Foreign exchange currency rate prediction using machine learning methods

Ananiya Deneke

Drexel University

ECEC-487

12/6/2022

Foreign exchange currency rate prediction using machine learning methods

Abstract

This paper explains the importance of Data Mining in decision making process. The analysis is designed to predict the forex rates based on the past data. The approaches studied are supervised learning and unsupervised learning algorithms. Machine learning can be used to predict the future exchange rates by analyzing the data. Since pinpointing the exact price of a future currency is near-impossible we'll be classifying our analyzed data as HIGH/LOW to minimize error and to be used as another indicator in technical analysis. HIGH class is for currency pairs/stocks that go up compared to the price at the previous time stamp while LOW class is for currency pairs/stocks that go down compared to the price at the previous time stamp. The use of these technical indicators can help traders make more accurate prediction and knowing when a currency pair is going to rise or fall can prevent one trader from losing money, as well as allow him or her to invest at or near the bottom of the market. In addition to this, this project also benefits traders because they'll be able to learn how neural networks work and how they are able to learn patterns and relationships between data such as random numbers and time series.

1. INTRODUCTION

A. Introduction

Since its inception in 1949, the International Monetary Fund (IMF) has made efforts to produce reports that capture global economic performance patterns on a regular basis so that policymakers can use them when making decisions about how best to maintain stability among participating states. These reports are produced on a regular basis, namely, quarterly and semi-annually. The recent financial crisis has prompted numerous other organizations, such as the World Bank and the Organization for Economic Co-operation and Development (OECD), to produce their own economic forecasts and reports [1].

In order to better understand the global economic environment, it is important to have access to reliable economic indicators. Exchange rates play an important role in this regard because they can provide clues about relative performance trends between two economies—and if a trend appears to be unfavorable for one economy, it could be an indicator that a decline in value is likely, which can then potentially lead to further decline. The purpose of this project is to create a model that can be used to predict the future value of currency exchange rates using simple machine learning techniques, such as linear regression, random forest, and support vector machine. This paper will explore how machine learning can be used to predict currency rates and make suggestions for ways in which one might improve the accuracy of predictions by adding external factors such as GDP growth, interest rate differentials, and others.

B. Background

One of the biggest issues in trading (Technical Analysis) is predicting future trends due to the high-frequency trading that is becoming widespread. Since the advent of high-frequency trading, it has become increasingly difficult to predict future trends from past patterns because the market is now dominated by computers that analyze price trends and make trades at a much higher rate than human traders can [2]. This high level of competition increases the volatility of the market and puts human traders at a disadvantage because computer programs can react to an event in just milliseconds while humans may take minutes or longer to respond.

The algorithmic trading is performed automatically by computer software so that once the strategy has been set up, traders can sit back and watch their money multiply.

C. Problem statement

The aim of this project is to use various machine learning algorithms to create a model that can predict future exchange rates using past exchange rate data. Analysis and calculations will be done in a Jupyter notebook. For data collection, I will be using TradingView.com for historical exchange rate data of 1 PKR to USD from January 2021 until August 2021.

Objectives

- Develop a model that can predict future currency exchange rates.
- Explore the benefits of adding external factors to the model, such as GDP growth and interest rates.
- Identify other potential factors that can be used to further enhance the prediction accuracy of the model.

2. Related work

A. A Review on Machine Learning for Asset Management

Mirete-Ferrer et al (2022) discuss how they use machine learning to perform analyses on investment portfolios, providing examples of their models in action. They also discuss what they have learned, and where they hope to go with the future of machine learning as it relates to asset management. They use a genetic algorithm (GA), which is a "machine learning meta-algorithm that can be used to generate multiple alternative solutions, each with different properties." as their model. The GA they use has three main properties: "decision trees, funds of funds, and portfolio optimization." The authors applied their model to the investment portfolios for their hedge fund by using it to quantify risk on certain positions. They then fed this data into the model and let it run in order to optimize each position.

B. Reinforcement Learning in Financial Markets

Meng and Kushi [5] discuss the use of artificial intelligence being used in trading in financial markets such as forex and stock. They state that artificial intelligence has the advantages of being low risk and high speed. The disadvantages are that the market might change drastically before the algorithm can be updated, as well as many other algorithms creating similar trades. The authors believe that AI can be used in financial markets to improve trading as well as optimize the profitability of several traders and proposed reinforcement Learning for stock picking and portfolio optimization problems. This was because RL has proven to be successful in several applications when applied to dynamic problems with uncertainty and stochasticity. The main issue with this approach is that if an appropriate reward function is not already defined, then you are forced to develop one.

C. US Dollar/Turkish lira exchange rate forecasting model based on deep learning Methodologies and Time Series Analysis

In this article, the authors discuss how to forecast the exchange rate of US dollar against Turkish lira. The authors state that any forecasting model should meet certain criteria to be considered for practical applications, such as accuracy and robustness. The accuracy of a forecasting model can be measured using Mean Absolute Error (MAE) or Mean Squared Error (MSE). The MAE is calculated by taking the average of all squared errors in a time series and dividing by their total weight while MSE is calculated by taking the mean of each squared error term in an un-centered data set with respect to its corresponding weighting factor. The authors examine three forecasting models based on neural networks, support vector regression and logistic regression.

METHODOLOGY AND RESULTS

A. Research design

This paper will use a quantitative research design to train a model that can predict future exchange rates using past exchange rate data. After predicting the future exchange rate, we classify our results as HIGH/LOW. HIGH class is for currency pairs/stocks that go up compared to the price at the previous time stamp while LOW class is for currency pairs/stocks that go down compared to the price at the previous time stamp. To do this, the research will first collect data from January to August 2021 of historic trading data available on foreign currency exchanges. It will then use regression analysis to make predictions of future values based on this historical information and then classify the value as being HIGH or LOW. Finally, it will assess the accuracy of these predictions by following up with another set of new observations in the next 1 year and comparing them with previously calculated values. It will aim to predict the price of 1 Pakistan Rupee in USD and account for the error in predicting this value to the classify our results. The research will use linear regression, random forest, and support vector machine regression to predict the exchange rate of Pakistan Rupee/USD.

B. Data collection and results

To get the data from Jan 2021- August 2021 for the price action of PKR/USD, trading view’s candle stick charts can be used

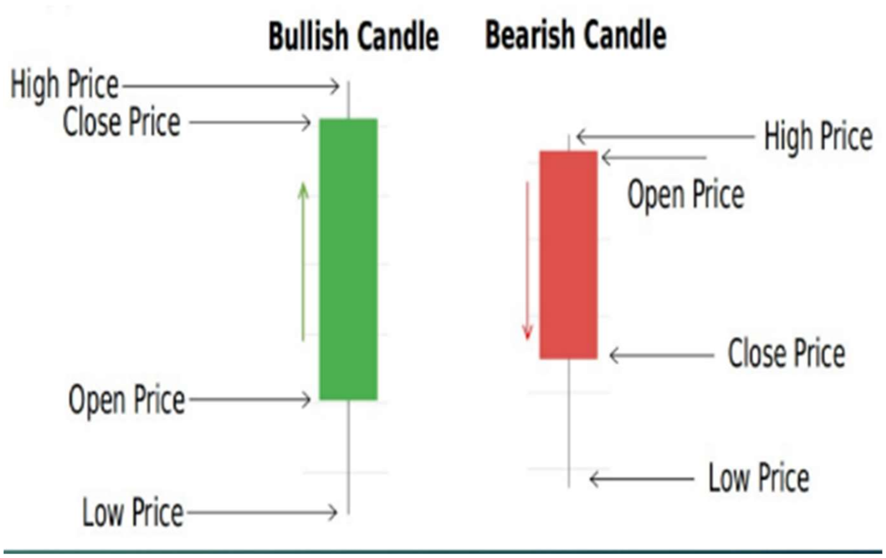

Figure-1 PKR/USD monthly chart-tradingview.com

After analyzing the structure of the data, it had 7 columns and 262 rows. The columns include; date, open(M=184.18, SD=17.38), high (M= 185.18, SD=17.75), low(M=183.46, SD=16.61), close(M=184.62, SD=17.21), adj close(M=184.62, SD=17.21), and volume(M=0, SD=0).

Figure 1: Monthly distribution (PKR/USD)

Figure 2 shows the monthly distribution of the Pakistan rupee prices in USD. The bar chart shows that the Pakistan rupee was most expensive on March and August. While the prices of the other months were average. A simple linear regression can be applied here since we just want to know the trend on the monthly one to better analyze the daily price action based on the previous' day's last price which will act as our threshold for determining our classes.

After determining our trend we go on the daily chart to analyze the previous months performance in terms of daily trends instead on the monthly one and finally determining our classes for the price action of the next day, to do this we need to find the best features

	Open	High	Low	Close	Adj Close	Volume	month	day
Open	1.000000	0.993287	0.996836	0.988099	0.988099	nan	-0.253059	0.017101
High	0.993287	1.000000	0.993165	0.990365	0.990365	nan	-0.259863	0.010028
Low	0.996836	0.993165	1.000000	0.987784	0.987784	nan	-0.255827	0.006669
Close	0.988099	0.990365	0.987784	1.000000	1.000000	nan	-0.271564	-0.000200
Adj Close	0.988099	0.990365	0.987784	1.000000	1.000000	nan	-0.271564	-0.000200
Volume	nan	nan	nan	nan	nan	nan	nan	nan
month	-0.253059	-0.259863	-0.255827	-0.271564	-0.271564	nan	1.000000	0.009461
day	0.017101	0.010028	0.006669	-0.000200	-0.000200	nan	0.009461	1.000000

Figure 2: Heatmap (correlation)

A heatmap was used to find how the variables correlated with each other and find the best features to use for our daily model. We can see that open, high, low, close, and adj close prices were strongly correlated

A. Modeling

The main aim was to explore how machine learning can be used to predict daily currency rates. Since the dataset was linear after we analyzed the heatmap, regression methods were considered in the analysis.

B. Evaluation

Evaluation was done using root means squared error (RMSE) and R-squared (R^2). The results were as follows;

Table 1: Model Evaluation Results

Model	**RMSE**	**R^2**
Linear regression	3.4969	0.96
Support vector Regression (SVR)	3.5043	0.95
Random forest	3.7237	0.95

From the results above, linear regression was the best performing model. Since it had the highest R-squared value and lowest root means squared error.

3. DISCUSSION/CONCLUSION

From the results obtained above, it is evident that past/historical exchange rate data can be used to predict future/daily exchange rate data. Therefore, as long as the relationships between past and present exchange rates can be identified, future exchange rates can be predicted and classified. This process can be applied to other exchange rates as well. We used a Jupyter notebook to analyze the data. After analyzing the structure of the data, we noted that it had 7 columns and 262 rows. The columns include; date, open(M=184.18, SD=17.38), high (M= 185.18, SD=17.75), low(M=183.46, SD=16.61), close(M=184.62, SD=17.21), adj close(M=184.62, SD=17.21), and volume(M=0, SD=0).

A heatmap was used to find how the variables correlated with each other. We saw that open, high, low, close, and ad close were strongly correlated. The machine learning methods which were considered include linear regression, support vector regression, and random forest regression. Evaluation was done using root means squared error (RMSE) and R-squared (R2). These results indicate that the results obtained were very promising in predicting future exchange rates, especially for valuing the volatility of the Pakistan economy.

Hence, we can see that machine learning techniques can be used to predict future exchange rates. Machine learning algorithms/artificial intelligence can be utilized to predict future exchange rates. The major challenge in predicting future exchange rates is that the financial markets are highly volatile and are influenced by economic factors such as interest rates, inflation rates, government policies, and exchange rates. One of the main reasons to use machine learning techniques is that they can handle a large amount of data and can find underlying relationship between variables. Furthermore, it can learn complex relationships which are difficult to model using traditional methods such as linear regression or simple machine learning algorithms (e.g., k-nearest neighbors). These challenges can be addressed by using non-linear methods such as Support vector regression and Random Forest regression.

Code-Supplement

```
#Loading the libraries
import os
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
#setting the directory

os.chdir("D:\Ananiya")
os.getcwd()
#loading the data

data =pd.read_csv("PKRX.csv")
data.head()


# shape of the dataset
print("Data Frame Dimensions:")

# rows
print("Lines:\t\t{}".format(data.shape[0]))
# columns
print("Columns:\t{}".format(data.shape[1]))

#checking the missing value
data.isnull().sum()

# Adding day and month to the train dataframe using datatime column
data['Date'] = data['Date'].apply(pd.to_datetime)

# datetime
data['month'] = data['Date'].apply(lambda x: x.month)
data['day'] = data['Date'].apply(lambda x: x.day)

# Monthly Distribution
plt.figure(figsize = (18, 6))
sns.countplot(data['month'])
plt.title('Monthly Distribution (PKR/USD)', size = 25)
plt.xticks(size = 15)
plt.yticks(size = 15)
plt.xlabel("Months (2021(Jan-Aug))", size = 20)
plt.ylabel("Frequency", size = 20)
plt.show()

#correlation
data.corr(method='spearman').style.background_gradient(cmap='coolwarm')

#split the data using 70/30 rule
X = data [["Open", "High", "Low"]]
Y = data["Close"]
```

```
x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size = 0.3, random_state = 7)
```

REFERENCES

[1] M.-Á. Galindo-Martín, M.-S. Castaño-Martínez, and M.-T. Méndez-Picazo, "Effects of the pandemic crisis on entrepreneurship and sustainable development," *Journal of Business Research*, vol. 137, pp. 345–353, 2021.

[2] C. J. Neely and P. A. Weller, "Technical analysis in the Foreign Exchange Market," *Handbook of Exchange Rates*, pp. 343–373, 2012.

[3] Y. Chihab, Z. Bousbaa, M. Chihab, O. Bencharef, and S. Ziti, "Algo-trading strategy for Intraweek foreign exchange speculation based on random forest and probit regression," *Applied Computational Intelligence and Soft Computing*, vol. 2019, pp. 1–13, 2019.

[4] P. M. Mirete-Ferrer, A. Garcia-Garcia, J. S. Baixauli-Soler, and M. A. Prats, "A review on machine learning for asset management," *Risks*, vol. 10, no. 4, p. 84, 2022.

[5] T. L. Meng and M. Khushi, "Reinforcement learning in Financial Markets," *Data*, vol. 4, no. 3, p. 110, 2019.

[6] H. Yasar and Z. H. Kilimci, "US dollar/turkish lira exchange rate forecasting model based on Deep Learning Methodologies and time series analysis," *Symmetry*, vol. 12, no. 9, p. 1553, 2020.