

Automated Web Application for predicting high impact Forex Economic News Release

by

Michael Dubem, Chibueze Louis, Abiye Danagogo and Michael George.

June 17, 2021

1. Introduction

1.1 Background

The awareness of foreign exchange (Forex) has grown significantly over the past few years amongst the youth and other age grades across the globe (including countries like Nigeria), and many persons (natural and artificial) have realised the potential in understanding and taking advantage of the fluctuation in the rate of hard currencies that occur on a daily basis. In addition to this, a lot of multinational companies also use these fluctuations and projections to their advantage when conducting business transaction across international borders, taking loan from banks (foreign and local), planning expansions and many others. Speaking from my experience as a forex trader, the entirety of the trading business is riddled with predictions and analysis, and those with relevant skills and experience in predictions (mathematical, economical, psychological etc.) or analysis whether through empirical or technical means tend to have an edge. There are a host of factors that influence the forex market on a daily basis, but all these factors can't move the price of a certain currency without being accompanied by traders'/market sentiment (this refers to the position of majority of traders in the market at any given time). There are about two major type of traders (with respect to type of analysis); Technical and Fundamental. Fundamental traders can further be broken down into economic traders (traders who trade based on economic factors over extended periods of time), periodic news traders (traders who take advantage of market sentiment at periodic news releases), and breaking news traders (traders who take advantage of the effects of unplanned impactful news releases).

1.2 Problem

All types of traders (as well as business and multinational firms) stated before now are affected by the periodic release of economic parameters, which give insight into the state of an economy. For traders like periodic news traders, this effect can be instantaneous, but for economic traders, technical traders and multinational firms, these news releases and their trends can have extended effects spanning for weeks, months or even years. This is where foresight becomes important, because being able to predict a news parameter beforehand could help a trader take an early position in the market, or being able to determine the trend of a certain economy using its parameters could be crucial during business projections

1.3 Interest

Any individual or firm directly affected by these fluctuations in currencies and the state of their respective economies, certainly have use for these insights and projection to further tap into the numerous opportunities provided by the forex market.

2. Data Acquisition and Cleaning

2.1 Data Sources

This project was majorly focused on the prediction of the periodic value of the bank rate set by the Federal Market Open Committee (FOMC) of the United States of America. The entire dataset was gotten from a website known as forex factory, which focuses on monthly economic news releases of different economies. This website was used due to its consistency and accuracy, as well as it's availability. However, the concept of

the project isn't limited to this particular website, as any other effective website can be used for the same purpose with a little change of code.

Date	Time	Currency	Impact	Detail	Actual	Forecast	Previous	Graph
Tue Jun 1	5:30am	AUD	High	RBA Rate Statement				
	All Day	All	High	OPEC-JMMC Meetings				
	3:00pm	USD	High	ISM Manufacturing PMI	61.2	60.8	60.7	
	4:00pm	GBP	High	BOE Gov Bailey Speaks				
Wed Jun 2	2:30am	AUD	High	GDP q/q	1.8%	1.5%	3.2% ↑	
Thu Jun 3	3:00pm	USD	High	ISM Services PMI	64.0	63.0	62.7	
Fri Jun 4	12:00pm	USD	High	Fed Chair Powell Speaks				
	1:30pm	CAD	High	Employment Change	-68.0K	-23.5K	-207.1K	
		CAD	High	Unemployment Rate	8.2%	8.2%	8.1%	
		USD	High	Average Hourly Earnings m/m	0.5%	0.2%	0.7%	
		USD	High	Non-Farm Employment Change	559K	645K	278K ↑	
		USD	High	Unemployment Rate	5.8%	5.9%	6.1%	

<https://www.forexfactory.com/calendar?month=this>

2.2 Data Extraction and Cleaning

The Federal Funds Rate set by the Federal Open Market Committee, from the research conducted in this project, can be predicted using a couple of other economic parameters (news releases). These independent parameters were extracted using the Selenium and Beautiful Soups APIs (for a historical period of 10 years), cleaned, transformed and fed separately into an IBM DB2 Database using the ibmdb API, with the intention of calling up these table from the database and using the data within to train and run predictions.

Due to the intention of automating the project, the web scraping section of the program is designed to extract, clean and update the initially created tables in the database from time to time.

2.3 Feature Selection

The correlation method (`df.corr(method='Pearson')`) was run on all the data extracted from the website, to check which independent parameters had the most statistical correlation to the parameter being predicted (in this case, the Federal Fund Rate periodically set by the Federal Open Market Committee). Only two parameters had strong correlations of over 60% with the dependent parameter, and these two were then used for the training and inference.

```
In [142]: dataset_train
```

Out[142]:	US_TRADE_BALANCE	US_NFP_RATES	US_FOMC_RATES
0	-71.1	0.080	0.0025
1	-68.2	0.082	0.0025
2	-68.1	0.087	0.0025
3	-63.1	0.087	0.0025
4	-63.9	0.089	0.0025
...
86	-42.8	0.096	0.0025
87	-49.9	0.095	0.0025
88	-40.3	0.097	0.0025
89	-39.7	0.097	0.0025
90	-37.3	0.097	0.0025

91 rows × 3 columns

The parameter being predicted wasn't a continuous variable hence classification methods were used, and the results from a couple of test runs showed no significant change in the accuracy of the prediction after applying polynomial features to the selected independent parameters, so this method of feature engineering didn't seem necessary for this project.

3. Methodology

In this section of the project, a supervised machine learning algorithm (Decision tree) was applied on the final selected features, and subsequently used for the inference. As earlier stated, the dependent parameter for this phase of the project was largely categorical, hence a classification model was most suited for the prediction, also due to the fact that slightest deviation in this type of parameter prediction nullifies its usefulness, hence getting a regression model with high accuracy would still be less useful than a classification model because of the residues that comes with regression models.

3.1 Selection of the Decision Tree Model

The algorithm was run across a range of tree depths/levels (hyper parameter) and the number of levels/depth with the maximum accuracy was recorded. This optimal hyper parameter was then used to run the inference.

This section of the program was written in such a way that it always takes new data into account and retrains the model with every schedule run to adjust the inference to the shift in data distribution. Although, there are standard ways to measure and counter skewness in data from original conditions in projects, these methods are mainly for projects involving large amounts of datasets which can become cumbersome to retrain with every schedule run. However, the dataset used for this project was relatively small, and would take a while before training and retraining the model used for inference starts becoming cumbersome. Due to this reason, the model is always retrained with every schedule run, and the optimal tree level is used for the next prediction. This makes the model reaction to skewness in data much quicker, and when it comes to dealing with Forex parameters, this level of accuracy and flexibility can be a real advantage.

Prediction

```
In [145]: from sklearn.tree import DecisionTreeClassifier
          from sklearn.model_selection import train_test_split
          import seaborn as sns
          %matplotlib inline

In [146]: N = dataset_train.iloc[:, :-1]
          Z = dataset_train.iloc[:, -1].astype(str)
          N_train, N_test, Z_train, Z_test = train_test_split(N, Z, test_size=0.1, random_state=1)

In [147]: md_scores = {}
          for md in range(1,31):
              loanTree = DecisionTreeClassifier(criterion="entropy", max_depth = md)
              loanTree.fit(N_train, Z_train)
              md_scores[md] = loanTree.score(N_test, Z_test)
          opt_md = list(md_scores.keys())[list(md_scores.values()).index(max(md_scores.values()))]
          print(opt_md)
          sns.lineplot(x = list(md_scores.keys()), y = list(md_scores.values()))
```

4. Results and Discussion

4.1 Results

The following classification algorithms; Decision Tree, K-Nearest Neighbours, and Logistic Regression, were tried on the dataset and the algorithm with the highest average predictive accuracy was the Decision Tree algorithm. After multiple experiments with a range of tree depth/levels (hyper parameters), the maximum predictive accuracy gotten from the Decision Tree model training was sitting between 90-100%. This value is

represented in a range instead of a specific value due to the fact that the model is retrained with every schedule run.

This optimal model that emerged from the series of experiments was then used to run the inference on the new set of data gotten for the month of June, and predicted the Federal Fund Rates accurately.

```
In [148]: loanTree = DecisionTreeClassifier(criterion="entropy", max_depth = opt_md)
loanTree.fit(N_train,Z_train)
ln_pred = loanTree.predict(dataset_pred)
ln_pred = ln_pred.tolist()
#ln_pred
df = pd.DataFrame({'Date':pred_date, 'FOMC Pred.':ln_pred})
df
```

```
Out[148]:
```

	Date	FOMC Pred.
0	2021-06-16	0.0025

Prediction before Federal Funds Rate was released for June 16th.

	Date	Time	US_FOMC_RATES
0	2007-01-31	00:00:00	0.0525
1	2007-03-21	00:00:00	0.0525
2	2007-05-09	00:00:00	0.0525
3	2007-06-28	00:00:00	0.0525
4	2007-08-07	00:00:00	0.0525
...
114	2020-12-16	00:00:00	0.0025
115	2021-01-27	00:00:00	0.0025
116	2021-03-17	00:00:00	0.0025
117	2021-04-28	19:00:00	0.0025
118	2021-06-16	19:00:00	0.0025

The actual released Federal Funds Rate for June 16th on the 118th row.

4.2 Further Discussion

This phase of the project focused solely on the prediction of the Federal Fund Rate for the US Dollar, despite the fact that there are a host of other high impact inter-dependent news releases that are based on the US Dollar, and each major and minor currency (like the Canadian Dollar, Euros, Pounds, Australian and New Zealand Dollar, and so on) have their respective similar high impact news releases, all being released periodically. This is because all these predictions can be done quite easily by adopting the established program used for the US Dollar, for the rest of them. So focusing on the US Dollar Federal Funds Rate was simply for the purpose of breaking down a relatively difficult project into digestible chunks and easily targetable milestones, before scaling the results. Therefore, a further study or extension of this current project could cover a host of other high impact news releases, one might simply need to adjust the base algorithm, or even stack algorithms for higher accuracy.

Another not so obvious extension of this project is real-time market prediction. This is no different from the project carried out in this report. The underlying concept still simply entails extracting a couple of independent parameters from market historical data, doing some wrangling and then applying some supervised or unsupervised machine learning algorithm on the dataset to predict whichever detail about the market we find useful, and this could be a number of things like, market direction for the day (categorical), total pips moved in the market for the day (continuous numerical), and so on.

With further brainstorming, one could come up with many more areas of further study, hailing from the core concept behind the project covered in this report.

5. Conclusion

The purpose of this project was to predict a high impact economic news parameter, before its release, to help directly affected or interested persons to be able to react quickly to the impact of this news release and to also help those who seek to make long term projection with this news release. This was done by extracting data for each independent parameter for this prediction from an online website, cleaning and transforming this data, carrying out feature selection by using the correlation method, and loading this data into a database. This dataset is then called up from the database and use to inference purposes.

In a similar manner, this approach could be used for predicting a variety of other economic or market related parameter, for a host of reasons. This is the overall aim of this project.