Machine Learning Engineer Nanodegree Program

Federal Reserve Interest Rates

Jovani Martinez Rico - June 11, 2017

Capstone proposal

1.Definition

- ### Project Overview

Any decision by the The Federal Reserve directly affects mortgage, credit card, or prices rising in the product that we buy. Therefore the Federal Reserve sets interest rates to promote conditions that achieve high employment, low and stable inflation, sustainable economic growth, and moderate long-term interest rates. The Fed monitors inflation indicators to manage inflation[1]. When indicators rise more than 3% a year, the Fed raises the federal funds rate to keep rising prices under control.

Higher interest rates mean higher borrowing costs, so consumers and businesses borrow less and spend less. Demand for goods and services drop, and inflation falls. On the other hand, falling interest rates result when the Fed lowers the federal funds rate. Borrowing becomes cheaper and people spend more. This can end a recession[2].

Knowing the interest rate could help companies to decide if they invest or not, help people to get a new mortgage or help people to manage better their money.

[1] What are the Federal Reserve's objectives in conducting monetary policy?(15 March 2017), Board of Governors of the Federal Reserve System.

[2] How does monetary policy influence inflation and employment?(16 December 2015), Board of Governors of the Federal Reserve System.

• Problem Statement

Looking at the country's unemployment data as well as gross domestic product and inflation rates you may try to predict the following Fed changes in the interest rate. This is a clear example of a regression problem, making a regression along of the variables we can try to predict the expecting interest rate.

Trying to predict the next interest rate can be beneficial for both a company and a person. This can help you make a better decision on their finances, they can invest their money or save it for a future.

• Solution Statement

To construct the model I will use a supervised learning algorithm, a regression is the best option to predict the next interest rate. Using the consumer price index, gross domestic product and the unemployment rate as independent variables the model will predict the interest rate.

Using different supervised algorithms will be evaluated to see which one make a better prediction, the algorithms that I will test will be KNeighborsRegressor, LinearRegression, Ridge Regression and Lasso.

• Metrics

R2 is the ratio of the explained variation compared to the total variation; thus, it is interpreted as the fraction of the sample variation in y that is explained by x.

If the data points all lie on the same line, OLS provides a perfect fit to the data. In this case, R2 = 1. A value of R2 that is nearly equal to zero indicates a poor fit of the OLS line: very little of the variation in the y_i is captured by the variation in the y^i (which all lie on the regression line). In fact, it can be shown that R2 is equal to the square of the sample correlation coefficient between y_i and y_hat_i.

To evaluate the performance of the model, I will use the R-squared to check how well the model is getting, since the highest value of the r-squared is 1 and the lowest is 0, if we got a value close to 1 that means our model is well defined and we have enough variables to predict our dependent variable.

Key Limitation of R-squared, *cannot* determine whether the coefficient estimates and predictions are biased, which is why you must assess the residual plots.

One trap of R-squared is that it can only increase as predictors are added to the regression model. This increase is affected when predictors are not actually improving the model’s fit.

2.Analysis

- ## Data Exploration

The data set for this project was obtained from Kaggle[3], but this data can be obtained directly by the Federal Reserve Bank of St. Louis' economic data portal. This dataset includes 904 observation with 7 variables, this has data on the economic conditions in the United States on a monthly basis since 1954.

	year	month	day	federalfun~e	fe~pertarget	fe~wertarget	effectivef~e	realgdppe~e	unemployme~e	inflationr~e
894	2016	7	1	.	.5	.25	.39	3.5	4.9	2.2
895	2016	8	1	.	.5	.25	.4	.	4.9	2.3
896	2016	9	1	.	.5	.25	.4	.	4.9	2.2
897	2016	10	1	.	.5	.25	.4	1.9	4.8	2.1
898	2016	11	1	.	.5	.25	.41	.	4.6	2.1
899	2016	12	1	.	.5	.25	.54	.	4.7	2.2
900	2016	12	14	.	.75	.5	.	.	.	.
901	2017	1	1	.	.75	.5	.65	.	4.8	2.3
902	2017	2	1	.	.75	.5	.66	.	4.7	2.2
903	2017	3	1	.	.75	.5	.	.	.	.
904	2017	3	16	.	1	.75	.	.	.	.

The federal funds rate is the interest rate at which depository institutions trade federal funds with each other overnight. When a depository institution has surplus balances in its reserve account, it lends to other banks in need of larger balances.

The rate that the borrowing institution pays to the lending institution is determined between the two banks; the weighted average rate for all of these types of negotiations is called the effective federal funds rate. The effective federal funds rate is essentially determined by the market but is influenced by the Federal Reserve through open market operations to reach the federal funds rate target.[4]

[3] (n.d.). Retrieved June 09, 2017, from http://blog.datadive.net/selecting-good-features-part-iii-random-forests/

[4] Board of Governors of the Federal Reserve System. (n.d.). Retrieved June 10, 2017, from http://www.federalreserve.gov/monetarypolicy/default.htm

The real gross domestic (GDP) is the value of the goods and services produced by the nation's economy less the value of the goods and services used up in production. This is calculated as the seasonally adjusted quarterly rate of change in the gross domestic product based on chained 2009 dollars

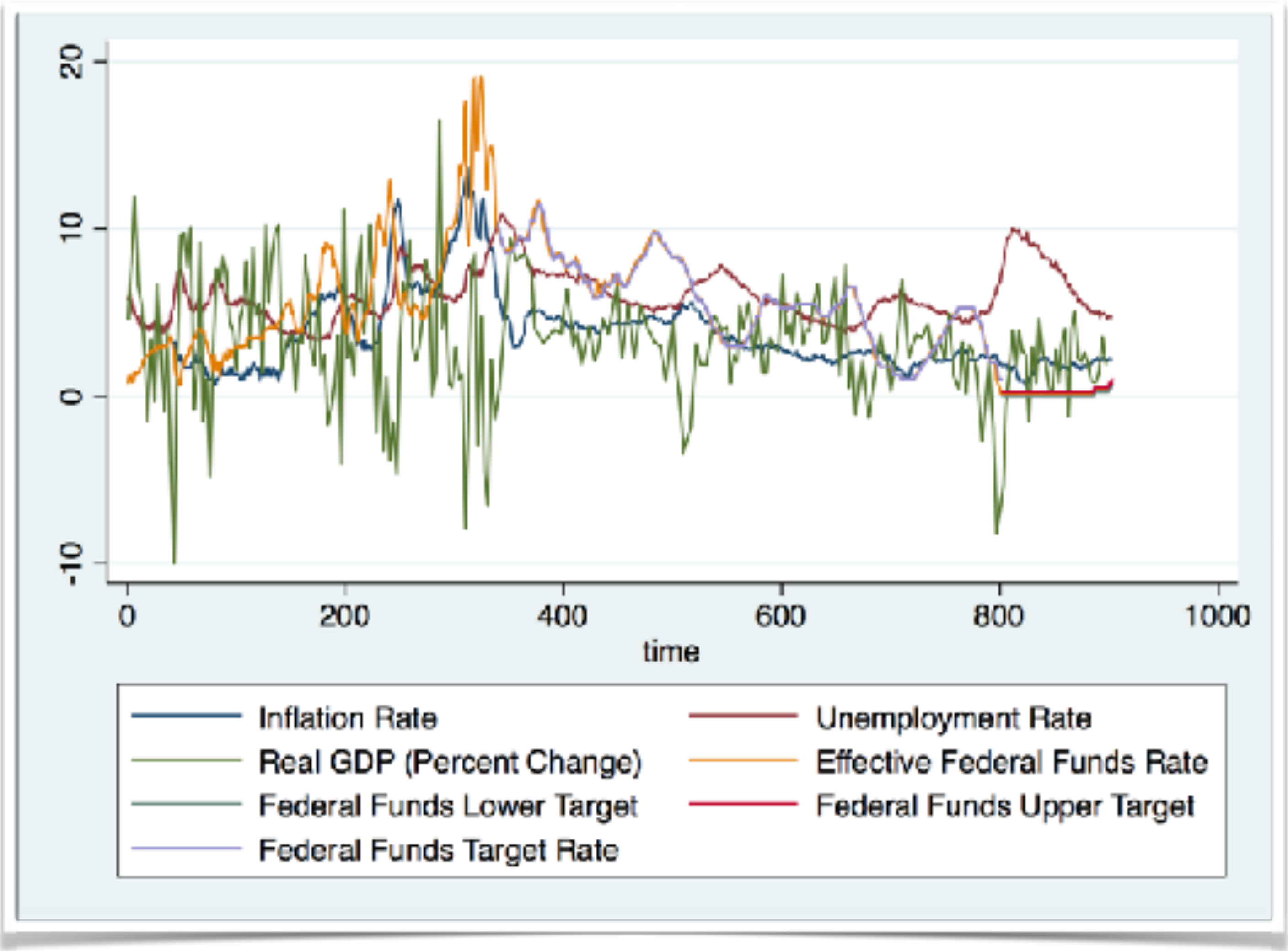
The unemployment rate represents the number of unemployed as a seasonally adjusted percentage of the labor force. The inflation rate reflects the monthly change in the Consumer Price Index of products excluding food and energy.

Below are some descriptive statistics to have as first insight into the shape of each attribute. It is worth taking some time to review observations initially, there might be the presence of "NA" values for missing data or surprising distributions of attributes.

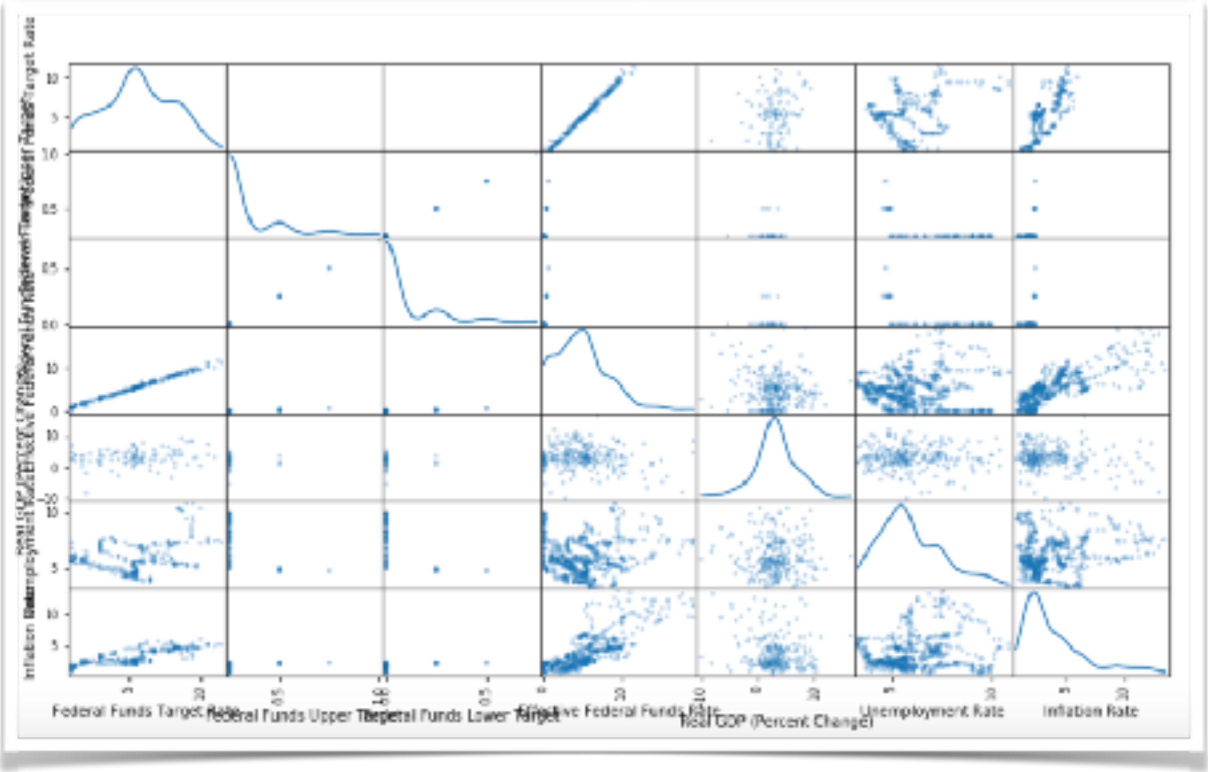
	Month	Day	Federal Funds Target Rate	Federal Funds Upper Target	Federal Funds Lower Target	Effective Federal Funds Rate	Real GDP (Percent Change)	Unemployment Rate	Inflation Rate
count	904.000000	904.000000	462.000000	103.000000	103.000000	752.000000	250.000000	752.000000	710.000000
mean	6.598451	3.598451	5.658415	0.309252	0.059252	4.910997	3.138000	5.979122	3.733380
std	3.471983	6.791079	2.553891	0.140917	0.140917	3.611431	3.598521	1.567641	2.573637
min	1.000000	1.000000	1.000000	0.250000	0.000000	0.070000	-10.000000	3.400000	0.600000
25%	4.000000	1.000000	3.750000	0.250000	0.000000	2.427500	1.400000	4.900000	2.000000
50%	7.000000	1.000000	5.500000	0.250000	0.000000	4.700000	3.100000	5.700000	2.800000
75%	10.000000	1.000000	7.750000	0.250000	0.000000	6.580000	4.875000	7.000000	4.700000
max	12.000000	31.000000	11.500000	1.000000	0.750000	19.100000	16.500000	10.800000	13.600000

- ## Exploratory Visualization

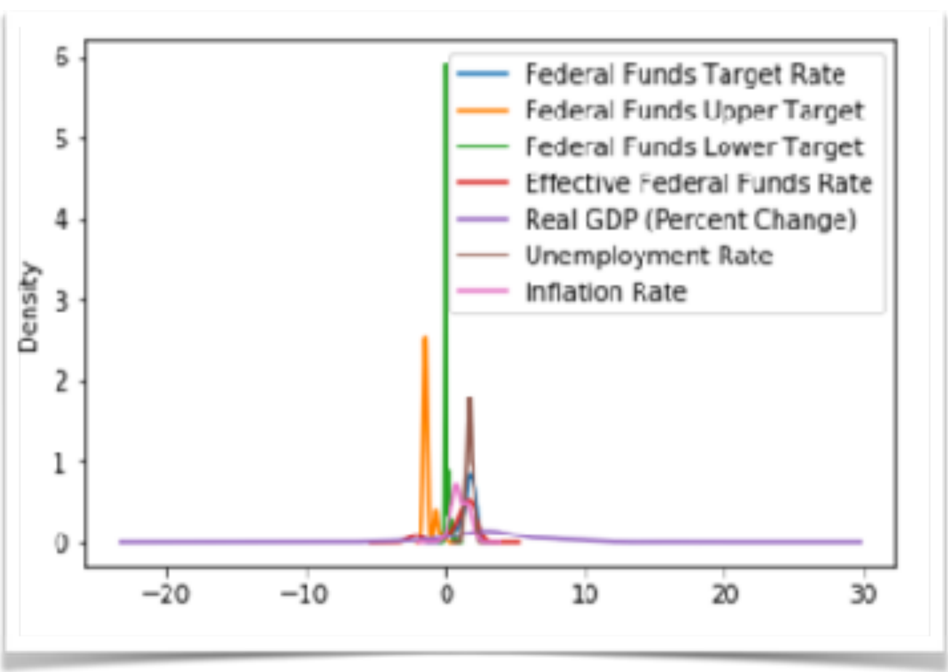
In the graph below we can see the representation of the data set, on the graph we can see the relationship between two or more variables.



To better understand the dataset, a scatter matrix was created. The graph shows each of the features present in the data. Although the graph below shows that there might be a correlation between “Federal Founds Target Rate” and “Effective Federal Founds Rate” features.The graph shows the skewness of the data, we can see this at the diagonal of the scatter matrix plot.



In the following graph, kernel density estimation (KDE) we can see the the probability density of the variables, this help us to know where inferences about the population are made, based on the data set.

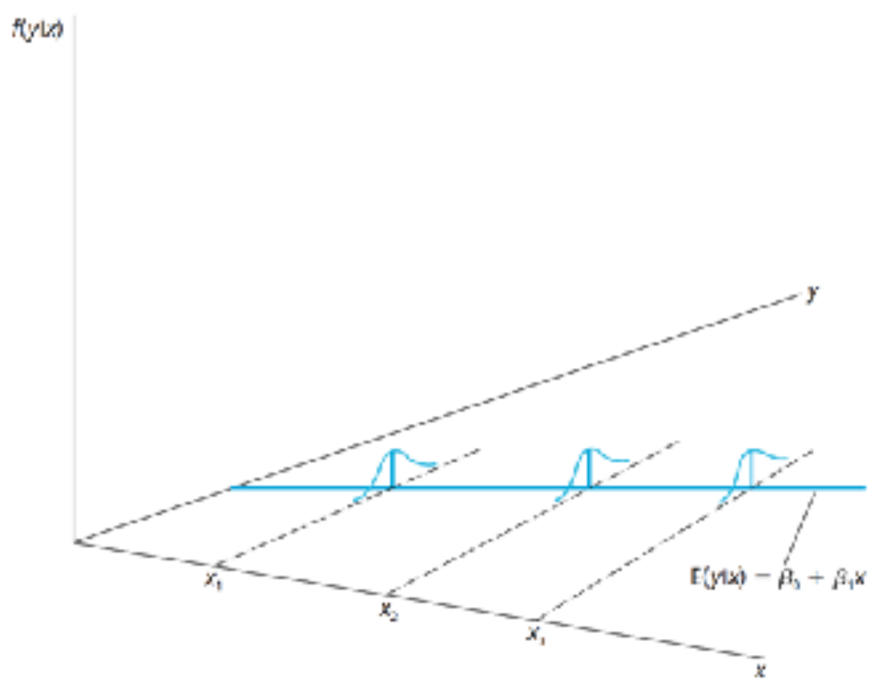


• Algorithms and Techniques

1. **Linear Regression**.

Linear Regression method uses Ordinary least squares (OLS) or linear least squares to minimize the residual sum of squares between the observed responses in the dataset, and the responses predicted by the linear approximation. This method is one of the most use in econometrics to studying how *y* varies with changes in *x. The least-squares estimate becomes highly sensitive to random errors (Heteroskedasticity) in the observed response, producing a large variance.*

The following graph represent a simple regression model under homoskedasticity (The error term has the same variance given any value of the explanatory variable).



2. HuberRegressor

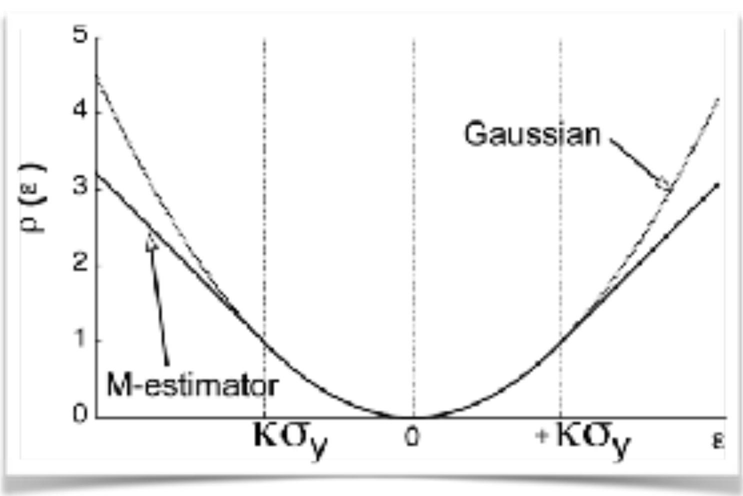
In 1973, Huber introduced M-estimation for regression. The M in M-estimation stands for "maximum likelihood type". The method is robust to outliers in the response variable, but turned out not to be resistant to outliers in the explanatory variables (leverage points). In fact, when there are outliers in the explanatory variables, the method has no advantage over least squares[5].

M-estimators are a broad class of estimators, which are obtained as the minima of sums of functions of the data. Least-squares estimators are a special case of M-estimators. The definition of M-estimators was motivated by robust statistics, which contributed new types of M-estimators. The statistical procedure of evaluating an M-estimator on a data set is called M-estimation[6].

$$\sum_{i=1}^{n} \rho(x_i, \theta),$$

where ρ is a function with certain properties. The solutions

$$\hat{\theta} = \arg\min_{\theta} \left(\sum_{i=1}^{n} \rho(x_i, \theta) \right)$$



[5] Robust regression. (2017, June 07). Retrieved June 10, 2017, from https://en.wikipedia.org/wiki/Robust_regression

[6] M-estimator. (2017, May 19). Retrieved June 10, 2017, from https://en.wikipedia.org/wiki/M-estimator

Because the data can has outliers I think this method is useful for the predictions.

3. RandomForestRegressor

Random forests are among the most popular machine learning methods thanks to their relatively good accuracy, robustness and ease of use. They also provide two straightforward methods for feature selection: mean decrease impurity and mean decrease accuracy.

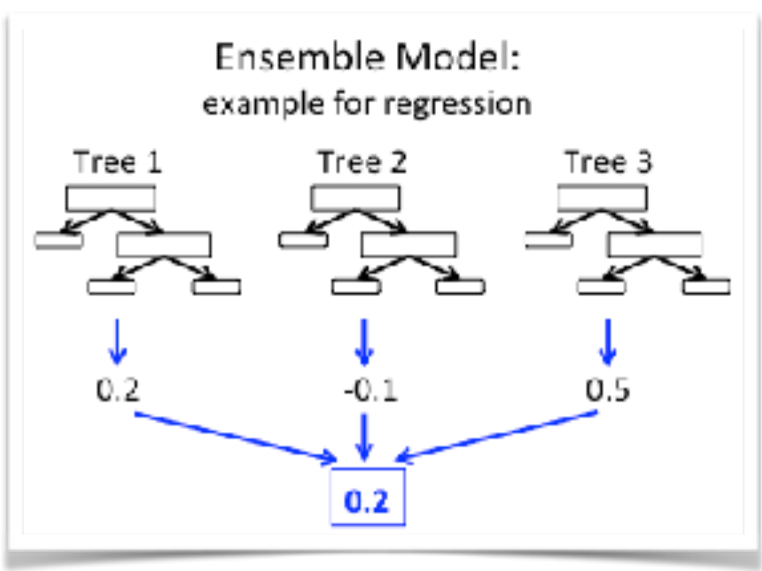
- Mean decrease impurity

Random forest consists of a number of decision trees. Every node in the decision trees is a condition on a single feature, designed to split the dataset into two so that similar response values end up in the same set. The measure based on which the (locally) optimal condition is chosen is called impurity.

- Mean decrease accuracy

Another popular feature selection method is to directly measure the impact of each feature on accuracy of the model.

But come with their own gotchas, With correlated features, strong features can end up with low scores and the method can be biased towards variables with many categories[7].



[7] (n.d.). Retrieved June 10, 2017, from http://blog.datadive.net/selecting-good-features-part-iii-random-forests/

• Benchmark

Sklearn provides a DummyRegressor which was used as a baseline. DummyRegressor is a regressor that makes predictions using simple rules.

DummyClassifier implements four simple rules of thumb for regression:

- mean always predicts the mean of the training targets.
- median always predicts the median of the training targets.
- quantile always predicts a user provided quantile of the training targets.
- constant always predicts a constant value that is provided by the user.

In all these strategies, the predict method completely ignores the input data[8].

Comparing the R-squared scores of the DummyRegressor with other algorithms like LinearRegression, HuberRegressor and RandomForestRegressor I tried to beat its performance. The R-squared score of this dummy regressor was obtained using the same set of training and testing sets as the real classifier.

In order to assume that the model is good, the $\mathbf{R}^2$ scores obtained from models created need to show that significantly outperform the DummyRegressor results.

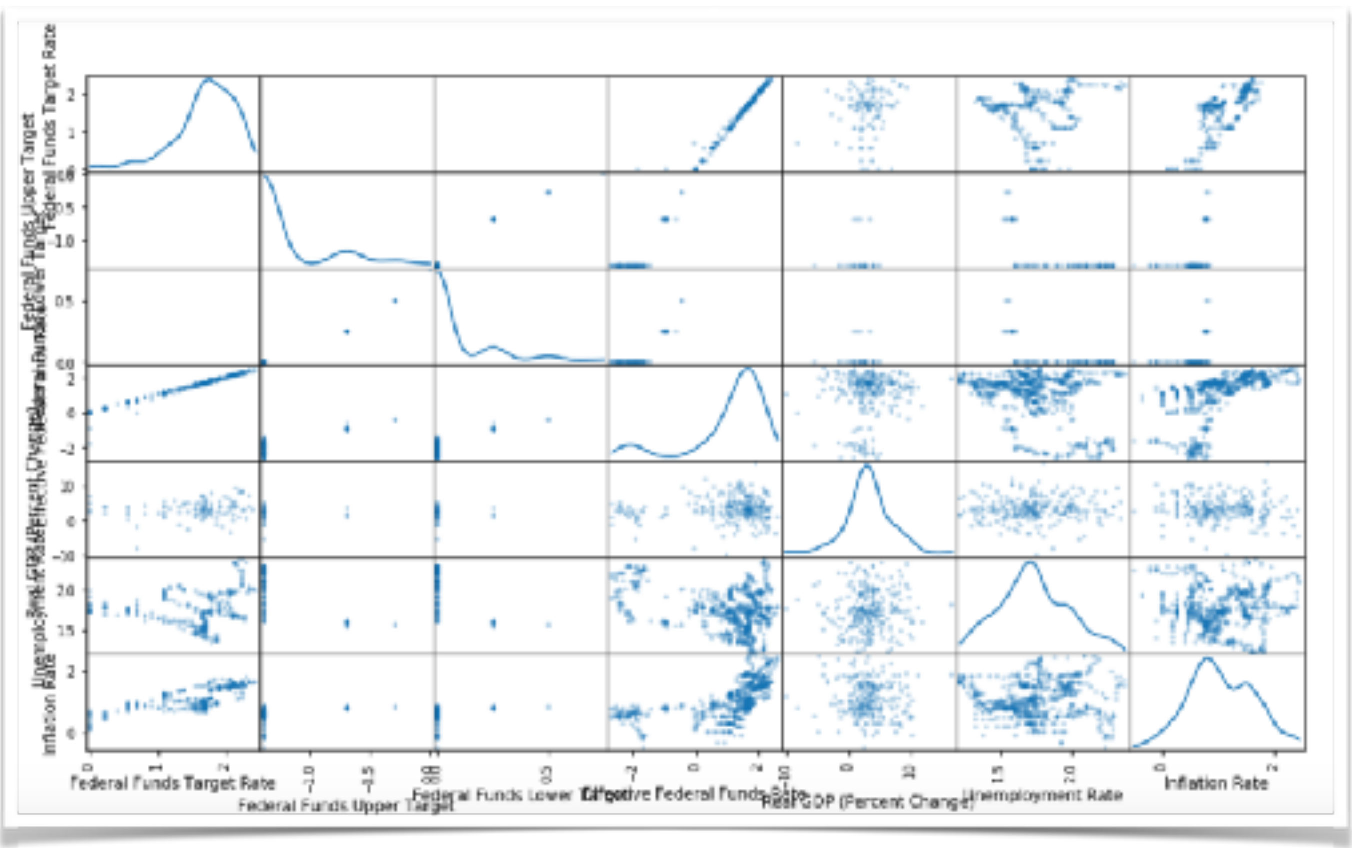
8 Sklearn.dummy.DummyRegressor¶. (n.d.). Retrieved June 10, 2017, from http://scikit-learn.org/stable/modules/generated/sklearn.dummy.DummyRegressor.html

3.Methodology

• Data Preprocessing

Once the dataset was obtained from the Kaggle the following steps were performed to prepare the data before creating the models:

- In the dataset, we used the ‘datetime’ variable as the index column for the dataset, I combined the Year, Month and Day variables to form this feature.
- Because of the positive skewed and to have a more normal distribution, I took log of some variables (Federal Funds Target Rate, Effective Federal Funds Rate, Unemployment Rate, Inflation Rate, Federal Funds Target Rate and Federal Funds Upper Target).



- To have a better model, I took the data after the year 1983, because in the 80’s it was a global economic recession[9]. This model won’t work under recession.
- Separate the target variable (Effective Federal Funds Rate) from the rest features.
- Replace all the NA values for zeros.
- Split the data into two different sets, training and testing datasets respecting the order of the entries in the dataset. The testing dataset consists of 33% of the size of the original dataset.

• Implementation

The implementation of the Interest Rate predictor model was performed executing the following steps:

1. The load the data from a file.
2. Split historical data collected into two datasets, training and testing data.
3. Training the model using the historical data.
4. Apply different supervised learning algorithms.
5. Measure the score and the performance of each algorithm.
6. Tune the model, if necessary, to get better scores results.
7. Select the algorithm that provides a better score.

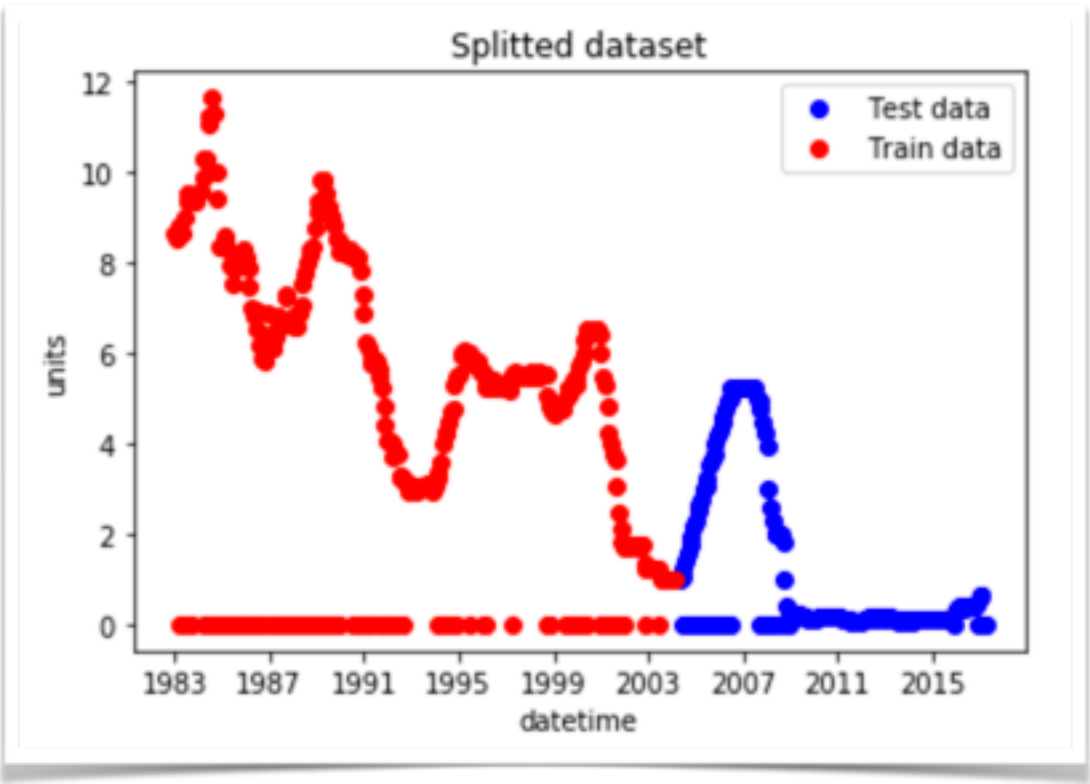
Supervised Learning Algorithms

Different algorithms were applied to select the algorithm that provides a better score. The following supervised learning models that were used are available in scikit-learn:

- Linear Regression.
- HuberRegressor
- RandomForestRegressor.

[9] Early 1980s recession. (2017, June 06). Retrieved June 10, 2017, from https://en.wikipedia.org/wiki/Early_1980s_recession

The nature of the model is the use of a model to predict future values based on previously observed values we have Time series forecasting. I used TimeSeriesSplit from Sklearn to split the data into training and testing data. The following graph represent the data divided, the red data is the training data and the blue points is the testing data.



- Refinement

After the default model of Linear Regression, Huber regressors and Random Forest regressor obtained a better R2 score in each one. The models were tuned using different parameters values of each regressor using grid search from Sklearn.

Linear Regression Fine Tune

For the Linear Regression model, also known as ordinary least squares (OLS), a few parameters were considered to improve the **R^2 score** of the model. The following table describe the parameters and the values that grid search identified.

Parameter	Description	Value
fit_intercept	Whether to calculate the intercept for this model. If set to false, no intercept will be used in calculations .	True,False
normalize	If True, the regressors X will be normalized before regression. This parameter is ignored when fit_intercept is set to False. When the regressors are normalized, note that this makes the hyperparameters learned more robust and almost independent of the number of samples.	True,False
copy_X	If True, X will be copied; else, it may be overwritten	True,False

After applying grid to the Linear Regression model the initial R2 score was slightly improved from 0.378437341433 to 0.378437341433.

HuberRegressor

Huber Regressor is a linear model that optimizes the squared loss for the samples. For this model, some parameters were considered to improve the R2 score of the model. The following table describe the parameters and the values that were applied to perform the grid search.

Parameter	Description	Value
epsilon	The parameter epsilon controls the number of samples that should be classified as outliers. The smaller the epsilon, the more robust it is to outliers.	1.2,1.35
alpha	Regularization parameter.	0.01, 0.001, 0.0001, 0.00001
warm_start	This is useful if the stored attributes of a previously used model has to be reused. If set to False, then the coefficients will be rewritten for every call to fit.	True,False
fit_intercept	Whether or not to fit the intercept. This can be set to False if the data is already centered around the origin.	True,False

After applying grid to the Linear Regression model the initial R2 score was slightly improved from 0.414764348592 to 0.414764348592.

RandomForestRegressor

RandomForest Regressor is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and use averaging to improve the predictive accuracy and control over-fitting. For this model, some parameters were considered to improve the R2 score of the model. The following table describe the parameters and the values that were applied to perform the grid search.

Parameter	Description	Values
random_state	If int, random_state is the seed used by the random number generator; If RandomState instance, random_state is the random number generator; If None, the random number generator is the RandomState instance used by np.random.	1,42
n_estimators	The number of trees in the forest.	200,100,50,10,150
warm_start	When set to True, reuse the solution of the previous call to fit and add more estimators to the ensemble, otherwise, just fit a whole new forest.	True,False

After applying grid to the Linear Regression model the initial R2 score was slightly improved from 0.770937114391 to 0.774603967025

.

4.Results

• Model Evaluation and Validation

This project explored further Random Forest Regression model because this model provided a far better **R^2 score** than the other models explored, Linear Regression, and Huber Regression, even before that the tuning process was applied.

A grid search was used to tune Random Forest Regression model and obtained the better estimators; we can see a summary in the following table.

Model	Parameters	R2 Score
Random Forest Regression.	random_state': [1,42],'n_estimators': [200,100,50,10,150], 'warm_start':(True,False)	0.774603967025

The difference between the two different models is minimal (default model and tuning model). The Tuning Random Forest Regression model **R^2 score** is slightly above the default Random Forest Regressor model by **~0.4%**. As the purpose of this project is to predict the stock price I believe that having a more precise model is a better idea. In this case, I decided to use the Tanning Random Forest Regressor because it produces a **R^2 score** that it closer to 1.0. This model was tested using different parameters of unemployment rate, Inflation rate and Real GDP.

I believe that this model needs to be improved to be used to make real life trading decisions. For instance, similar rates of unemployment, Inflation and GDP could be grouped together to create a model that would discover new features that can impact the interest rate. Interest rate predictions are so complex that we need to build a more robust dataset to be able to build a model that it can predict the rates more accurate. Having a machine learning model like this is a good starting point, however, it needs improvement.

• Justification

After tuning both models, Linear Regression, Huber Regression and Random Forest Regression, the R2 score obtained significantly outperform the R2 score obtained by the DummyRegressor.

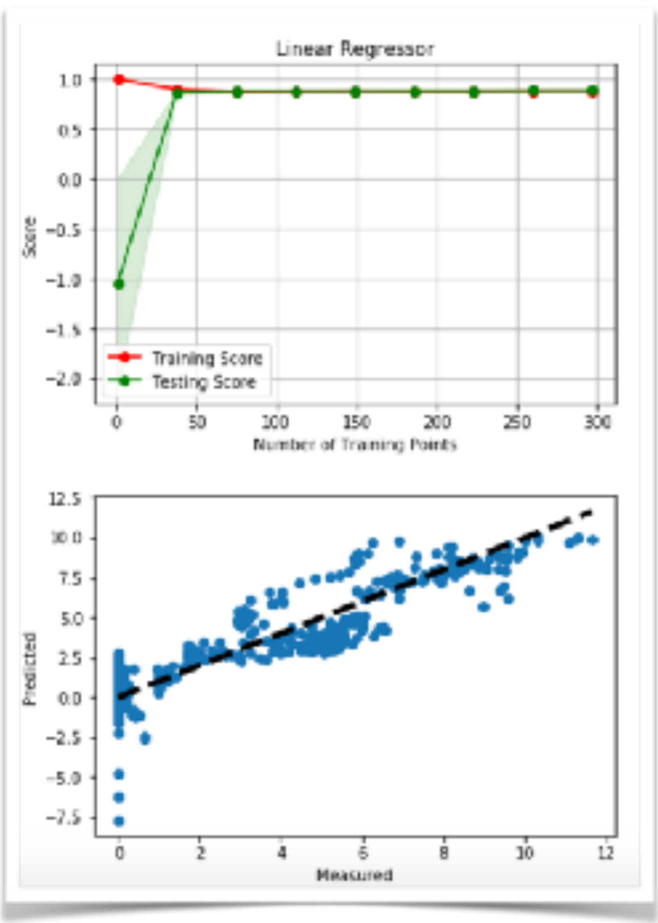
Model	Parameters	R2 Score
Dummy Regressor	constant=None, quantile=None, strategy='mean'	-2.56281297338
Linear Regression.	fit_intercept': (True, False), 'normalize':(True, False), 'copy_X':(True, False)	0.378437341433
Huber Regression.	epsilon':[1.2,1.35],'alpha': [0.01, 0.001, 0.0001, 0.00001],'warm_start': (True,False), 'fit_intercept': (True, False)	0.414764348592
Random Forest Regression.	random_state': [1,42],'n_estimators': [200,100,50,10,150], 'warm_start':(True,False)	0.774603967025

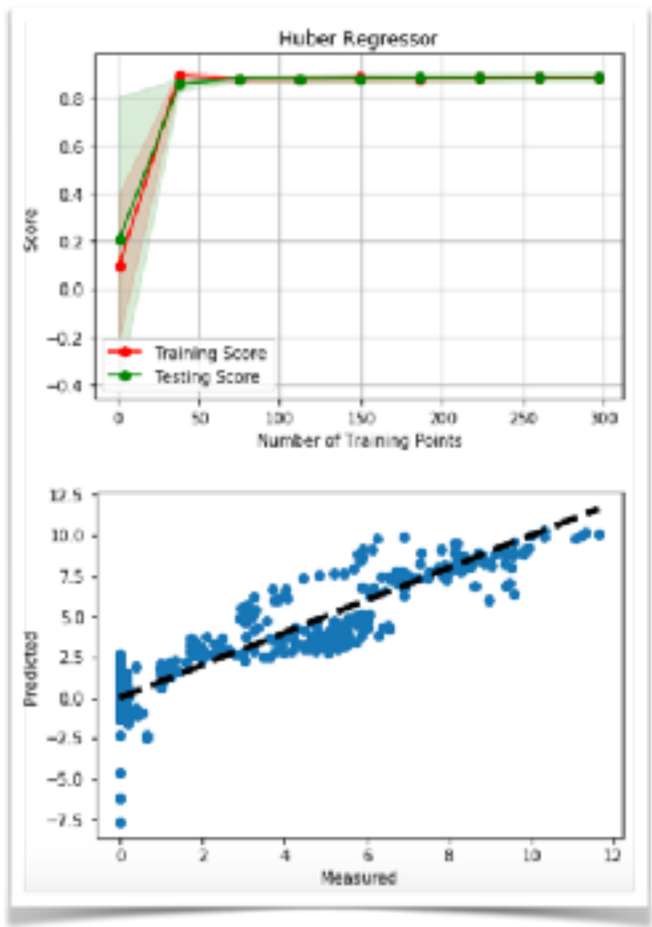
As we can see in the table above, the models were able to outperform the DummyRegressor used for the benchmark for this project. However, while building this project, it was noted that to create a model that predicts the interest rate more accurately, it is necessary to obtain a more robust dataset to captures other important features that help to predict the Fed interest rate.

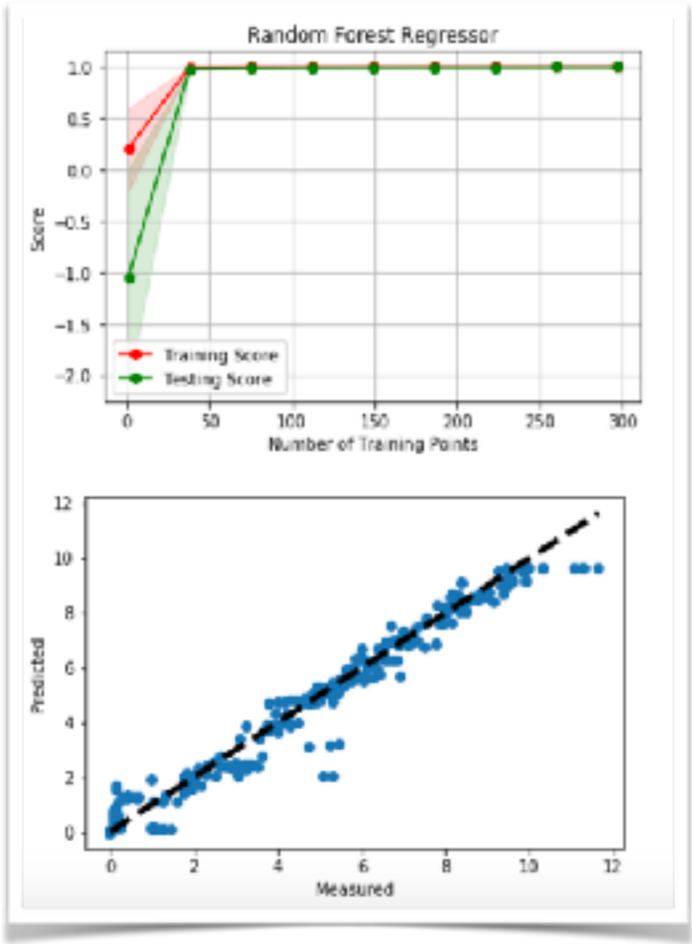
5.Conclusion

- Free-Form Visualization

After tuning both models, Linear Regression, Huber Regression and Random Forest Regression models, the **R^2 score** obtained was close to 1.0, which the possible score for the **R^2** score.







We can see in the graphs above that after approximately 50 training points the R2 score were 1.0 for the Random Forest Regressor, and closed to 50 training points the R2 score were closed to 1.0 for the Linear Regressor and Huber regressor . However, this high R2 score does not mean that the model has a good fit. After obtaining these scores for the models, I believe that scores values were good because we can see the distribution of the points along the regression line for each model. In the Reflection section I will explore some possibilities that can be causing these scores.

- # Reflection

Creating a regression to predict the Fed Interest rate is a hard problem to solve. This project only shows a first step in the process of creating a more robust model. Also it's important to realize that the split process in different for this type of regressions where time series data is present. Splitting the data respecting the order of the entries is a key important task to solve the interest rate. For this type of problem we can not split it randomly using sklearn train_test_split. When dealing with time series data, we have to split the data maintaining the order in the data, and split at a certain point in time. In this case we used the last 33% of the data for testing. Is important to noticed that I only used the data since the year 1983 and after, I tried to avoid the data of Early 1980s recession because this won't help to the model, this would give us heteroskedasticity (common in time series data).

The steps taken to create the model are the following:

- Explored different Supervised Machine Learning algorithms.
- Obtain the datasets from Kaggle.
- Explore the dataset using different plots such as scatter matrix plot.
- Dataset preprocessing:
    1. Extract the target variable from the original dataset.
    2. Perform feature scaling using natural logarithm.
- Split the dataset into training and testing datasets respecting the ordered list of observations.
- Use and calculate the benchmark for the dataset, DummyRegressor from sklearn.
- Evaluate the following regressor models from sklearn.
    1.Linear Regression.
    2.Huber Regressor.
    3.Random Forest Regression.
- Calculate the **R^2 scores** and determine the top model.
    1.Random Forest Regression.
- Tune the top model using Grid Search.
- Determine the best model base of the score results.
    1.Random Forest Regression.

As we discuss earlier high **R^2 scores** doesn't necessarily mean that the model is more precise. Some factors that could be causing these low **R^2 scores** in the Linear and Huber regression are:

•**Recession**: Because under recession, the mean of the error term is not the same across all point along the regression line, therefore for some point the variance is greater, which give us different result on the predictions.

• **Multicollinearity**: is a phenomenon in which two or more predictor variables in a multiple regression model are highly correlated, meaning that one can be linearly predicted from the others with a substantial degree of accuracy. The precision of the estimated regression coefficients decreases as more predictors are added to the model.

The low R-squared that even noisy, high-variability data can have a significant trend. The trend indicates that the predictor variable still provides information about the response even though data points fall further from the regression line.

The table below shows the correlation between the all the features and the target variable in the original dataset used to train the models described earlier.

	Federal Funds Target Rate	Federal Funds Upper Target	Federal Funds Lower Target	Effective Federal Funds Rate	Real GDP (Percent Change)	Unemployment Rate	Inflation Rate
Federal Funds Target Rate	1.000000	NaN	NaN	0.996450	0.207719	0.282460	0.761163
Federal Funds Upper Target	NaN	1.000000	1.000000	0.937782	0.016265	-0.557939	0.429528
Federal Funds Lower Target	NaN	1.000000	1.000000	0.937782	0.016265	-0.557939	0.429528
Effective Federal Funds Rate	0.996450	0.937782	0.937782	1.000000	-0.101943	0.036626	0.784376
Real GDP (Percent Change)	0.207719	0.016265	0.016265	-0.101943	1.000000	-0.025781	-0.181342

	Federal Funds Target Rate	Federal Funds Upper Target	Federal Funds Lower Target	Effective Federal Funds Rate	Real GDP (Percent Change)	Unemployment Rate	Inflation Rate
Unemployment Rate	0.282460	-0.557939	-0.557939	0.036626	-0.025781	1.000000	0.206522
Inflation Rate	0.761163	0.429528	0.429528	0.784376	-0.181342	0.206522	1.000000

By looking at the table above we can see that the features “Federal Funds Upper Target”, “Federal Funds lower Target”, and “Effective Federal Funds Rate” are highly correlated. Since multicollinearity causes imprecise estimates of coefficient values in a regressor, the resulting predictions will also be imprecise. These are the reasons why I believe the current dataset needs to be improved.

- # Improvement

Fed interest rate forecasting is getting common and indeed San Francisco Fed has announced that companies can predict their next interest rate[10]. Finding the right time to get a new credit (credit card, mortgage, any other credit) for a person or a company can be a hard problem to solve.Movements in the federal funds rate are passed on to other short-term interest rates that influence borrowing costs for firms and households. If the interest rates go down, it becomes cheaper to borrow, so households are more willing to buy goods and services and firms are in a better position to purchase items to expand their businesses, such as property and equipment.

After the creation of the model for this project, it was clear to see that the original dataset needs to be extended to capture more features that can help to create a more realistic model. The Federal Reserve has set a 2% inflation target but there are many measures for the inflation rate, there are eight time series data that receive a lot of attention in the context of

[10] Turner, Matt. "The Fed Just Admitted That a Startup Can Predict Its Upcoming Interest Rate Forecasts." *Business Insider*. Business Insider, 08 Sept. 2016. Web. 11 June 2017.

policy[11]. Another important thing to consider for the model is the recession of the country, this can make the prediction to be not as good as expected.

Taking all this in consideration, there is possible to make a close prediction for the next interest rate of the Fed and know the best time to invest.

[11] "The FRED® Blog." *FRED Blog*. Web. 11 June 2017.

6.References

- What are the Federal Reserve's objectives in conducting monetary policy?(15 March 2017), Board of Governors of the Federal Reserve System.
- How does monetary policy influence inflation and employment?(16 December 2015), Board of Governors of the Federal Reserve System.
- (n.d.). Retrieved June 09, 2017, from http://blog.datadive.net/selecting-good-features-part-iii-random-forests/
- Board of Governors of the Federal Reserve System. (n.d.). Retrieved June 10, 2017, from http://www.federalreserve.gov/monetarypolicy/default.htm
- Robust regression. (2017, June 07). Retrieved June 10, 2017, from https://en.wikipedia.org/wiki/Robust_regression
- M-estimator. (2017, May 19). Retrieved June 10, 2017, from https://en.wikipedia.org/wiki/M-estimator
- (n.d.). Retrieved June 10, 2017, from http://blog.datadive.net/selecting-good-features-part-iii-random-forests/
- Sklearn.dummy.DummyRegressor¶. (n.d.). Retrieved June 10, 2017, from http://scikit-learn.org/stable/modules/generated/sklearn.dummy.DummyRegressor.html
- Early 1980s recession. (2017, June 06). Retrieved June 10, 2017, from https://en.wikipedia.org/wiki/Early_1980s_recession
- FRED Graph. (n.d.). Retrieved June 10, 2017, from https://fred.stlouisfed.org/graph/?g=8dGq&utm_campaign=graph&utm_content=&utm_medium=Referral&utm_source=fred.stlouisfed.org&utm_term=related_resources
- Civilian Unemployment Rate. (2017, June 02). Retrieved June 10, 2017, from https://fred.stlouisfed.org/series/UNRATE
- Real Gross Domestic Product. (2017, May 26). Retrieved June 10, 2017, from https://fred.stlouisfed.org/series/A191RL1Q225SBEA
- Effective Federal Funds Rate. (2017, June 01). Retrieved June 10, 2017, from https://fred.stlouisfed.org/series/FEDFUNDS
- (n.d.). Retrieved June 10, 2017, from https://www.federalreserve.gov/faqs/money_12856.htm
- (n.d.). Retrieved June 10, 2017, from https://www.federalreserve.gov/faqs/money_12856.htm

- Brownlee, Jason. (2016). Metrics To Evaluate Machine Learning Algorithms in Python. Retrieved June 8, 2017, from http://machinelearningmastery.com/metrics-evaluate-machine-learning-algorithms-python/
- Multicollinearity. (2017). In Wikipedia, The Free Encyclopedia. Retrieved June 5, 2017, from https://en.wikipedia.org/w/index.php?title=Multicollinearity&oldid=776794468
- Turner, Matt. "The Fed Just Admitted That a Startup Can Predict Its Upcoming Interest Rate Forecasts." Business Insider. Business Insider, 08 Sept. 2016. Web. 11 June 2017.
- "The FRED® Blog." FRED Blog. Web. 11 June 2017.