

FOREX DATA ANALYSIS USING WEKA

Luciana Abednego and Cecilia Esti Nugraheni

Department of Informatics, Parahyangan Catholic University, Indonesia

ABSTRACT

This paper conducts some experiments with forex trading data. The data being used is from kaggle.com, a website that provides datasets for machine learning and data scientists. The goal of the experiments is to know how to design many parameters in a forex trading robot. Some questions that want to be investigated are: How far the robot must set the stop loss or target profit level from the open position? When is the best time to apply for a forex robot that works only in a trending market? Which one is better: a forex trading robot that waits for a trending market or a robot that works during a sideways market? To answer these questions, some data visualizations are plotted in many types of graphs. The data representations are built using Weka, an open-source machine learning software. The data visualization helps the trader to design the strategy to trade the forex market.

KEYWORDS

forex trading data, forex data experiments, forex data analysis, forex data visualization.

1. INTRODUCTION

When planning a forex trading system, a trader needs to carefully design the system and extensively test it. Besides the help of some technical indicators and fundamental analysis [1][2], a trading system needs to set many risk management parameters, such as stop loss and take profit [3]. These parameters play an important rule to determine the trader's target profit and also limit the loss risk of each open trade.

To investigate the ideal level for risk management parameters and the trading system, this research tries to find the answer to those questions. Some experiments are conducted in the H1 timeframe, which updates the price hourly. Two main currency pairs with different time ranges are used in this paper: EUR/USD (1 year) and USD/JPY (20 years). The experiments use the dataset from Kaggle, a website that provides many kinds of datasets for machine learning and data scientists [4]. Some data visualization techniques are used to represent the result of these data using Weka. Weka is open-source software that provides many machine learning techniques and data visualization tools [5].

2. PROBLEM DESCRIPTION AND ANALYSIS

This section describes the problem domain and data that want to be investigated in this research.

2.1. Forex Trading

Forex (foreign exchange) is a global marketplace where the banks, corporations, investors, and individual traders exchange foreign currencies for a variety of reasons. The fluctuations of these currencies are the target for traders for making some profit. But at the same time, a trader risks their account when the market moves against his open position. The currencies are traded in

pairs. The four major currency pairs are EUR/USD, USD/JPY, GBP/USD, and USD/CHF. Figure 1 shows the approximate volume breakdown per currency pair [6].

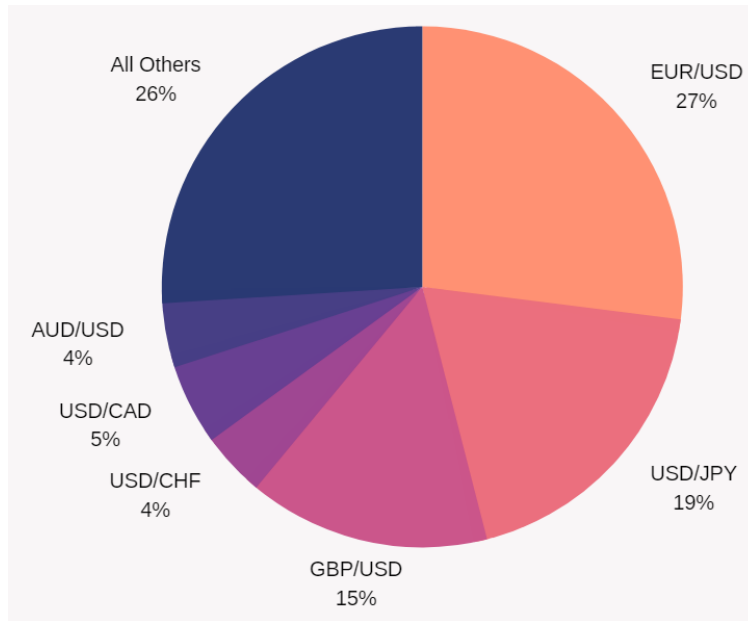


Figure 1. Estimated Trading Volume by Currency Pair

The forex market works 24 hours a day, 5 days a week. Table 1 shows the opening and closing times [6].

Table 1. Global Trading Hour Schedule

Time Zone	New York	GMT
Tokyo Open	7:00 p.m.	00:00
Tokyo Close	4:00 a.m.	09:00
London Open	3:00 a.m.	08:00
London Close	12:00 p.m.	17:00
New York Open	8:00 a.m.	13:00
New York Close	5:00 p.m.	22:00

2.2. Forex Risk Management

When a trader opens a position in the forex market, two actions can be taken: buy or sell. If the trader thinks that the price will go upward, he is supposed to open a buy position. On the contrary, if the trader considers that the price will go downward, he is supposed to open a sell position. After a trader chooses one of that action, but unluckily the market moves against its open position, the trader will lose. In this case, he has to protect his account by limiting the loss he's suffered. There are many types of risk management strategies [3]. Some parameters that can be set to limit the loss of any open trade are stop loss and target profit. In this paper, some experiments are conducted to investigate the ideal level to set these parameters.

2.3. Data Preparation and Mining

This research uses past forex data that is gained from Kaggle, a website that provides many kinds of datasets for machine learning and data scientists [4]. We choose the H1 timeframe of the two

top biggest volume currencies traded in the global market: EUR/USD & USD/JPY [6]. This raw data is then cleaned, transformed, and represented in some visualizations charts by using Weka. Weka is an open-source data mining and visualization framework. Weka was developed at the University of Waikato, New Zealand. Figure 2 shows the user interface of Weka. This paper uses Weka as a tool for data visualization and mining.

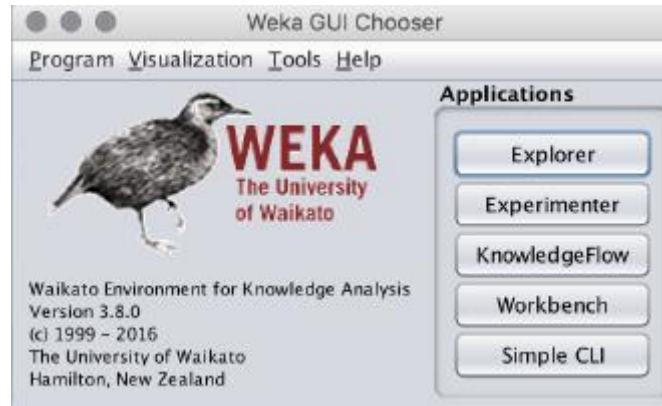


Figure 2. Weka Interface

3. EXPERIMENT SETUP AND RESULT

Experiments are conducted to the top three biggest volume traded currency pairs: EUR/USD USD/JPY, and GBP/USD. The H1 timeframe for 20 years (1999-2019) is used for all the experiments. As mention before, the datasets that are used in these experiments are from Kaggle.com [3], a website that provides many kinds of datasets for machine learning and data science purposes. These datasets use pip (price in percentage), which is the smallest value by which a currency may fluctuate in the forex market [5]. The goals of these experiments are explained in the following sections.

3.1. Experiment with Information Gain

The Kaggle dataset consists of 7 attributes (see Table 2). These attributes are then evaluated based on the information gain value. Information gain is a method to score all attributes and then sort these attributes based on the most important attributes to the price change above 10 pips. Table 2 shows the sorted sequence of attributes from the highest influence attribute to the less.

Table 2. Information Gain Experiment.

No.	Attribute	Information Gain
1	Date	0.1438
2	Volume	0.125
3	Low	0.0661
4	Close	0.0661
5	High	0.0657
6	Open	0.0656
7	Hour	0.0599

The two top attributes are date and volume. This shows that in some certain dates, there are some trends happened in the forex market. The price change above 10 pips and the number of volumes influences this trend. Many fundamental news affects this trend.

3.2. Experiment to Determine the Level of Some Risk Management Parameters

When design a trading system, a trader needs to define a strategy of how to trade the foreign currency pairs. Beside the technical and fundamental analysis, a trader needs to define the risk management system to control the loss risk of each open trade. Stop loss and take profit are the tools in risk management system. To determine the best level of these tools, this research investigates a forex dataset for 20 years (from 1999 until 2019) of the three most traded currency pairs: (1) EUR/USD, (2) USD/JPY, (3) GBP/USD.

Based on the first experiment, the dataset is then categorized based on the pip change that shows whether the market is on the condition of trending or sideways. To cluster the dataset into groups, in this experiment a new attribute, **Class 10 Pip Change**, was added based on the open price of the next candle minus the close price of the previous candle. This attribute has three possibilities of value:

- **-10pips**: the price decrease more than or equals to 10 pips
- **ranging**: the price change below 10 pips
- **+10pips**: the price increase more than or equals to 10 pips

Table 3-4 and Figure 3 show the result of this experiment.

Table 3. Experiment with Class 10 Pips Change

No.	Attribute Value for Class 10 Pip Change	Number of records		
		EUR/USD	USD/JPY	GBP/USD
1	-10pips	21,180	19,494	25,972
2	ranging	86,522	89,618	76,555
3	+10pips	21,098	19,688	26,273
Total		128,800	128,800	128,800

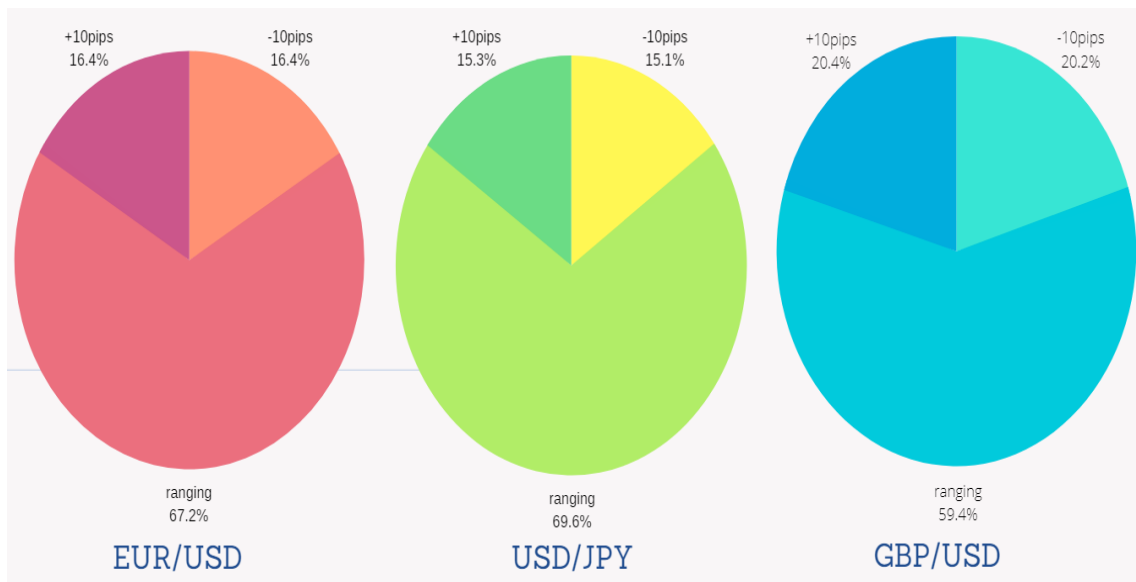


Figure 3. Experiment with Class 10 Pips Change

Table 4. Experiment with Class 10 Pips Change (in percent)

No.	Attribute Value for Class 10 Pip Change	Number of records (in percent)		
		EUR/USD	USD/JPY	GBP/USD
1	-10pips	16.4 %	15.1 %	20.2 %
2	Ranging	67.2 %	69.6 %	59.4 %
3	+10pips	16.4 %	15.3 %	20.4 %
Total		100 %	100 %	100 %

Table 4 represents the number of records in percent. This data can be used to design the trading system:

- About 60 % of all the records for each main currency pairs fluctuate below 10 pips. A trader can consider this fact to determine many trading strategies. A trading system that can be dealt with this ranging market is more preferable than those which can't.
- The number of downtrend and uptrend are comparable for each currency pair:
 - EUR/USD: 16.4 % : 16.4 %
 - USD/JPY: 15.1 % : 15.3 %
 - GBP/USD: 20.2 % : 20.4 %

3.3. Experiment with 25 Pips of Currency Fluctuation

Similar with the previous experiment, the dataset is categorized based on the 25 pips change that shows whether the market is on the condition of trending or sideways. In this experiment, a new attribute, **Class 25 Pip Change**, was added based on the open price of the next candle minus the close price of the previous candle. This attribute has three possibilities of value:

- **-25pips**: the price decrease above 25 pips
- **ranging**: the price change below 25 pips
- **+25pips**: the price increase above 25 pips

Table 5. Experiment with Class 25 Pips Change

No.	Attribute Value for Class 10 Pip Change	Number of records		
		EUR/USD	USD/JPY	GBP/USD
1	-25pips	5,150	4,125	7,813
2	ranging	118,554	120,865	113,347
3	+25pips	5,096	3,810	7,640
Total		128,800	128,800	128,800

Table 6. Experiment with Class 25 Pips Change (in percent)

No.	Attribute Value for Class 10 Pip Change	Number of records (in percent)		
		EUR/USD	USD/JPY	GBP/USD
1	-25pips	4 %	3.2 %	6.1 %
2	Ranging	92 %	93.8 %	88 %
3	+25pips	4 %	3 %	5.9 %
Total		100 %	100 %	100 %

Table 5-6 and Figure 4 show the result of this experiment.

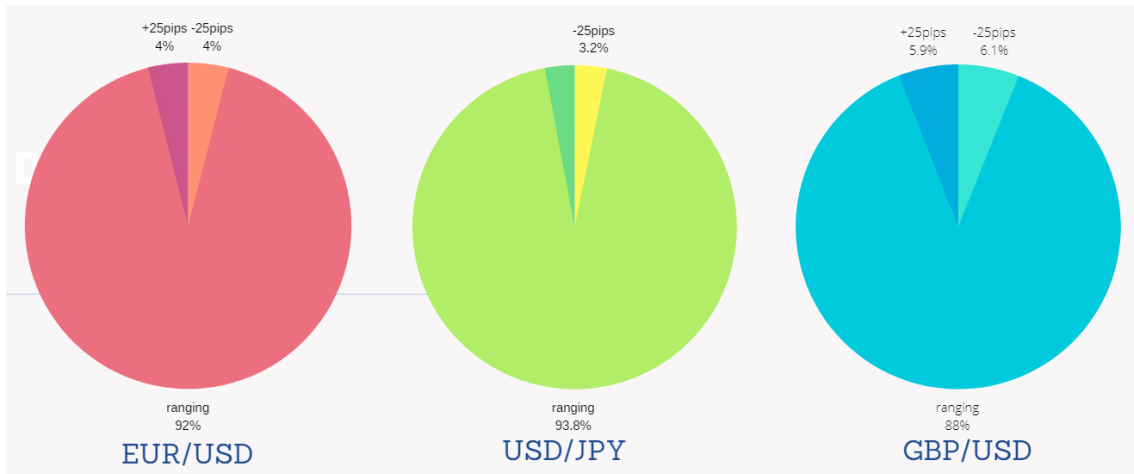


Figure 4. Experiment with Class 25 Pips Change

As can be seen from these results:

- About 90 % of all the records for each main currency pairs fluctuate below 25 pips. A trader can consider this fact to determine many trading strategies. For example, to determine the level of stop loss and take profit. A trading system that can be dealt with this ranging market is more preferable than those which can't.
- The number of downtrend and uptrend are comparable for each currency pair:
 - EUR/USD: 4 % : 4 %
 - USD/JPY: 3.2 % : 3 %
 - GBP/USD: 6.1 % : 5.9 %

From Table 3-4 and Table 5-6, we derive Table 7-8 and Figure 5.

Table 7. Experiment with Class X Pips Change

No.	Attribute Value for Class Pip Change	Number of records		
		EUR/USD	USD/JPY	GBP/USD
1	(-..., -25) pips	5,150	4,125	7,813
2	[-10, -25] pips	16,030	15,369	18,159
3	(-10, 10) pips	86,522	89,618	76,555
4	[+10,+25] pips	16,002	15,878	18,633
5	(25, ...) pips	5,096	3,810	7,640
Total		128,800	128,800	128,800

Table 8. Experiment with Class X Pips Change (in percent)

No.	Attribute Value for Class Pip Change	Number of records		
		EUR/USD	USD/JPY	GBP/USD
1	(-..., -25) pips	4 %	3.2 %	6.1 %
2	[-10, -25] pips	12.4 %	11.9 %	14.1 %
3	(-10, 10) pips	67.2 %	69.6%	59.4 %
4	[+10,+25] pips	12.4 %	12.3%	14.5 %
5	(25, ...) pips	4 %	3 %	5.9 %
Total		100 %	100 %	100 %

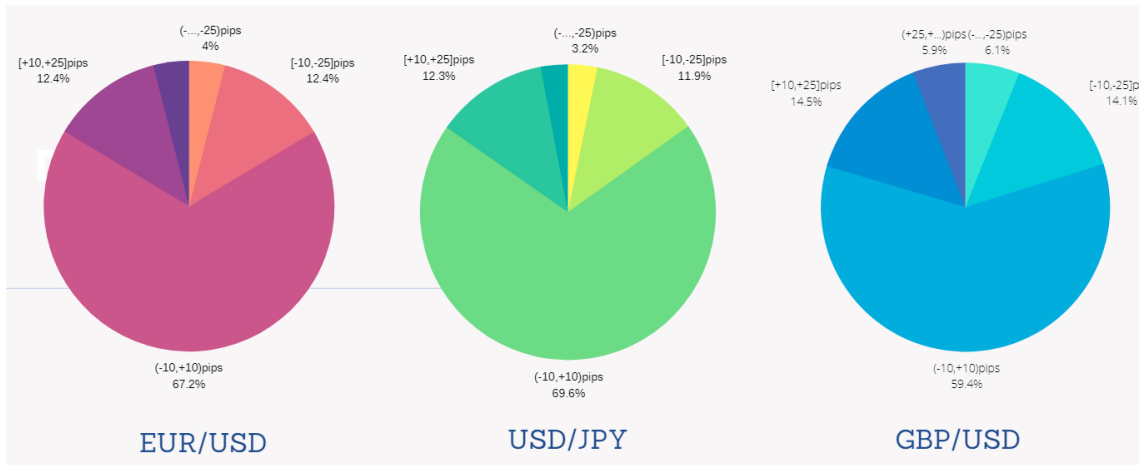


Figure 5. Experiment with Class X Pips Change

3.4. Experiment Comparison of The Up Prices to The Down Prices

The goal of this experiment is to know the comparison of the up prices to the down prices in the top three major pairs. A new attribute **Class Price Up** was added in this experiment, with two possibilities of value: TRUE or FALSE. TRUE means the next close price is higher than the previous close price. FALSE means the contrary.

Table 9 and Figure 6 show the result of this experiment.

Table 9. Experiment based on Price Up Class

No.	Attribute Value for Class Price Up	Number of records		
		EUR/USD	USD/JPY	GBP/USD
1	TRUE	65,135	63,650	63,418
2	FALSE	63,665	65,150	65,382
Total		128,800	128,800	128,800

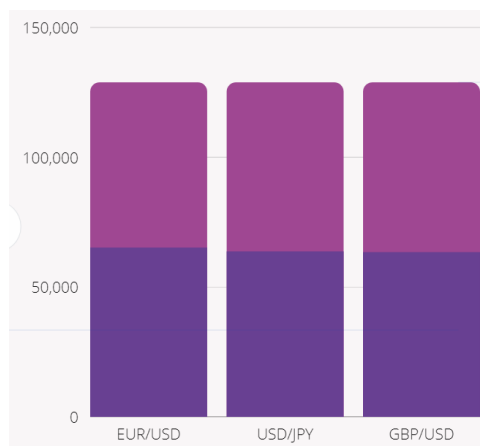


Figure 6. Experiment with Class Price Up

This experiment shows that the number of up prices more or less equals the number of down prices. From this experiment, the trader has a 50:50 percent chance to buy or sell decisions.

3.5. Comparison of Up Trend to Down Trend

The goal of this experiment is to know the comparison between the uptrend to the downtrend of the three top currency pairs: (1) EUR/USD, (2) USD/JPY, (3) GBP/USD. Some new attributes are added to the Kaggle dataset to gain the pattern of the uptrend and downtrend. An uptrend can be indicated by higher high and a higher low. On the other hand, a downtrend can be indicated by lower high and lower low. Based on that fact, Table 10 and Figure 7 show the result of this experiment.

Table 10. Experiment with Class 10 Pips Change.

No.	Attribute Value Class Up/Down Trend	Number of records		
		EUR/USD	USD/JPY	GBP/USD
1	Up Trend	32.58 %	32.81 %	32.62 %
2	Down Trend	32.1%	31.89 %	31.96 %

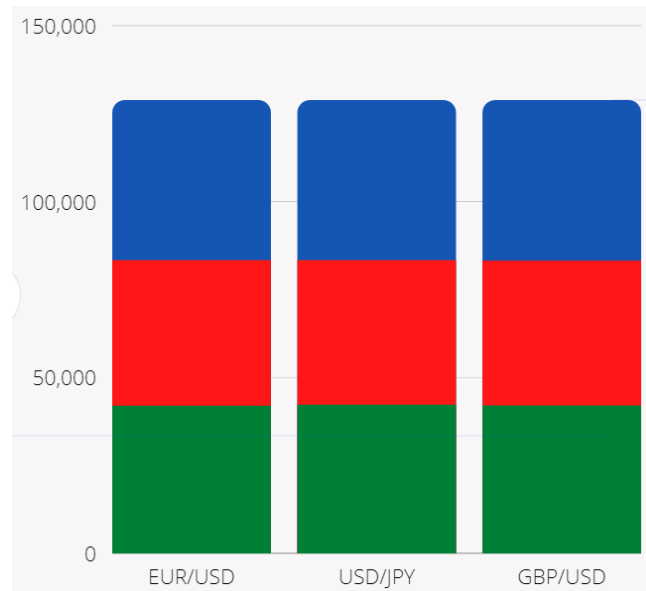


Figure 7. Data Composition of Up Trend (Green Bars), Down Trend (Red Bars), and Ranging (Blue Bars)

From this data, it can be seen that the proportion of uptrend, down, and ranging for these three main pairs are about 30 % : 30 % : 40 %.

3.6. Experiment Time of EUR/USD Trending Market

The goal of this experiment is to know the tendency of the time when the EUR/USD is trending during a day. The price of each transaction to the time of a day is plotted in the chart below (see Figure 8).

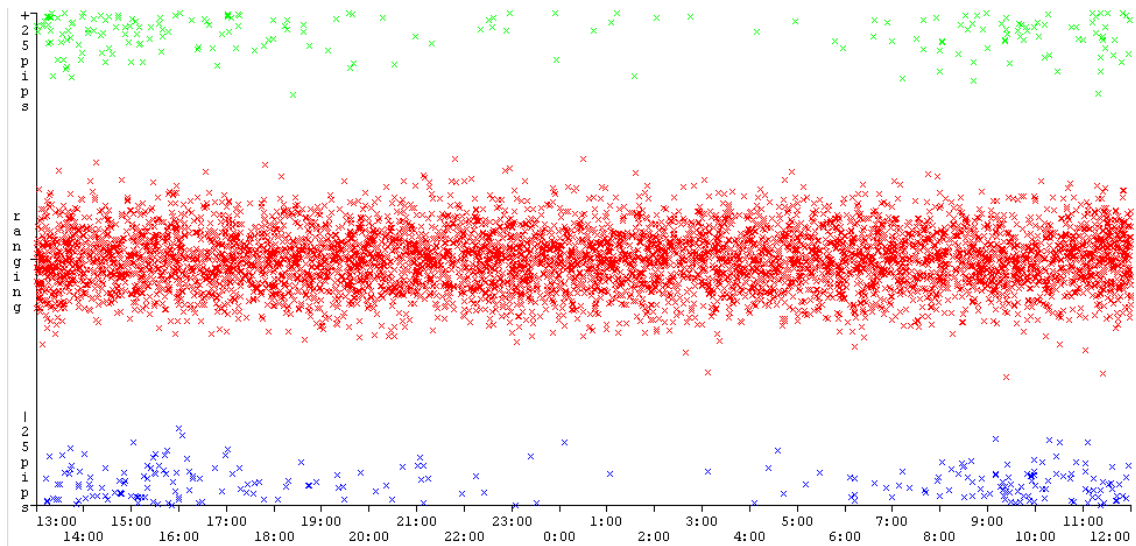


Figure 8. Market price to the time chart

The X-axis shows the time of the days and the Y-axis shows the attribute value of Class 25 Pips Change: -25pips, ranging, or +25pips. The red dots show the ranging market that happens most of the time of any day. The green dots represent the up-trending market that moves above 25 pips. From the chart, it can be seen that most of the trending market happened during office hours (7 a.m. to 5 p.m.). Outside that time, the trend rarely happened.

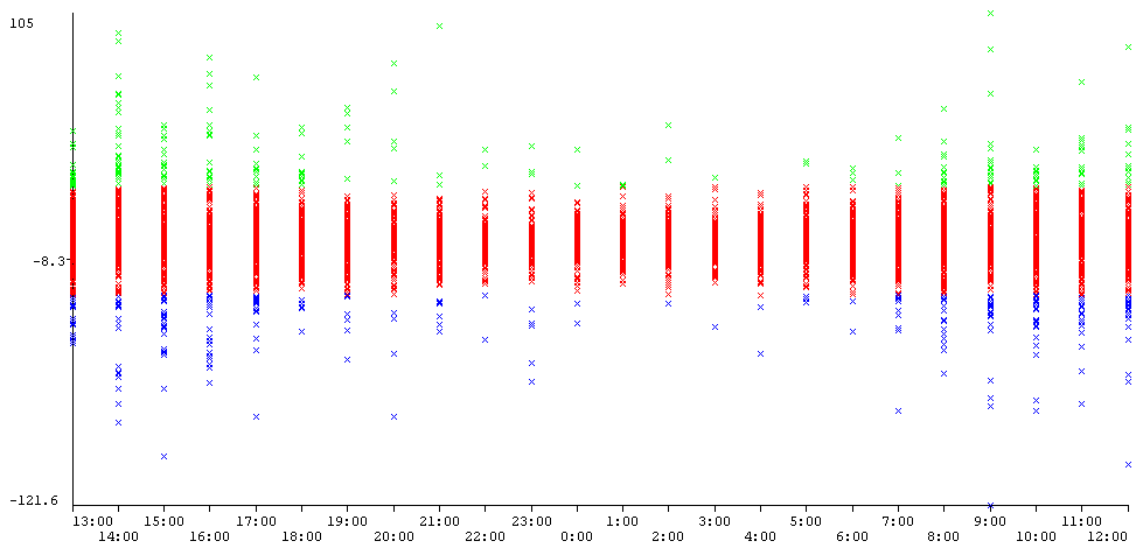


Figure 9. EUR/USD Fluctuation Range (in pips)

From this experiment, if the trader's used the trending algorithm, it would be better to apply it during office hours. On the other hand, if the trader uses an algorithm that can be dealt with ranging markets, it can be applied most of the time of the day. The trader can set the forex parameters, such as stop loss and take profit below 25 pips to gain more profit or reduce the risks. Figure 9 shows the pip change range to time in the EUR/USD forex market. From this figure, it can be seen that the most trending market happened at about 14:00 - 15:00. If some of this data is selected (see Figure 10), it can be seen that when the market starts to open, the possibility of the downtrend is more often than the uptrend.

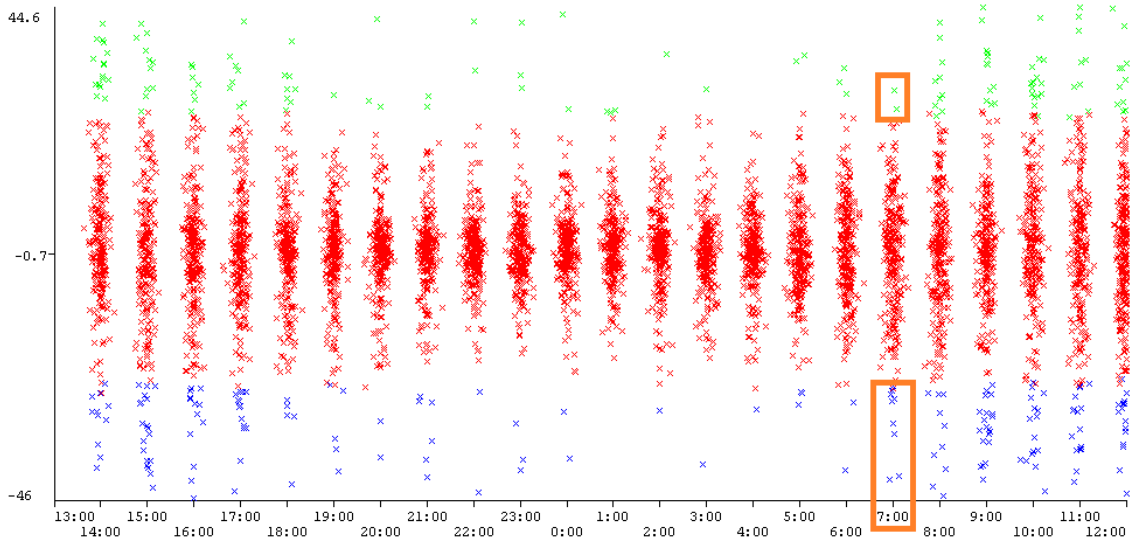


Figure 10. The Possibility of the Downtrend When The Market Starts to Open

3.7. Experiment with USD/JPY Trending Time

The goal of this experiment is to know the best time to trend USD/JPY if a trader uses an algorithm that counts on-trend. Figure 11 shows the transactions plotted against time. The red dots show the upward trend above 25 pips. The green dots show downward trends of more than 25 pips. The blue dots represent the ranging market. The trend not only happened during the office hours (7 a.m. - 5 p.m.) but also during midnights (11 p.m. - 3 a.m.).

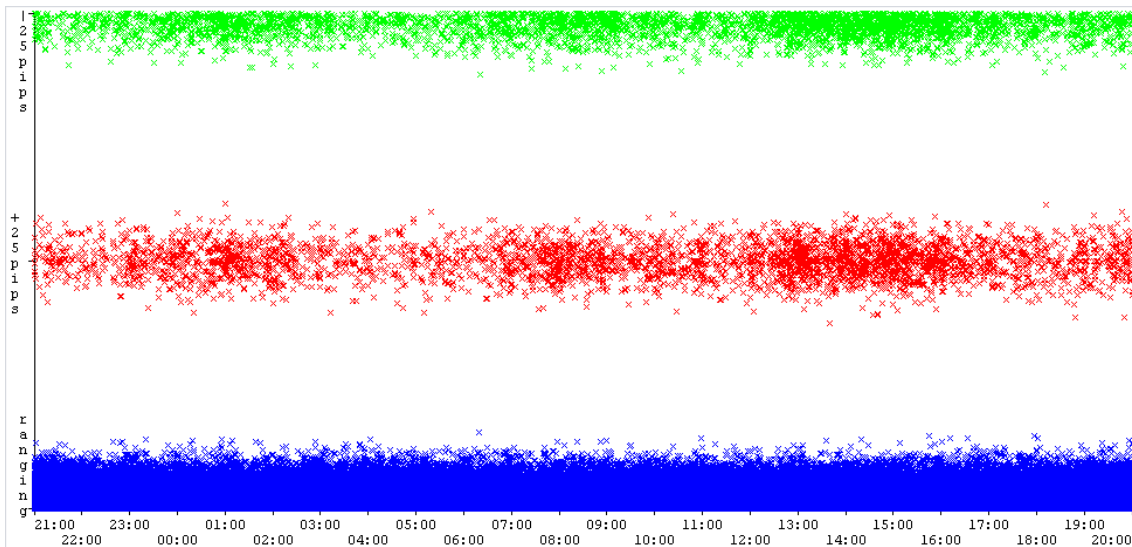


Figure 11. Experiment with Class 25 Pips Change in USD/JPY

If we decrease the threshold to 10 pips, the chart will look like Figure 12 . This data can be used to determine the level of stop loss and take profit. Most of the time, the market fluctuates between -10 pips to +10 pips.

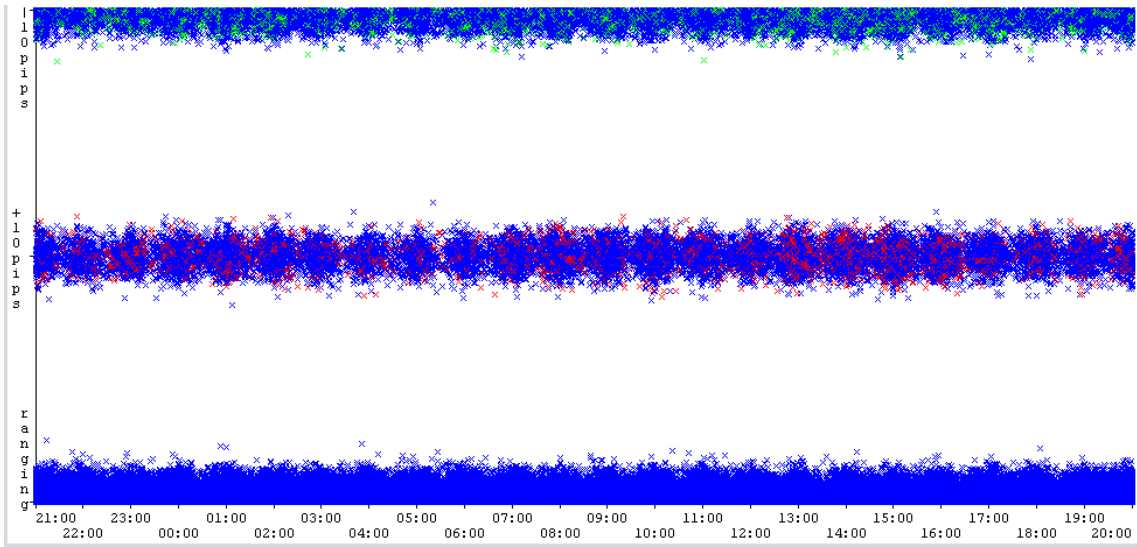


Figure 12. Price movement (in pip) plotted to time

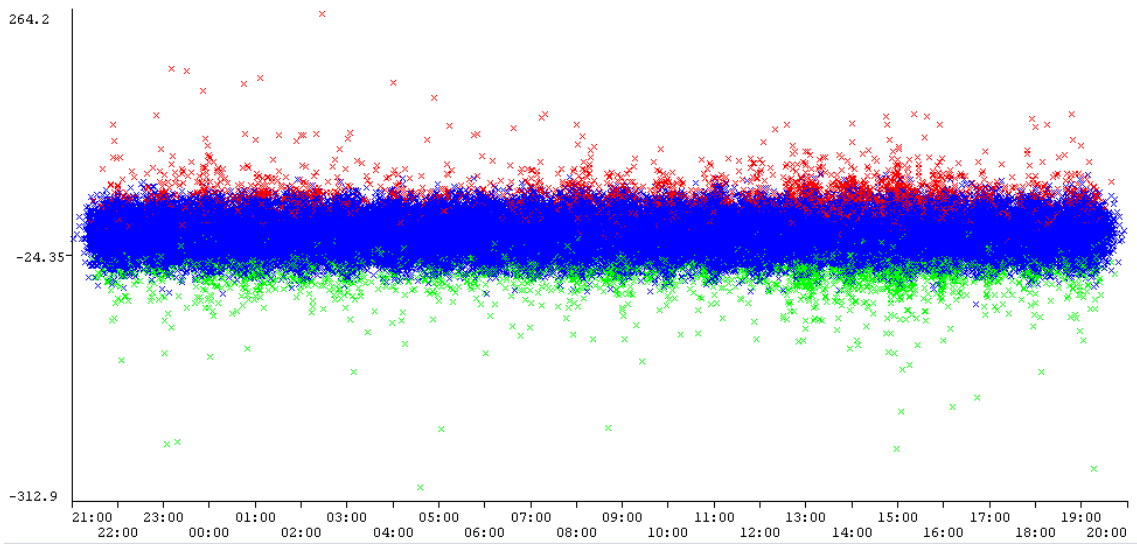


Figure 13. Price Movement (in pip) Plotted to Time (USD/JPY)

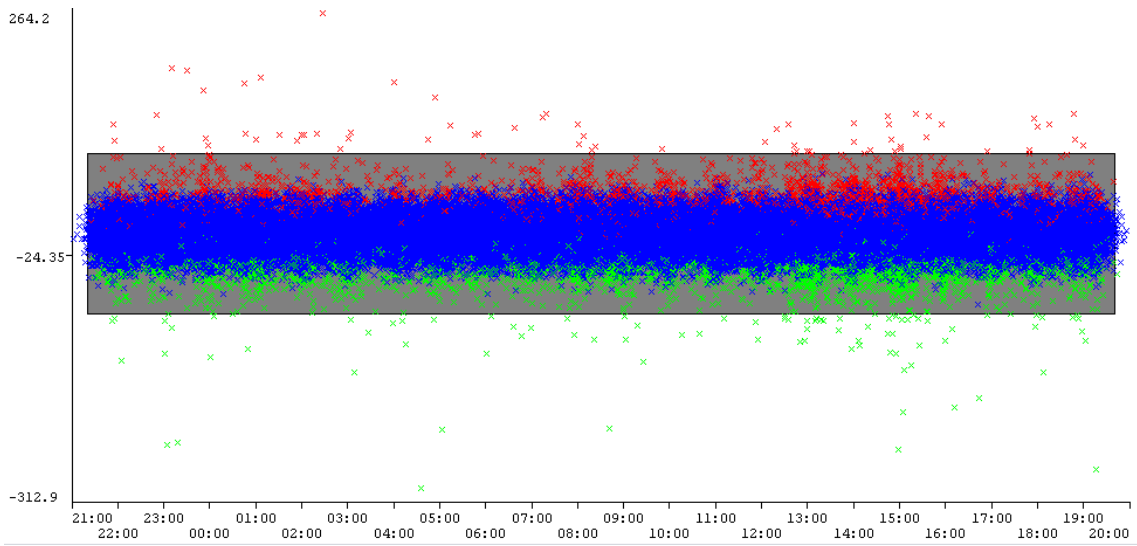


Figure 14. Choose only The Gray Area of this Data

Figure 13 shows the pip change plotted to the time of a day. From this plot, it can be seen that most of the trending market happened at about 1 p.m. to 4 p.m. If we ignore the outliers of Figure 13 (see Figure 14), we get the chart that is shown in Figure 15. The red dots show the upward trend above 25 pips, while the green dots show the downward trend.

Figure 16 shows the uptrend fluctuation range and Figure 17 shows the downtrend fluctuation range (both in pip).

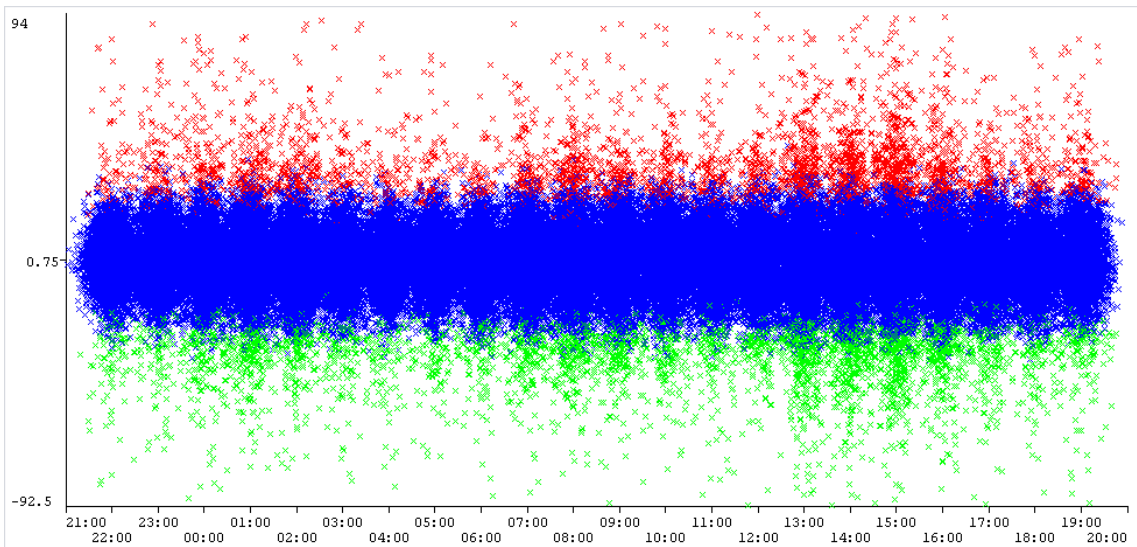


Figure 15. The Data without The Outliers

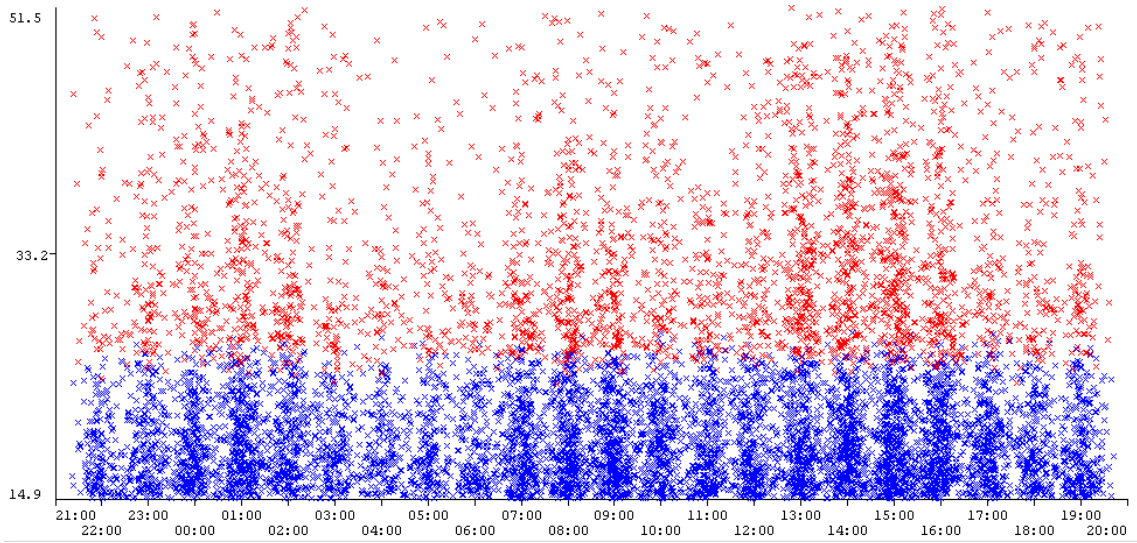


Figure 16. Range of the UpTrend of USD/JPY Pair (in Pip)

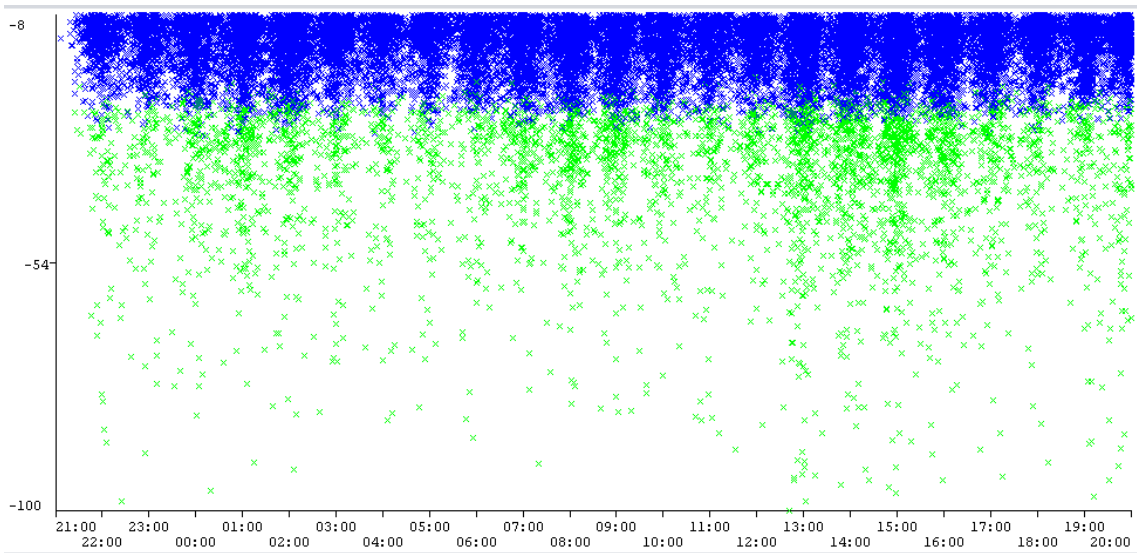


Figure 17. Range of the DownTrend of USD/JPY Pair (in Pip)

4. CONCLUSIONS

From the experiments, it can be concluded that most of the time, the forex market is ranging below 10 pips. This can be used to determine how a trading algorithm works. A forex trading robot that can deal with ranging markets is preferable than the one which only waits for the trending market. Most of the market trends happened during office hours (7 a.m. to 5 p.m.) for EUR/USD, but almost all the time for USD/JPY. The possibilities of winning between buy and sell actions are comparable for the three major currency pairs.

ACKNOWLEDGEMENTS

The authors would like to thank LPPM Parahyangan Catholic University for the research grant and the Department of Informatics Parahyangan Catholic University which supports the research.

REFERENCES

- [1] L. Abednego, C. E. Nugraheni (2015). Development of Forex Robot in MetaTrader 4. Prosiding International Congress on Engineering and Information.
- [2] L. Abednego, C. E. Nugraheni, I. Rinaldy (2018). Forex Trading Robot with Technical and Fundamental Analysis. Journal of Computers JCP 2018 Vol.13(9): 1089-1097 ISSN: 1796-203X. doi: 10.17706/jcp.13.9.1089-1097
- [3] L. Abednego, C. E. Nugraheni (2018). Development of Forex Trading Robot with Money Management. Proceeding of Higher Education, Sydney, Australia.
- [4] D. F. Jimenez (2020). Forex currencies M1,M5,M15,M30,H1,H4,D1. <https://www.kaggle.com/lehomme/forex-currencies-m1m5m15m30h1h4d1/notebooks>
- [5] I. Witten, E. Frank, M. Hall, C. J. Pal (2016). Data Mining Practical Machine Learning Tools and Techniques. Morgan Kaufmann. Fourth Edition.
- [6] J. Norris, T. Bell, A. Gaskill (2010). Mastering the Currency Market: Forex Strategies for High- and Low-Volatility Markets. McGraw-Hill.

AUTHORS

Luciana Abednego, M. T. received her bachelor degree from Dept. of Informatics, Parahyangan Catholic University, Bandung, Indonesia. She has done her Master in Informatics from Bandung Institut of Technology, Bandung, Indonesia. Currently, she is working as a lecturer at the Dept. of Informatics, Parahyangan Catholic University. Her research interest includes machine learning and intelligent systems.



Dr. Cecilia E. Nugraheni received her bachelor degree (1993) and master degree (1995) from Dept. of Informatic Engineering, Bandung Institute of Technology (ITB), Bandung, Indonesia. She has received PhD Degree (2004) from Dept. of Informatics, Ludwig Maximilians Universität, Munich, Germany. Her research interest includes formal methods, intelligent systems, machine learning, meta-heuristic and hyper-heuristic techniques.

