UNDERSTANDING JOB LOSS CONCERNS IN MALAYSIA: A SENTIMENT ANALYSIS APPROACH

Nur Hurriyatul Huda Abdullah Sani[1]

ABSTRAK

Kemajuan pesat teknologi dan automasi yang didorong oleh revolusi industri keempat telah menimbulkan kebimbangan mengenai kehilangan pekerjaan di seluruh dunia, termasuk di Malaysia. Kajian ini dijalankan bertujuan mengkaji sentimen terhadap kehilangan pekerjaan di Malaysia dengan menganalisis artikel berita dalam talian menggunakan teknik analisis sentimen. Lima artikel dipilih berdasarkan kaitannya dengan topik dan dianalisis menggunakan perisian R dengan pakej tidytext. Analisis sentimen menggunakan tiga leksikon: afinn, bing, dan nrc, untuk mengklasifikasikan sentimen sebagai positif, negatif, atau neutral. Penemuan menunjukkan sentimen yang kebanyakannya negatif terhadap kehilangan pekerjaan, dengan kebimbangan khusus mengenai pengangguran di kalangan graduan universiti. Hasil kajian menyoroti isu seperti kekurangan kemahiran insaniah dan masalah sikap di kalangan graduan, mencadangkan keperluan untuk program latihan yang dipertingkatkan dan semakan kurikulum. Kajian ini menekankan kepentingan memahami sentimen awam terhadap keselamatan pekerjaan sebagai input bagi pembuatan keputusan dasar dan menangani cabaran yang ditimbulkan oleh kemajuan teknologi.

Kata kunci: Analisis sentimen, Kehilangan pekerjaan, Lexicon

ABSTRACT

The rapid advancement of technology and automation, driven by the fourth industrial revolution, has raised concerns about job losses globally, including in Malaysia. This study investigates the sentiment towards job loss in Malaysia by analysing online news articles using sentiment analysis techniques. Five articles were selected based on their relevance to the topic and analyzed using R software with the tidytext package. The sentiment analysis employed three lexicons: afinn, bing, and nrc, to classify sentiments as positive, negative, or neutral. The findings indicate a predominantly negative sentiment towards job loss, with specific concerns about unemployment among university graduates. The results highlight issues such as a lack of soft skills and attitude problems among graduates, suggesting the need for enhanced training programs and curriculum revisions. This study underscores the importance of understanding public sentiment on job security to inform policy decisions and address the challenges posed by technological advancements.

Keywords: Sentiment analysis, Job loss, Lexicon

---

[1] Nur Hurriyatul Huda Abdullah Sani is currently Senior Assistant Director of Core Team Big Data Analytics (CTADR), Department of Statistics Malaysia.

1. INTRODUCTION

Data and information are easily accessible online for government service delivery, businesses and even personal needs and self-expression. The magnitude of data generated and shared by various parties such as business organisations, public services, various industries, non-profit organizations and scientific research, has been increasing substantially (Agarwal & Dhar, 2014). Dobre and Xhafa (2014) reported that every day, the world produces around 2.5 quintillion bytes of data (i.e., 2.5 exabytes or 2.5 billion gigabytes), of which 90.0% of the data is unstructured data.

The emergence of technology and automation under the wave of the fourth industrial revolution raises concerns about potential job losses. The rapid advancement of technology can transform entire industries, potentially replacing many existing employees with machines. Malaysia must address these issues to ensure its economic activities remain competitive in the global market.

Various parties provide insights and statements on job security challenges resulting from the fourth industrial revolution in Malaysia. These include Khazanah Research Institute (KRI), the Organization for Economic Co-operation and Development (OECD), official statistics from the Department of Statistics, Malaysia (DOSM), online newspapers and even casual conversations and writings in these issues in social media such as blog, Twitter, Facebook and others. For example, a Google search for "job loss in Malaysia" yields over 1.4 million results.

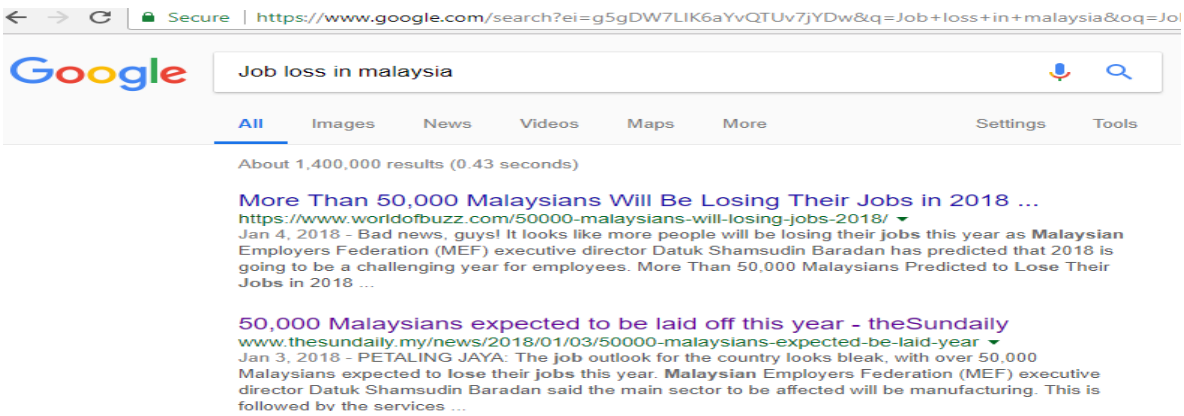

Figure 1.1: Example of Search Results on Google using Keywords Job Loss in Malaysia

Similarly, a Twitter search for “unemployment Malaysia” generates numerous narratives, as shown in Figure 1.2 below.

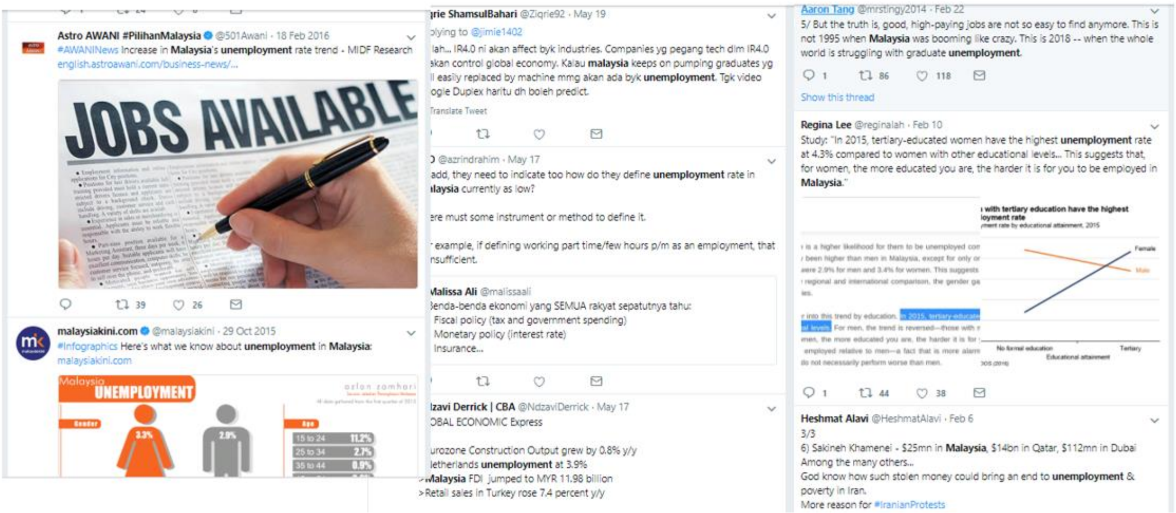

Figure 1.2: An Example Search on Twitter using Keyword “Unemployment Malaysia”

2. SENTIMENT ANALYSIS

Writings on online platforms can easily reach a variety of comments and views from anyone. Some people agree that industrial and technological advancements increase job loss, while others disagree, providing arguments and facts. To have some basic ideas of the public’s feelings and views expressed in this issue more clearly, sentiment analysis was conducted.

According to Bing Liu (2012), sentiment analysis, is also called opinion mining. It is the field of study that analyses people’s opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. It is one of the most active research areas in natural language processing and is also widely studied in data mining, web mining, and text mining. This research has expanded beyond Computer Science into management and social sciences due to its importance to business and society. The growth of sentiment analysis coincides with the rise of social media, including reviews, forum discussions, blogs, micro-blogs, Twitter, and social networks.

Several challenges need to be addressed before performing sentiment analysis, such as the usage of slang, sarcasm and the absence of build-in lexicon in most of our local languages. However, due to promising and useful information that might be generated from this analysis, people are interested in exploring this field further.

By using sentiment analysis of job losses in Malaysia through online news articles, the main objectives of this case study are as follows:

- To identify the sentiment and emotion words used;
- To generate the sentiment score through the words used;
- To classify positive, negative or neutral sentiment towards job loss in Malaysia.

3. METHODOLOGY

3.1 Source of Data

For this study, online news articles were selected using the keywords "job loss in Malaysia". Five articles were selected as follows:

- "Unemployment among graduates needs to be sorted out fast", published in www.thestar.com.my on 25 March 2017;
- "At Industry 4.0 forum, panellists say automation not about losing jobs", published in www.malaymail.com on 20 November 2017;
- "Young and jobless", published in www.thestar.com.my on 27 March 2017;
- "Skilled workers ensure excellence", published in www.nst.com.my on 1 May 2018;
- "Labour-intensive job loss normal in fourth industrial revolution", published in www.theedgemarkets.com on 7 March 2018.

3.2 Text Processing

The sentiment analysis was carried out using R software. The packages used in this analysis are tidyverse, tidytext, stringr, textstem, dan wordcloud. Before further analysis can be carried out, the text from the websites was imported into R and converted into a data frame format. The data was tokenized into one token per row using the unnest_tokens() function and cleaned of stop words (e.g., 'to', 'be', 'a', 'the', etc.) using the stop_word() function in the tidytext package. There were 1,149 stop words listed in stop_word() function within tidytext packages. Numbers were also removed. Subsequently, sentiment analysis was performed.

3.3 Sentiment Analysis

Sentiment analysis can be performed at three main levels; document-level, sentence-level and entity & aspect-level. Document-level analysis classifies whether an entire opinion document expresses a positive or negative sentiment (Pang et al., 2002; Turney, 2002). Meanwhile, sentence-level analysis is used to determine whether each sentence expressed a positive, negative, or neutral opinion. Neutral usually means no opinion (Wiebe et al., 1999). Entity and Aspect level, both the document-level and sentence-level analyses do not discover what exactly people liked and did not like. Aspect level performs a finer-grained analysis. The aspect level was earlier called feature level (feature-based opinion mining and summarization) (Hu and Liu, 2004).

A list of sentiment words, be it positive or negative is called a sentiment lexicon (Bing Liu, 2012). Three general sentiment lexicons from tidytext package are used to do sentiment analysis. This study used three general sentiment lexicons from the tidytext package: afinn lexicon, bing lexicon from Bing Liu and collaborators and nrc lexicon from Saif Mohammad and Peter Turney (Silge and Robinson, 2018). The afinn lexicon generates sentiment scores between -5 and 5, while bing lexicon classifies words as positive (score 1) or negative (score -1). Meanwhile, nrc lexicon lists words with eight basic emotions i.e., anger, fear, anticipation, trust, surprise, sadness, joy, and disgust and also two sentiments i.e., positive and negative. There are 2,476 words in afinn lexicon, 6,788 words in bing lexicon and 13,901 words in nrc lexicon. However, there are repeated words in nrc lexicon because the same word can be classified as emotion and also as sentiments.

4. RESULT

4.1 Text Analysis and Word Cloud for Five Articles

The five articles from this online news were analyzed separately by comparing the articles and then looking at the results as a whole. Figures 4.1 to 4.5 below are the word cloud used to illustrate the words and their frequency in each article (Refer to Source of Data). The highest words in article 1 as shown in Figure 4.1 below are ‘graduates’ followed by ‘unemployment’ and ‘youths’. Article 2 is about ‘jobs’, ‘data’, ‘automation’, ‘programs’, ‘work’ and ‘training’. Article 3 is similar to article 1 about ‘youth’ and ‘unemployment’. Article 4 is about ‘skilled’, ‘workers’, ‘training’, ‘government’, ‘tvet’ which refer to technical and vocational education and training, while article 5 shows the words ‘job’, ‘industry’, ‘intensive’, ‘labor’ and ‘created’ as among the highest words used.

Figure 4.1: Article 1

Figure 4.2: Article 2

Figure 4.3: Article 3

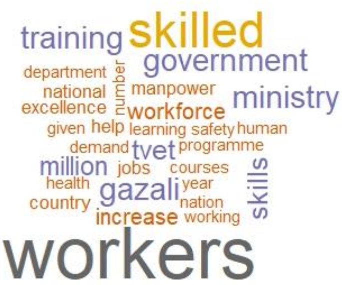

Figure 4.4: Article 4

Figure 4.5: Article 5

Based on these five wordclouds, we generated an overview of the issue especially concerns about job and unemployment among the youth as well as possible initiatives being taken by the government in education, training and industry sectors to address the issue. The combined illustration by wordcloud for articleas 1 to 5 is shown in Figure 4.6 below.

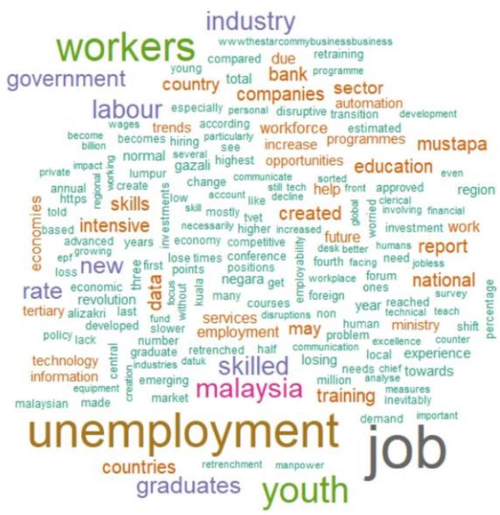

Figure 4.6: Articles 1 to 5

4.2 Analysis using Afinn Sentiment Lexicon

Sentiment analysis using afinn lexicon showed that four out of five articles a have negative sentiment. Afinn lexicon used sentiment scores between -5 to 5. Words like 'unemployment' appeared 10 times in article 1 with a score of -2, so when the word 'unemployment' emerge 10 times, it gives a total score of -20. The total score sentiment for article 1 using afinn lexicon is -11, which is 18 positive scores and 29 negative scores as shown in Figure 4.7 below.

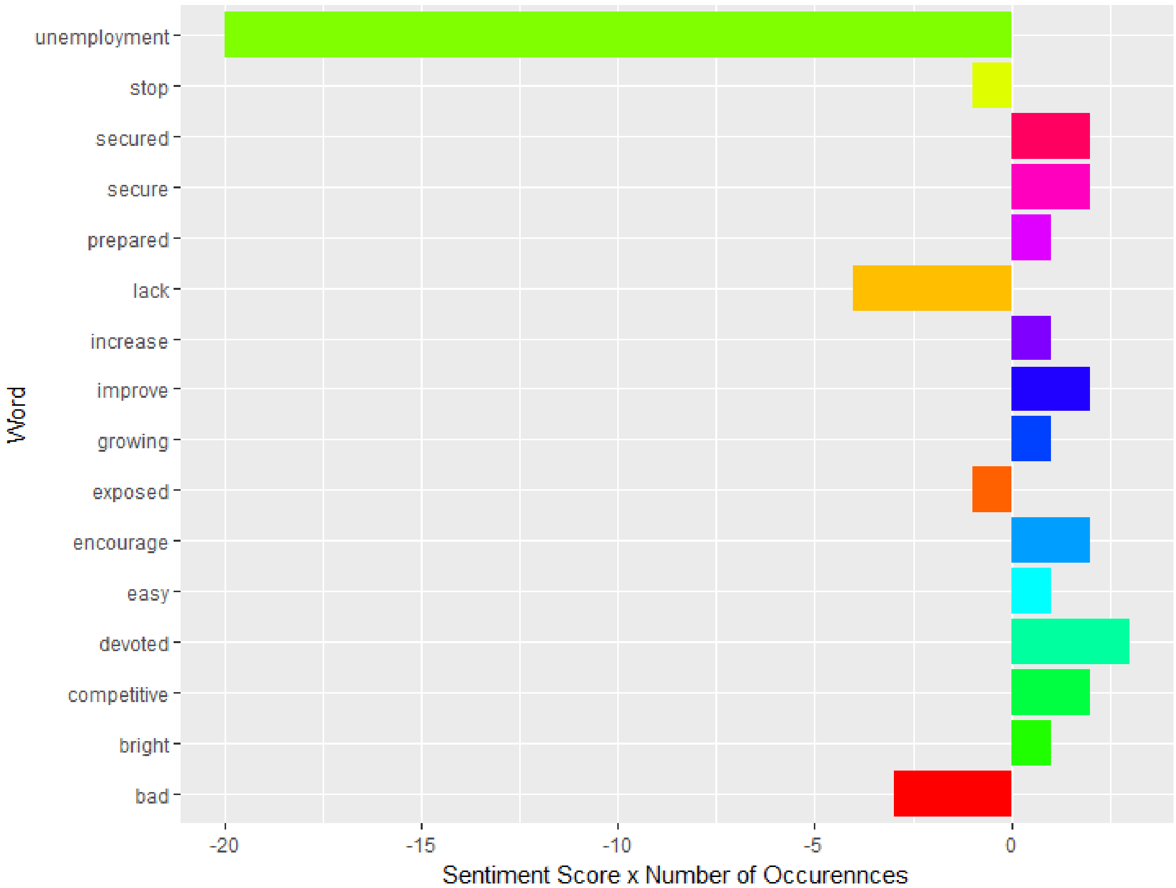

Figure 4.7: Afinn Lexicon Sentiment Score for Article 1

In article 2, there are 20 words listed by afinn lexicon which gives a total score of -32 with a positive score of only 9 compared to a negative score of -41 as shown in Figure 4.8 below.

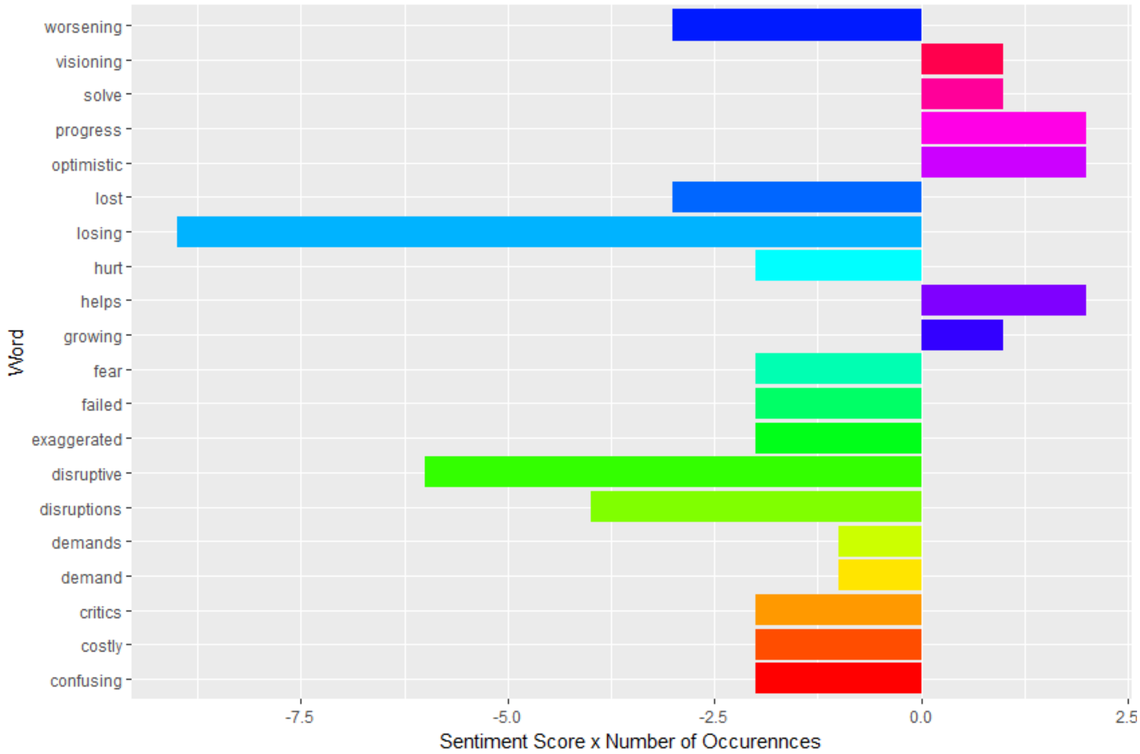

Figure 4.8: Afinn Lexicon Sentiment Score for Article 2

In article 3, there are 9 words listed by afinn lexicon which gives a total score of -21 with a positive score of only 7 compared to a negative score of -28 as shown in Figure 4.9 below.

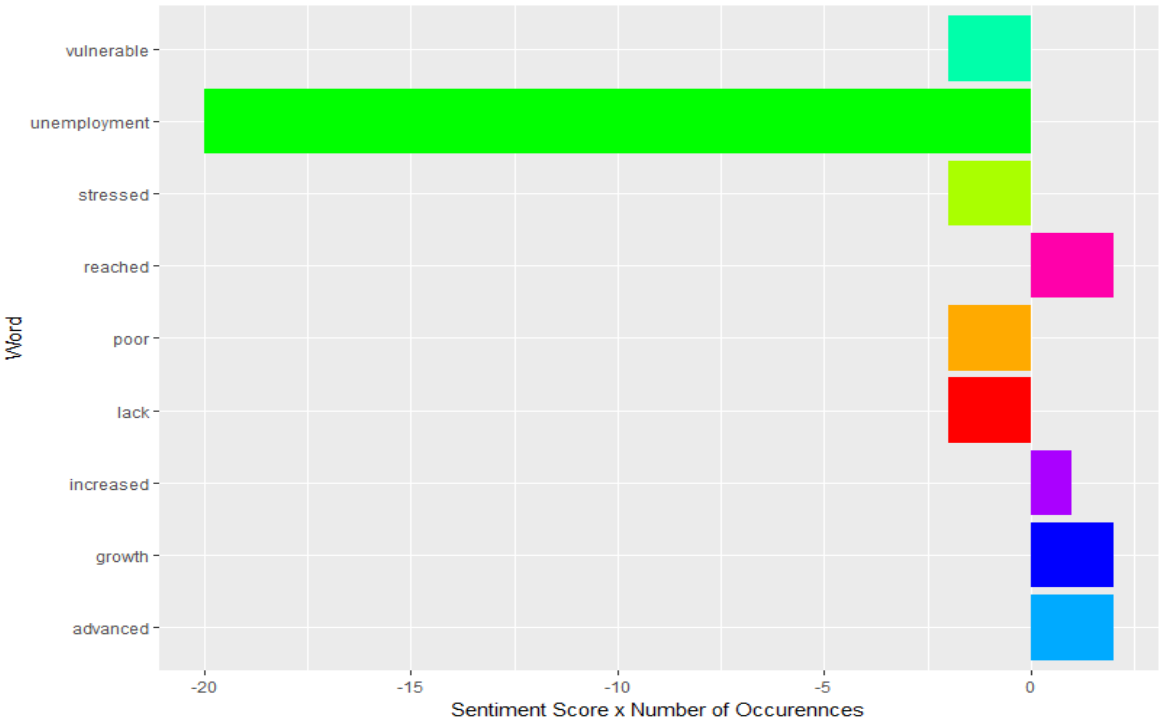

Figure 4.9: Afinn Lexicon Sentiment Score for Article 3

Meanwhile, in article 4, there are 20 words listed by afinn lexicon which gives a total score of 41 with the highest positive score of 58 compared to the rest of the other articles. In article 4, the negative score is quite low i.e. -17.

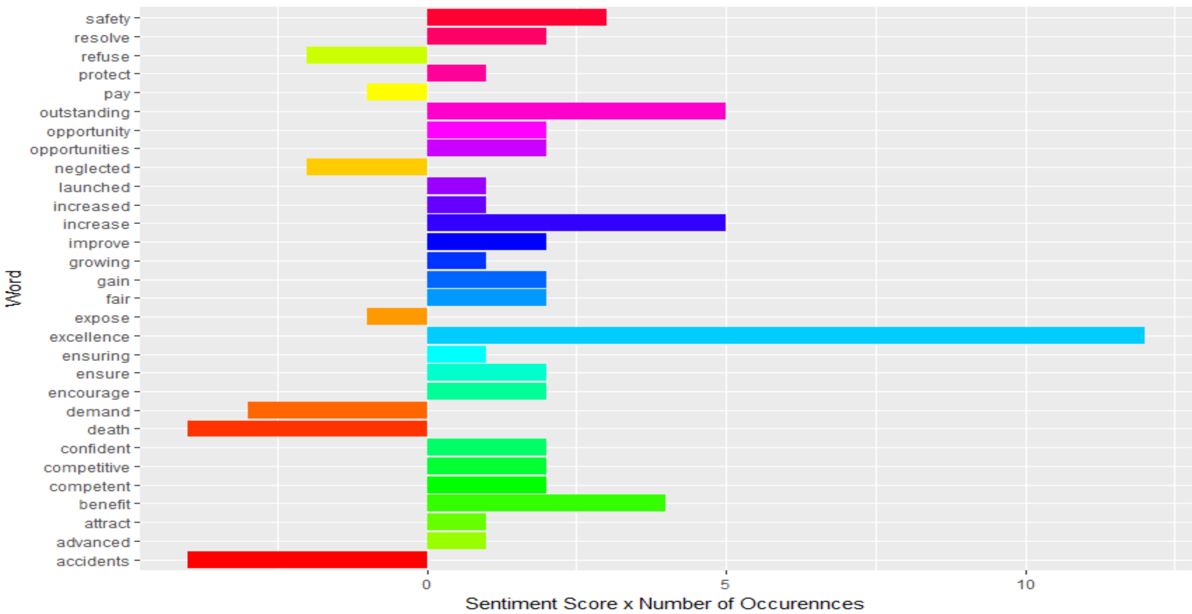

Figure 4.10: Afinn Lexicon Sentiment Score for Article 4

In article 5, there are 9 words listed by afinn lexicon which gives a total score of 2 with a positive score 17 and a negative score of -15 as shown in Figure 4.11 below.

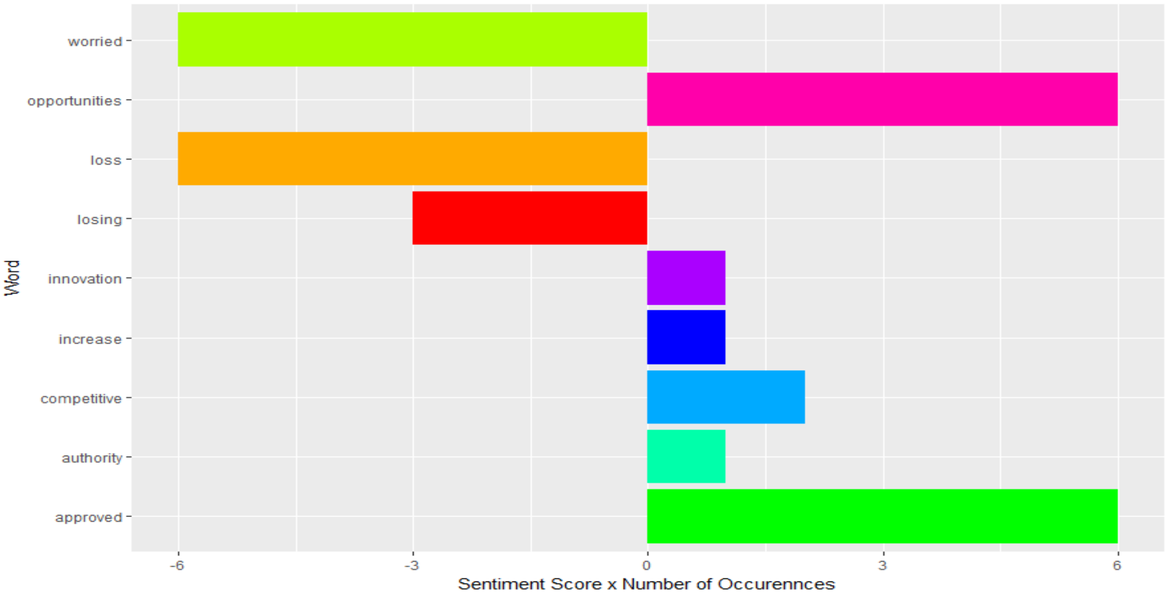

Figure 4.11: Afinn Lexicon Sentiment Score for Article

Overall, all five articles give negative sentiments (positive score: 160, negative score: -251) with a total sentiment score of -91 by using afinn lexicon.

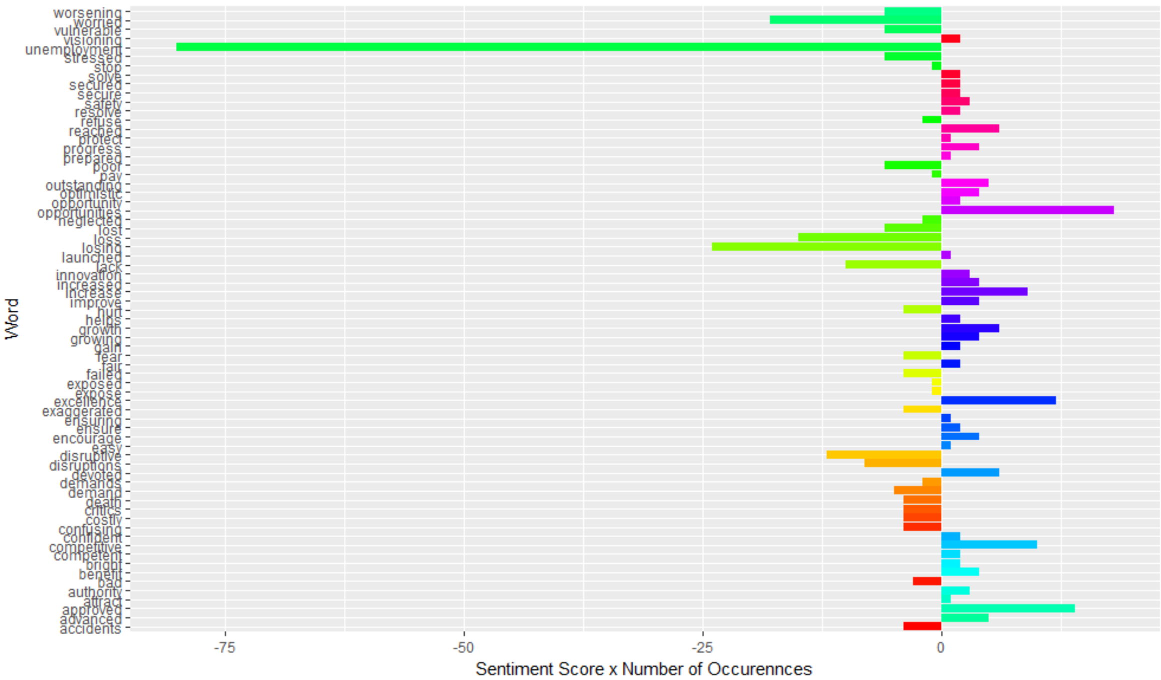

Figure 4.12: Afinn Lexicon Sentiment Score for All Five Articles

4.3 Analysis using Bing Sentiment Lexicon

Using Bing lexicon also shows similar results as afinn lexicon. Bing lexicon has only two sentiments which are positive and negative. Article 1 to 3 gives negative score as shown in Figure 4.13 below. Meanwhile, article 4 gives a positive sentiment score, while article 5 gives almost the same score between positive and negative as shown in Figure 4.14.

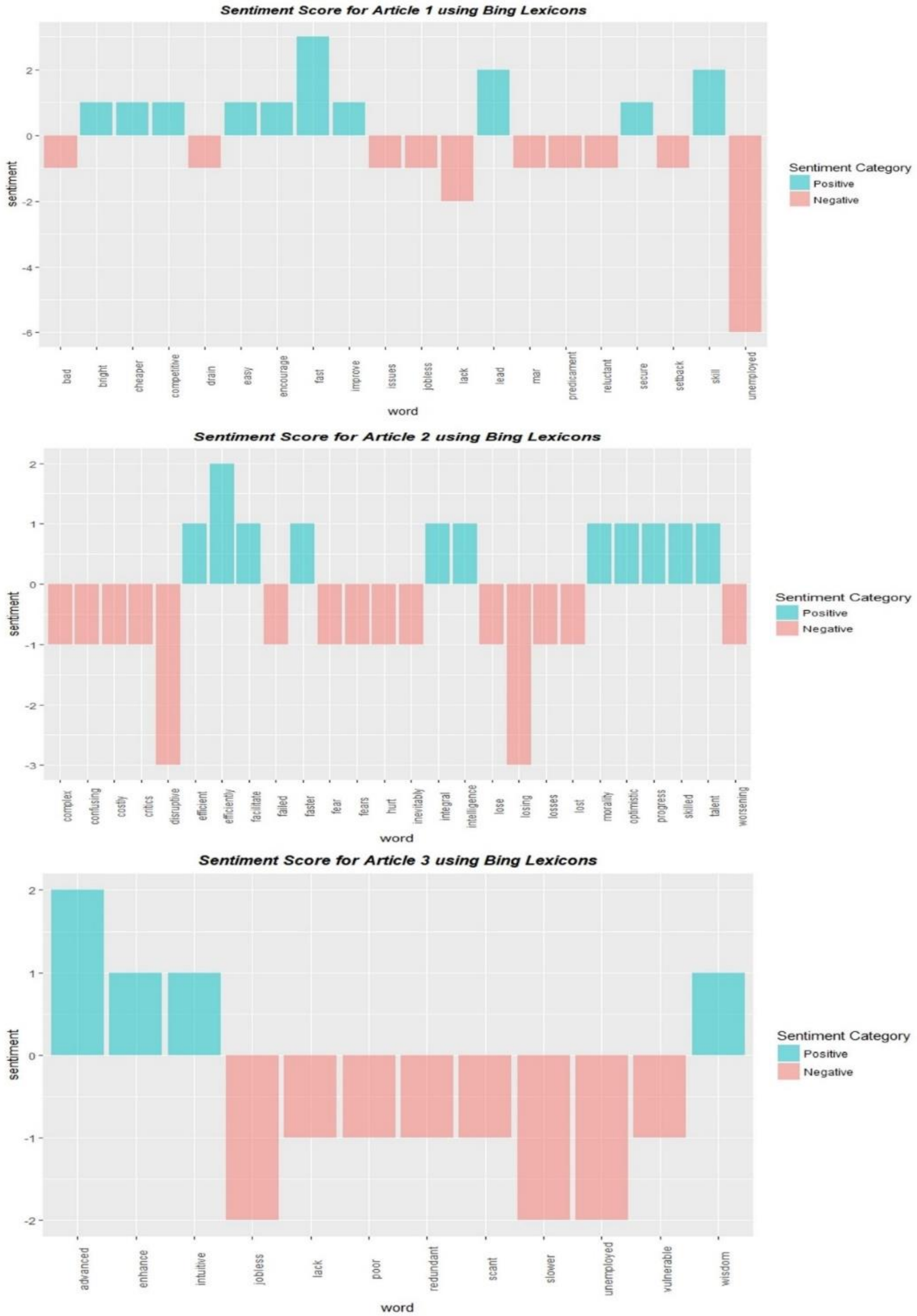

Figure 4.13: Bing Lexicon Sentiment Score for Articles 1, 2 and 3

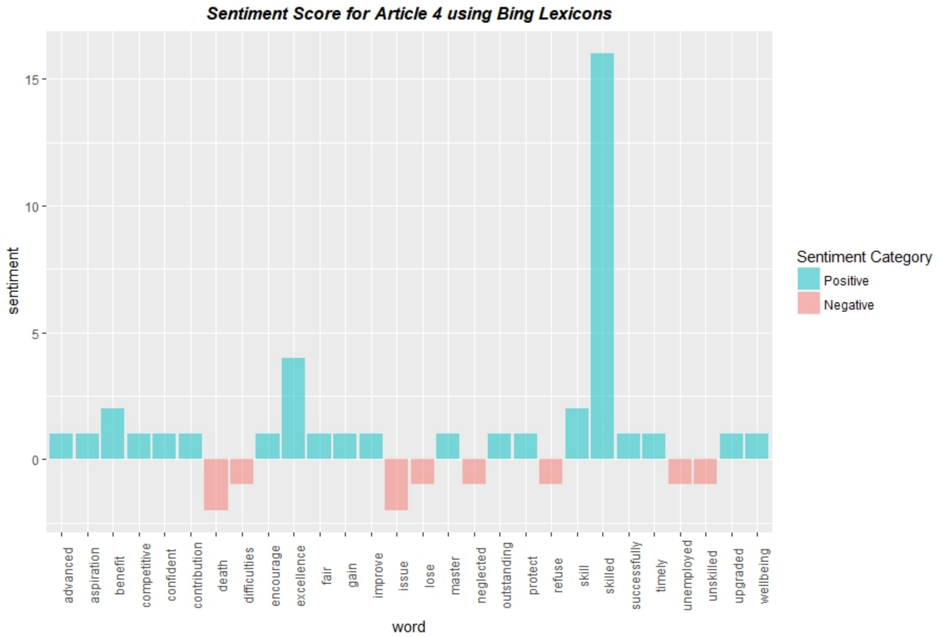

Figure 4.14: Bing Lexicon Sentiment Score for Articles 4

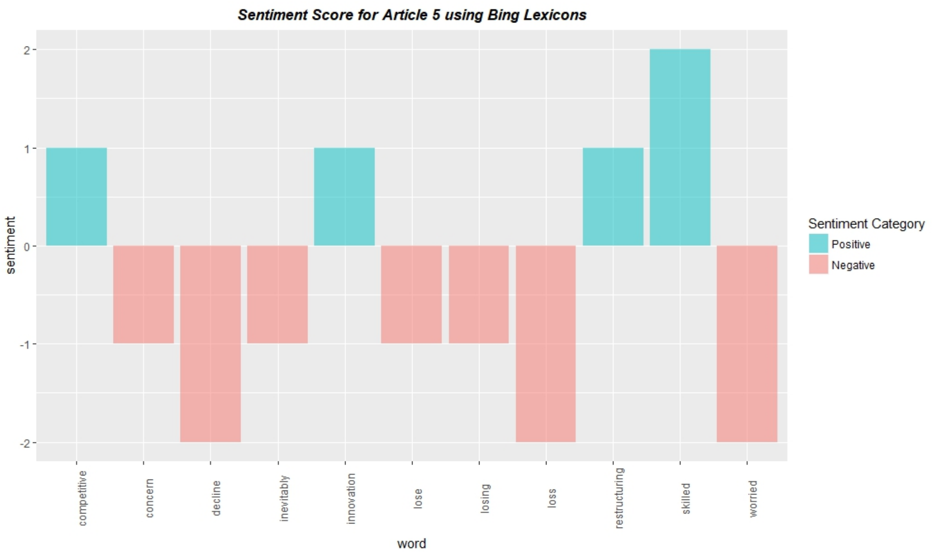

Figure 4.15: Bing Lexicon Sentiment Score for Article 5

Figure 4.16: Wordcloud using Bing Lexicon for Five Articles

4.4 Analysis using NRC Sentiment Lexicon

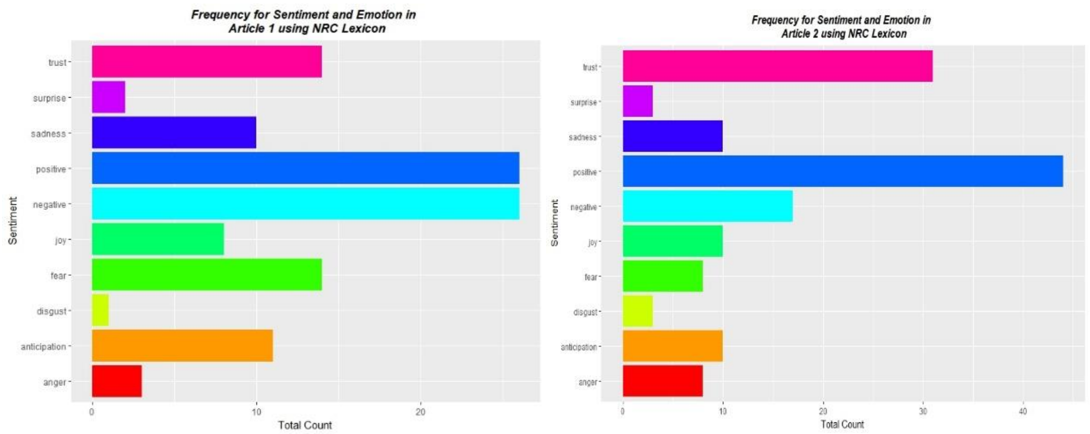

Figure 4.17: Frequency of Sentiment and Emotion Score for Article 1 and 2

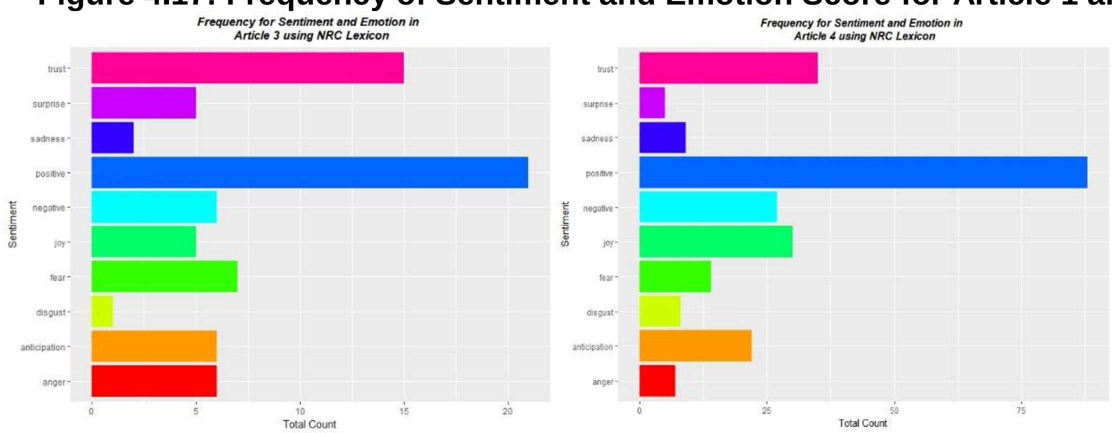

Figure 4.18: Frequency of Sentiment and Emotion Score for Article 3 and 4

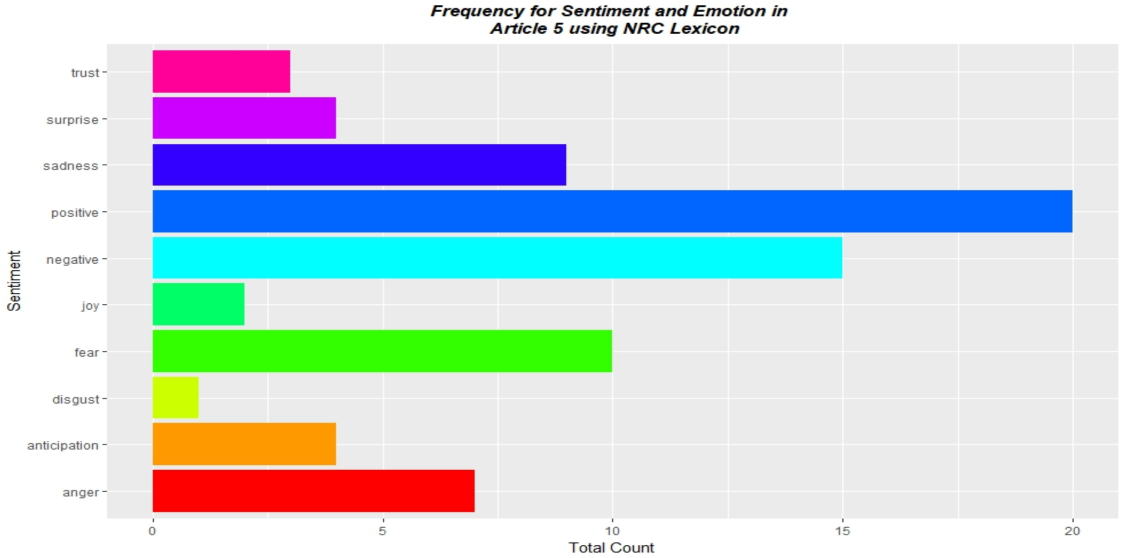

Figure 4.19: Frequency of Sentiment and Emotion Score for Article 5

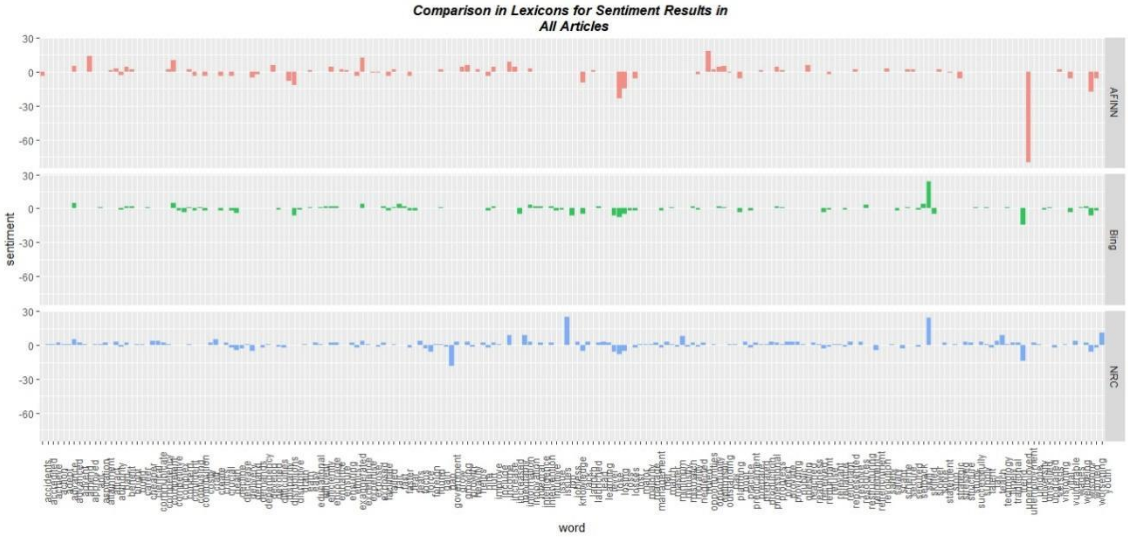

Figure 4.20: Comparison in Lexicon for Sentiment Results in All Articles

5. CONCLUSION

In conclusion, using 'tidytext' in R, sentiment analysis could be conducted. However, when comparing the three lexicon used, it is found that these three lexicon did not list the same sentiment word. For example, the word ‘unemployment’ exists in afinn lexicon with a score of -2, but this word does not exist in bing and nrc lexicon. Based on the afinn and bing lexicon mentioned above, it is found that the issue of job loss in Malaysia raises concern and negativity. NRC lexicon shows positivity as a result of the list of words found.

Findings from all five articles on job loss issue in Malaysia raises concern about unemployment issues, especially among university graduates. The main concern is related to the lack of soft skills and some attitude problems among graduates which contribute to the issue. However, a comprehensive study needs to be done by obtaining official or right figures on supply and demand labour in Malaysia. This is because if we look at a quick search at least from two large job search sites i.e. Jobstreet and LinkedIn, there are indeed many jobs offered including postgraduate not only looking for experienced workers. If the lack of skills may be the cause, perhaps the training for graduates can be given or enhanced. Subjects and syllabus at the university should be viewed and reviewed to suit the current and future needs of the country. In the event of a job mismatch, the salary offered compare to the qualification causes the graduates to be more interested in not working, the job matching policy should be taken seriously. But if further studies show that it is an attitude problem, it would be a more difficult issue to address to curb the unemployment problem among graduates in the country.

REFERENCES

Agarwal, R., & Dhar, V. (2014). Editorial – big data, data science, and analytics: the opportunity and challenge for is research. *Information Systems Research*, *25*(3), 443–448.

Acemoglu, D. (2002). Technical change, inequality, and the labor market. *Journal of economic literature*, *40*(1), 7-72.

Dobre, C., & Xhafa, F. (2014). Intelligent services for big data science. *Future Generation Computer Systems*, *37*, 267–281.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, *5*(1), 1-167.

J. Silge & D. Robinson. (2018). *Text mining with R*. https://www.tidytextmining.com/sentiment.html [12 Jun 2018].

Mudinas, A., Zhang, D., & Levene, M. (2012, August). Combining lexicon and learning based approaches for concept-level sentiment analysis. *In Proceedings of the first international workshop on issues of sentiment discovery and opinion mining*. 5. ACM.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. *In Proceedings of the ACL-02 conference on Empirical methods in natural language processing. 10*, 79-86. Association for Computational Linguistics. content [28 May 2019]