Name: Hamdan Ahmed

Background

Forex Exchange is the process of converting one currency into another. According to a recent triennial report from the Bank for International Settlements (a global bank for national central banks)[1], the average was more than $5.1 trillion in daily forex trading volume. Due to the worldwide accessibility of the currency trade, forex markets tend to be the largest and the most liquid asset markets in the world.

When trading forex, you are always trading a currency pair – selling one currency while simultaneously buying another. Each currency in the pair is listed as a three-letter code, which tends to be formed of two letters that stand for the region, and one standing for the currency itself. For example, USD stands for the US dollar and JPY for the Japanese yen. In the USD/JPY pair, you are buying the US dollar by selling the Japanese yen. Some of the most frequently traded Forex pairs are the euro versus the US dollar (EUR/USD), the British pound against the euro (GBP/EUR), and the British pound versus the US dollar (GBP/USD).

Most forex transactions are carried out by banks or individuals by seeking to buy a currency that will increase in value against the currency they sell. However, if you have ever converted one currency into another, for example, when traveling, you have made a forex transaction. Institutional forex trading takes place directly between two parties in an over-the-counter (OTC) market. Meaning there are no centralized exchanges (like the stock market), and the institutional forex market is instead run by a global network of banks and other organizations.

Transactions are spread across four major forex trading centers in different time zones: London, New York, Sydney, and Tokyo. Since there is no centralized location, you can trade forex 24 hours a day. Most traders speculating on forex prices do not take delivery of the

currency itself. Instead, traders will make exchange rate predictions to take advantage of price movements in the market. In a currency pair exchange, the first currency listed in a forex pair is called the base currency, and the second currency is called the quote currency. The price of a forex pair is how much one unit of the base currency is worth in the quote currency.

There are multiple platforms that can be used to trade currencies in the forex market with MetaTrader Trading Platform being one of the most commonly used worldwide. There are multiple time frames that can be used to buy and sell the currencies ranging from 1 minute to 1 month. To gain access to the Foreign Exchange market you need a broker and a trading account. As retail traders we need a broker to act as an intermediate and provide us access to the financial markets. To start trading you need a trading platform. Once you have downloaded your trading software you can log in with your trading account and start trading the instruments available. Which instruments and markets you can trade is dependent on your broker's offering.

Today the brokers compete to reduce the latency of their program to the minimum so that they (and their clients) can get the forex data before other brokers which gives them a headstart to make their appropriate trades. The only difference between people trading at home compared to the firms trading is the scale. Financial firms trade on a big scale and have more power in shaping the market with their investments.

There are many different strategies that are being used today to trade currencies and yield profit. However, those strategies are being implemented by humans who read the past data and make predictions about the future prices. In a forex market, the exchange rate depends on the many different factors with the economic situation of the two countries being one of the major contributors. If we compare that to a stock market, we only have to consider how the company is

doing and is it worth investing in it. That is the reason, the forex market is harder to predict compared to the stock market.

However, we can somewhat predict the forex market using the past exchange rates of those currencies. Using only the past data to determine the trend of the market is called technical analysis and is one of the most used strategies used by traders. The goal of this project is to learn about different machine learning algorithms and create a model to predict future prices automatically through technical analysis. Since there are many factors that affect the price of a currency pair, you can only make an estimation of future prices and not an accurate prediction. Previous prices can help us recognize different trends in a pair, bounds (supports/resistance), etc., but they cannot predict the economic situation of a country at a given time. As mentioned before, the exchange rate of a pair depends on the economic situation of those countries. For example, the release of unemployment data of a country can have a major effect on its currency. If the unemployment rate decreases, it means that the economy is strong and thus the currency will strengthen and vice versa. Therefore, we can improve the model and identify different events in the respective countries which would affect its currency positively or negatively.

Forex Trading Strategies

As mentioned above, there are many different strategies/indicators that humans use to predict the future prices and trends. We are going to discuss some of the most commonly used ones. Each indicator tells us about the market type i.e Bullish or Bearish. As the name suggests, in a bullish market, people are looking to invest money since they expect the prices to go up. On the other hand, in a bearish market, people expect the prices to go down so they take the money

out of the market. In a bullish market, traders are looking to enter the market when prices are rising so that they can sell once they believe the market has reached its peak. In a bearish market, traders are looking to exit the market when prices are falling so that they can buy again once they believe that market has reached its bottom.

Bollinger Bands are popular with technical analysts and traders in all markets, including forex. Since traders of currency look for very incremental moves to profit, recognizing volatility and trend changes quickly is essential. Bollinger Bands help by signaling changes in volatility. For generally steady ranges of a security, such as many currency pairs, Bollinger Bands act as relatively clear signals for buying and selling. There is an upper and lower band, each set at a distance of two standard deviations from the security's 21-day simple moving average. Therefore, the Bands show the volatility of the price in relation to the average, and traders can expect movements in price anywhere between the two bands. Forex traders can use the bands to place sell orders at the upper band limit and buy orders at the lower band limit. This strategy works well with currencies that follow a range pattern, but it can be costly to a trader if a breakout occurs. Since Bollinger Bands measure deviation from the average, they react and change shape when price fluctuations increase or decrease. Increased volatility is nearly always a sign that new normals will be set, and traders can capitalize using Bollinger Bands. When the Bollinger Bands converge on the moving average, indicating lower price volatility, it is known as "the Squeeze." This is one of the most reliable signals given by Bollinger Bands, and it works well with forex trading. A Squeeze was seen in the USD/JPY currency pair on Oct. 31, 2014. News that the Bank of Japan would be increasing its stimulus bond-buying policy sparked the trend change. Even if a trader did not hear about this news, the trend change could be spotted with the Bollinger Band Squeeze. In the image below, you can see that the exchange rate stayed

mostly between the upper band and the lower band. Whenever it tried to cross either of the

bands, it balanced itself back between the two bands.

The next indicator is MACD. MACD stands for Moving Average Convergence

Divergence. MACD shows the relationship between two moving averages of a currency pair.

The MACD is calculated by subtracting the 26-period Exponential Moving Average (EMA)

from the 12-period EMA. An exponential moving average (EMA) is a type of moving average

(MA) that places a greater weight and significance on the most recent data points.

The result of that calculation is the MACD line. A nine-day EMA of the MACD called the

"signal line," is then plotted on top of the MACD line, which can function as a trigger for buy

and sell signals. Traders may buy the currency pair when the MACD crosses above its signal line

and sell - or short - the security when the MACD crosses below the signal line. The image below

shows the exchange rate of the pair along with its MACD line (blue) and signal line (red). We

can see that whenever the MACD line crossed above the signal line, the trend became bullish

(buying point) and vice versa. MACD is often displayed with a histogram which shows the

difference between the MACD line and the signal line. If the MACD line is above the signal line,

then the histogram bars will be above 0 and vice versa. The histogram is used by traders to identify the strength of the bullish/bearish trend; the higher the bars are, the stronger the trend is.

The next indicator is RSI. RSI stands for Relative Strength Index and is one of the most popular indicators used in the forex market. The RSI is a basic measure of how well a currency pair is performing against itself by comparing the strength of the up days versus the down days. This number is computed and has a range between 0 and 100. A reading above 70 is considered bullish, while a reading below 30 is an indication of bearishness. Reading above 70 represents that the currency pair has been overbought, therefore, to get the market back in balance, the price will drop. Similarly, a reading below 30 shows that the pair has been oversold, therefore, to balance the market, the price will increase. Hence, using that information, we should enter the market when the reading is below 30 and sell before or at the 70 mark. Overbought describes a period of time where there has been a significant and consistent upward move in price over a period of time without much pullback. The term Oversold describes a period of time where there has been a significant and consistent downward move in price over a period of time without much pullback. RSI is used by traders to confirm the trend and the reversal. Depending on the

time frame you are using to trade, you can adjust the RSI's length. The length (e.g. 14 days) will give you an estimate of how the pair has been doing compared to itself in that period. In the graph below, you can see that exchange rate of a currency pair. The long range (14 day) RSI was mostly stable but we can get a lot of more information from the short range (5 day) RSI. Whenever the pair was overbought, the RSI crossed the 70 mark period and is expected to come down to balance it out. Similarly, whenever the RSI crosses the 30 mark, it is expected to come back up. As shown in the graph, RSI can be used to identify the points of entry and exiting the market.

https://i.ytimg.com/vi/QDwRIAixsDU/hqdefault.jpg

The last indicator is more of a strategy. Naked Graph is one of the most popular ways to determine the future trends. This can be used to see pivot points, points of interest along with trends. Point of interest occurs when there was a previous point where the pair gained support or resistance. Support for a pair in the x amount of period is the lowest point the pair has reached in that period. We expect the pair to bounce back up once it reaches that point again. Similarly, resistance for a pair in the x amount of period is the highest point that pair has reached in that period. We expect that pair to come down once it has reached that point again. If the pair crosses the resistance or support, it is known as breaking of the trend and is a warning sign to get out of

the market since the market just became unpredictable. The strength of a resistance and a support point can be measured by how many times the market has tried to cross those points but could not. In the image below, we can see multiple resistance and support points/zones. We can tell that the market tried to cross the second resistance zone (at $1.2533) multiple times. Even though it did cross it once and went to $1.2740, those resistance points are still considered strong and we can expect the pair to stay below it. If it did cross it and is still going up, then we can expect it to stop or reflect back after the resistance at $1.2740. One thing to be very careful about is sometimes the market tests itself. Therefore, it is very common that we will see that the pair actually crossed the resistance or support a little bit before coming back.

https://www.babypips.com/learn/forex/support-and-resistance
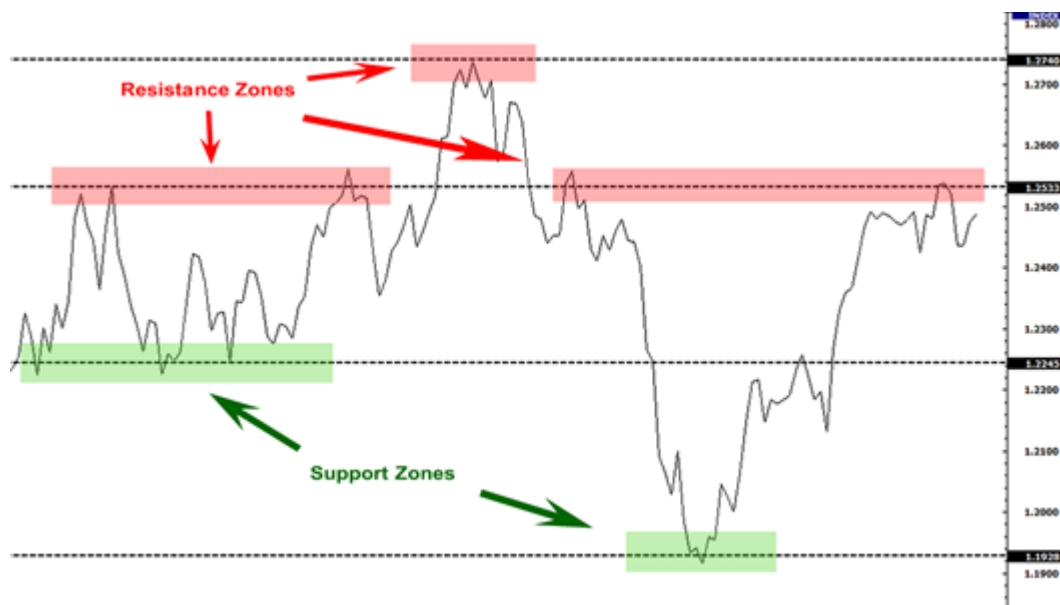
If we combine all the indicators above, it gives us a pretty good idea about the future trend and helps us determine prices.

Below is my forex chart (from forex.com) where I have used all the indicators mentioned above. To make a buy/sell decision from this chart, we can just look at the RSI. We can see that the

market was not doing a lot for a while and picked up the pace and went up consistently (overbought). RSI confirms the trend and we can see that in March, it crossed the 70 mark period and right after that the market crashed to balance itself back. So using RSI, we know that we should sell as soon as the RSI crosses the 70 mark period since the market is going to go down to balance itself back.

www.forex.com

As mentioned in the beginning, technical analysis is used by humans to predict the market. However, what if we can teach a machine to perform the analysis and determine the future prices, trend and buying/selling points? That's where machine learning comes in handy. We are going to look at some machine learning algorithms that can be used to perform technical analysis.

Machine Learning

According to Netflix's Research (research.netflix.com), more than 80% of the shows people watch on netflix come from Netflix recommendation. That means the recommendations

must be very accurate to the taste of the user. Ever wondered how? Netflix uses machine learning to gather large amounts of data about its users, find patterns in them and determine the results (recommendations). Today, machine learning is being used in almost every major service we use; Netflix, Spotify, Google, Facebook, Amazon, Siri, Alexa, etc. These services collect information about its users and use that information to determine the best outcome for each specific user.

Within Machine Learning, there are two different types of tasks; Supervised Learning and Unsupervised Learning. Supervised learning is simply a process of learning algorithms from the training dataset. It is where you have input variables and an output variable and you use an algorithm to learn the mapping function from the input to the output. The aim is to approximate the mapping function so that when we have new input data we can predict the output variables for that data. Unsupervised learning is modeling the underlying or hidden structure or distribution in the data in order to learn more about the data. Unsupervised learning is where you only have input data and no corresponding output variables.

Machine Learning Algorithms

There are many machine learning algorithms but in this section we are going to look at some which are more relevant to this project and the forex market.

The first algorithm is Linear Regression. Before knowing what is linear regression, let us get ourselves accustomed to regression. Regression is a method of modelling a target value based on independent predictors. This method is mostly used for forecasting and finding out the cause and effect relationship between variables. Regression techniques mostly differ based on the number of independent variables and the type of relationship between the independent and dependent variables.

Simple linear regression is a type of regression analysis where the number of independent variables is one and there is a linear relationship between the independent variable (x) and dependent variable (y). The red line in the graph below is referred to as the best fit straight line. Based on the given data points, we try to plot a line that models the points the best. The line can be modelled based on the linear equation shown:

$y = a\_0 + a\_1 * x$
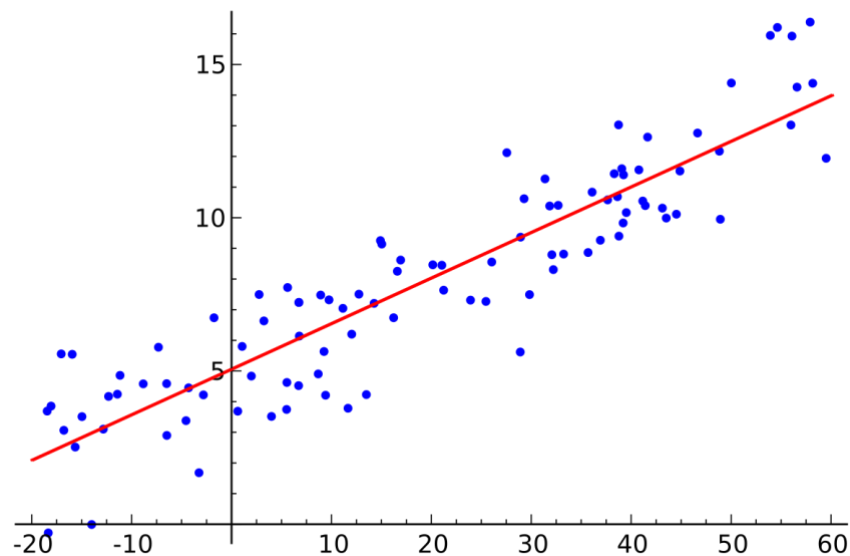
The motive of the linear regression algorithm is to find the best values for a_0 and a_1.

The forex data is in the form of numeric data (prices) from previous year. This data can also be categorized as time-series data. We can use linear regression on the past forex data to determine the future trend of the prices.

Next is a group of algorithms known as classification algorithms or classifiers. Classification is the process of predicting the class of given data points. Classes are sometimes called targets/ labels or categories. Classification predictive modeling is the task of

approximating a mapping function (f) from input variables (X) to discrete output variables (y). A very basic example of classification algorithm in use is for detecting spam emails. A classification algorithm can be trained on a large amount of spam and non-spam emails such that it recognizes patterns in each category. Then it can be used on new emails to determine if the email is spam or not. Classification belongs to supervised learning where we have an idea of what the output of our model should be. Many currency pairs like EUR/USD repeat patterns which is why they are one of the most predictable pairs and are very popular amongst traders. These patterns allow us to identify different points in the graph which can be classified as buying or selling points. We can use classification algorithms and teach the machine to identify these points.

There are several different classification algorithms that can be used for this purpose. One of them is Decision Trees. To determine a buying or selling point, we check different indicators. This process can be converted into a Decision Tree, therefore, I believe, Decision Trees are very suitable to classify buying and selling points. For an example, we can ask if the RSI value is above the 70 mark. If yes, then it is a selling point. Similarly, if the RSI value is below the 30 mark, it is a buying point. Usually, we will check multiple conditions/indicators before confirming a buying or a selling point. For example, if RSI is above 70 and the exchange rate of the pair is above the upper bound of the boilenger band, then it is definitely a selling point. A decision tree is a flowchart-like structure in which each internal node represents a test on a feature, each leaf node represents a class label (decision taken after computing all features) and branches represent conjunctions of features that lead to those class labels. The paths from root to leaf represent classification rules.

In general, we can use machine learning algorithms in the forex market to determine the future prices and identify buying (entering) or selling (exiting) points. Therefore, those tasks can be categorized as regression and classification tasks.

As mentioned before, the past exchange rate data for a currency pair can be categorized as time-series data. Time series data is a sequence of data points, measuring the same thing over time, stored in some order. Time series data of an element shows how that element changed over a time. The exchange rate data of a pair tell us how the pair was doing over time and help us determine the trend. We are going to use Weka to perform time-series analysis and extract meaningful information from the data.

Using Weka to perform Time-Series Analysis

There are many ways to perform time-series analysis on the data to make predictions. We use Weka to perform the analysis. The steps to perform time-series analysis and forecast are:

1) Using Tools → Package Manager, download the "Time Series Forecasting" package. You will see a new tab "forecast"

2) Open the .arff data file in the pre-process tab and move to the forecast tab.

3) Select the attribute you want to predict in the target selection box.

4) Select the number of time units you want to forecast (at the right side of the window)

5) Time-Stamp should be the data field in your data.

6) In Advanced Configuration → Base Learner, you can select the algorithm you want to use.

7) In Advanced Configuration → Evaluation tab, you can select to evaluate on held out training data. You can select the amount of data you want to put aside for testing. We use 0.3 to ensure we have enough data to train and test on.

8) Click Start.

Results

We were able to get forex data of EUR/USD from November 29, 2003 till May 8, 2020 from Yahoo Finance. The data of the currency pair included the open, high, low, close, adj. close, and volume values. The more relevant fields are open (the opening price of that pair on that day), high (highest the pair reached that day), low (the lowest the pair reached that day, and close (the closing price of that pair for that day). Usually we want to predict the opening and the closing prices for a pair which tells us the overall trend of that pair for that day. We are going to use Weka to train a model and test it on the EUR/USD pair.

We will be forecasting 7 units. Since each unit represents the exchange rate for that day, our forecasted data will represent the exchange rate for a week. We will be using supervised learning and will be splitting the data in training and testing sets. We can either split the data by percentage or by instances. For this experiment, we will set aside the data of the last 30 days for testing and evaluate our model on it. We can evaluate our model in Weka through calculating different errors given below:

1. Mean absolute error (MAE): sum(abs(predicted - actual)) / N

2. Mean squared error (MSE): sum((predicted - actual)^2) / N

3. Root mean squared error (RMSE): sqrt(sum((predicted - actual)^2) / N)

4. Mean absolute percentage error (MAPE): sum(abs((predicted - actual) / actual)) / N

5. Direction accuracy (DAC): count(sign(actual_current - actual_previous) == sign(pred_current - pred_previous)) / N

6. Relative absolute error (RAE): sum(abs(predicted - actual)) / sum(abs(previous_target - actual))

7. Root relative squared error (RRSE): sqrt(sum((predicted - actual)^2) / N) / sqrt(sum(previous_target - actual)^2) / N)

Since forex data usually do not have much fluctuation in it, a numerical difference between the predicted and the actual value would not give us a lot of information. Therefore, we will be using Mean Absolute Percentage error which would give us percentage error between the actual and the predicted value.

Below is the output received. It shows the result of evaluation on the testing data (last 30 days).

| inst#	actual	predicted	error
4261	1.1139	1.106	−0.0079
4262	1.103	1.1149	0.0119
4263	1.1027	1.1026	−0.0001
4264	1.0954	1.1019	0.0065
4265	1.0847	1.0957	0.011
4266	1.0807	1.0841	0.0034
4267	1.0804	1.0805	0.0001
4268	1.0895	1.0808	−0.0087
4269	1.0864	1.0903	0.0039
4270	1.0927	1.0869	−0.0058
4271	1.0933	1.0927	−0.0006
4272	1.0923	1.094	0.0017
4273	1.0985	1.0922	−0.0063
4274	1.0905	1.0989	0.0084
4275	1.0858	1.0907	0.0049
4276	1.0867	1.0852	−0.0015
4277	1.0865	1.0871	0.0006
4278	1.0856	1.0869	0.0013
4279	1.081	1.0858	0.0048
4280	1.0777	1.081	0.0033
4281	1.0824	1.0776	−0.0048
4282	1.0825	1.0828	0.0004
4283	1.0837	1.0831	−0.0005
4284	1.0877	1.0837	−0.004
4285	1.0945	1.0882	−0.0064
4286	1.096	1.0952	−0.0007
4287	1.0904	1.0964	0.006
4288	1.0839	1.0902	0.0063
4289	1.0799	1.0834	0.0036
4290	1.0837	1.0798	−0.0039

The evaluation results with MAPE are:

```
=== Evaluation on test data ===
Target                          1-step-ahead  2-steps-ahead  3-steps-ahead  4-steps-ahead  5-steps-ahead  6-steps-ahead  7-steps-ahead
=====================================================================================================================================
Close
  N                                       30             29             28             27             26             25             24
  Mean absolute percentage error      0.3951         0.5689         0.7589          0.901         0.9716         0.9232         0.8903

Total number of instances: 30
```

From the output above, we can see as we try to predict the values in future, the error increases.

Below is the graph of the actual and predicted values of the last month:
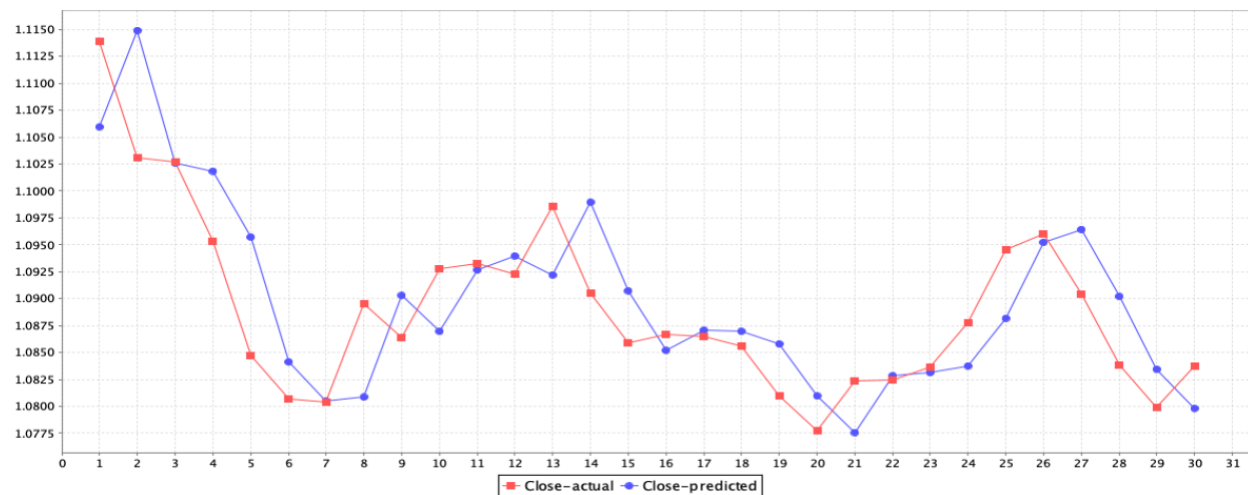
This model might give us an idea of how the market will do overall in the long run but it is very hard to get an idea of what the exact daily closing price would be for a specific day. However, from this project we did get to learn a lot about machine learning, different machine learning tasks, and how to implement them in Weka.

Sources:

https://www.bis.org/statistics/rpfx19_fx.htm

https://www.quora.com/How-do-you-read-Bollinger-Bands

https://www.babypips.com/learn/forex/macd

https://towardsdatascience.com/introduction-to-machine-learning-algorithms-linear-regression-14c4e325882a

https://www.youtube.com/watch?v=cv3FitkI4KQ

https://archive.ics.uci.edu/ml/datasets/ISTANBUL+STOCK+EXCHANGE

https://www.forex.com/en-us/

https://www.babypips.com/learn/forex/support-and-resistance

https://wiki.pentaho.com/display/DATAMINING/Time+Series+Analysis+and+Forecasting+with+Weka#TimeSeriesAnalysisandForecastingwithWeka-3.2.5Evaluation