Mira Daya
Due: February 6, 2021
ID: 47486312
BANA 290: Forecasting

Super Bowl 55 Prediction: Kansas vs. Tampa Bay

Below is a brief bulleted description of the approach taken to predict the superbowl outcome which includes the thought process behind it, which methodologies to use, and what kind of model to use. Following that section is a formal summary.

Thought Process:

- Need to come up with a model to get the numerical value of the score vs probability of a team winning
- Need to come up with a way to select the best variables to include in the model
- Wait.. need to actually have SB55 variables → use ETS to predict possible outcomes
- Use predicted outcomes, use features selected to create models and plug and chug

Methodologies

1. Feature Selection using forward stepwise regression
2. ETS to predict each variable value for each team
3. Separate linear regression models for each team

Model(s):

1. Linear Regression and ETS using feature selection

Summary:

Since we are trying to predict the actual outcome of the SuperBowl, we can't just use a logistic regression model, that would only tell us if a team will win or not. We have to be creative in building the necessary models that will give us that score. The steps taken here included having to tackle the following issues:

1. Finding a good amount of data
2. Figuring out what to do with missing values
3. Feature selection
4. Predicting variable outcomes selected features for game day
5. Building a model that will predict the final outcome

Data

For each team, I obtained data from 2018-2020 NFL seasons. There are 54 observations for the Tampa Bay Buccaneers and 51 observations for the Kansas City Chiefs

Variables for each team include:

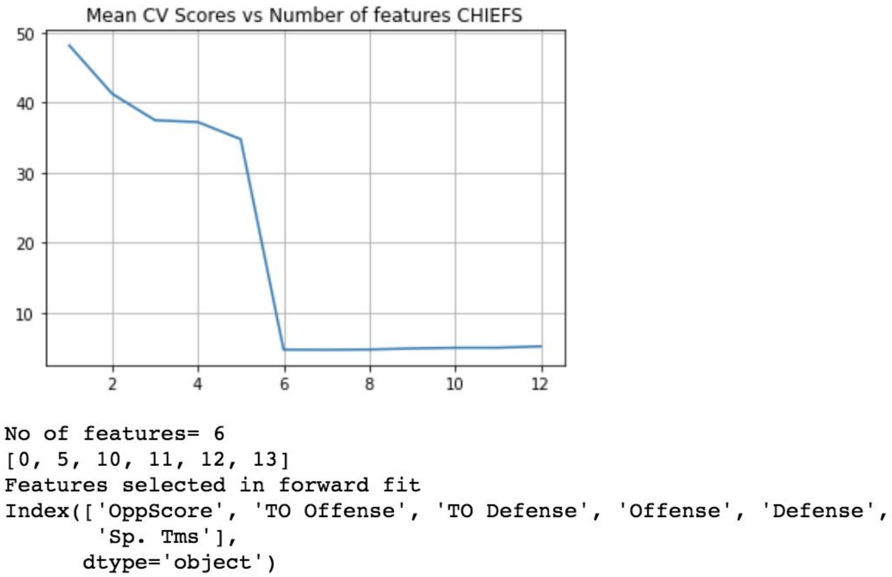
- TeamScore
- OppScore
- 1stD Offense
- TotYd Offence
- PassY Offence
- PushYOffense
- TO Offenes

- 1stD Defense
- TotYd Defense
- PassY Degense
- RushY Defense
- TO Defense

- Offense
- Defense
- Sp. Tms
- Season
- Game
- Week

There were a few scores missing for the Kansas City Chiefs which were simply replaced with the average of each respective column.

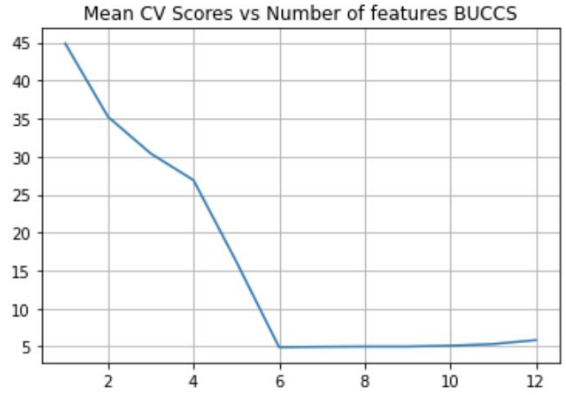
The next step was to determine which of the variables in the data set would help accurately predict the Superbowl outcome. Using the mlxtend package in Python, SequentialFeatureSelector was used to find how many and which variables would help give the final model the best performance. Figure 1 shows that for the Chiefs, 6 variables were deemed to be important in the performance of the Linear Regression model.



```
No of features= 6
[0, 5, 10, 11, 12, 13]
Features selected in forward fit
Index(['OppScore', 'TO Offense', 'TO Defense', 'Offense', 'Defense',
       'Sp. Tms'],
      dtype='object')
```

Figure 1

Figure 2 shows that there were 5 variables that were deemed to be important in the performance of the Linear Regression model.



```
No of features= 5
[0, 9, 10, 11, 12]
Features selected in forward fit
Index(['Week', 'PassY Defense', 'RushY Defense', 'TO Defense', 'Offense'], dtype='object')
```

Figure 2

The variables selected will later be used to create Linear Regression models for each team. With those models, we will feed in forecasted numbers, which are discussed next, and determine the final scores for each team.

In order to provide Superbowl score results, it is necessary to know what the outcomes would be for each team in terms of input values/variables in the final model. Here we will use the ETS, or Error, Trend, Seasonal model to forecast the parameters needed for the final Linear Regression model. We will use a Simple Exponential Smoothing model. This was chosen over an ARIMA model because logically and domain knowledge wise, team performance can improve over the course of the NFL season. The forecasts produced through this process are weighted averages of past observations with more weight placed on more recent observations.

From feature selection the following variables determined and were predicted through ETS models:

Bucks 2: ● OppScore ● Rush YDefense ● TO Defense ● Offense ● Defense	Chiefs 2: ● OppScore ● TO Defense ● Offense ● Defense ● Sp. Tms

Mira Daya
Due: February 6, 2021
ID: 47486312
BANA 290: Forecasting

Included in Regression	
● Week	● Week

After forecasting the selected features, we need to build our final Linear Regression model for each team. Below are the equations that were produced through the Linear Regression. Overall, we can see that both teams have different factors that may affect their final scores. The Chiefs have more emphasis on their Opponents Score while the Buccs have more emphasis on turnovers gained by defense defense. This is to say that each team is very unique and there can be millions of factors that help predict their score.

Regression Models:

$$Score_{Chiefs} = 0.052Week + 0.914OppScore + 0.253TOOffense + 0.411TODefense + \\ + 0.919Offense + 0.939defense + 0.899Sp.Tms$$

$$R^2 = 0.918 \ MSE = 14.370 \ MAPE = 6.067$$

$$Score_{Buccs} = -0.09Week + 0.799OppScore + 0.0007RushYDefense + 2.273TODefense \\ + 0.829Offense + 0.591Defense$$

$$R^2 = 0.9241 \ MSE = 10.747 \ MAPE = 23.49$$

The regression models have very high R^2 which would be good news if it were not for potential overfitting. There needs to be further refinement of the variables used and the relationship between all independent variables to each other and to the dependent variable. Additionally, the MAPE, mean absolute percent errors are listed above as well. For the chiefs the MAPE is approximately 6, meaning, on average, the forecast is off by 6%. For the Buccs it is higher as off by 23.49%. This is interesting to note that the Buccs model has a higher R^2 but also less accuracy according to the MAPE. Further analysis is needed!

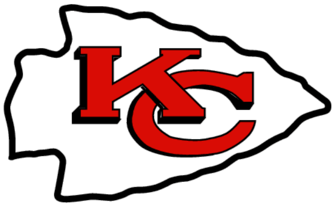
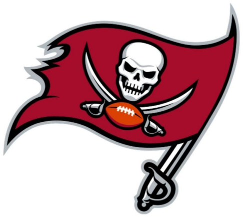
Future Steps:

1. Further analyze what variables are truly significant in predicting a team's score
2. Gather more data for analysis accordingly
3. Fine tune models and understand the statistics behind them better
4. Create better model, perhaps consider home/away variable and somehow find team ranking statistics
5. Watch SB55!

Final Score Below!

Mira Daya
Due: February 6, 2021
ID: 47486312
BANA 290: Forecasting

After plugging in ETS results for the variables the following outcomes are predicted.

	32
	26