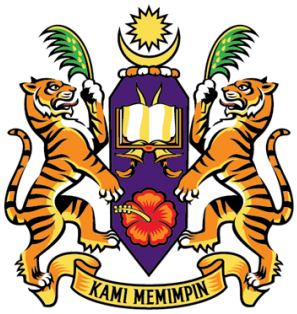

UNIVERSITI SAINS MALAYSIA

SCHOOL OF COMPUTER SCIENCES

SEMESTER 1: ACADEMIC SESSION 2020/2021

CMT423 DECISION SUPPORT SYSTEMS & BUSINESS INTELLIGENCE

ASSIGNMENT 1

Name: Firdaus Bin Mohd Nazri (137067), Muhammad Adam Fikri Bin Anuar (137109), Gosvindraj A/L Munusamy (137072)

Title: Consumers' Sentiment Analysis on PayPal and Cryptocurrency Integration News

Lecturer: Dr. Noor Farizah Binti Ibrahim

Submission Date: 04/06/2021

Table of Contents

1. Introduction

The dataset that was chosen for this assignment was mined from one of the most popular social media application -Twitter using RapidMiner software. The focus of this assignment lingers around PayPal. PayPal is an American company which offers online payment solutions to many countries that supports online money transfers while serving as an electronic alternative to traditional finance methods [1]. Bank account, debit card or credit card can be added to one's online account which then can be used to make transactions both locally and globally.

Recently with the boom in mainstream adoption of cryptocurrency, PayPal is one of the few top firms that have integrated cryptocurrency transactions consequently enabling their 360 million active users to hold or spend the money. Although one could hold or sell cryptos, PayPal users were not able to withdraw them directly. Users had to trade it first for money (dollars or other FIAT currencies) to be able to withdraw their assets or make purchases.

However, in the past few days, PayPal has announced that they have plans to add the crypto withdrawal feature [2]. This news took the trending spot on Twitter. As part of the data analyst team for PayPal, we have decided to extract data from Twitter that includes the keyword "PayPal Crypto". The objective of this assignment is to analyse multiple key information especially in terms of user sentiments caused by this crypto withdrawal announcement.

The motivation behind this is to firstly, preserve a good reputation of the company. As we all know, social media platforms are where people have the freedom to voice out their opinion. News and rumours get viral within minutes and can reach millions of users globally. Many relevant negative tweets regarding a company could be detrimental in both the short-term and long-term. For PayPal, a company that is fundamentally relying on its consumers to run the business, maintaining a good reputation is vital. Mining data from users on Twitter during the trend of news on future plans, for instance, introduction of crypto withdrawal feature in PayPal, will allow us to detect the negative sentiment tweets and try to reduce them whilst increasing positive sentiments by providing better solutions where possible before it is too late.

The second motivation is to maintain a quality customer relationship which closely relates with the first. When a company starts to listen to their consumers and acts accordingly, the customer relationship could be greatly improved. Mining data from tweets could help find out consumers' issues, problems, suggestions, criticisms and many more which then can be utilized to be implemented in business decisions of the company. This will in turn improve the customer relationship allowing a win-win situation for the company to maximize profitability. In our case, many users would definitely have reacted and given their two cents for the recent news on plans of crypto

withdrawal feature. Mining these user tweets will allow us to gather a lot of data that can be translated and used to influence (if justifiable) the planning decisions before going live with the feature.

2. Background and Related Works

The background and logic behind our reasoning is not without a backbone of credible research. Instead, the notion that customer satisfaction is vital in the growth of a business has been tested as true in the real-world situation [3]. In conjunction with big data, real time opinions of customers are important in ensuring that the right decisions could be made by management of a business. The opinions of customers or in other words, the customers' demands are satisfied by business through collaborative planning, forecasting and replenishment (CPFR) which could all be achieved more effectively with data mining and business intelligence [3].

An interesting concept that should also be adhered to by businesses of any size is by categorizing the different types of products with their level of importance. Generally speaking, inventory should be classified into four different categories which are vital and expensive, desirable and expensive, vital and inexpensive and desirable and cheap. Table below demonstrates the significance of categorizing inventory with this approach [3].

Inventory Type	Inventory Supply at a Given Time
Vital and Expensive	Inventory supply must be reliable. Quality of delivery must be smooth.
Desirable and Expensive	Inventory can be kept at minimum. Continuously review inventory policy.
Vital and Inexpensive	Supply must be at maximum at all times. Weekly review on inventory policy.
Desirable and Cheap	Inventory can be kept at maximum but purchased least frequently out of all the products.

Table 1: Inventory Type and their Supply at a Given Time in a Business.

The reason that the classification above is important for a business to identify is because by being aware of their inventory type, business owners are able to make informed decisions in terms of their operating costs. This in turn would translate into a higher profit margin for the business since operating costs could be kept at a minimum [3]. Thus, it is important for inventories of a business to be analyzed in detail which could be accomplished by incorporating business intelligence in the management level of a company.

In addition to inventory keeping which are mostly applicable to businesses that rely on sales as their main source of income, healthcare organizations can also improve with the help of business intelligence methodologies. A study was conducted on healthcare employees and it was discovered that users of business intelligence systems are more likely to exhibit better performance compared to those that are not exposed to business intelligence systems [4].

This finding is definitely crucial and paints a clearer picture of the importance of business intelligence systems, even in the healthcare industry. This finding also tells us that with the right implementation of business intelligence methodologies in any organization, productivity of employees could also be improved which indirectly would translate to better performance and higher customer engagement. For instance, in the healthcare industry, it is shown that organizations that equip their employees with a business intelligence system exhibit better performance as compared to employees in an organization that does not incorporate a business intelligence system [4].

One factor of this increase in productivity is due to the real-time data that business intelligence systems provide to the users of the system. If an employee is able to see that the work being done by them is affecting the company in a positive way, they would be more motivated in maintaining their good performance. This is closely related to the reward system (the mesocorticolimbic circuit) that is present in our brain which is responsible for making us feel good when performing a task.

In addition to the points mentioned above, a separate study was also conducted on 2 industry partners, PricewaterhouseCoopers (PWC) and the Canadian Advanced Technology Association (CATA). The findings from the surveys of the study also indicates that there is a strong relationship between Business Intelligence system implementation effectiveness and the effectiveness of Business Analytics of these 2 organizations [4]. The findings of this study indicates that the concept of media richness needs to be revisited by re-evaluating the advanced analytics capabilities in modern organization.

This is due to the large amount of raw data that is present in modern organization that could be falsely interpreted in the hands of human intervention [4]. Instead, what needs to be done is that data mining and artificial intelligence algorithms need to be normalized in modern organizations in order to reduce the ambiguity of the raw data present in a company’s data warehouse.

3. Data Exploration

The data was mined from Twitter with both recency and popularity as criteria using RapidMiner software. RapidMiner extracted a total of 9550 tweets from Twitter with the keyword “paypal crypto”. In order to remove duplicate tweets, unnecessary tokens, hyperlinks and irrelevant columns, and replacing missing values, data pre-processing was done to clean the data using RapidMiner.

Upon cleaning the data, we ended up with a modest but extremely vital 1288 datasets. In other words, only 13.5% of the initial dataset were unique. This goes to show how much spam and redundancy exists on the internet and if it is not pre-processed well, data analysis results could be significantly different than the reality.

Initially, RapidMiner had 12 distinct columns for each dataset. We reduced the number of columns to just 4 for the purpose of this analysis. The columns that were maintained are date it was created at, username of the person who tweeted, the retweet counts and the text of the tweet. These 4 columns are the most important to fulfill our objectives.

The first column, date created, is vital as it shows the recency and relevancy of the tweet against the time when the news broke out. The second, username, could be useful to determine and differentiate whether the tweet was made by a company/firm, individual/retail or bots. Third column, retweet counts, shows the exposure level and virality of a tweet and the impact it could carry. Lastly, the text content column is important to analyse sentiments behind the tweet.

4. Descriptive Analytics

The information collected from the tweets fetched about PayPal is presented in the form of a dashboard. For a clearer view, the dashboard can be seen here.

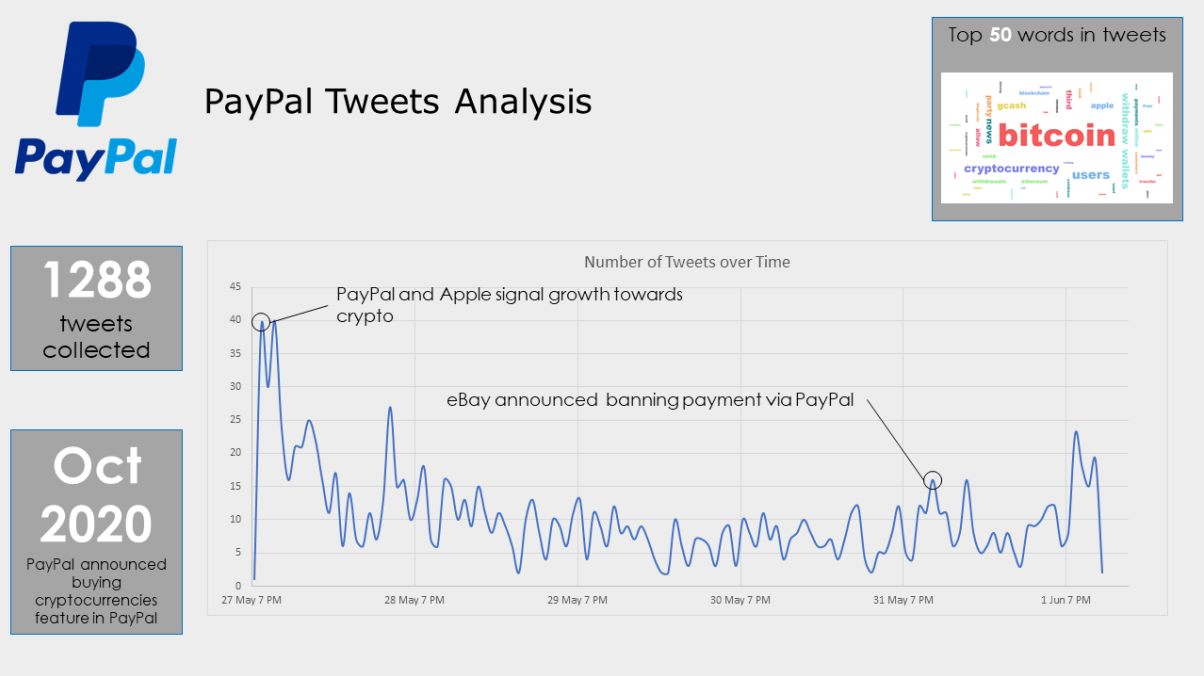

Figure 1: PayPal tweets descriptive analysis dashboard.

From the dashboard, a lot of insight can be gained. Firstly, the number of tweets collected is 1288 tweets. Next, the time when PayPal announced a feature that allows users to buy cryptocurrencies is in October last year. There is also a word cloud that represents the top 50 frequent words in the tweets fetched. The word cloud could give a big picture of people tweeted about. The most people tweeted about PayPal is cryptocurrencies, wallet, third-party, Apple and withdrawal.

There is also the time series that shows the number of tweets posted over time. The interval between each point is 1 hour. The tweets are from 27th May 2021 at 8PM until 2nd June 2021 at 1AM. From here, the distribution and trend of tweets posted can be observed. The highest number of tweets posted is on 27th May at 11PM with 40 tweets and the least tweet posted is on 27th May at 8PM with 1 tweet. The downward trend can be seen from 27th May to 30th May and the upward trend can be seen from 31st May until 2nd June.

There are 2 events that are pointed out in the time series. Firstly, Apple and PayPal are growing toward crypto. This event makes the most tweets posted in the time series for 2 consecutive days. After 28th June the trend started decreasing and went flat. The next event is on 1st June which is when eBay announced banning PayPal as payment options. The increase of tweets can be observed from this day.

The second event needs to be observed as the banning of PayPal payment on eBay could be future challenges/problems. The future risks need to be determined and calculated. For now, the future problems could increase on negative sentiment/perception towards PayPal and another platform could be a competitor for PayPal such as ApplePay (since Apple also signals growth towards crypto). A more detailed analysis is needed to understand the impact of the banning of PayPal payment on eBay.

5. Data Analysis

For this analysis, it is also presented in a dashboard but more detailed. From the dashboard, the sentiments and topics covered in tweets are presented. For a clearer view, the dashboard can be seen here.

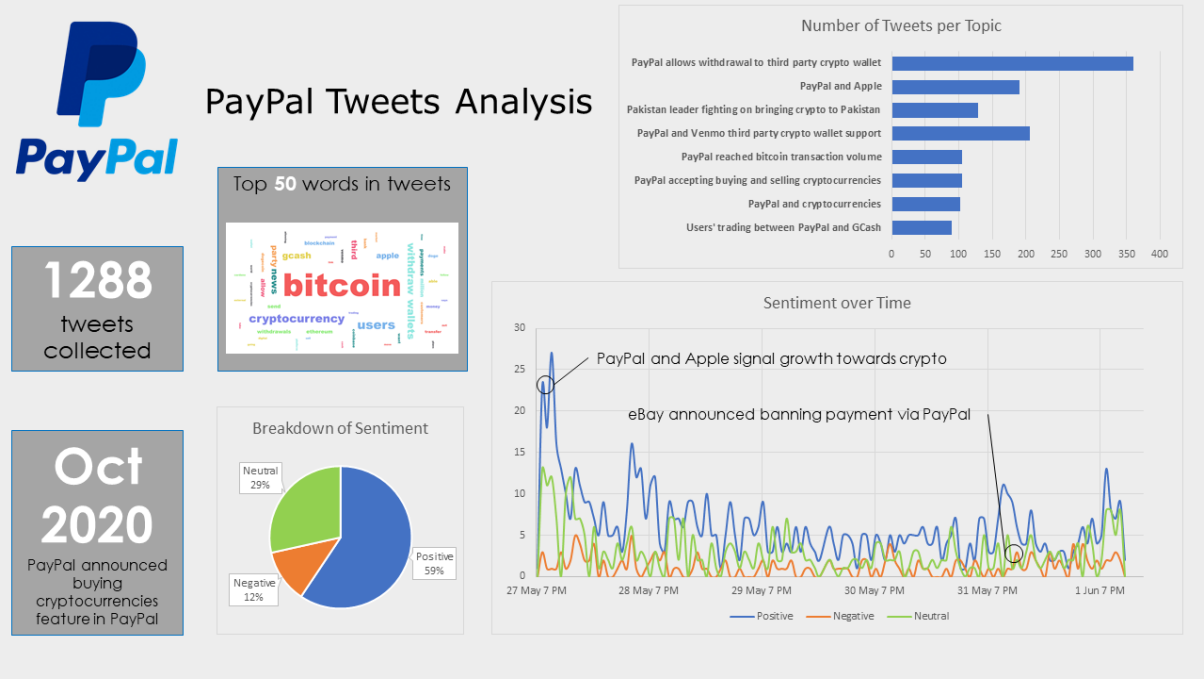

Figure 2: PayPal tweets data analysis dashboard.

Compared to the last dashboard, this dashboard is more detailed and has more charts. The first chart, breakdown of sentiments gives the insight of volume of tweets based on its sentiment. The sentiments are gained by using the Vader sentiment analysis algorithm. From the algorithm, each tweet is labeled with a score. Then the score is labeled with positive sentiment where score more than 0, negative sentiment where score less than 0 and neutral where the score is 0. From the sentiment breakdown, it can be seen that the most of the tweets are positive with 59%, followed by neutral sentiment with 29% and lastly with negative sentiment with 12%.

Next, the time series is now differentiated by the sentiments.The interval between each point is still 1 hour. The tweets are from 27th May 2021 at 8PM until 2nd June 2021 at 1AM. Now, the users' perception over time about PayPal can be observed. From the first event which is Apple and PayPal growth towards crypto brought a lot of positive sentiment from the users.

From the second event on eBay banning PayPal payment method just brought a little increase in negative sentiment. PayPal should not worry too much about this because the positive sentiment is still higher. Still the little increase of negative sentiment should be observed as negative sentiment could rise in the future.

There are 8 topics that can be observed in the number of tweets per topic chart. The topics are gained by using the topic modeling algorithm Latent Dirichlet Allocation (LDA). Topic modeling will give topics of what people are talking about in Twitter. The results from LDA are 8 word clouds with 15 words each and tweets labeled with topics. From the word cloud, we need to use our interpretation to identify the topic. For example, if the word cloud consists of words such as paypal, crypto, wallets, withdraw and third-party, we can interpret that as PayPal allows withdrawal to third-party crypto wallets. Reading news also will help you in interpreting.

Figure 3: Word cloud of topic PayPal allows withdrawal to third-party crypto wallet.

The most popular topic users tweeted is about the feature of PayPal that allows withdrawal to third-party crypto wallets. This is such a good feature and probably gives a positive sentiment to users since this feature gives consumers accessibility to withdraw their cryptocurrencies. This topic is also related to PayPal and Venmo third-party crypto wallet support. This shows that people like features of withdrawal to third-party crypto wallets. The third most tweeted topic is PayPal and Apple. This must be related to the first event. Since PayPal shows interest in crypto and people liked the feature withdrawal to third-party crypto wallet, **PayPal should hire more crypto experts** to extend the crypto wallet features. PayPal also needs to **maintain and upkeep the third-party crypto wallet feature regularly**.

One of the interesting topics and a lot of people tweeted about is the Pakistan leader fighting on bringing crypto to Pakistan. This could be an opportunity for PayPal to introduce PayPal itself and cryptocurrencies to Pakistan by **supporting acts of the Pakistan political leader** and try to **do marketing campaigns in Pakistan**.

6. Predictive Analytics

From the analysis performed on the data collected, we were able to see the sentiment of the general public, as far as Twitter is concerned. We were also able to pinpoint the exact time as to when the sentiments of the public increase in our favor (positive sentiment) as well as when the public is dissatisfied with our company's new implementation or policy (negative sentiment).

For example, in the case of our company announcing the new withdrawal feature, we were able to see from the data analyzed in Rapidminer that users are generally happy about this feature. Generally speaking, when a company adds a new feature or service which is accepted and celebrated by existing customers, this would attract new customers since they would be compelled as well by the good news.

Thus, in the future, the method of acquiring sentiments through mining Twitter data could be adopted by our company which in turn would help the Managerial Level of our organisation to make better decisions in terms of launching new services or products [5]. Extending to the benefits of sentiment analysis, our company could also use this strategy to test the market before deploying a new product to the market.

This means that whenever the board of directors agree to the launch of a new product, we could use social media, specifically Twitter to announce the launch of the new product. We would not have to actually launch the product in conjunction with the public announcement, instead use the approach of mining public's data to get an overview of the user's sentiment. If the public exhibits positive sentiment analysis, we would be able to launch the product with confidence. Instead, if the general public is dissatisfied with our new product announcement, we could retract them and perform a deeper analysis as to why the user's are not happy with our new idea.

Decision makers could make better plans for the company since they are equipped with data that are close to real-time. Real-time data is one of the most important fundamentals of business intelligence since there is a distinguishable difference between real-time data and past data [6]. Past data (data of users collected in the past dating up to several years) is no longer relevant in the current world that we live in. Business owners that are able to acquire real-time data would have an advantage over business owners who do not possess them.

7. Conclusion

Business Intelligence (BI) is a vital part of business for any company or firm that wants to be successful and compete at their highest level. Gathering data from users directly and transforming them into useful information without much effort is a beautiful feature that is provided by BI. Social media platforms such as Twitter acts as an indirect database filled with millions of data that can be transformed into statistics that influences business decisions.

As we all know, cryptocurrency is a fairly new “pseudo” class of asset built on blockchain technology that is heavily debated over its legitimacy. It is a dilemma for big firms currently to decide on whether to adopt cryptocurrencies or not. If one decides to be an early adopter of crypto, then they need to risk the possible uncertainty of it being legit and other consequences that may follow. On the other hand, if one decides to wait it out for the crypto space to be established, then they will have to risk their business performance in terms of competitiveness as they will be left miles behind by their competitors who adopted it early.

For finance related companies such as PayPal, who are trying their best to stay ahead of the herd by introducing new features such as cryptocurrency integration before their competitors, having precise,accurate and useful information is extremely crucial to avoid or decrease the potential risks as aforementioned. For cases as such, data of the public sentiment is pivotal as sentiments of users can act as an useful parameter for risk management and analysis.

Solutions for multiple business problems can be constructed with the aid of data analysis. The two main problems that can be solved or reduced for PayPal with sentiment analysis of its users are preserving a good reputation and maintaining a great customer relationship. Getting to know the consumers’ sentiments regarding plans or proposals of business decisions through news and rumours will significantly portray the overall reaction of the business decision. The crypto integration (withdrawal) feature plan resulted in massive reactions and with the help of BI, we have managed to extract intrinsic information based on thousands of users’ sentiments. With this information, we can influence and modify the business decision and consequently try to solve the problems stated above.

Models that were derived from the unsupervised learning method, topic modelling, can be used to determine potential following tweets in the future to figure out its classification based on topics. This will be extremely helpful especially if the topic lingers around cryptocurrency which we can confidently say it will as we are still in the early days of blockchain technology adoption. With data analysis, predictions in terms of possible user sentiments on similar future news and rumours of plans, especially relating to cryptocurrency can be constructed as well. Other data analysis techniques

such as decision trees are not feasible as the nature of this topic (its business problems) are mainly on sentiment analysis hence; it was not included within this report.

In a nutshell, having great insights about users' opinions, thoughts, criticisms and comments are key in making business decisions. BI, data analysis and data analytic tools, helps to not only make business decisions but also to improve the outcome significantly. We can say that PayPal will be able to put the gathered information to good use not only before going live with the crypto withdrawal feature that will be introduced soon but also future business decisions.

8. References

[1] "PayPal - Wikipedia", *En.wikipedia.org*. [Online]. Available: https://en.wikipedia.org/wiki/PayPal. [Accessed: 01- Jun- 2021].

[2]"PayPal Plans To Add Crypto Withdraw Feature", *Pymnts.com*, 2021. [Online]. Available: https://www.pymnts.com/cryptocurrency/2021/paypal-plans-crypto-withdraw-feature/. [Accessed: 01- Jun- 2021].

[3] S. Malik and R. Jeswani, "Literature Review and Techniques of Machine Learning Algorithm Used in Business Intelligence for Inventory Management", *International Journal of Engineering Sciences & Research Technology*, vol. 7, no. 1, pp. 230-241, 2018. Available: 10.5281/zenodo.1135987 [Accessed 3 June 2021].

[4] R. Gaardboe and P. Svejvig, "Better and more Efficient Treatment: The Individual and Organizational Impacts of Business Intelligence Use in Health Care Organizations", *Selected Paper of the IRIS*, vol. 7, no. 9, 2019. Available: https://aisel.aisnet.org/iris2018/7. [Accessed 3 June 2021].

[5] G. Richards, W. Yeoh, A. Chong and A. Popovič, "Business Intelligence Effectiveness and Corporate Performance Management: An Empirical Analysis", *Journal of Computer Information Systems*, vol. 59, no. 2, pp. 188-196, 2017. Available: 10.1080/08874417.2017.1334244.

[6] M. Llave, "Business Intelligence and Analytics in Small and Medium-sized Enterprises: A Systematic Literature Review", *Procedia Computer Science*, vol. 121, pp. 194-205, 2017. Available: 10.1016/j.procs.2017.11.027 [Accessed 3 June 2021].