International Journal of Advanced Trends in Computer Science and Engineering

The Visualization of Sentiment Analysis for Malaysia Financial News Headlines

Shuhaida Mohamed Shuhidan[1,2]*, Nur Aini Shuhada Fadzil[2], Saidatul Rahah Hamidi[2]

[1]Accounting Research Institute
[2]Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.

* shuhaida@tmsk.uitm.edu.my

ABSTRACT

Visualization goal is to provide the background of the story, and also to provide a powerful medium to discover the story. Thus, it can be simplifying that visualization techniques are not just a tool for generating images, but it is also have strongly connected to the information that have been visualized. With the development of the algorithm trading system, human decision-making has reached its limits and unable to follow the processing and execution of orders and decisions. Analytical methods apply to different categories of groups that are expected to hold information about polarity of sentiment. The financial headline is often sharing the price movements, which show what is trendy in the current economy. Negative headlines could discourage people from spending on share price and it can hamper some firm's capability to raise finance on the stock market. The research objectives for this research are to analysed the different techniques of Malaysia financial news headlines sentiment and to demonstrate the visualization of sentiment analysis for Malaysia financial news headlines dashboard. This research are can be used for stakeholders who want to know about the financial news and seek knowledge or data in financial world. Besides, it also can classify the sentiment in Malaysia financial news headlines into the polarity intelligently. The results show the data had been analyzed and it is strongly proved that the 'sentiment' technique is more accurate based on the calculated sentiment score. As for the recommendation, it is hoped that this tested technique can be applied for future data analysis.

Key words : IoT, Sentiment analysis, Visulization.

1. INTRODUCTION

As for now, in Malaysia, KLSE (Kuala Lumpur Stock Exchange) or formerly known as Bursa Malaysia Berhad, is the medium or tool that focus on various initiatives aimed at improving the supply of its products and services, increasing its liquidity and market velocities, enhancing its business efficiency and achieving economies of scale in its operations. On 18 March 2005, Bursa Malaysia was listed on the Main Board of Bursa Malaysia Securities Berhad with a premium of 17% or RM0.50 as compared to the retail price of RM3.00. Additionally, the prediction of financial mar-kets based on online sentiment detection has attracted a lot of attention recently. However, most results in this pop-up domain depend on a combination of unique data and unique sentiment detectors. This makes it difficult to divide the effects of measurements and instruments from factors that are actually involved in the obvious relationship between online sentiments and market [1]. Studies have shown that business and financial news has a strong correlation with future stock performance[2]. People only use basic analysis and technical analysis to predict their stock prices, but most of them are unaware of the dependence of sentiment analysis on market trends. More precisely, fluctuations in the stock market are analyzed and forecasted to gain knowledge that can be used for interested parties interested in financial news and financial data search [3].

According to [4], market participants have recently raised interest in the field of sentiment analysis. With the development of the algorithm trading system, human decision making has reached its limits and cannot follow the processing and execution of orders and decisions. In addition, the amount of information that cannot be handled without support. To manage all information, analytical methods apply to different categories of groups that are expected to hold information about polarity of sentiment. Forums, blogs and wikis are seen as the most representative type of text data in which users can express their opinions. Also news, research reports and financial related content produced by firms are analysed. The economy of Malaysia is the third largest in Southeast Asia after the populated Indonesia and Thailand. It is also ranked 35th largest in the world based on the gross domestic product (GDP). However, in 2015 unrestrained deviations have been seen across the whole range of the Malaysian politic and economy. The Malaysia's economy

seems to be quite rocky with the rising inflation, continued high levels of capital flight, contracting growth, and declining consumer and investor confidence. Society like to think and talk about the economic problems that affect them as consumers, workers, investors, citizens and else. By this, it is important for people to always know and aware about their country economy state and what is going on in the economy world.

2. THE VISUALIZATION

Data growth rates have increased rapidly over time due to several factors such as Internet Articles (IoTs), sensors in our environment, and copying all offline records such as our medical history and Big Data has proven its importance to this world in very little time now almost everything IT companies and non IT stores all the data they produce. Today's business struggles to save a lot of money data while analyzing, interpreting and presenting it. The Big Data main challenge is located in capturing, storing, analyzing, sharing, searching, and visualizing data.

Data visualization is the technique of integrating data into one collective, representative graphic. Traditionally, data visualization has been used for quantitative work such as info graphics, but ways to represent qualitative work have shown to be equally as powerful. There are a lot of benefits from data visualization and one of them is improvement in decision making which are helps the organization to view their position and the process carried by the organization. According to the visualization and analyses of data the company can take better decision and they can also change their business flow according to it [5]. Apart from that, information sharing is also important so that the organizations can gain knowledge about their previous and present business flow.

According to [6], the techniques of data visualization assist in spreading the journalism information of massive amount of data. Text and statistical data can now be turned into diagram and charts to convey a message in terms of simplified version. In addition, visualization goal is to provide the background of the story, and also to provide a powerful medium to discover the story. Thus, it can be simplifying that visualization techniques are not just a tool for generating images, but it is also have strongly connected to the information that have been visualized.

Figure 1 and 2 exhibits the example of visualization techniques. The most prevalent used technique is based on the static method of informational representation, but also an interactive tool available which allows the user to modify the graph by changing some variables
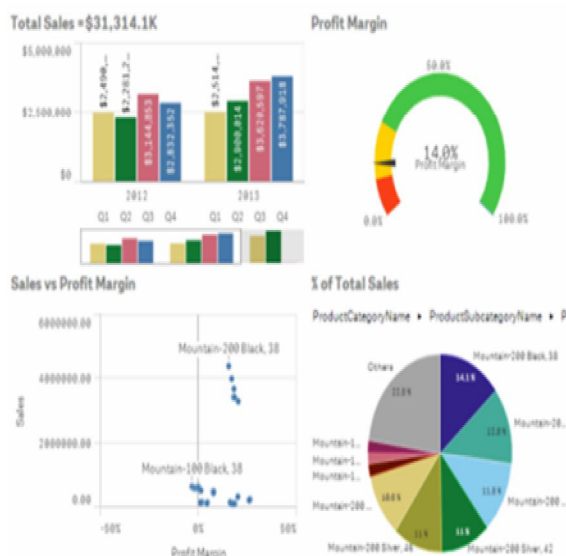

Figure 1: Graphs of visualizations techniques

Figure 2: Example of visualizations techniques

3. SENTIMENT ANALYSIS

People share their knowledge, experiences and thoughts with the world by using social media through blogs, forums, wikis, studies websites, social networks, twitter and others. It has changed the way people communicate and also influence the social, political and economic behaviour of others on Web 2.0.

Web 2.0 lets everyone to have a voice, including improving human collaboration capabilities at a scale throughout the world and allowing individuals to share their views by reading write and user generated contents. An opinion is positive or negative sentiment, views, attitudes, emotions, or judgments about entities or aspects of entities of the opinion-holder [7] at the specified time. It can be a product or service, event, person, organization, or topic that consists of aspects that represent both the components and attributes of the entity [8].

3.1 Classification of Sentiment Analysis Approaches

In the survey on the different methods of analysis of sentiments found in the literature relating to product reviews (such as machine learning, semantic orientation, opinion polls, and holistic lexicon-based approaches) were carried out. This survey outlines analysis of match overlays and considerations playing an important role in making decisions about products or services.

Few approaches used for sentiment analysis are:

- Lexicon approach: a list of words for wizards scored which indicates their nature in positive, negative or objective terms;
- n-gram modelling approach: application of uni-gram, bi-gram, tri-gram or combination for the classification of sentiments;
- Machine learning approach: implementation of semi or supervised learning through the extraction of features from texts and learning models.

Table 1 indicated the summarized the detail of sentiment analysis approach based on previous research, including existing tools developed to analyze sentiment.

Table 1: Summarization of Sentiment Analysis

Definition	● positive or negative sentiment, including views, attitudes, emotions, or judgments about entities or aspects of entities of the opinion holder at a certain time. ● can be a product / service, event, person, organization, or topic that consists of aspects (Features / properties) that represent both the components and attributes of the entity.
Sentiment Analysis Steps	● Data collection ● Text preparation ● Sentiment detection ● Sentiment classification ● Presentation of Output
Sentiment Classification Approaches	● nGram ● Lexicon Based ● Machine Learning
Existing Tools for Sentiment Analysis	● EMOTICONS ● LIWC ● SentiStrengh ● Senti WordNet ● SenticNet ● Happiness Index ● AFINN

3.2 Existing Sentiment Analysis Tools for Financial Market

Financial news circulates daily on the Web and financial market are continuously changing and growing. This section discusses the existing sentiment analysis tools specifically relates to financial market.
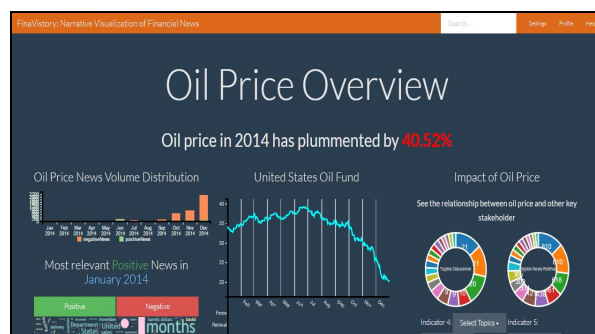

Figure: 3: Price Overview

Figure 3 represented interface of Finavistory (http://finavistory.herokuapp.com/). Finavistory represent the positive and negative sentiment focusing on its reflection to oil price review in the United States [9].

Other than that, StockFluence (https://www.stockfluence.com) is also an existing tools that provides financial sentiment analysis for investors to discover, react and respond to market opinions. It can monitor social media channels and analyse the overall sentiment with algorithms. Based on the sentiment, the system make predictions with accuracy level of 70%. Figure 4 below show the screenshot of StockFluence.
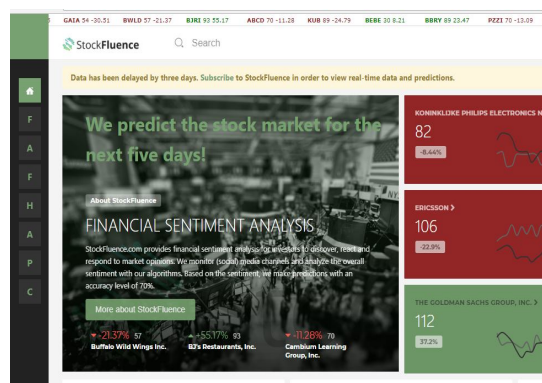

Figure 4: StockFluence.

As shown in Table 2 below, the comparison of the sentiment analysis tools discussed.

Table 2. The existing sentiment analysis tools

Tools	Techniques	Approaches
FinaVistory	• Word Cloud • Network Visualization • Summary Plot • Bubble Chart	• Wordle and Tag Cloud are popular tools visualizing the significance of certain keywords by using different font sizes, colors etc • Force-directed layout can be used to visualize modularity clusters
StockFluence	• Network Visualization	• Visualize the analyses and predictions in a user-friendly and personal dashboard.

4. METHDOLGY

4.1 Research Model

Research model is a sample for the research project that indicated the variables involved and analogy on how the variables relates to each other. Apart from this re-search study, Big Data Processes has been used to achieve the second objective. The overall process of extracting insights from big data can be broken down into five stages as per shown in Figure 5 below.
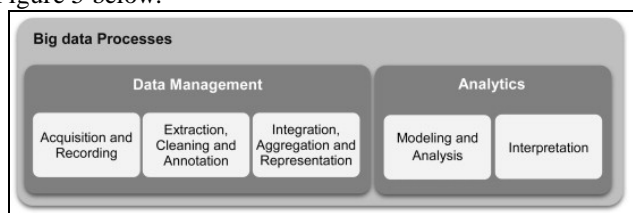
Figure 5: Big Data Processes

4.1.1. Data Management

The data collected were retrieved from online newspaper: New Straits Times (NST) that address only financial news headlines. The tools used were Web Scraper. The next step is to identify and eliminate non-textual content from collected data. Data pre-processing steps including stop words removal, stemming. Dictionary based approach been choosing for text mining. In order to build the polarity dictionary, two types of words collection needed; i.e. positive words and negative words. Then the article's is match words against both these words list and count numbers of words appears in both the dictionaries and calculate the score of that document.

4.1.2 Analytics

- Created the polarity words dictionary using general words with positive and negative polarity.

- Conducted the appropriate requirement to design the tool to visualize the Malaysia financial news headlines sentiment dashboard.
- After completion of analysis, the text result will be displayed on graphs through the dashboard.

5. RESULTS AND DISCUSSIONS

5.1 Data Management
The gathered data retrieved is from 2015 to 2017, which contain 4779 total of sample data extracted randomly, from NST specifically that are related to financial headlines.

5.2 Analytics
The data was analyzed using 'sentiment' and 'sentimentr' library packages.

According to the score based on 'sentiment' library package's results finding, it is indicated that most of the primitive packages handling sentiment analysis perform a simple dictionary lookup and calculate a final composite score based on the number of occurrences of positive and negative words. But that often ends up in a lot of false positives, with a very obvious case being 'happy' vs 'not happy'.

Figure 6 and 7 below shows the result of sentiment analysis on financial news headlines after perform the score function. The result of scoring function:
- If the score '2' means it highly positive
- If the score '0' means it's a neutral sentence
- If the score '-1' means it has no impact or negative sentence

	Headlines	Score
166	UEM Edgenta remains confident of prospects this year	1
167	Malaysian household income improving, says Khazanah ...	1
168	CIMB sees 6 to 7 pct loan growth this year	0
169	FGVH registers 34.9 pct increase in net profit for Q2 2016	0
170	UDA settles Tanjung Tokong land issue	-1
171	Ringgit at 4.04 against USD	0
172	Normal to shut, relocate branches: Maybank	0
173	RM likely to strengthen against USD next week	0
174	GTSB push for YFG board removal	0
175	PNB declares 6.30 sen dividend payment for ASW 2020	0
176	Malaysia-GCC to expedite FTA framework	0
177	Axiata revises upwards capex to near RM6b	0
178	Ringgit at 4.01 against USD	0
179	Lotte Group says vice chairman dies amid criminal probe	-3
180	Maybank Group posts slight decline in net profit for firs...	-1

Figure: 6: The Result of Scoring based on 'sentiment' library package technique

5.2 The results using 'sentimentr' technique library package

Figure 8 below the results after implementing different technique which is 'sentimentr' package.

Figure 7: The Result of Scoring based on 'sentimentr' library package technique

After performed this package, the results shown that the express of the sentiment headlines was enhanced which is the score is changed to 0, that indicated neutral statement. As for the second statement, the score is 0.45 which is positive sentence and the third sentence calculated as 0.2 which is positive sentence as compared from the first technique, the score is more accurate.

The 'sentimentr' attempts to balance accuracy and speed. The valence shifters affect the polarized words. In the case of negators and adversative conjunctions the entire sentiment of the clause may be reversed or overruled. Therefore, the valence shifters occur fairly frequently a

simple dictionary lookup may not be modeling the sentiment appropriately.

Table 2: Result comparison between two packages

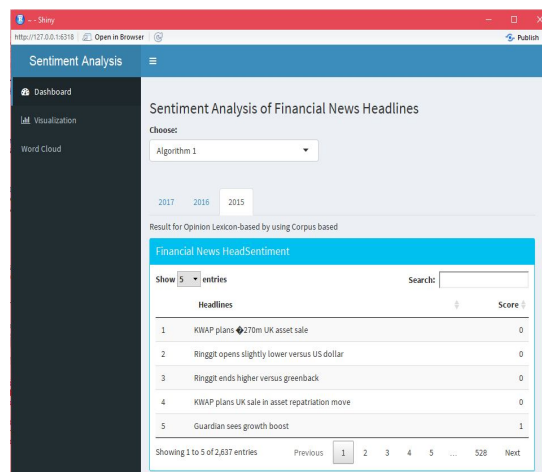| Headlines	sentiment library package	Sentimentr library package
Score UDA settles Tanjung Tokong land issue	-1	0
FGV optimistic of stronger H2	2	0.45
No increase in GST percentage for 2017	0	0.2

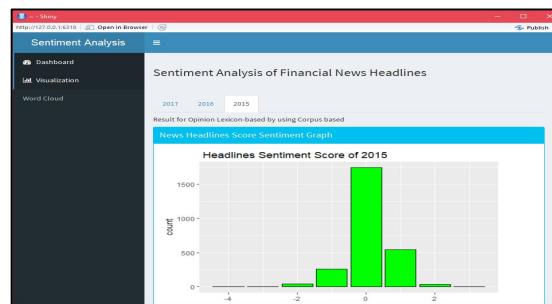**Figure 8**: Main Interface of Dashboard

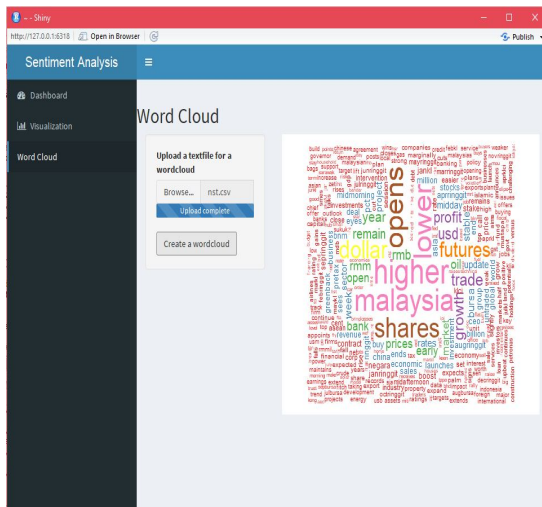**Figure 9**: Interface in Visualization Tab Item

Figure 10: Interface of Word Cloud Tab Item

5.4 Discussion

From the results, the sentiment analysis of financial news has been compared by using different techniques. It is clearly stated that techniques by using lexicon based which is the one that consists of valence shifters are perform better as to evaluate and distinguishing between positive, negative, and neutral news. It is proved that the sentiment in financial news is very useful. Sentiment analysis can be applied to explain the market returns and how the investors can make better decision for their investment.

5. CONCLUSION AND RECOMMENDATION

The news headlines can be used in conjunction with othe techniques as a provider to the top management of the company, shareholder or investor to analyze the impact of their business units from certain regional sentiments and demands. News has a direct impact on the finance and it can easily change a good day into a bad day and vice versa. Market sentiments are important factors and news about company decisions, government policies, political interests, inclusion or exclusion from the index, mergers and acquisitions and so on. From the results, it shows that the technique applied are one of the element that can influence the pattern of the headlines sentiment accuracy which is every news faction can be classified as containing positive or negative sentiments in it. R tools are used as a potential calculation software for data analysis and the visualization are displayed through the dashboard as it came out with the sentiment score table and graphs.

As for the future enhancement of this research project, to get better results, the news headlines for larger periods is needed for an efficient data analysis and pattern from the different kind of online news as an aggregate source for financial news and used by key industrialists and financial experts. On the other hand,
clear and detail. The study might contribute to the new findings and improve for the future research.

REFERENCES

1. Mao, H., Counts, S., & Bollen, J. (2011). Predicting financial markets: Comparing survey, news, twitter and search engine data. arXiv preprint arXiv:1112.1051.
2. Im, T. L., San, P. W., On, C. K., Alfred, R., & Anthony, P. (2013). Analysing market sentiment in financial news using lexical approach. In Open Systems (ICOS), 2013 IEEE Conference on (pp. 145-149)
   Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert Systems with Applications, 42(1), 259-268
3. Uhr, P., Zenkert, J., & Fathi, M. (2014, October). Sentiment analysis in financial markets: A framework to utilize the human ability of word association for analyzing stock market news reports. In Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on (pp. 912-917)
4. Liu, S., Cui, W., Wu, Y., & Liu, M. (2014). A survey on information visualization: recent advances and challenges. The Visual Computer, 30(12), 1373-1393.
5. Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. IEEE transactions on visualization and computer graphics, 16(6), 1139-1148.
6. Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.
7. Zhang, L., & Liu, B. (2014). Aspect and entity extraction for opinion mining. In Data mining and knowledge discovery for big data (pp. 1-40). Springer, Berlin, Heidelberg.
8. Y. Chan and H. Qu, "FinaVistory: Using Narrative Visualization to explain social and Economic relationships in financial news," 2016 International Conference on Big Data and Smart Computing (BigComp), Hong Kong, 2016, pp. 32-39. doi: 10.1109/BIGCOMP.2016.74257