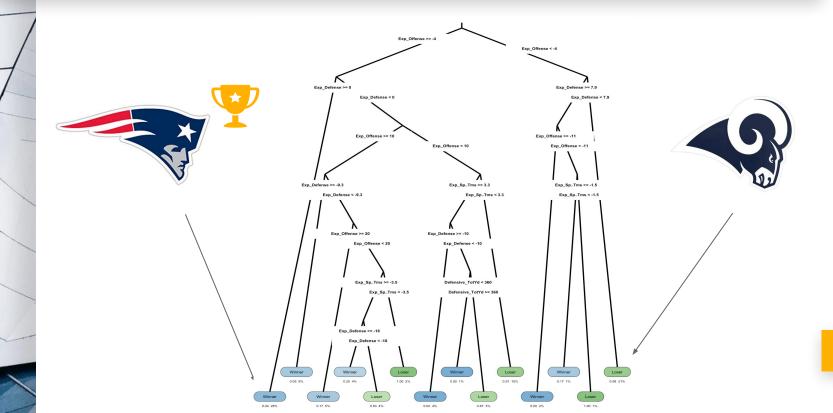
- SARIMA

Logistic Regression

```
Coefficients:
                Estimate Std. Error z value   Pr(>|z|)
(Intercept)     -5.483943   3.111608  -1.762    0.07800 .
Offensive_1stD  -0.027344   0.082027  -0.333    0.73886
Offensive_PassY  0.002299   0.005574   0.412    0.67998
Offensive_RushY  0.003231   0.007571   0.427    0.66959
Offensive_TO    -1.090652   0.355081  -3.072    0.00213 **
Defensive_1stD  -0.014008   0.097577  -0.144    0.88585
Defensive_PassY  0.020888   0.007038   2.968    0.00300 **
Defensive_RushY  0.014649   0.008345   1.755    0.07919 .
Defensive_TO     0.704976   0.296384   2.379    0.01738 *
Exp_Offense      0.303312   0.061266   4.951 0.00000073928 ***
Exp_Defense      0.480864   0.083725   5.743 0.00000000928 ***
stadium_type    -0.359854   0.518060  -0.695    0.48729
Favorite_or_not  1.130869   0.620120   1.824    0.06821 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 422.43  on 315  degrees of freedom
Residual deviance: 114.68  on 303  degrees of freedom
  (397 observations deleted due to missingness)
AIC: 140.68

Number of Fisher Scoring iterations: 8
```
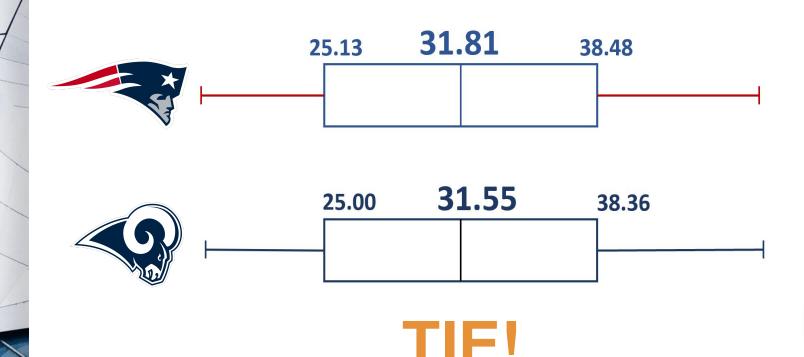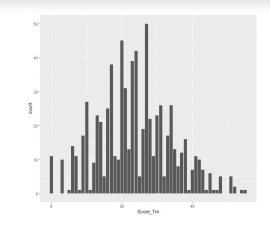
```
> predict(fit5, newdata = NFL_rams, type = "response")
        1
0.9473962
> predict(fit5, newdata = NFL_patriots, type = "response")
        1
0.9156235
```

Decision Tree

Poisson Regression


```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.4515  -0.8754  -0.0321   0.9112   4.0932

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)      1.8224452  0.1187001  15.353  < 2e-16 ***
Offensive_1stD   0.0131034  0.0049083   2.670  0.00759 **
Offensive_PassY  0.0021406  0.0003197   6.695 2.16e-11 ***
Offensive_RushY  0.0025864  0.0004032   6.414 1.42e-10 ***
Offensive_TO    -0.1064921  0.0192804  -5.523 3.33e-08 ***
Defensive_1stD   0.0101540  0.0061990   1.638  0.10142
Defensive_PassY  0.0002940  0.0003463   0.849  0.39587
Defensive_RushY -0.0010887  0.0004623  -2.355  0.01854 *
Defensive_TO     0.0682428  0.0155828   4.379 1.19e-05 ***
stadium_type     0.1225009  0.0388557   3.153  0.00162 **
Spread          -0.0165678  0.0031499  -5.260 1.44e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)
```


26 VS 32

Seasonal | Autoregressive | Integrated | Moving Average

[Period]   [Prior Values]   [Differencing]   [Regression Error]

SARIMA with Exogenous Variables

Super Bowl LIII Prediction Dataset

Team	Year	Score	Internal		External		
			Offensive	Defensive	Temp	Weather	Bet
Patriots	2013 - 2018	...	...	...	...	...	...
Rams	2013 - 2018	...	...	...	...	...	...
...	...	...	...	...	...	...	...
Other Teams	2018	...	...	...	...	...	...
...	...	...	...	...	...	...	...

- # SARIMA (p, d, q, P, D, Q)
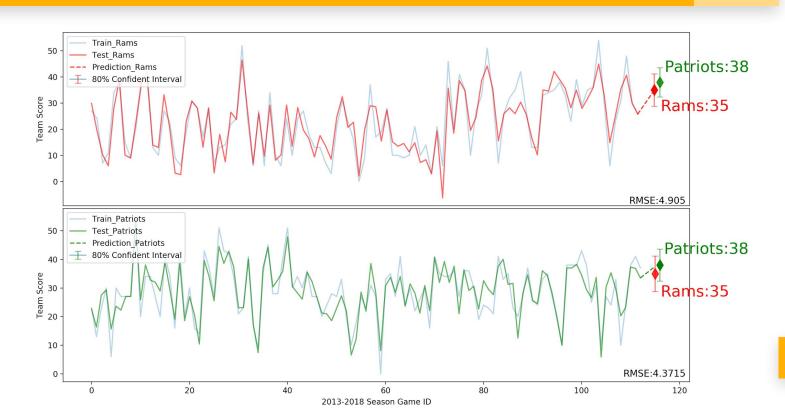
  - p: the order of non-seasonal AR term
  - d: the order of non-seasonal differencing
  - q: the order of non-seasonal MA term
  - P: the order of seasonal AR term
  - D: the order of seasonal differencing
  - Q: the order of seasonal MA term

from **pmdarima.arima** import **auto_arima**

Author: Taylor G. Smith

SARIMA with Exogenous Variables

3. Results

Workflow

Dataset	Model	Predict	Score
- 2013-2018 Match Data - Weather - Las Vegas Betting Line - 717 x 22	- Logistic Regression - Decision Tree - Linear Regression with Betting - Poisson Regression - SARIMA	- Patriots - Rams - Input = Avg. data of previous two games: Division & Conf. Champ	- Weight Avg. - 20% - Linear - 20% - Poisson - 60%- SARIMA

Summary

Logistic Regression	Decision Tree	Linear Regression with Betting	Poisson Regression	Seasonal ARIMA
8 Sig. Var.	First node: Exp.Offensive	8 Sig. Var.	9 Sig. Var.	P(0,1,1,0,1,1) R(0,1,1,1,1,2)
AUC = 0.979	ACC = 0.933	Adj. R square = 0.7895	Residual Deviance = 319.1 (d.f.=149)	Avg(RMSE) = 4.639
		Weight: 20%	Weight: 20%	Weight: 60%
Rams	Patriots	P(31) : R(31)	P(26) : R(32)	P(38) : R(35)

P = Patriots, R = Rams

34 : 33

Who will win?

Stay tuned...

20

Reference

[1]. Data source,
https://www.pro-football-reference.com
[2]. NFL Predictions Using Machine Learning,
http://pjoos.github.io/2015/11/11/nfl_ml.html
[3]. Forecasting: Principles and Practice,
https://otexts.com/fpp2/
[4]. 7 methods to perform Time Series forecasting,
https://www.analyticsvidhya.com/blog/2018/02/time-series-forecasting-methods/
[5]. An Introduction to Vector Autoregression (VAR),
https://www.r-econometrics.com/timeseries/varintro/
[6]. Create an ARIMA Model for Time Series Forecasting in Python,
https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/
[7]. Detailed NFL Play-by-Play Data 2009-2018,
https://www.kaggle.com/maxhorowitz/nflplaybyplay2009to2016