EUR/USD Prediction Final Report

George Vojvodic
Department of Computer Science
University of Illinois at Chicago
dvojvo2@uic.edu

Nikita Mashchenko
Department of Computer Science
University of Illinois at Chicago
nmashc2@uic.edu

David Serrano
Department of Computer Science
University of Illinois at Chicago
dserra22@uic.edu

Yuqing Wu
Department of Computer Science
University of Illinois at Chicago
ywu239@uic.edu

Abstract

The purpose of this project was to predict the state of the future Euro-Dollar (EUR/USD) exchange rate using various indicators and to evaluate which indicators were most impactful. Our implementation of a Random Forest Model was the main resource in ascertaining that future. This model yielded results with an accuracy up to 80% when looking at the rate movement one week ahead. Our analysis also showed a strong correlation between the price and macroeconomic factors, rather than technical factors. We conclude that the future exchange rate of the Euro-Dollar and potentially other global currencies can be predicted accurately through the use of a Random Forest Model.

1 Summary

Many small to medium size businesses in the US need to import products and services from abroad. With examples of said businesses being: Ethnic food stores, Manufacturers, Resellers, etc. Businesses such as these rely on stable currencies and their exchange rates in order to conduct business. A European food store may need to import goods from Greece or any Euro zone country. A manufacturer may rely on parts that are imported from multiple countries which use different currencies. Unfortunately, exchange rates can fluctuate a lot throughout any given year due to factors such as war, pandemics, stagnant economic growth, government intervention, and banks' appetite for risk. Businesses want to know if they should pay for imported products / services sooner or later. If the exchange rate goes in favor of the dollar in the future, a business may decide to hold off on purchases until later. On the other hand, if the exchange rate goes against the dollar in the future, a better strategy would be to make purchases as soon as possible. Decisions following this rationale would allow businesses to save money on purchases, which in turn increases their profit margins.

2 Approach

It is very hard to quantify the effects of global events on the exchange rates. However, one can see changes in certain macroeconomic indicators whenever these events occur. The goal of this project was to check how accurately the future Euro-Dollar (EUR/USD) movement can be predicted using various macroeconomic and technical indicators. We were only interested in predicting whether the exchange rate has gone up or down after a certain number of days.

Macroeconomic indicators used:

- Interest rates in the US and Euro zone reflected through the treasury yield rates.
- Inflation rates in the US and the Euro zone.
- Dollar Index (DXY), which is a measure of the Dollar's strength against a basket of major currencies.

Technical indicators being:

- Moving averages
- Relative Strength Index (RSI), which finds potentially overbought and oversold areas.
- Moving Average Convergence/Divergence (MACD), which looks for momentum.
- Bollinger Bands, which indicate trends and deviations from the average rate.
- Previous days' closing rates (to see whether previous days rate has any impact)

There are two popular machine learning models that have been used for forecasting exchange rates: Long Short-Term Memory (LSTM) models and Random Forests. An LSTM model is a variation of recurrent neural networks (RNN) that is capable of learning long-term dependencies in problems such as sequence prediction. According to online references, LSTMs do not provide a significant performance boost, therefore we have chosen to implement a Random Forest Classifier. A random forest is an ensemble method where multiple decision trees are used with varying features, where each tree can result in a different prediction. The final prediction is decided by a majority vote between all the generated trees. The classification would indicate whether a currency will be up/down in the future.

2.1 Background

Background is necessary to understand the project as it delves into the deeper details of global currencies and their interactions with one another. Contrary to the United States' general public's understanding, the penny is not the smallest fractional unit of currency as there is a need to have a more standardized unit to hold true for all currencies. The smallest fractional unit of any given currency is called a pip. For the US Dollar, a pip is the fourth decimal place; for the Japanese Yen, a pip is the second decimal place. Whenever currency is being transferred or converted, the value is being assessed at the pip level.

Interest rates of different central banks play a major role in the exchange rate movements. These interest rates are set by the government and are used as a general lever to control economic activity at a large scale. If an economy experiences rapid growth, the interest rates will typically rise to curb the expansion. Likewise, if an economy is underperforming, the interest rates are typically lowered. Since economies are impacted by global events such as wars and pandemics, they in turn also affect interest rates. Generally, when a central bank's interest rate rises, it attracts more investors to that currency, therefore affecting the exchange rate, favoring that currency. The opposite is also true. However, since exchange rates depend on two economies, the interest rates of both currencies need to be taken into consideration. When the interest rate of the central bank of government A grows in relation to the central bank of government B, the currency of country A tends to strengthen against B, and vice versa. Otherwise, if the interest rate increases or decreases uniformly for both central banks, there may not be much movement in the exchange rate. Therefore, the interest rate differential between the two economies is important.

Another important macroeconomic indicator is the inflation rate, which can be measured in a variety of ways. Inflation is a helpful indicator when it comes to predicting interest rate movements. Usually when inflation rises, the interest rate tends to rise as well after the inflation data is released. There is a positive correlation.

Lastly, the dollar index (DXY) measures the US dollar's strength against a basket of major currencies exchanged in the foreign exchange market. These include the Euro, Japanese yen, Pound sterling, Canadian dollar, Swedish krona, and the Swiss franc. A lot of the uncertainties in the world and risk assessments from major investors are baked into this index, and therefore we believe that it can be used to help predict the future movement of the Euro Dollar rate.

2.1 Data Collection

Our data collection ranged from various sources to accumulate all the necessary features for the model. Yahoo Finance [1] was used to obtain the EUR/USD rates, DXY, and US treasury yields. We used the treasury yields as opposed to the pure interest rates because these rates only update every 6 weeks, while treasury yields vary day to day, and therefore are more sensitive to daily events. The ECB (European Central Bank) dataset [2] was used to acquire the Euro treasury yield rates. US Inflation Calculator by CoinNews Media Group Company [3] was used to obtain historical US inflation data. Rate Inflation by RI [4] was used for historical Euro area inflation data.

2.2 Data Preparation

The data was prepared in Google Collab, allowing for seamless project collaboration between group members. First, the data was imported via API, online CSV and HTML tables. Then the data was formatted into a date index and value columns for each feature. Once we had organized the data with common axes, they were then combined into a single data frame. The null values were handled by trimming top and bottom rows, as well as forward filling inflation rates as they only update once per month. Finally, labels were added for the next day's price movement (up/down) and the 7 days later (coded as 1 or 0).

	Close	rsi	sma_10	sma_20	cross	MACD_12_26_9	MACDh_12_26_9	MACDs_12_26_9
2005-01-03	1.347001	NaN	NaN	NaN	0	NaN	NaN	NaN
2005-01-04	1.328198	NaN	NaN	NaN	0	NaN	NaN	NaN
2005-01-05	1.328004	NaN	NaN	NaN	0	NaN	NaN	NaN
2005-01-06	1.318305	NaN	NaN	NaN	0	NaN	NaN	NaN
2005-01-07	1.306097	NaN	NaN	NaN	0	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...
2023-04-14	1.105461	65.794414	1.092541	1.086524	0	0.007220	0.001300	0.005919
2023-04-17	1.098660	59.745408	1.094374	1.088063	0	0.007230	0.001049	0.006181
2023-04-18	1.092538	54.856839	1.094505	1.089094	0	0.006668	0.000389	0.006279
2023-04-19	1.097538	57.887880	1.094657	1.090100	0	0.006550	0.000217	0.006333
2023-04-20	1.095218	56.008882	1.095146	1.090509	0	0.006198	-0.000108	0.006306

Figure 1: Final Data Frame

3 Evaluation

We ran the classifier on two labels: next day's close status, and the close status 7 days ahead. By status, we mean whether the close was above or below the current day's close. The classifier was ran multiple times on two different hyperparameters: maximum tree depth, and number of estimators/trees to use. Maximum tree depth ranged from 5 to 50 in increments of 5. Number of estimators/trees used ranged from 5 to 60 in increments of 5. There were around 4500 rows of training data used with each row representing data for a single day, ranging from August 2005 to April 2023. The training and test data was split 80 / 20.

4 Results

We first ran the classifier to try to predict the next day's movement, whether the exchange rate ended above the current day or below. The accuracy of this test was 50-56%. The classifier was only slightly better on average as the max depth and # of trees increased. This result was in line with our prediction as there is a lot of daily market noise, which can interfere with the prediction. We expected an average of 50-53%, so on occasion the classifier outperformed our prediction however only slightly.

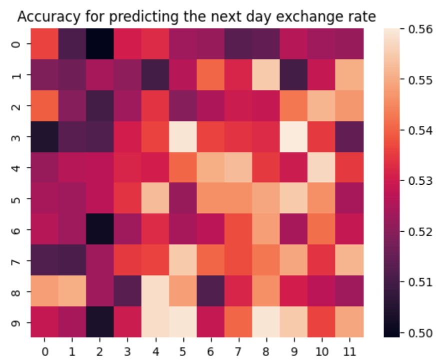

Figure 2: Next Day Prediction Accuracy

Whereas when we ran the classifier to predict the movement 7 days ahead compared to the current day's rate, we have seen an accuracy of up to 80%. The confusion matrix showed that there wasn't a big difference in performance between the two classifications. After inspecting the model and confusion matrix we became curious as to why the accuracy was so high, since the accepted accuracy in the professional world is 60%. Higher accuracy was expected nonetheless since daily noise would become less of an issue.

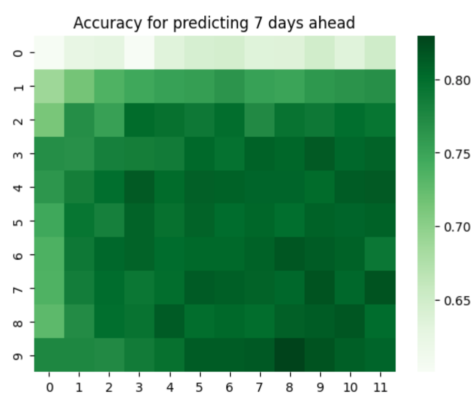

Figure 3: 7 Days Ahead Prediction Accuracy

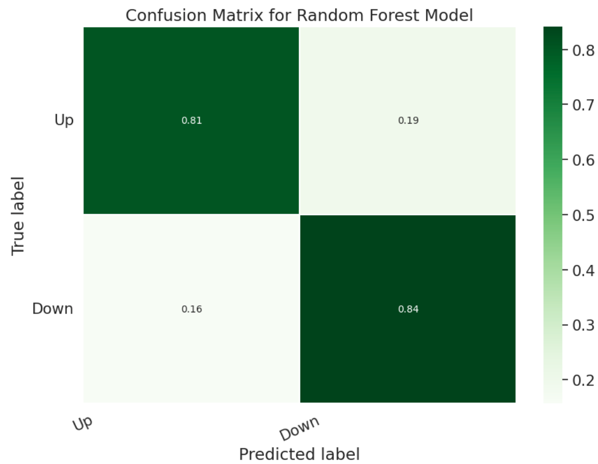

Figure 4: Confusion Matrix for the Random Forest Model

```
Accuracy score: 0.8242491657397107
Classification score:
              precision    recall  f1-score   support

           0       0.84      0.81      0.83       462
           1       0.81      0.84      0.82       437

    accuracy                           0.82       899
   macro avg       0.82      0.82      0.82       899
weighted avg       0.83      0.82      0.82       899
```

Figure 5: 7 Day Ahead Classification Score

Following this, the features were inspected to see what features were mostly responsible for the result. The most impactful features were the: Euro 10-year treasury yield, US 5-year treasury yield, US 10-year treasury yield, MACD, and Dollar Index (in 11th place out of 27 features). Our conclusion was that macroeconomic factors may play more of a role than anticipated but will require further testing to validate this finding. One somewhat surprising finding is that the inflation rate had much lower importance, but that may be because it did not update often.

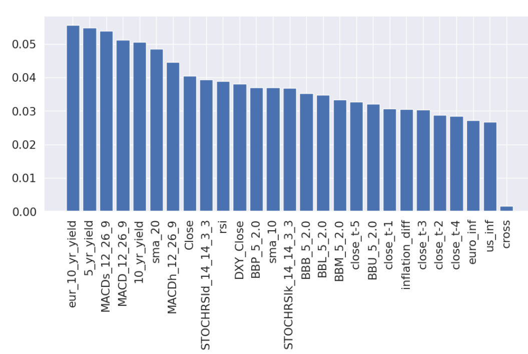

Figure 6: Feature Importance bar chart

1)	eur_10_yr_yield	0.055604
2)	5_yr_yield	0.054927
3)	MACDs_12_26_9	0.053875
4)	MACD_12_26_9	0.051221
5)	10_yr_yield	0.050543
6)	sma_20	0.048610
7)	MACDh_12_26_9	0.044604
8)	Close	0.040435
9)	STOCHRSId_14_14_3_3	0.039329
10)	rsi	0.038963
11)	DXY_Close	0.038086
12)	BBP_5_2.0	0.037033
13)	sma_10	0.036967
14)	STOCHRSIk_14_14_3_3	0.036915
15)	BBB_5_2.0	0.035210
16)	BBL_5_2.0	0.034847
17)	BBM_5_2.0	0.033400
18)	close_t-5	0.032734
19)	BBU_5_2.0	0.032141
20)	close_t-1	0.030694
21)	inflation_diff	0.030594
22)	close_t-3	0.030335
23)	close_t-2	0.028865
24)	close_t-4	0.028514
25)	euro_inf	0.027189
26)	us_inf	0.026668
27)	cross	0.001699

Figure 7: Feature Importance decimal representation

5 Project Retrospective

We have learned which features are more impactful on the future Euro-Dollar (EUR/USD) exchange rate as shown in the figure above. In addition, we discovered why the LSTM model may not offer a significant performance boost over the Random Forests model. This is due to the features named “close_t-2”, “close_t-3”, “close_t-4”, etc., which are the close prices for the past 5 days and are what the LTSM models depend on. During our feature inspection, we saw that those features were some of the least important, showing that the historical prices for evaluation didn’t matter as much as other data did. However, this conclusion may not hold for other currencies or other financial instruments.

5.1 Problems

Throughout the development of this project, a few problems were encountered though they were quickly resolved. However, some problems remain whether it be with the model or the data. The model did not go through cross validation, so the performance may change in the future. In addition, the exchange rate movement prediction did not take into account how the rate moved between the current day and 7 days ahead. Even though the model might accurately predict that the exchange rate went down, the rate could have fluctuated wildly up and down in the meantime. The problems regarding the data slightly affect the performance of specific features though there is no countermeasure as they are due to original data collection from our sources. The inflation rates are released about once per month, so there was no variation within each month, only between months, making it a suboptimal feature. Price movement may also have been negligible, anywhere between 5-50 pips, which is not a strong indication of an actual movement.

5.2 Future

Future development to this project would focus on model validation and refactoring for efficiency. More macroeconomic indicators will be added to further increase accuracy and to gauge which are impactful on the dataset and by what scale. After all features have been tested and evaluated, we will remove any that are unnecessary to further the efficiency of the model. Cross validation will be implemented to gain a more consistent and precise prediction. In addition to the cross validation, this model will be compared to a LSTM model. Once the model has been refined for EUR/USD interaction, it will also be tested on other global currencies such as the Japanese Yen. The currency selection process is yet to be decided, however we are prone to selecting currencies with less government and bank manipulation. Prior to testing two completely different currencies against each other, we will be testing one against an already tested currency such as the Euro or Dollar. The model will then be extrapolated to predict movement on other scales such as 5, 10, and 14 days in advance.

Acknowledgments

We are grateful to our classmates and professor for their contributions in expanding our knowledge and interest in the machine learning field. Thanks, should also go all of the individuals interested in machine learning as their myriad of papers and videos regarding the topic, assisting us with our studies into this topic.

References

[1] “Bitcoin USD (BTC-USD) Price, Value, News & History.” *Yahoo! Finance*, Yahoo!, 28 Apr. 2023, https://finance.yahoo.com/quote/BTC-USD?p=BTC-USD&.tsrc=fin-srch.

[2] *ECB Statistical Data Warehouse*, European Central Bank EuroSystem, https://sdw.ecb.europa.eu/.

[3] “Historical Inflation Rates: 1914-2023.” *US Inflation Calculator*, CoinNews Media Group Company, 12 Apr. 2023, https://www.usinflationcalculator.com/inflation/historical-inflation-rates/.

[4] “US Inflation Rate (I:USIR).” *US Inflation Rate*, YCharts, https://ycharts.com/indicators/us_inflation_rate.

[5] Jansen, Stefan. “11 Decision Trees Random Forests.” Machine Learning for Algorithmic Trading, Second ed., Packt Publishing, Birmingham, 2020.

[6] O. Schumann, C. Wöhler, M. Hahn and J. Dickmann, "Comparison of random forest and long short-term memory network performances in classification tasks using radar," 2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF), Bonn, Germany, 2017, pp. 1-6, doi: 10.1109/SDF.2017.8126350.

[7] Chen, James. “What Is the U.S. Dollar Index (USDX) and How to Trade It.” *Investopedia*, Investopedia, 30 Sept. 2022, https://www.investopedia.com/terms/u/usdx.asp#toc-what-is-the-us-dollar-index-usdx.

[8] Foot, Patrick. “What Is the US Dollar Index and How Can You Trade It?” *What Is the US Dollar Index and How Can You Trade It?*, FOREX US CD, 17 May 2022, https://www.forex.com/en-us/news-and-analysis/what-is-the-dollar-index/.

[9] Heron, Tina. “How Do Interest Rates and Inflation Affect Forex?” *IG*, IG, https://www.ig.com/us/trading-strategies/how-do-interest-rates-and-inflation-affect-forex-230117.

[10] “Inflation and Consumer Price Index (CPI) in Trading - Forex.com Us.” *Inflation and Consumer Price Index (CPI) in Trading - FOREX.com US*, FOREX US CD, https://www.forex.com/en-us/trading-academy/courses/fundamental-analysis/inflation-and-cpi/.

Google Drive Link to the Python notebooks and Word Document of this report:

https://drive.google.com/drive/folders/1UH4gIn2_OWc9fnye0PdQpfZLguhrrMHG?usp=sharing

(I prefer the school Google Drive over Github to make it easier to maintain privacy.)