Predicting the Federal Fund Effective Rates: A Machine Learning Approach

Derick Hanscom, Jharold Jose Montoya Villalta,

Tristan Ko, Kaiyu Wen, Yuqiang Zhang, Leyi Hu

**Professor Chris Kelliher & Professor Eugene Sorets**

MF 728 Fixed Income

May 1st, 2023

**Table of Contents**

Introduction……………………………………………………………….………………....1

Data Cleaning and Preparation…….……………………………………………….………..1

Machine Learning Models Building……………………………..…………………………..4

Result…….……………………………………………….………………………………….7

Improvement……………………………………………………………………………..….10

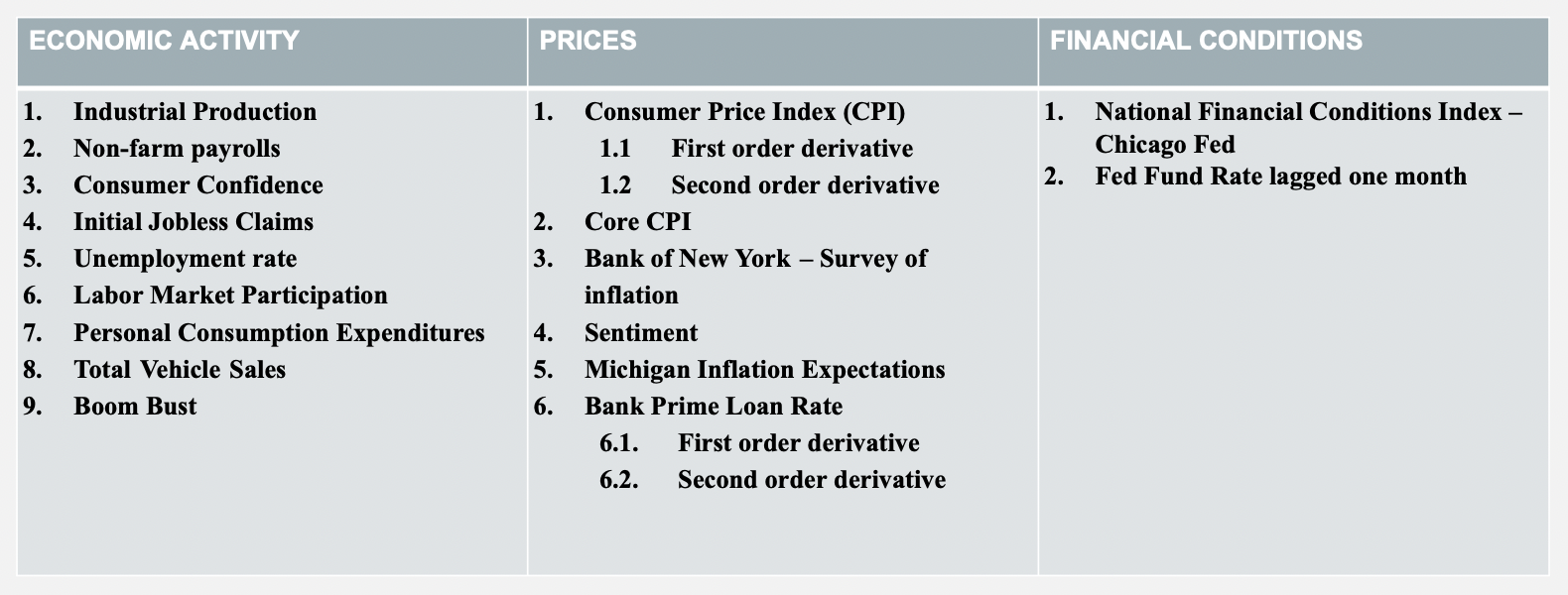
References……………………………………………………………………..…….…....…11

**Introduction**

The Federal Reserve Banks adjust rates throughout the year depending on the financial landscape observed by monitoring key indicators ranging between real activity indicators, pricing indicators, and financial conditions. Throughout time the federal reserve shifts regime and changes in policy based on new information and future financial outlooks. This problem at first seems straightforward but when observing from a technical perspective it appears that the regime/policy can cause a blurry boundary between outcomes of decreasing, maintaining, and increasing rates. This gives the machine learning algorithms a difficult time classifying movements over time periods distant from the training set. Our goal is to use all of this information to search for parameters that best optimize our models' predictive capabilities and to apply feature selection to determine if a subset of parameters helps to reduce any possible noise-blurring class boundaries.

**Data Cleaning and Preparation**

1. **Indicators Selection**

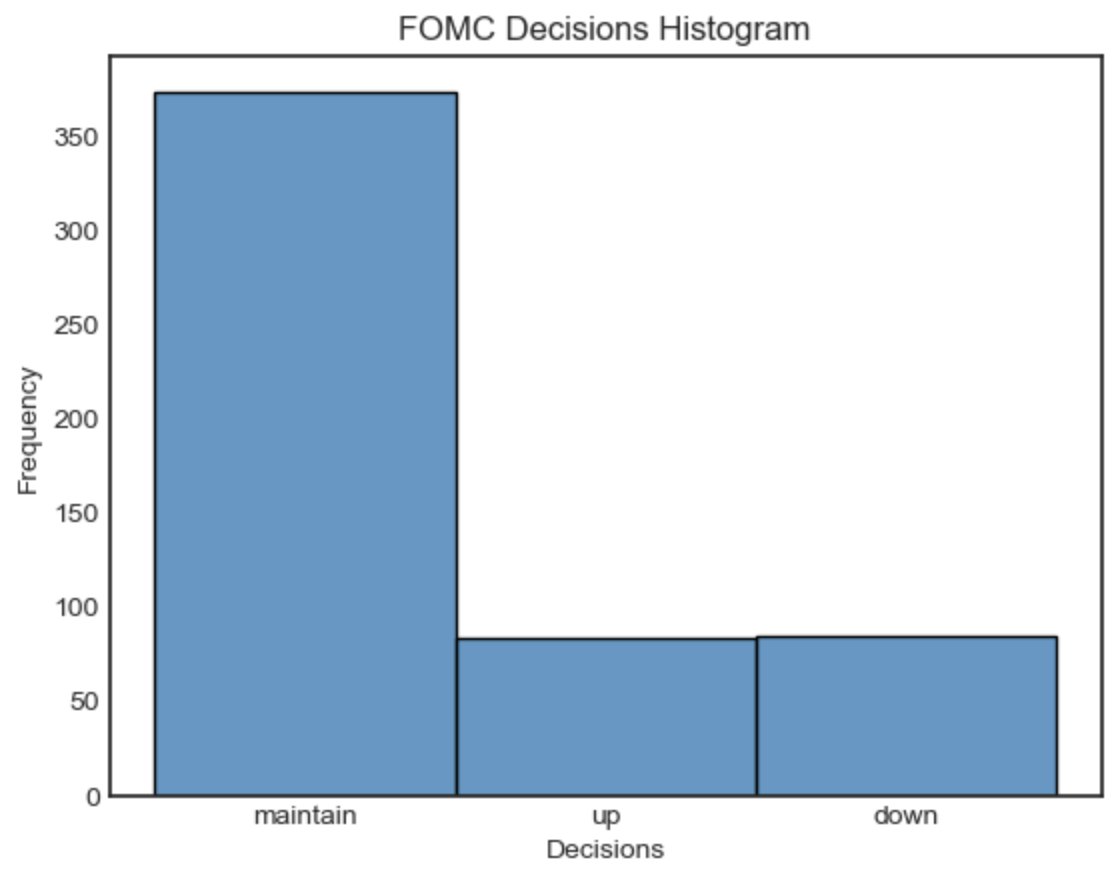


\*Deleted 'Consumer Sentiment', 'Inflation Expectation' because of high correlation with other data; Added Fed Fund Rate lagged one month

For the y data, we chose the processed percentage change of Fed Fund effective rate. The standard we used in the procession is: if the percentage change goes up more than 20 bps, we define it as ‘up’; if the percentage change goes down more than 20 bps, we define it as ‘down’; if the absolute value of the percentage change is within the range, we define it as ‘maintain’. The reason we do this is that we couldn't get the historical data of the Fed target rate change based on the Fed decision, so we tried to use the market traded rate, Fed Fund effective rate, to reflect the actual decision. The chosen threshold of +/- 20 bps is based on the assumption that the market-traded rate volatility would not typically exceed this range if there were no interventions or decisions made by the FED. In other words, the threshold is a reasonable level to indicate the impact of the FED's decisions on the market, as anything beyond that range would likely be attributed to other factors.

The advantage is that the Fed Fund effective rate is market traded, and we can retrieve the periodic monthly data which aligns with the frequency of our features data.

From the graph below we can clearly see it is imbalanced because based on our procession, most of the time the Fed Fund Rate is just maintaining a similar level as before.



**1.1 Real Activity Indicators**

| Industrial Production | The IP index measures the output of manufacturing, mining, and utilities in the US, regardless of ownership. It's a key economic indicator for assessing the country's overall economic health. |
| --- | --- |
| Non-farm payrolls | The Total Nonfarm Payroll measures the number of paid workers in the US economy, excluding certain groups. It reflects the overall health of the economy and indicates job creation or destruction. The Bureau of Labor Statistics releases two versions: the seasonally adjusted and non-seasonally adjusted, allowing for long-term trend analysis and short-term fluctuations. |
| Consumer Confidence | Historical data on Total Nonfarm Payroll before January 1978 is available with a one-month delay. The Bureau of Labor Statistics provides access to this data for researchers and analysts interested in examining historical trends. |
| Initial Jobless Claims | An initial claim is the first step in applying for Unemployment Insurance benefits. It is filed by an unemployed individual who has been separated from their employer and is used to determine basic eligibility for the program. The claim includes work history and reason for separation. The state then uses this information to determine the individual's eligibility, amount, and duration of benefits. |
| Unemployment rate | The unemployment rate is a percentage that represents the number of unemployed people out of the total labor force. The labor force includes people aged 16 or older living in one of the 50 states or the District of Columbia who are not residing in institutions or on active duty in the Armed Forces. |
| Labor Market Participation | The Labor Force Participation Rate is the percentage of the civilian noninstitutional population that is working or actively seeking employment. It is a monthly economic indicator used to analyze current labor market trends and assess the overall health of the economy. It is an essential tool for government agencies, financial markets, and researchers to gauge unemployment levels and economic activity. |
| Personal Consumption Expenditures | Real Personal Consumption Expenditures is a measure of the total value of goods and services purchased by individuals and households in the United States, adjusted for changes in prices over time. PCE is used to track trends in consumer spending and is an essential component of GDP. It includes a wide range of goods and services and is calculated quarterly by the Bureau of Economic Analysis. |
| Total Vehicle Sales | Total Vehicle Sales is the total number of new vehicles sold in the United States, including cars, trucks, and other motor vehicles. It is an important economic indicator that reflects the strength of the automotive industry and is used to track consumer demand and competitiveness within the market. |
| (10y - 2y Treasury Constant Maturity) | The interest rate spread is the difference between the yields on the 10-Year Treasury and the 2-Year Treasury securities and is a widely used economic indicator. It reflects the overall health of the economy and is calculated using data obtained directly from the U.S. Treasury Department. A larger spread suggests a stronger economy, while a smaller spread can indicate weaker economic conditions. |

These macro factors are indicators of the overall economic environment and include factors such as unemployment levels and GDP growth. By monitoring and analyzing these factors, the Federal Reserve can make informed decisions about adjusting interest rates to stabilize the economy. Changes in interest rates can affect the cost of borrowing, spending, and investing, and can impact various sectors of the economy differently. Therefore, it is essential for the Federal Reserve to consider these common factors when making monetary policy decisions.

**1.2 Price Indicators**

| CPI  First & Second Order Derivative | It measures the average change in prices of goods and services consumed by households. Meanwhile, the first derivative captures the trend of inflation, while the second derivative indicates the momentum of inflation. |
| --- | --- |
| Core CPI | Excludes the prices of food and energy, which can be more volatile than other prices and can lead to a distorted inflation signal. |
| Bank of New York – Survey of inflation | It is a survey of consumer inflation expectations conducted by the Bank of New York, which can provide insight into how people expect prices to change in the future. |
| Consumer Sentiment | It refers to the overall mood or attitude of consumers, businesses, or investors towards the economy or a particular asset, which can influence their behavior and decision-making. |
| Michigan Inflation Expectations | It is a survey of consumer inflation expectations conducted by the University of Michigan, which can provide insight into how people expect prices to change in the future. |
| Bank Prime Loan Rate  First & Second Order Derivative | The interest rate that commercial banks charge their most creditworthy customers for loans, which can indicate the general level of interest rates in the economy. |

In summary, the factors mentioned are all related to inflation and have different implications for forecasting. The CPI and Core CPI are widely used indicators of inflation. The first and second-order derivatives of CPI are also considered to capture the trend and momentum of inflation. The Bank of New York Survey of Inflation and Michigan Inflation Expectations are both surveys that measure the public's expectation of future inflation. The Bank Prime Loan Rate reflects the cost of borrowing for banks, which can affect the overall economy. The sentiment is a more abstract factor that can reflect consumers' confidence in the economy.

Among these factors, the first and second-order derivatives of CPI are particularly important because they can provide additional information about the momentum of inflation. By measuring the change in the rate of inflation over time, the derivatives can help to capture trends and turning points that may not be apparent in the raw CPI data.

**1.3 Financial Conditions**

| National Financial Conditions Index – Chicago Fed | Provides a comprehensive weekly update on U.S. financial conditions in money markets, debt and equity markets, and the traditional and “shadow” banking systems. |
| --- | --- |
| Fed Fund Rate lagged one month | It is a variable that takes into account the past interest rate actions of the Federal Reserve. |

The National Financial Conditions Index by the Chicago Fed and the Fed Fund Rate lagged one month are innovative factors that have been shown to provide valuable information for predicting inflation.

In particular, the lag in Fed Fund Rates helps capture the time-series features of inflation, while the National Financial Conditions Index offers a comprehensive measure of the overall health of financial markets.

**Machine Learning Models Building**

1. **Model Building**

**1.1 Multinomial Logistics Regression**

Multinomial Logistics regression is a type of logistic regression used for multiclass classification problems. The algorithm estimates the probabilities of each category, given the independent variables. These probabilities are calculated using a softmax function, which ensures that the probabilities sum up to 1.

In this case, We implement a logistic regression model on a given dataset. The accuracy of the predictions is calculated using the accuracy\_score function and the area under the ROC curve is calculated using the roc\_auc\_score function.

**1.2 LDA**

Linear Discriminant Analysis (LDA) is a technique used for supervised classification, which involves identifying patterns in data by maximizing the separation between classes. LDA is a method for dimensionality reduction, where a set of predictors is transformed into a lower-dimensional space to simplify the analysis. LDA assumes that the data are normally distributed and that the covariance matrices for each class are equal.

In this case, we perform a parameter tuning skill with a grid search over different values of shrinkage and solver using the Pipeline object to find the best set of hyperparameters, based on the highest accuracy score. After finding the best set of hyperparameters, we then fit the LDA model on the training data, and the accuracy, F1 score, and balanced accuracy score are computed on the validation set.

**1.3 QDA**

Quadratic Discriminant Analysis (QDA) is a classification algorithm that belongs to the family of linear and quadratic discriminant analysis methods. It is a classification technique that is widely used in statistics and machine learning for solving multiclass classification problems. QDA models assume that each class is normally distributed with a distinct covariance matrix. In contrast to Linear Discriminant Analysis (LDA), QDA allows the variance-covariance matrix to be different for each class.

We perform hyperparameter tuning using grid search with time series cross-validation to find the optimal values for the regularization parameter in QDA. The regularization parameter controls the degree of regularization or shrinkage applied to the covariance matrix to prevent overfitting.

**1.4 K-Nearest Neighbors (KNN)**

KNN predicts the label of a data point based on the labels of its k nearest neighbors in the training dataset. The algorithm calculates the distance between the test data point and all the training data points and selects the k closest data points based on the distance metric chosen (e.g., Euclidean distance or Manhattan distance). Then, the algorithm assigns the most common class among the k-nearest neighbors as the predicted class for the test data point.

KNN is a non-parametric algorithm, meaning that it does not assume any specific functional form for the data distribution. This makes it very flexible and suitable for a wide range of datasets. However, KNN can be computationally expensive for large datasets, as it requires calculating the distance between each test data point and all the training data points.

We are interested in how much K can result in the best result in our dataset, so we used a for-loop to find which number of K can give us the highest accuracy and AUC. We have also plotted out the results. Turned out it performed the best when K = 16, so we keep K = 16.

**1.5 Support vector machines (SVM)**

SVM can find a hyperplane in a high-dimensional space that best separates the different classes of data. The hyperplane is chosen such that it maximizes the margin between the closest points of the different classes. These closest points are called support vectors, and they play a critical role in determining the hyperplane. SVMs are particularly useful when the number of features (i.e., dimensions) is large, and the number of samples is small. SVMs can handle both linearly and nonlinearly separable data, by using kernel functions to transform the data into a higher-dimensional space, where it is more likely to be linearly separable.

We hyperparameter tuned the regularization parameter (“C”), kernel coefficient(“gamma”), degree of the polynomial kernel function (“degree”), and independent term in kernel function (“coef0”) using grid search with time series cross-validation.

**1.6 Random Forest**

Random Forest is an ensemble learning algorithm used for classification, regression, and other machine learning tasks. It is based on the idea of creating multiple decision trees, each of which is built using a random subset of the training data and a random subset of the features. It works by creating a large number of decision trees and aggregating their predictions to make a final prediction. When a new input is given to the model, each tree in the forest makes a prediction, and the most common prediction is chosen as the final output.

Although the use of multiple trees and random subsets of data and features helps to reduce overfitting and improve the accuracy of the model, when we do the hyperparameter tuning, we still pay attention not to put too many estimators or make the tree too deep to avoid overfitting, so we hyperparameter tuned the number of estimators and the maximum depth of the tree using random search with time series cross-validation.

**1.7 LSTM (long short-term memory networks)**

LSTM networks are a type of recurrent neural network (RNN) that is designed to overcome the vanishing gradient problem that occurs in traditional RNNs. The main advantage of LSTM networks is their ability to selectively remember or forget information from previous time steps, which makes them well-suited for processing sequences of data with long-term dependencies. This is achieved through the use of specialized memory cells, which are able to store information for an extended period of time. Each LSTM unit consists of three gates: the input gate, output gate, and forget gate. These gates control the flow of information into and out of the cell, allowing the LSTM to selectively store or discard information based on its relevance to the current task. The input gate determines which information from the current input should be stored in the cell, the forget gate determines which information from the previous time step should be discarded, and the output gate determines which information should be output from the cell.

For our input, we set the ‘time steps’ equal to one, and didn’t select the batch size, because the original data shape was 325\*19 and 108\*19, and we need to reshape them into three-dimensional data. We have already assigned one dimension as 19, and for 325 and 108, the greatest common factor is 1, so we can only choose ‘timesteps’ = 1, and the batch is left unchoose to give flexibility to our data.

We did not use hyperparameter tuning because the complexity of the model and the features of our data didn’t give us too much flexibility for doing this. However, we did try to manually modify the model in order to lead to the best result.

**1.8 Gradient Boosting**

Gradient boosting builds predictive models by iteratively adding weak learners to a model, with each new learner correcting the errors made by the previous ones. It is a type of ensemble learning method that combines multiple weaker models into a stronger one. The key idea behind gradient boosting is to fit a sequence of models, each of which is trained to minimize the residual errors of the previous models. The learning process involves adjusting the weights of the data points based on their misclassification rate, with more weight given to the misclassified samples.

We hyperparameter-tuned the number of estimators, maximum depth of the tree, gamma, alpha, lambda, and minimum sum of instance weight (hessian) needed in a child using random search with time series cross-validation.

1. **Results**

For the results, we picked three standards to estimate the performance of the models. The first one is ROC AUC, a.k.a. area under the ROC Curve, which is basically the matrix of model efficiency, the larger the area under the ROC curve, the better the model is. The second one is the F1 weighted score, which is the weighted average of the F1 scores in each class in our dataset. The third one is imbalance accuracy, which is the weighted average of the accuracy scores in each class in our dataset. For the second and third methods, since they are weighted averages, they take the size of the sample into the calculation, and this can help solve the imbalance problem in our dataset. Below is the performance before the LASSO feature selection.

|  | ROC AUC  (Train | Validation) | F1 weighted score  (Train | Validation) | Imbalanced accuracy  (Train | Validation) |
| --- | --- | --- | --- |
| Logistic | 85.53 | 62.46 | 69.66 | 75.76 | 60.42 | 33.33 |
| LDA | 84.68 | 75.20 | 68.91 | 74.36 | 59.99 | 32.22 |
| QDA | 82.39 | 80.75 | 66.56 | 75.83 | 56.65 | 37.17 |
| KNN | 86.79 | 66.56 | 68.32 | 80.44 | 59.74 | 44.34 |
| SVM | 74.22 | 51.72 | 62.91 | 73.98 | 52.33 | 52.46 |
| Random Forest | 84.60 | 56.30 | 50.68 | 79.66 | 38.96 | 39.39 |
| **XGBOOST** | **87.88 | 85.70** | **68.81 | 84.15** | **59.03 | 56.09** |
| Sequential Neural Network | 86.09 | 71.89 | 67.36 | 77.02 | 57.57 | 42.49 |
| LSTM | 84.02 | 67.68 | 66.11 | 76.43 | 54.66 | 37.54 |

Then we used LASSO feature selection to select the top 10 Features, those features being: Fed Rate Lag, Bank Loan Rate, Labor Market Participation, CPI Acceleration, Core CPI, Consumer Opinion Survey, Personal Consumption Expenditure, Sentiment, CPI, and Industrial Production.

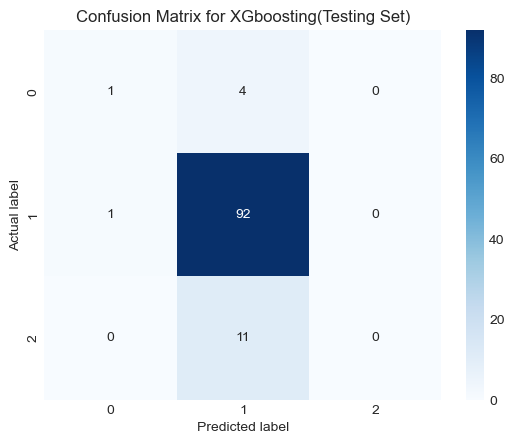
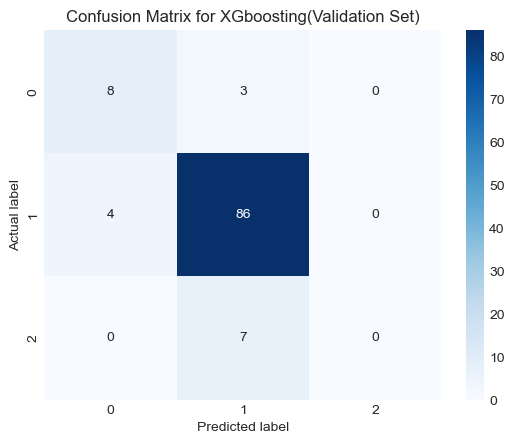
Below is the result after LASSO feature selection:

|  | ROC AUC  (Train | Validation) | F1 weighted score  (Train | Validation) | Imbalanced accuracy  (Train | Validation) |
| --- | --- | --- | --- |
| Logistic | 82.53 | 68.31 | 68.03 | 80.44 | 58.09 | 44.34 |
| LDA | 82.27 | 62.89 | 67.25 | 79.24 | 57.10 | 43.60 |
| QDA | 77.09 | 59.01 | 62.44 | 75.76 | 52.97 | 33.33 |
| KNN | 84.04 | 79.96 | 66.19 | 77.84 | 56.68 | 36.36 |
| **SVM** | **75.12 | 65.60** | **64.48 | 80.60** | **58.13 | 56.90** |
| Random Forest | 84.12 | 52.92 | 63.47 | 79.66 | 51.17 | 39.39 |
| XGBOOST | 99.67 | 63.88 | 96.91 | 79.08 | 95.67 | 48.18 |
| Sequential Neural Network | 71.54 | 54.50 | 59.55 | 75.30 | 48.04 | 32.96 |
| LSTM | 85.65 | 69.73 | 69.77 | 85.58 | 60.73 | 54.18 |

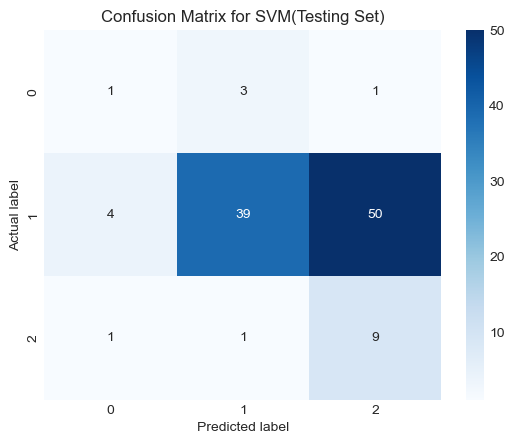
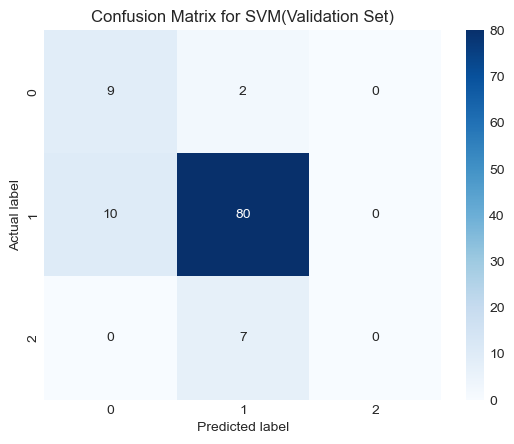
From the results above, we can see that the top model is SVM with a polynomial kernel using the LASSO top ten features. The second model is XGBoosting using all the features.

SVM is one of the most popular classification algorithms due to its ability to be represented through many kernels. Using our optimized SVM model, this was the only model to correctly classify any upward predictions. XGBoosting is another popular classification algorithm that uses a set of decision trees to develop a generalized model. Using our optimized XGBoosting algorithm we were able to finely define the border between 0 and 1 (down and maintain)

We have seen the training-validation performances of the models with different sets of features. Now we are going to examine the top 2 models by looking at the testing set performance.

****

The above figures show the confusion matrix of the validation set(left) and testing set(right) for the XGboosting. We can see the model could not catch upward movement in the federal fund rate for both datasets.



The above figures show the confusion matrix of the validation set(left) and testing set(right) for the SVM with a polynomial kernel. This model can actually correctly catch the up movements 9 out of 11 times. However, this does not suggest the SVM model’s testing set performance is better than the XGboosting model. Because the prediction for the maintain class is worse; the SVM model fails to find a clear border between the maintain(1) and up(2) movements.

**Improvement**

1. **Collect a longer history of data for the features:** As we have discussed previously, the dataset we used started after 1978. Most of the historical up movements happened between 1960 - 1980. Hence, in the future, we can find higher-quality data that contains longer historical observations. This can help to provide a more comprehensive understanding of the patterns and trends in the data, which can ultimately improve the accuracy of our prediction.
2. **Incorporate NLP techniques towards the FOMC Minutes:**
   1. **Extract important features which FOMC considers:** Based on the analysis of the FOMC Minutes using NLP techniques, this step involves extracting the features that are most relevant to the FOMC's decision-making process. This can help to create a more focused set of features for the predictive model.
   2. **Use the Minutes to the actual prediction:** We can use NLP techniques to create knowledge bases related to the Fed decision, measure the distance between the Minutes to the knowledge bases, and use classification models to implement prediction
3. **Ensemble multiple ML models toward the final prediction:** Ensemble modeling involves combining the predictions of multiple machine learning models in order to improve accuracy and reduce errors. This step involves using multiple models to make predictions, and then combining their results to arrive at a final prediction.
4. **Introduce a rolling window to localize the regimes in time:** A rolling window is a technique used to analyze data over a moving time period. In this step, a rolling window is used to analyze the data over a specific time period in order to identify specific "regimes" or periods of time with similar economic conditions. By localizing the analysis to these specific periods, the predictive model can potentially provide more accurate predictions.

**Reference**

[1] Ellyn Boukus and Joshua V. Rosenberg. The Information Content of FOMC Minutes. *SSRN Electronic Journal*, 2011.

[2] Boyu Wu, Amina Enkhbold, Asawari Sathe & Qian Wang. How Does the Fed Make Decisions: A Machine Learning Augmented Taylor Rule. *The Journal of Fixed Income Winter 2023, 32 ( 3)49 - 60*

[3] Michael T. Kiley. Recession Signals and Business Cycle Dynamics: Tying the Pieces Together. *Federal Reserve Board, Washington, D.C. ISSN 2767-3898 (Online)2023-008*

*[4] Grishchenko, Olesya, Zhaogang Song, and Hao Zhou (2022). "Term Structure of Interest Rates with Short-run and Long-run Risks," Journal of Finance and Data Science, vol. 8, pp. 255-295.*

*[5] Rungcharoenkitkul, Phurichai, and Fabian Winkler (2022). "The Natural Rate of Interest Through a Hall of Mirrors," Finance and Economics Discussion Series 2022-010. Washington: Board of Governors of the Federal Reserve System.*

*[6] Pauwels, Laurent L. and Vasnev, Andrey L., Forecast Combination for Discrete . Choice Models: Predicting FOMC Monetary Policy Decisions (May 2012).*

*[7] Karnaukh, Nina, The Dollar Ahead of FOMC Target Rate Changes (January 31, 2020). Fisher College of Business Working Paper No. 2018-03-014, Charles A. Dice Working Paper No. 2018-14, 31st Australasian Finance and Banking Conference 2018.*

*[8] Kenneth Petersen, 2007. "*[*Does the Federal Reserve Follow a Non-Linear Taylor Rule?*](https://ideas.repec.org/p/uct/uconnp/2007-37.html)*,"* [*Working papers*](https://ideas.repec.org/s/uct/uconnp.html) *2007-37, University of Connecticut, Department of Economics.*

*[9]Leo Breiman (2001), “Random Forests” Machine Learning, 45, 5–32, 2001*

*[10]Boukus, E., and J. V. Rosenberg (2006) “The Information Content of FOMC Minutes”, mimeo, Federal Reserve Bank of New York, New York, July.*

*[11]Nunes, R. (2013) “Do Central Banks’ Forecasts Take Into Account Public Opinion and Views?”, International Finance Discussion Paper No. 1080, Board of Governors of the Federal Reserve System, Washington, D.C., May.*

*[12]H. Kauppi Predicting the Direction of the Fed’s Target Rate J. Forecast., 31 (2012), pp. 47-67*

*[13]Grove, A. & Schuurmans, D. (1998). “Boosting in the limit: Maximizing the margin of learned ensembles.” In Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98).*

*[14]L.L. Pauwels, A.L. Vasnev Forecast combination for discrete choice models: predicting FOMC monetary policy decisions Empir. Econ., 52 (2017), pp. 229-254*

*[15]Sepp Hochreiter, Jürgen Schmidhuber. “Long short-term memory” Neural computation 1735-1780*