Kolokium Statistik dan Scientific Poster 2019
Jabatan Perangkaan Malaysia

Sentiment Analysis on Non-citizen Workers in Malaysia

Muhammad Shafiq Harun & Nur Huriyatul Huda Abdullah Sani
Core Team Malaysian Bureau of Labour Statistics (MBLS)

26 September 2019

Outline

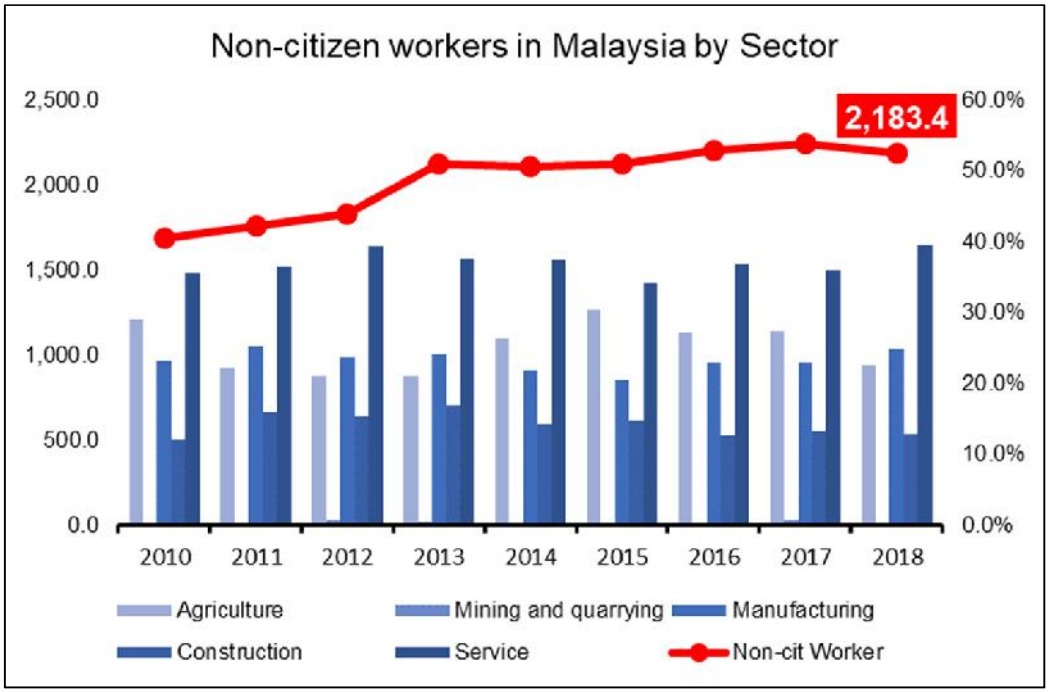
- Introduction
- Sentiment Analysis
- Methodology
- Results & Findings
- Conclusion

Introduction

Malaysia has experienced a rise in foreign labour flows in response to **steady economic expansion** and **demographic changes**.

Based on Labour Force Surveys (LFS) by the Department of Statistics Malaysia (DOSM), the non-citizen workers have been hovering around **15% of the total employed person in 2018** equivalent to 2.18 mil person.

Most of the Malaysians strongly believe the country's dependence on foreign workers endure serious long term consequences.



Introduction [cont]

It is undeniably true that there are many endless complaints and sentiments from Malaysians on the issue of increasing non-citizen workforce

Type in the search box on the Google portal with the keyword "**foreign workers in Malaysia**", there are over 34.9 million posts related to this keyword.

When searching with the keyword "**foreign workers in Malaysia**" on media social (Twitter, Facebook), a number of narrative texts in the app appeared as shown…

Introduction [cont]

This study consider an alternative approach to measuring sentiments, with a focus on the foreign workers in Malaysia embodied in the online news articles.

By using **sentiment analysis** pertaining to foreign workers in Malaysia through online news articles, the main objectives of this case study are to identify the sentiment and emotion words used:.

- to classify positive and negative sentiment words toward non-citizen workers in Malaysia
- to calculate the overall sentiment score for each articles through words used

Sentiment Analysis

Synonymous & interchangeable name

- Subjective analysis, review mining, **opinion mining**, appraisal extraction

What is it?

- is a computational study of applying Natural Language Processing (NLP) and Text Analysis technique to identify and extract subjective information from expressed plain text.
- It is an emerging field that attempts to analyze and measure human emotions and convert it into hard facts

Studies on labour market have used sentiment analysis, see…

1. Shahid, S, Phoong, S., Yeong, W.C., Ainin, S., Noor, I.J., & Shamshul, B.Z (2017) Social Media Sentiment Analysis on Employement in Malaysia
2. Bailliu, J., Han, X., Kruger, M., Liu, Y. H., & Thanabalasingam, S. (2018). Can media and text analytics provide insights into labour market conditions in China?. Shahid, S, Phoong, S., Yeong, W.C., Ainin, S., Noor, I.J., & Shamshul, B.Z (2017) Social Media Sentiment Analysis on Employement in Malaysia

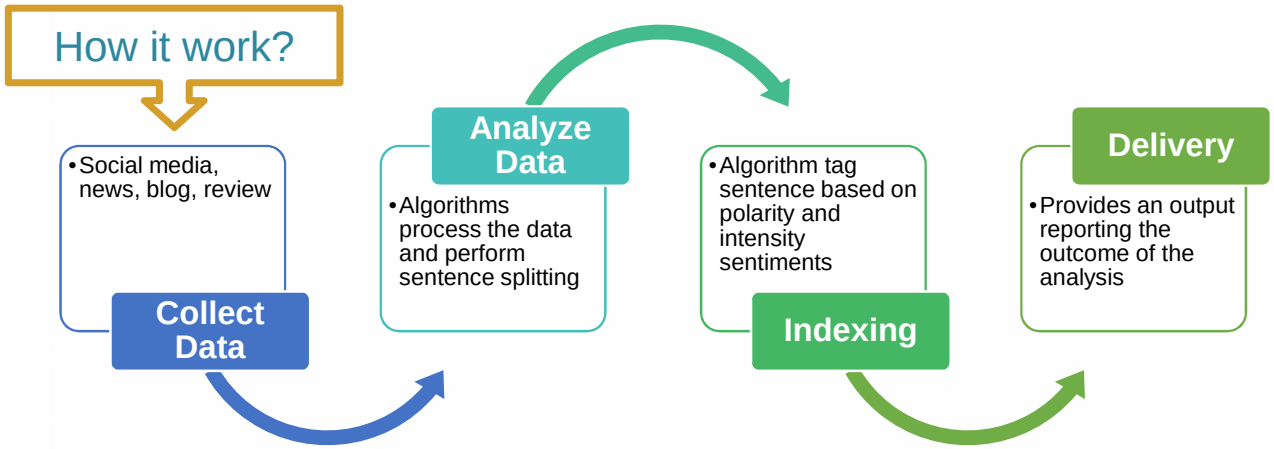
Sentiment Analysis [cont]

Sentiment analysis can be performed at three main levels

- Document-level, sentence-level and aspect-level

The methodology for sentiment analysis:

- Lexicon based analysis, machine learning technique, hybrid modeling



The challenges faced in conducting sentiment analysis

- Usage of slang, sarcasm and the absence of build-in lexicon library in most our local languages processing

Methodology

- Data source
- Text Processing & Sentiment Analysis

Data Source

Online news articles search by keyword '**foreign worker in Malaysia**'.
Ten articles (8,137 word) were selected randomly as follows:

1. Cover story: do we need so many foreign workers?
2. Don't demonise foreign workers, Malaysian told
3. Low skilled tech, foreign workers to slow down Malaysia's march to advanced economy
4. Better welfare for foreign workers
5. Foreign worker sector brace for a tough 2019
6. Do foreign workers really steal jobs
7. Foreign Workers: Nepal slams the door, unhappy with company monopoly
8. Economic Impact of Foreign Workers in Malaysia: An Objective Review
9. Place for migrant workers in our economy
10. Economist: Manage labour issues to achieve high-income economy

Article Date:

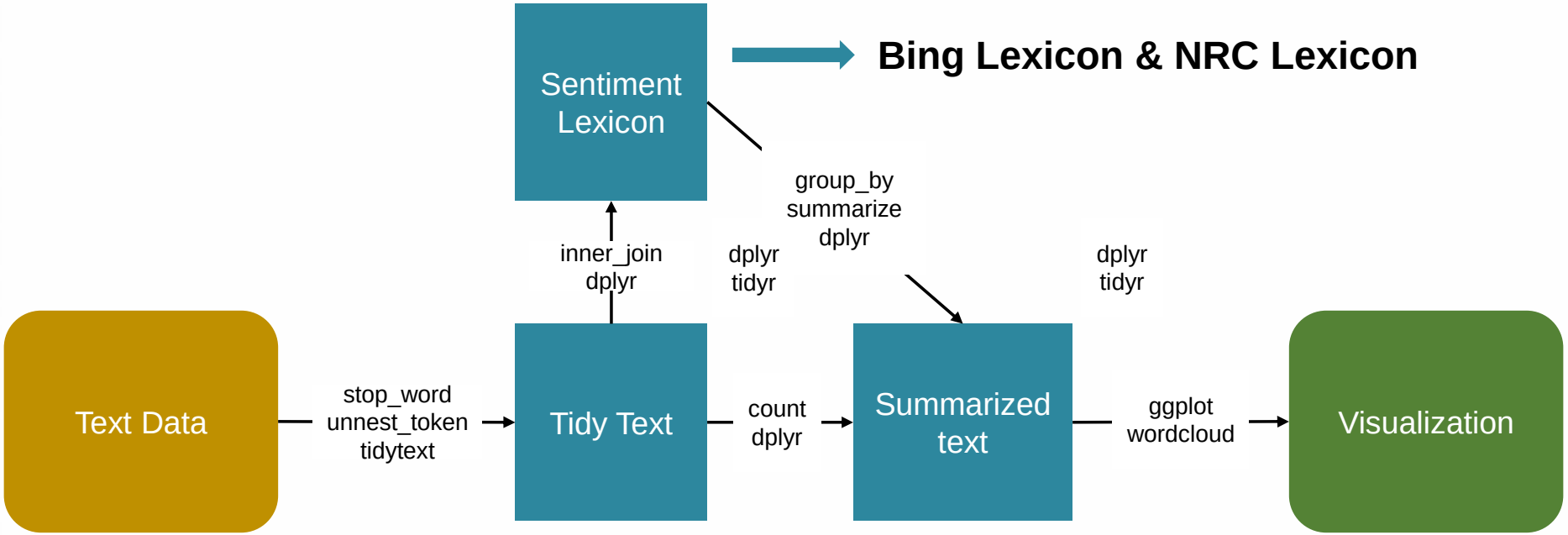
Sept 2018 – Aug 2019

Publisher:

The Edge Market
The Star Online
The News Strait Time
The Malaysian Reserve
The Free Malaysia Today

Text Processing & Sentiment Analysis

R programming software:
Packages **(rvest, NLP, tidytext, dplyr, ggplot, wordcloud)**



Flowchart of text analysis that uses tidytext for text processing and sentiment analysis.

Results and Findings

Most Common Words

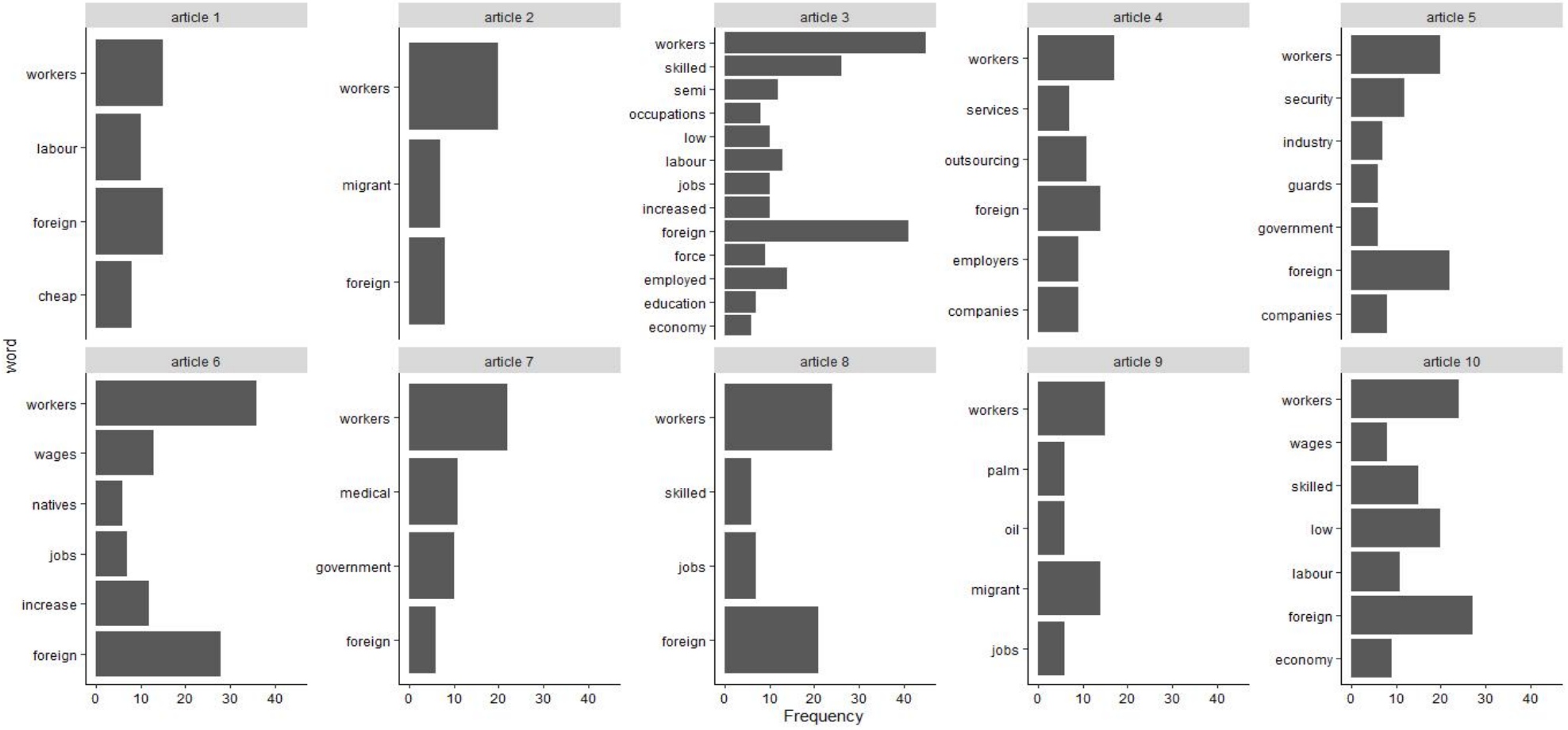
Word count for all articles…

The word cloud indicated that words like "**foreign", "workers", "skilled", "labour", "job", "low"** and **"services"** among the highest number of word.

- An overview of opinion expressed by writers in media regarding issue especially concerns about the **increase in employment of foreign worker in services sector** as well as the **level of foreign labour skill.**

Most Common Words

Word count for each articles…



Most common words

+ve and –ve word count for all articles…

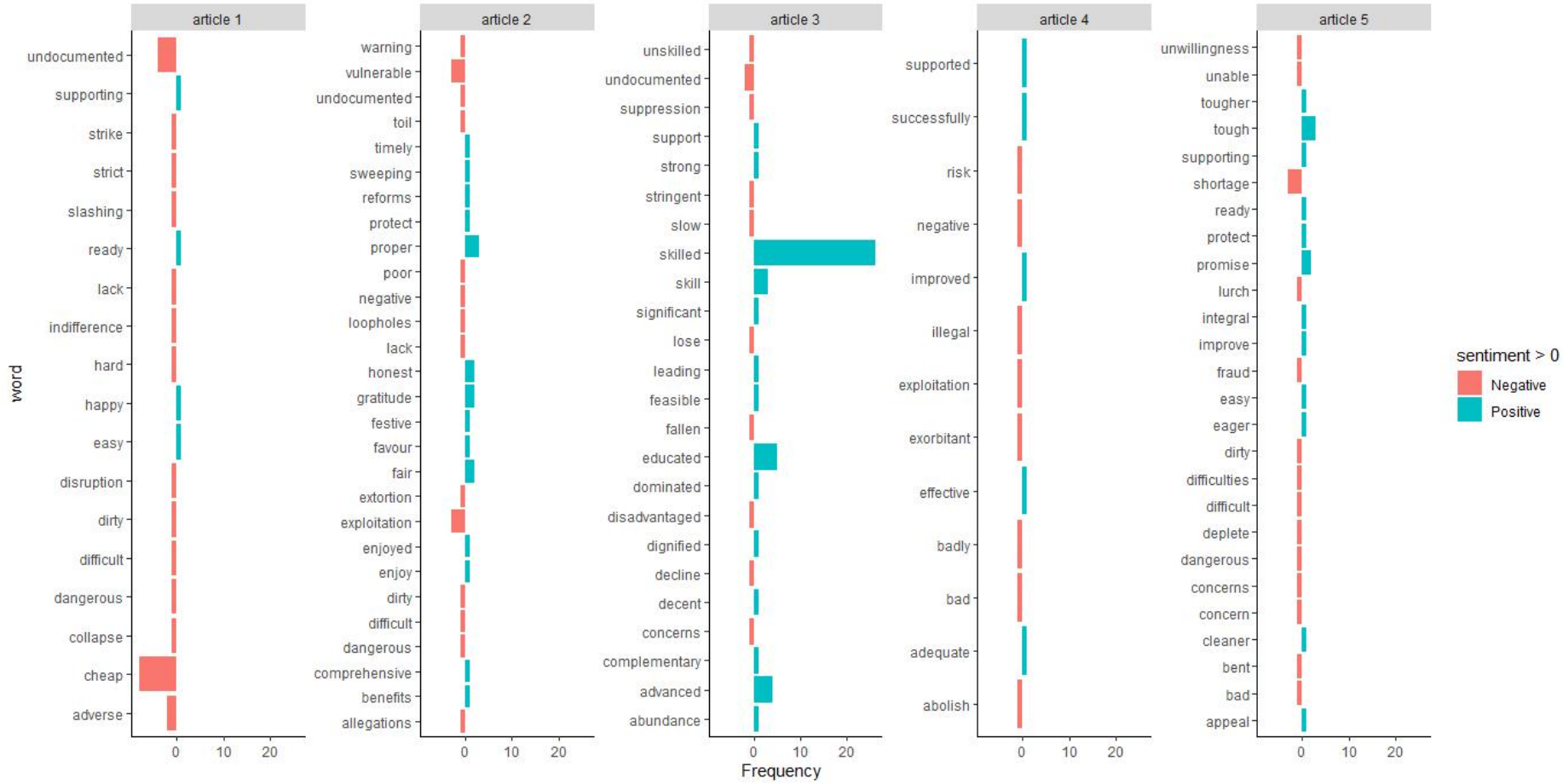
The word cloud visualized the most common positive and negative words based on the Bing and NRC lexicon.

The most frequently used:

- Positive: **“skilled”, “educated”** and **“benefits”**
- Negative: **“cheap”**, **“unskilled”** and **“undocumented”**.

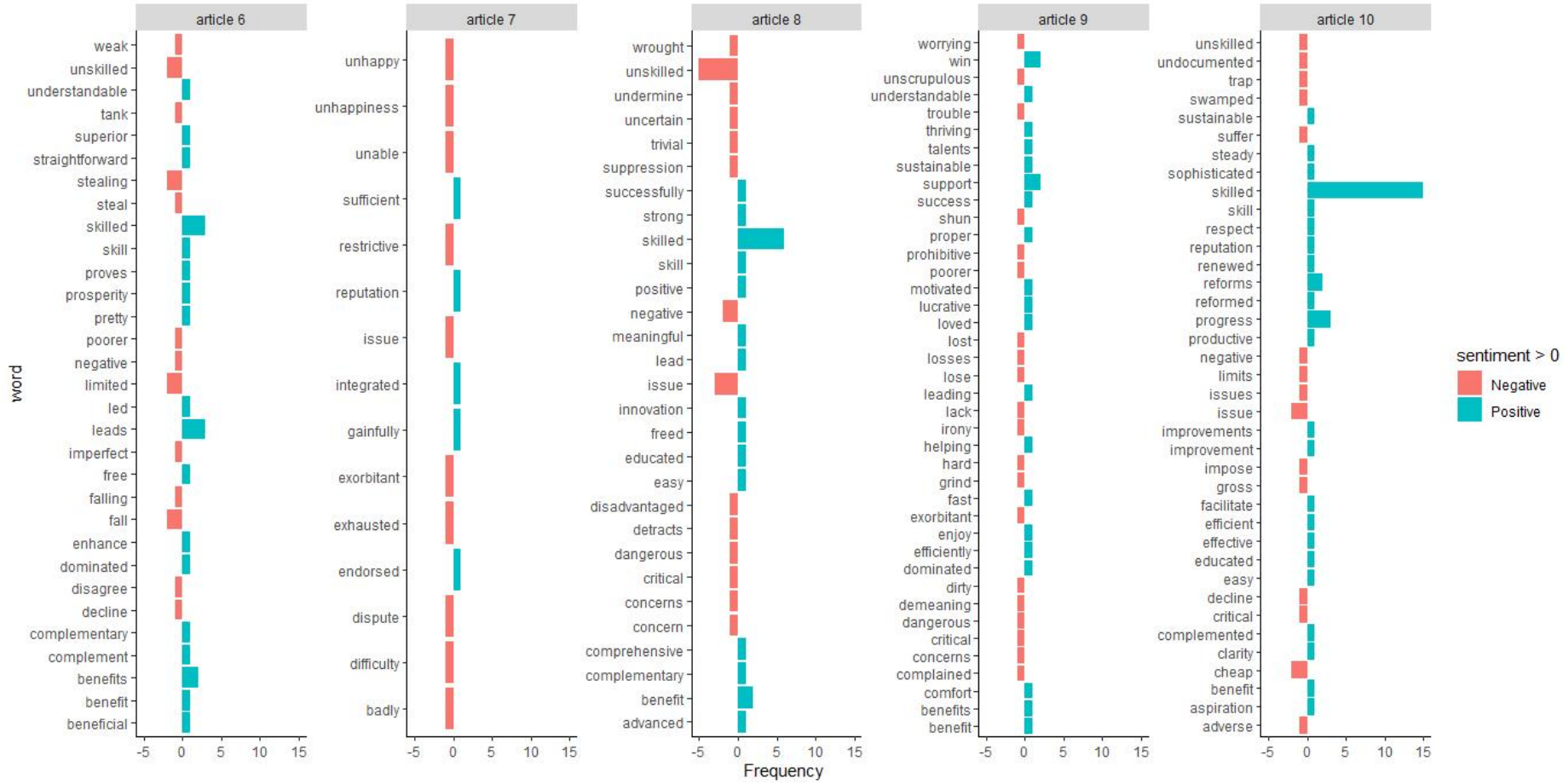
Most common words

+ve and –ve word count for each articles…



Most common words

+ve and –ve word count for each articles…

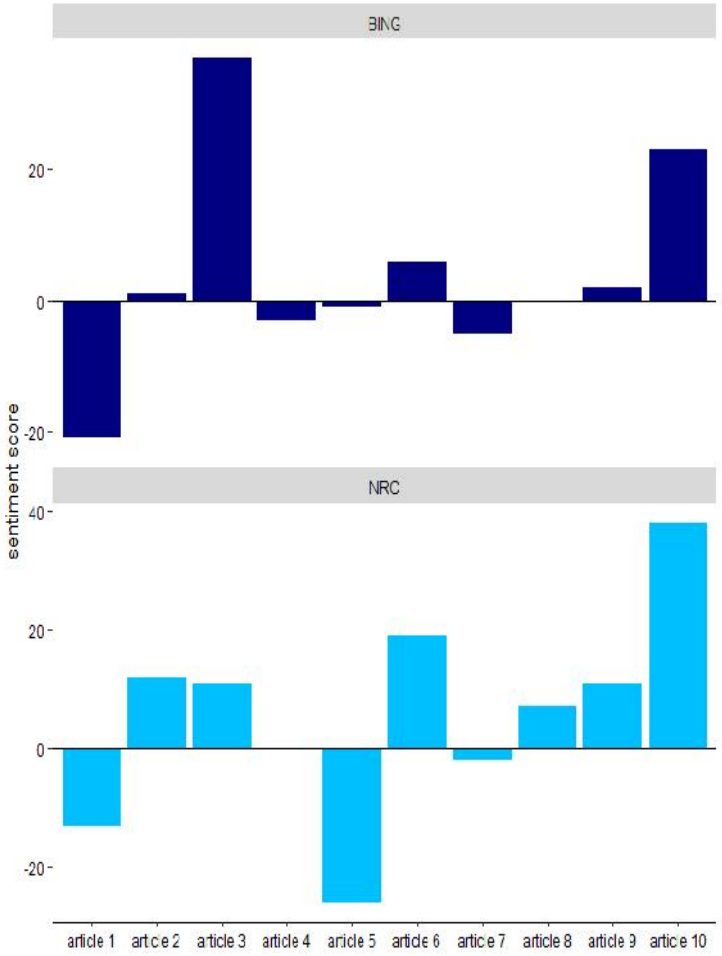


Sentiment analysis score

There are similar dips and peaks in sentiment at about the same articles but the magnitude of the sentiment score for each article are significantly different.

For both lexicon indicated Article 2, 3, 6, 9 and 10 are positive sentiment while article 1, 5, and 7 is negative sentiment deriving based on the opinion or attitude of the writer.

Article	Method: Bing			Method: NRC		
	Positive Word (n)	Negative Word (n)	Sentiment Score	Positive Word (n)	Negative Word (n)	Sentiment Score
1	4	25	-21	29	42	-13
2	19	18	1	33	21	12
3	48	11	37	74	63	11
4	5	8	-3	27	27	0
5	15	16	-1	33	59	-26
6	23	17	6	64	45	19
7	5	10	-5	25	27	-2
8	21	21	0	41	34	7
9	22	20	2	26	15	11
10	40	17	23	92	54	38



Conclusion

- Sentiment analysis provides a way to understand the attitudes and opinions by people expressed in texts.

- This study explored how to approach sentiment analysis using tidy data principles to understand how a narrative arc changes throughout its course or what words with emotional and opinion content are important for a particular text.

- Further classifying positive, negative or neutral sentiment need to be implemented in the future to provide more information for us especially government to monitor toward issue regarding foreign workers in Malaysia.

TERIMA KASIH & THANK YOU

20 OCT	18 - 23 AUG 2019	JULY 2020 (ACTUAL MYCENSUS)	JAN – DEC 2019	MAR – SEPT 2019	2015 - 2030

#StatsMalaysia | #MyStatsDay | #ISIWSC2019 | #MyCensus2020 | #HIES2019 | #MyRetailCensus2019| #LeaveNoOneBehind

References

Agarwal, R., & Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for IS research.

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

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177). ACM.

Ludvigson, S. C. (2004). Consumer confidence and consumer spending. *Journal of Economic perspectives*, *18*(2), 29-50.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, *5*(4), 1093-1113.

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* (p. 5). ACM.

Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. " O'Reilly Media, Inc.".

Silge, J., & Robinson, D. (2016). tidytext: Text Mining and Analysis Using Tidy Data Principles in R. *J. Open Source Software*, *1*(3), 37.

Wawre, S. V., & Deshmukh, S. N. (2016). Sentiment classification using machine learning techniques. *International Journal of Science and Research (IJSR)*, *5*(4), 819-821.

Zhao, J., Liu, K., & Xu, L. (2016). Sentiment analysis: mining opinions, sentiments, and emotions. 595-598.