

Identifying The Key Words Of Online Economic News In Malaysia Through Text Analytics



By:

Faiza Rusrianti Bt Tajul Arus

Capstone Project

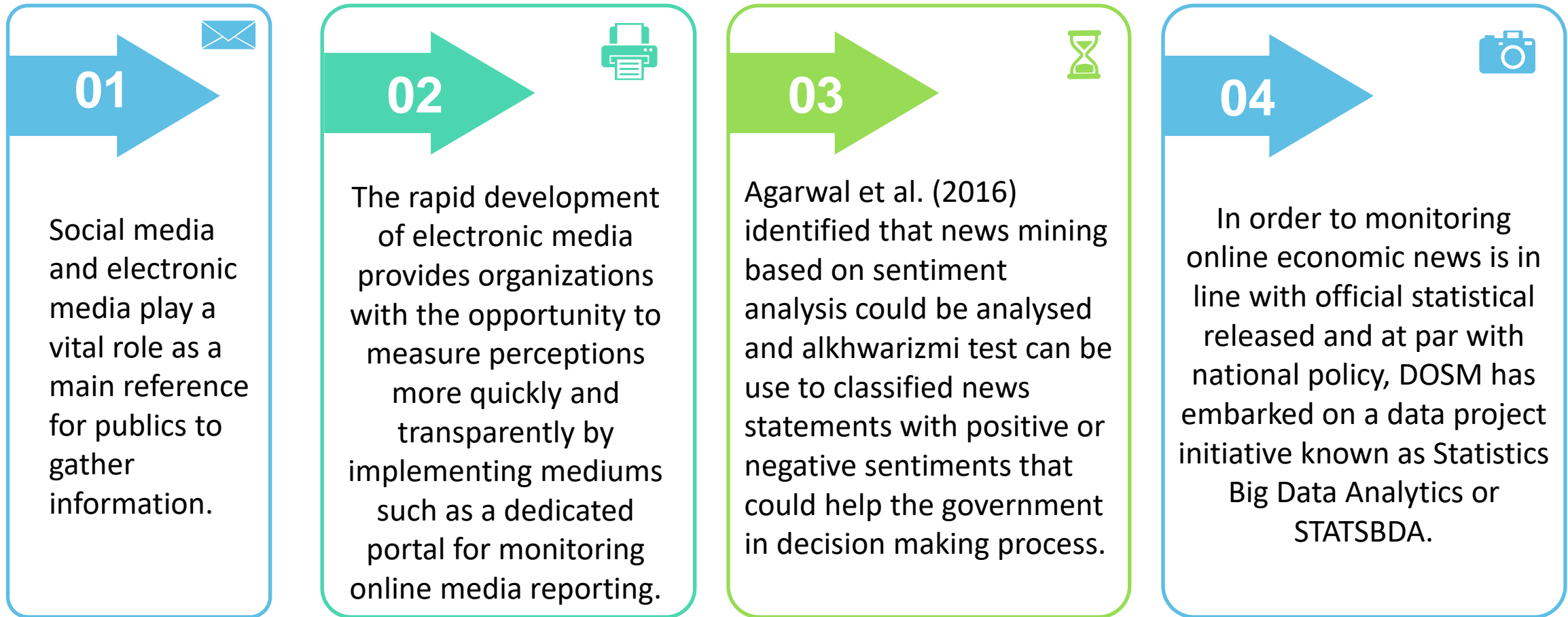
School of Mathematical Sciences, UKM

Presentation Outline



- 1 Introduction
- 2 Literature Review
- 3 Methodology
- 4 Data Collection
- 5 Statistical Analysis
- 6 Result
- 7 Conclusion

1. Introduction





Public Maturity Assessment on Official Statistics (PMAOS) and Real Time News on Official Statistics (RTOS) were developed to provide DOSM with holistic and comprehensive insight of perception developed by media.


At a glance of PMAOS & RTOS Analysis


mmcloud.ebdesk.com

Media Intelligence










It tells what you don't know
...

Media Monitoring



Media Campaign

Media Analysis

Media Intelligence | my topic 1 | my topic 2 | resume | map | edit |


H

DISCOVER

ANALYZE


ACTION

DOSM2




KES sindiket melibatkan 47 rakyat Malaysia ditipu ditawarkan pekerjaan sebanyak US\$1,500 (RM6,000) sebulan di Kemboja mencerminkan keterdesakan rakyat untuk...
3 months ago, Sarawak Voice

Population1




Jumlah populasi kaum wanita di negara ini adalah 16 juta daripada 33.3 juta populasi keseluruhan yang diunjurkan 2019 dengan nisbah sebanyak 103 lelaki dan...
3 months ago, Sinar Harian

GDP1




In the eyes of the world today, Malaysia is seen as a relatively affluent upper-middle income nation. According to World Bank projections, we are expected to...
3 months ago, Din Merican

Labour Force2




KES sindiket melibatkan 47 rakyat Malaysia ditipu ditawarkan pekerjaan sebanyak US\$1,500 (RM6,000) sebulan di Kemboja mencerminkan keterdesakan rakyat untuk...
3 months ago, Astro Awani

Salary & wages4




KES sindiket melibatkan 47 rakyat Malaysia ditipu ditawarkan pekerjaan sebanyak US\$1,500 (RM6,000) sebulan di Kemboja mencerminkan keterdesakan rakyat untuk...
3 months ago, Sarawak Voice

CPI-BPHPP1




KES sindiket melibatkan 47 rakyat Malaysia ditipu ditawarkan pekerjaan sebanyak US\$1,500 (RM6,000) sebulan di Kemboja mencerminkan keterdesakan rakyat untuk...
3 months ago, Astro Awani

DOSM Industry0



KES sindiket melibatkan 47 rakyat Malaysia ditipu ditawarkan pekerjaan sebanyak US\$1,500 (RM6,000) sebulan di Kemboja mencerminkan keterdesakan rakyat untuk...
3 months ago, Astro Awani

HIES0



Kita berada dalam zaman yang tidak menentu. Sungguhpun saban hari kita dihubungkan dengan jaringan sosial yang luas melalui Facebook, Whatsapp dan sebagainya....
Over a year ago, Ar Risalah

New Topic

Add new topic and related keywords to monitor and analyze subject.

eBdesk Media Intelligence

4

Media Intelligence Analysis

Media Intelligence | my topic 1 | my topic 2 | resume | map | edit |

H DISCOVER ANALYZE

DOSM

3

Population

4

An NGO today proposed a separate division or agency to handle senior citizen affairs, saying their number is growing significantly. Third Age Media...

14 hours ago, The Sun Daily

An NGO today proposed a separate division or agency to handle senior citizen affairs, saying their number is growing significantly. Third Age Media...

14 hours ago, The Sun Daily

Labour Force

3



Menteri Besar Johor Datuk Dr Sahrudin Jamal hari ini menasihati para graduan masa kini supaya tidak hanya mempertaruhkan kelulusan akademik semata-mata untuk...

Salary & wages

0

Young people, who are considered prime contributors to economic growth, are spending way beyond their means, and this has brought many of them to the brink of...

GDP



International Finance Centre (IFC) has revised its gross domestic product (GDP) for Malaysia from 1.4% to 1.2% for the first quarter...

1 month ago, The Sun Daily

CPI-BPH

Malaysia's inflation rate rose to 1.5% in August from 1.4% in July, according to the latest Consumer Price Index (CPI) ratings. The...

Select Topic

Dashboard

Timeline

Top Issues

Top Persons

Influencers

Media Share

Maps

Sentiments

Ontology

Reporting


Trending


People Connection


Top Persons & Top Issues


Top Persons - Population1


News


**Najib Razak**
93 news


**Mohd Uzir Mahidin**
90 news


**Mahathir Mohamad**
55 news


**Donald Trump**
37 news


**Kim Jong Un**
21 news

**Abdul Rahman Dahlan**
19 news

**Rex Tillerson**
19 news

**Susan Thornton**
18 news

**Azizah Ismail**
17 news

**Johari Abdul Ghani**
16 news

732 Persons

Prev Next

Mentioned in News

MORE



FDI figures based on approved investments

Dec 11th, 2018 - Daily Express

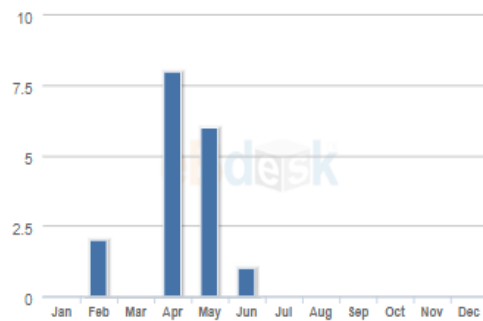
The spike in foreign direct investments (FDIs) into the country to RM49 billion in the first nine months of this year from RM14 billion in

the corresponding period last year was based on investments already approved by the National Committee on Investment (NCI), says Deputy International Trade and Industry Minister Dr Ong Kian Ming (pic). He said the FDIs into Malaysia jumped almost four-fold, a figure higher than that announced by Bank Negara Malaysia (BNM). "The figure shown by BNM is...



News

2019



Media Share | Pie | Bar |



News

MORE

To-do list of law reforms awaiting new...

Jun 11th, 2018 - The Malay Mail Online

Tommy Thomas's appointment as the new Attorney-General has been hailed as boding well for the new government's planned institutional reforms, but what are these promised ...

WILL TOMMY THOMAS START TO PLAY...

Jun 11th, 2018 - Malaysia Chronicle

Tommy Thomas's appointment as the new Attorney-General has been hailed as boding well for the new government's planned institutional reforms, but what are these promised ...

To-do list of law reforms awaiting new...

Jun 10th, 2018 - Yahoo News Singapore

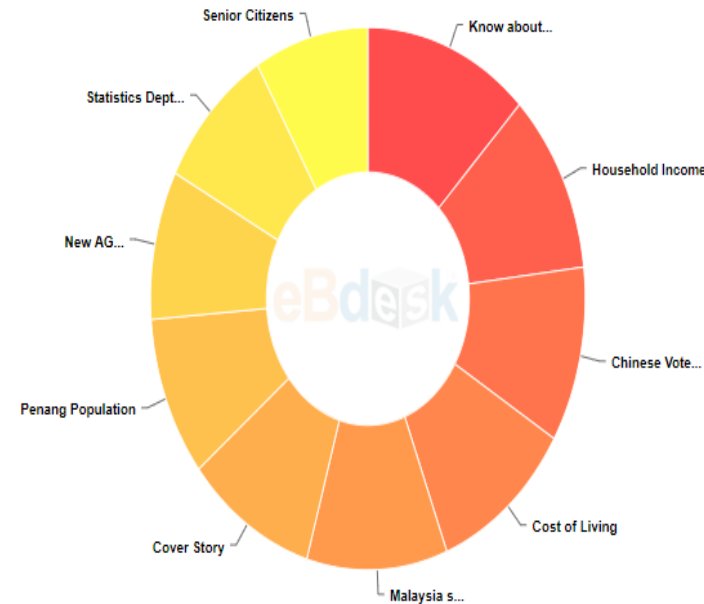
Tommy Thomas's appointment as the new Attorney-General has been hailed as boding well for the new government's planned institutional reforms, but what are these promised ...

Top Issues - Population1 | actual | history | comparison |

News | Twitter | Facebook

Top issues

All issues



Found 39 Related News :



Who is plucking Penang population figures from the air

Nov 27th, 2018 - Anilnetto

This statement has just been released by Penang Forum: Penang Forum would like to respond to Penang Chief Minister Chow Kon Yeow's statement that the Penang population, as projected...



Who is plucking Penang population figures from the air

Nov 27th, 2018 - Ailran

Penang Forum would like to respond to Penang Chief Minister Chow Kon Yeow's statement that the Penang population, as projected by the project delivery partner for the RM46bn transport...

Cover Story: Factors of locality must not be overlooked

Aug 23rd, 2018 - The Edge Markets

Affordable housing has been talked about for many years now but ask anyone about it and it is likely he or she will not be too clear about what that really means. What actually is an ...



Empowering senior citizens to be independent, live with dignity

Aug 2nd, 2018 - The Borneo Post

Malaysia's senior citizen demography is trending upwards, with the percentage of older persons nearing almost 20 per cent of the population by 2040, according to estimates by ...

Media Share & Sentiment Analysis (News Title)

Media - Population1

| index | sentiments |

H DISCOVER ANALYZE ACTION

News

T W M Q Y Adv

The Edge Markets
150 news



MY Stock 118
96 news



The Sun Daily
94 news



The Malay Mail Online
69 news



Bernama
50 news



MSN Malaysia
50 news



The Borneo Post
43 news



Free Malaysia Today
26 news



Malaysia Digest
25 news



New Strait Times
25 news

71 Media

Prev Next

Mentioned in News

MORE



Malaysia on track to achieve Sustainable...

Dec 28th, 2018 - The Edge Markets

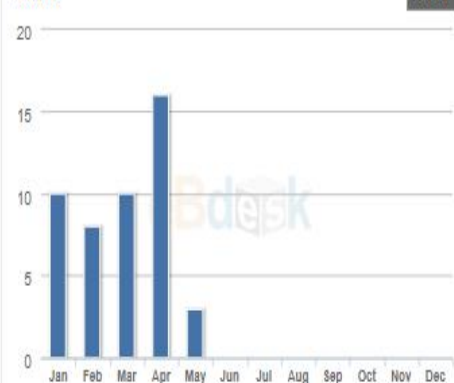
Malaysia is on the right track in reaching the Sustainable Development Goals (SDGs) Indicators 2030, said the Department of

Statistics Malaysia. In a statement today, it said the governments commitment to the sustainable development agenda is transpired through the Mid-Term Review of the 11th Malaysia Plan (11th MP) presented on Oct 18. In this regard, the commitment to Agenda 2030 for Sustainable Development has been aligned with the strategies and initiatives of the 11th Mala..



News

2019



Related Influencers



Statements

MORE

SDGs indicators cover 17 goals, 169 targets and 244 indicators to measure the achievement of universal development which comprise social, economic and...

Mohd Uzir Mahidin

Dec 28th, 2018 - The Edge Markets



"Out of the 244 indicators, 118 indicators (48%) are Available, 55 indicators (23%) Partially Available, 36 indicators (15%) Partially Available Need Further..."

Mohd Uzir Mahidin

Dec 28th, 2018 - The Edge Markets



"SDGs with a theme Leaving No One Behind focuses on the development of 5Ps namely

Media - Population1

| index | sentiments |

H DISCOVER ANALYZE ACTION

News

T W M Q Y Adv

The Edge Markets
150 news



MY Stock 118
96 news



The Sun Daily
94 news



The Malay Mail Online
69 news



Bernama
50 news



MSN Malaysia
50 news



The Borneo Post
43 news



Free Malaysia Today
26 news



Malaysia Digest
25 news



New Strait Times
25 news

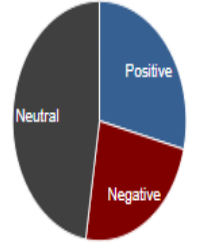
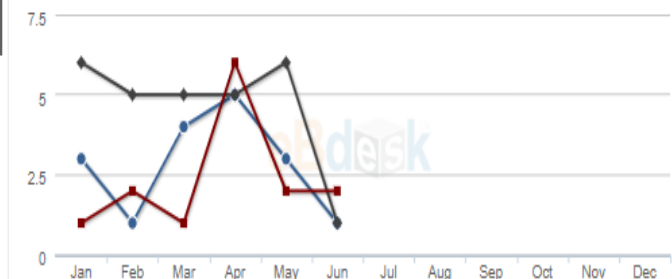
71 Media

Prev Next

Sentiment Analysis

News Title

2019



Sentiment for 2019

Positives MORE Neutrals MORE Negatives MORE

Malaysia on track to achieve Sustaina... Dec 28th, 2018 - The Edge Markets	Total supply and use in Malaysia up 5... Dec 26th, 2018 - The Edge Markets	November inflation growth slows mor... Dec 20th, 2018 - The Edge Markets
Ringgit volatility drops on global trad... Dec 12th, 2018 - The Edge Markets	Malaysia's economy likely to grow in ... Dec 24th, 2018 - The Edge Markets	Cover Story: Missing funds will have l... Sep 13th, 2018 - The Edge Markets
FDI figures based on approved invest... Dec 10th, 2018 - The Edge Markets	Framework for sustainable retirement ... Dec 23rd, 2018 - The Edge Markets	Concern over rising inequality in Sabah Sep 6th, 2018 - The Edge Markets
Malaysia's economic growth to ease i... Nov 22nd, 2018 - The Edge Markets	Malaysia Nov CPI up 0.2% on-year Dec 19th, 2018 - The Edge Markets	Inflation to rise to 1% in July, says RAM Aug 23rd, 2018 - The Edge Markets
CPI eases to 0.5% in Q3 Nov 16th, 2018 - The Edge Markets	The Week Ahead: November CPI, cent... Dec 17th, 2018 - The Edge Markets	The Week Ahead: Keeping an eye on ... Aug 20th, 2018 - The Edge Markets
The Week Ahead: Post-budget rally, in... Nov 5th, 2018 - The Edge Markets	Malaysia October IPI up 4.2% on-year Dec 12th, 2018 - The Edge Markets	Cover Story: Sizeable land income for... Aug 16th, 2018 - The Edge Markets
My Say: Increase wages, not retireme...	Malaysia Oct manufacturing sales up ...	Malaysia's IPI grows at a slower rate o...

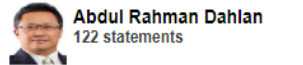
Media Share & Sentiment Analysis - Influencers

Influencers - Population1 | index | sentiments |

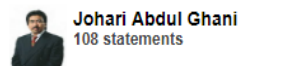
News | Twitter | Facebook



Mohd Uzir Mahidin
301 statements



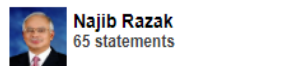
Abdul Rahman Dahlan
122 statements



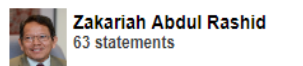
Johari Abdul Ghani
108 statements



Foo Gee Jen
74 statements



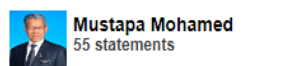
Najib Razak
65 statements



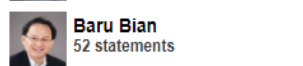
Zakariah Abdul Rashid
63 statements



Azizah Ismail
58 statements



Mustapa Mohamed
55 statements



Baru Bian
52 statements



Yeah Kim Leng
51 statements

368 Influencers

Prev Next

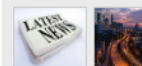
Mentioned in News



Malaysia on track to achieve Sustainable...

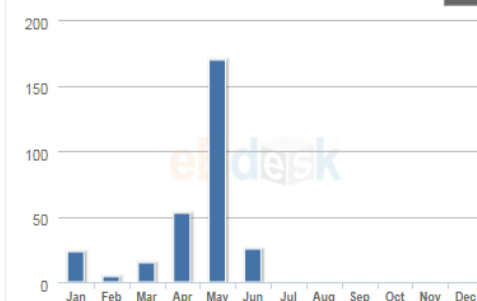
Dec 28th, 2018 - The Edge Markets

Malaysia is on the right track in reaching the Sustainable Development Goals (SDGs) Indicators 2030, said the Department of Statistics Malaysia. In a statement today, it said the government's commitment to the sustainable development agenda is transpired through the Mid-Term Review of the 11th Malaysia Plan (11th MP) presented on Oct 18. In this regard, the commitment to Agenda 2030 for Sustainable Development has been aligned with the strategies and initiatives of the 11th Malaysia Plan....

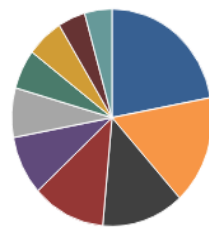


Statements

2019



Media Share | Pie | Bar |



Statements

MORE

SDGs indicators cover 17 goals, 169 targets and 244 indicators to measure the achievement of universal development which comprise social, economic and...

Mohd Uzir Mahidin
Dec 28th, 2018 - The Edge Markets

"Out of the 244 indicators, 118 indicators (48%) are Available, 55 indicators (23%) Partially Available, 36 indicators (15%) Partially Available Need Further...

Mohd Uzir Mahidin
Dec 28th, 2018 - The Edge Markets

"SDGs with a theme Leaving No One Behind focuses on the development of 5Ps namely People, Planet, Prosperity, Peace and Partnership...

Influencers - Population1 | index | sentiments |

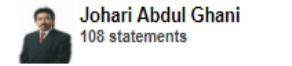
News | Twitter | Facebook



Mohd Uzir Mahidin
301 statements



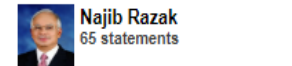
Abdul Rahman Dahlan
122 statements



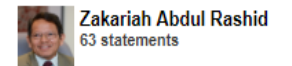
Johari Abdul Ghani
108 statements



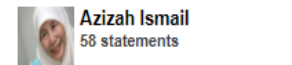
Foo Gee Jen
74 statements



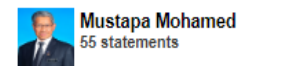
Najib Razak
65 statements



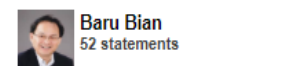
Zakariah Abdul Rashid
63 statements



Azizah Ismail
58 statements



Mustapa Mohamed
55 statements



Baru Bian
52 statements



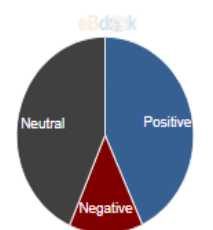
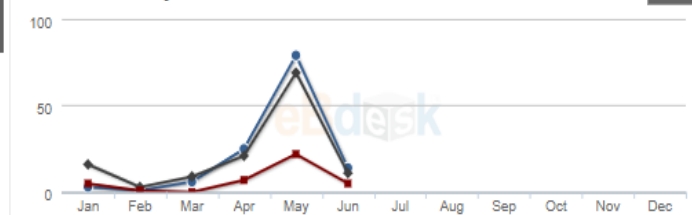
Yeah Kim Leng
51 statements

368 Influencers

Prev Next

Sentiment Analysis

2019



Sentiment for 2019

Positives

MORE

Neutrals

MORE

Negatives

MORE

Mohd Uzir Mahidin
Dec 28th, 2018 - The Edge Markets
SDGs indicators cover 17 goals, 169 targets and 244 indicators to measure the achievement of universal development which comprise social, economic and environmental aspects.

Mohd Uzir Mahidin
Dec 12th, 2018 - The Edge Markets
the total number of employees engaged in the manufacturing sector in October 2018 was 1.076 million persons, an increase of 2.2% from 1.053 million persons in October 2017.

Mohd Uzir Mahidin
Dec 28th, 2018 - The Edge Markets
"Out of the 244 indicators, 118 indicators (48%) are Available, 55 indicators (23%) Partially Available, 36 indicators (15%) Partially Available Need Further Development, 23..."

Mohd Uzir Mahidin
Dec 28th, 2018 - The Edge Markets
"SDGs with a theme Leaving No One Behind focuses on the development of 5Ps namely People, Planet, Prosperity, Peace and Partnership. Thus, the availability of these SDGs..."

Mohd Uzir Mahidin
Nov 22nd, 2018 - The Malay Mail Online
the annual change of the Leading Index (LI) stated that the index decreased 1.7 per cent in the reference month.

Mohd Uzir Mahidin
Dec 28th, 2018 - MSN Malaysia
"Out of the 244 indicators, 118 indicators (48%) are Available, 55 indicators (23%) Partially Available, 36 indicators (15%) Partially Available Need Further Development, 23..."

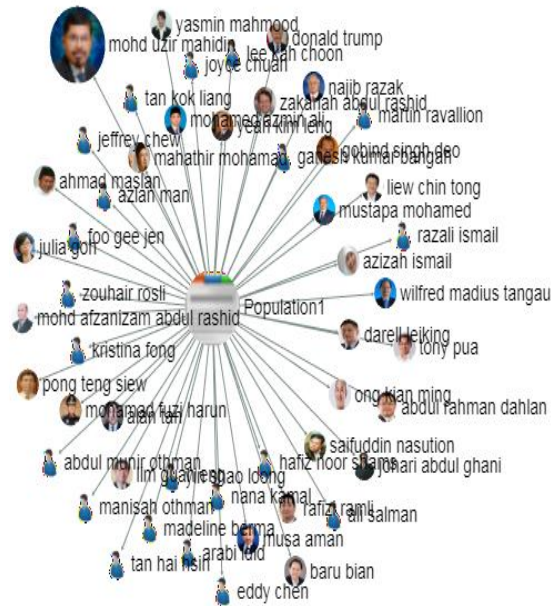
Ontology Analysis & Cluster Issue

Ontology Analysis

ch

50 ▼ Go

Total Node :51
Total Edge :50
Total Topic :1
Total Influencer :50



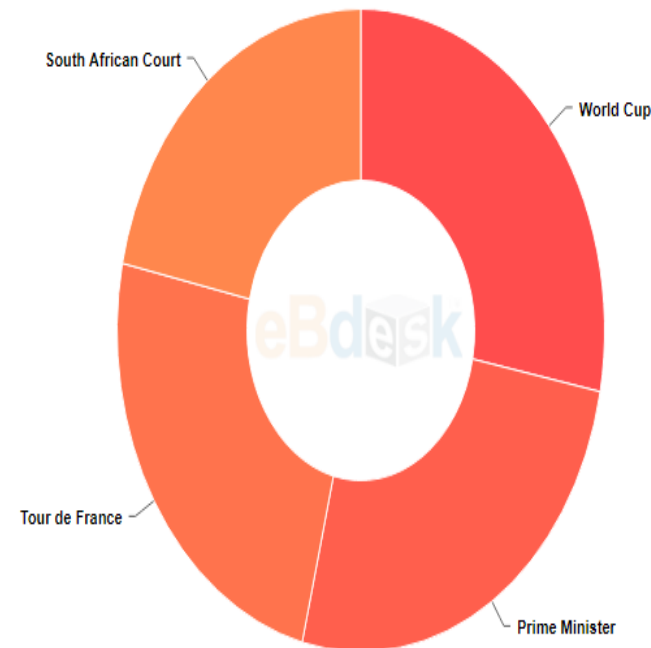
Cluster Issue | actual | timeline |

H DISCOVER ANALYZE ACTION

Hours 1 4 8 16 24 T W M Y

Top issues

All issues



Found 12 Related News :



Neymar admits to 'brat' behaviour as he reflects on World Cup campaign in TV advert

Jul 30th, 2018 - The National

Neymar has used a television advert to attempt to tackle critics who derided his play-acting at the World Cup, admitting he overreacts and blaming his brat like behaviour on his ...



South African court overturns immunity for Grace Mugabe in assault case

Jul 30th, 2018 - The Morning Express

South Africa's High Court on Monday overturned a decision by the government to grant Zimbabwe's former first lady Grace Mugabe diplomatic immunity after she was accused ...

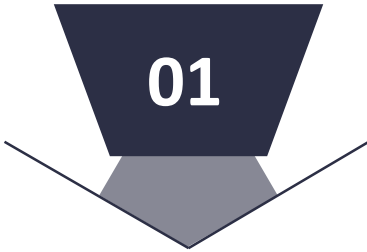


Opposition says Cambodian election was death of democracy

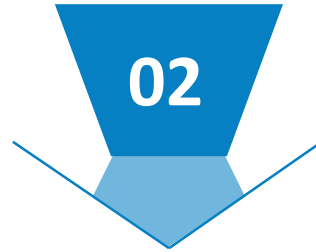
Jul 30th, 2018 - Manila Bulletin

The ruling party of Cambodian Prime Minister Hun Sen congratulated itself Monday on its election victory, while the opposition party unable to contest the polls said they marked the ...

Problem Statement



Requirement to monitor online economic news is in line with official statistics released as well as current national policy implementation;



Increasing trend in using the catchy topics as news headlines for capturing the reader's attention; and



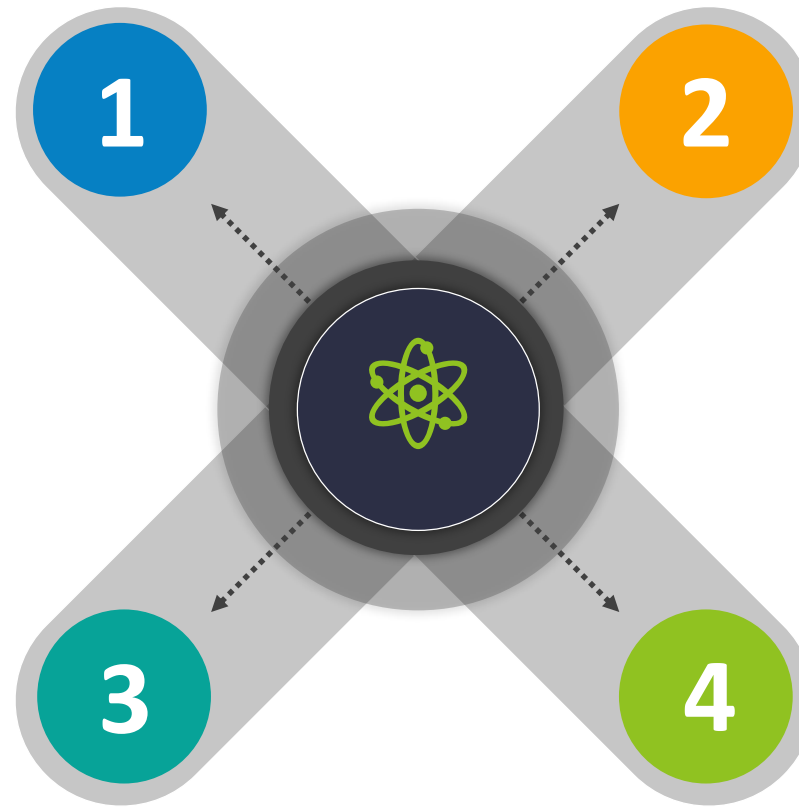
Efforts to attract readers attention is needed for enhance credibility of online news at par with written or printed news.

Generally, this study aims to identify and define the important of key words in layman for public attractions to read the entire economic report.

Objectives

Identify the key words used by media practitioners for online statistical and economic news on **population** issues in Malaysia;

conduct cluster analysis by using lexicon and corpus approaches and develop topics model by using Latent Dirichlet Allocation; and



measure perception developed by media to publics readers towards statistical information released by DOSM through news online (sentiment analysis);

develop list of layman key words as guidelines for publication write up and editorial in Department of Statistics Malaysia.

Scope Of The Research

Data

- Media Intelligence Portal
- Topic: Population1

Time

- 1 Jan 2017 – 31 Dec 2018
- News Online

Media - Population1 | index | sentiments |

News

THE EDGE The Edge Markets
150 news

M MY Stock 118
96 news

theSundaily The Sun Daily
94 news

M The Malay Mail Online
69 news

B Bernama
50 news

msn MSN Malaysia
50 news

THE BORNEO POST The Borneo Post
43 news

FMT Free Malaysia Today
26 news

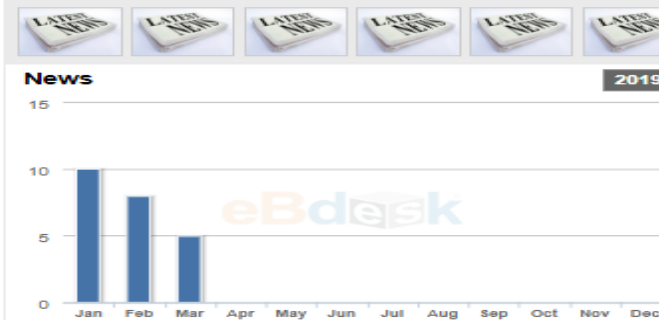
ME Malaysia Digest
25 news

NEW STRAITS TIMES New Strait Times
25 news

71 Media

Mentioned in News

Malaysia on track to achieve Sustainable...
Dec 26th, 2018 - The Edge Markets
Malaysia is on the right track in reaching the Sustainable Development Goals (SDGs) Indicators 2030, said the Department of Statistics Malaysia. In a statement today, it said the governments commitment to the sustainable development agenda is transpired through the Mid-Term Review of the 11th Malaysia Plan (11th MP) presented on Oct 18. In this regard, the commitment to Agenda 2030 for Sustainable Development has been aligned with the strategies and initiatives of the 11th Mala..



Related Influencers

Statements
SDGs indicators cover 17 goals, 1 244 indicators to measure the ach universal development which con economic and...

Mohd Uzir
Dec 28th, 2018 - The Edge

Mohd Uzir
Dec 28th, 2018 - The Edge

Statements

SDGs indicators cover 17 goals, 1 244 indicators to measure the ach universal development which con economic and...

Mohd Uzir
Dec 28th, 2018 - The Edge

Mohd Uzir
Dec 28th, 2018 - The Edge

"SDGs with a theme Leaving No One Behind focuses on the development of 5Ps namely

H DISCOVER ANALYZE ACTION

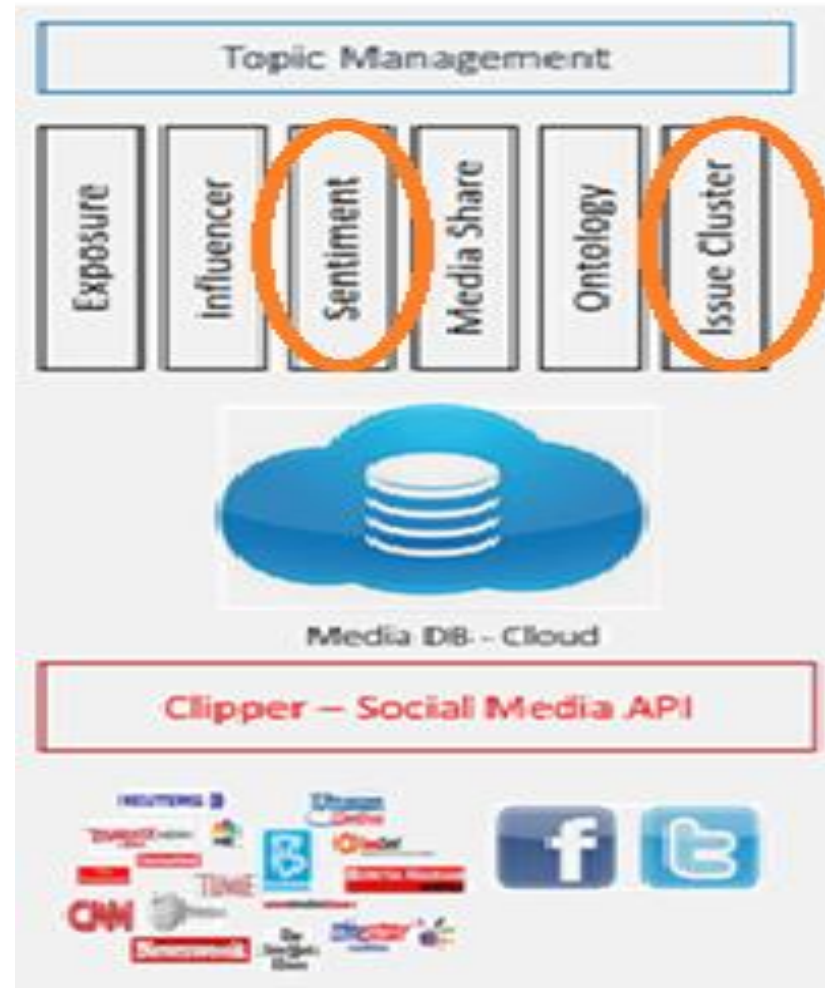
H DISCOVER ANALYZE ACTION

ADVANCED TIME FRAME

ALL 2017-01-01 to

Dec 2018						
Su	Mo	Tu	We	Th	Fr	Sa
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

Scope Of The Research vs PMAOS RTOS





The Significant Of The Research

Strengthening the Role as a **THRUST**
Producer of Statistics **STRATEGIC 4**

STRATEGY 1

Strengthening the Statistical Awareness, Understanding and
Usage in Evidence-Based Decision Making

Programme
2

Strengthening Statistical Communication

1. Strengthening media communications plan through:
 - a) expanding media partnerships;
 - b) creating special statistical column in the media;
 - c) increasing participation in journalistic writing;
 - d) highlighting the statistical products in the premier media;
 - e) developing a team with expertise in statistical writing in the field of economy, social, demographic and environment; and
 - f) creating a talk show on current issues "Statistics and You".
2. Using portal as the main medium of dissemination of statistical information with a user-friendly concept

1

to produce a list of layman words that can be used for readers better understanding;

2



to assist the Public Relations Officer in DOSM to share the list with media practitioners to avoid the use of unclear, indescribable for technical terms or negative tone which will cause confusion among news online readers; and

3



to assist respective officer or editorial team as guidelines in any write-up or report to be published by DOSM.

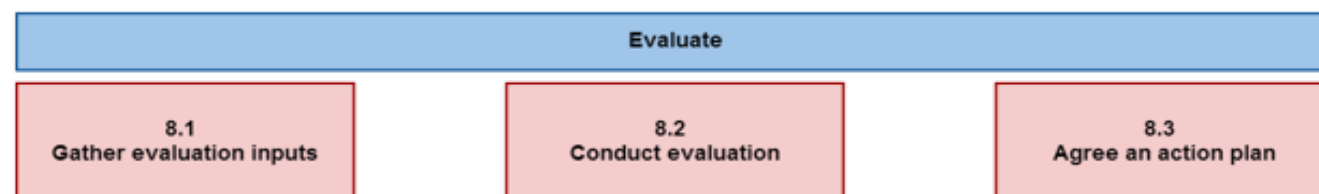


The Significant Of The Research



The Generic Statistical Business Process Model

Evaluate Phase



Eleventh Malaysia Plan 2016-2020

UN NQAF Principles

Level A: Managing the national statistical system (NSS)

- 1: Coordinating the NSS
- 2: Managing relationships with data users, data providers and other stakeholders
- 3: Managing statistical standards

Level B: Managing the institutional environment

- 4: Assuring professional independence
- 5: Assuring impartiality and objectivity
- 6: Assuring transparency
- 7: Assuring statistical confidentiality and data security



2. Literature Review



**Author: Ordenes
el al. (2014)**

**Title: Analyzing
Customer
Experience
Feedback Using
Text Mining: A
Linguistics-
based Approach.**

Findings: Text mining is the process of analyzing textual information in an attempt to discover structure and implicit meanings “hidden” within text.



**Author: Nanli et
al. (2012)**

**Title: Sentiment
Analysis: A
Literature
Review**

Findings: An analysis of sentiments or opinion aiming to investigate, analyse, extract subjective texts that involve user opinion, preferences and sentiments.



**Author:
Zamahsyari and
Nurwidyantoro
(2016)**

**Title: Sentiment
Analysis of
Economic News
in Bahasa
Indonesia Using
Majority Vote
Classifier**

Findings: Online news has the ability to influence public readers on the opinion of an article.



**Author:
Allahyari et al.
(2017)**

**Title: A Brief
Survey pf Text
Mining:
Classification,
Clustering and
Extraction
Techniques**

Findings: Clustering is the task of finding groups of similar documents in a collection of documents.

Literature Review



**Author: Yatim et al.
(2016)**

**Title: A Corpus-
Based Lexicon
Building in
Indonesian
Political Context
Through
Indonesian Online
News Media.**

Findings: The sentimental lexicon is a vocabulary that has a polarity whether the text is positive or negative



**Author: Kamps et al.
(2004)**

**Title: Using
WordNet to
Measure Semantic
Orientations of
Adjectives.**

Findings: Lexicon-based analysis approach is a semantic-oriented that is a non-supervisory approach in sentiment analysis.



**Author: Blei et al.
(2003)**

**Title: Latent
Dirichlet Allocation**

Findings: LDA is a generative probabilistic model for collections of discrete data such as text corpora. LDA is a three-level hierarchical Bayesian model.



Author: Liu (2013)

**Title: High
Performance
Latent Dirichlet
Allocation for Text
Mining**

Findings: LDA model is one of the most popular probability topic models and it has more comprehensive assumptions of text generation than other models.

3. Methodology

Tools

- R
- Package:
- tm, NLP, wordcloud, wordcloud2, ggplot2, tidyverse, tidytext, topicmodels, cluster, NbClust, RColorBrewer

Statistical Analysis

- Text Mining
- Cluster Analysis
- Topic Model – LDA
- Sentiment Analysis

Result

- Structured data
- Frequency of words
- Group of similar contents (distance)
- Suitable topic
- Positive and negative words

4. Data Collection



News

Twitter

Facebook

Keyword By

Query

Query Keyword

"Population" or "death" or "birth" or
"migration" or "life expectancy" or "penduduk"
or "kelahiran" or "kematian" or "jangkaan
hayat" or "kesuburan" or "gender" or "Orang

Media Tag

OR

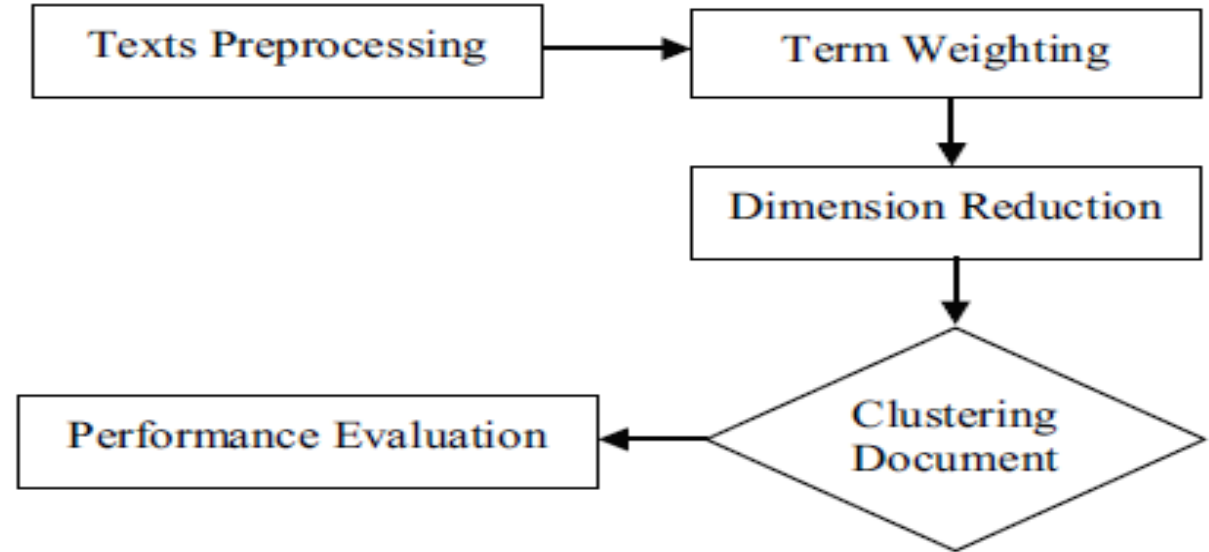
Add

News on Malay Language are excluded (6 news from blog i.e. Blog Kamaluddins, The Patriots, Generasi Pejuang Bangsa

No.	Folder	Media	Total News	Grand Total	No.	Folder	Media	Total News	Grand Total
1	Bernama	Bernama	50	130	20	Others	ABC	1	86
		Malay Mail	69				ABS CBN	1	
		Mtimes	11				Aliran	3	
2	Borneo	Borneo	43	Aninetto			1		
		Sabahtime	8	Asia One			3		
3	Business Time	Business Time	18	Astro Awani			1		
4	Chronicle	Chronicle	19	Bangkok Post			1		
5	Daily n Digital	Daily Express	9	Biz Insider			1		
		Digital News Asia	5	Biz Today			1		
6	Focus	Focus	18	Blogs Wsj			1		
7	Free M	Free M	26	Çhannel			6		
8	M Digest	M Digest	25	CoverBureau			4		
9	M Insight	M Insight	23	Express			1		
10	MSN	MSN	50	Ipoh Echo			2		
11	MyStock 118	MyStock 118	96	Jakarta Post			1		
12	NST	NST	25	Khmertimeskh			1		
13	Others Mainstream	Cilisos	6	38			KTemockConsiders	1	
		Dap	6				Lim Sian See	2	
		Din Merican	5				LKS Blog	1	
		kini	14				M Reserve	9	
		M MustKnow	1				Mole	10	
		M Today	4				My Sinchew	6	
		People's Parliament	1				News Xin	6	
		Umno	1				Reuters	3	
14	Singapore	Btimes SG	6	32			Says M	4	
		S BizReview	4				Srwk Voice	1	
		Yahoo S	22				Stop	1	
15	The Edge	The Edge	149	149			Strait Times	6	
16	The Star	The Star	12	12			Taiping	1	
17	The Sun Daily	The Sun Daily	92	92			Tech In Asia	1	
18	Vietnam	V Plus	5	8			Today Online	4	
		VietNam Net	3				Xinmsn	1	
19	Yahoo M	Yahoo M	20	20			TOTAL NEWS		932

5. Statistical Analysis – Text Mining

Kadhim et al. (2014) proposed system approach that can be used for this study. The proposed system is divided into four main stages such as: preprocessing stage, term weighting stage, dimensional reduction and performance evaluation as shown in Figure



- 01 upload data as corpus
- 02 start pre-processing
- 03 remove stopword by use standard stopword list and build-in
- 04 format vector
- 05 create document-term matrix

- 06 convert dtm to matrix
- 07 collapse matrix by summing over columns
- 08 inspect most frequently occurring terms
- 09 change to data frame format

5. Statistical Analysis – Cluster Analysis

Aims to grouping the text based on distance between document vectors, evaluated on the similarity by using the distance concept



5. Statistical Analysis - Sentiment Analysis

The main two methods of sentiment analysis are lexicon-based method and machine learning based approach. The lexicon-based approach, uses sentiment lexicons - dictionaries of words with labels specifying their sentiments - to identify the sentiment of text



This research is based on lexicon-based method with aims to classify the term either positive or negative and determined the sentiment polarity score at word level analysis based on the formula used by Kolchyna et al. (2015) as stated below:



Type of sentiment lexicon

```
> get_sentiments("afinn")
# A tibble: 2,476 x 2
  word      score
  <chr>    <int>
1 abandon     -2
2 abandoned   -2
3 abandons    -2
4 abducted    -2
5 abduction   -2
6 abductions  -2
7 abhor       -3
8 abhorred    -3
9 abhorrent   -3
10 abhors     -3
# ... with 2,466 more rows
```

```
> get_sentiments("bing")
# A tibble: 6,788 x 2
  word      sentiment
  <chr>    <chr>
1 2-faced negative
2 2-faces negative
3 a+      positive
4 abnormal negative
5 abolish negative
6 abominable negative
7 abominably negative
8 abominate negative
9 abomination negative
10 abort  negative
# ... with 6,778 more rows
```

```
> get_sentiments("nrc")
# A tibble: 13,901 x 2
  word      sentiment
  <chr>    <chr>
1 abacus   trust
2 abandon  fear
3 abandon  negative
4 abandon  sadness
5 abandoned anger
6 abandoned fear
7 abandoned negative
8 abandoned sadness
9 abandonment anger
10 abandonment fear
# ... with 13,891 more rows
```

The AFINN lexicon categorizes words using a score, with negative scores indicating a negative sentiment. The Bing lexicon categorizes words into a positive or negative flag ; the nrc lexicon categorizes words into emotions like anger, disgust, fear, joy and others.

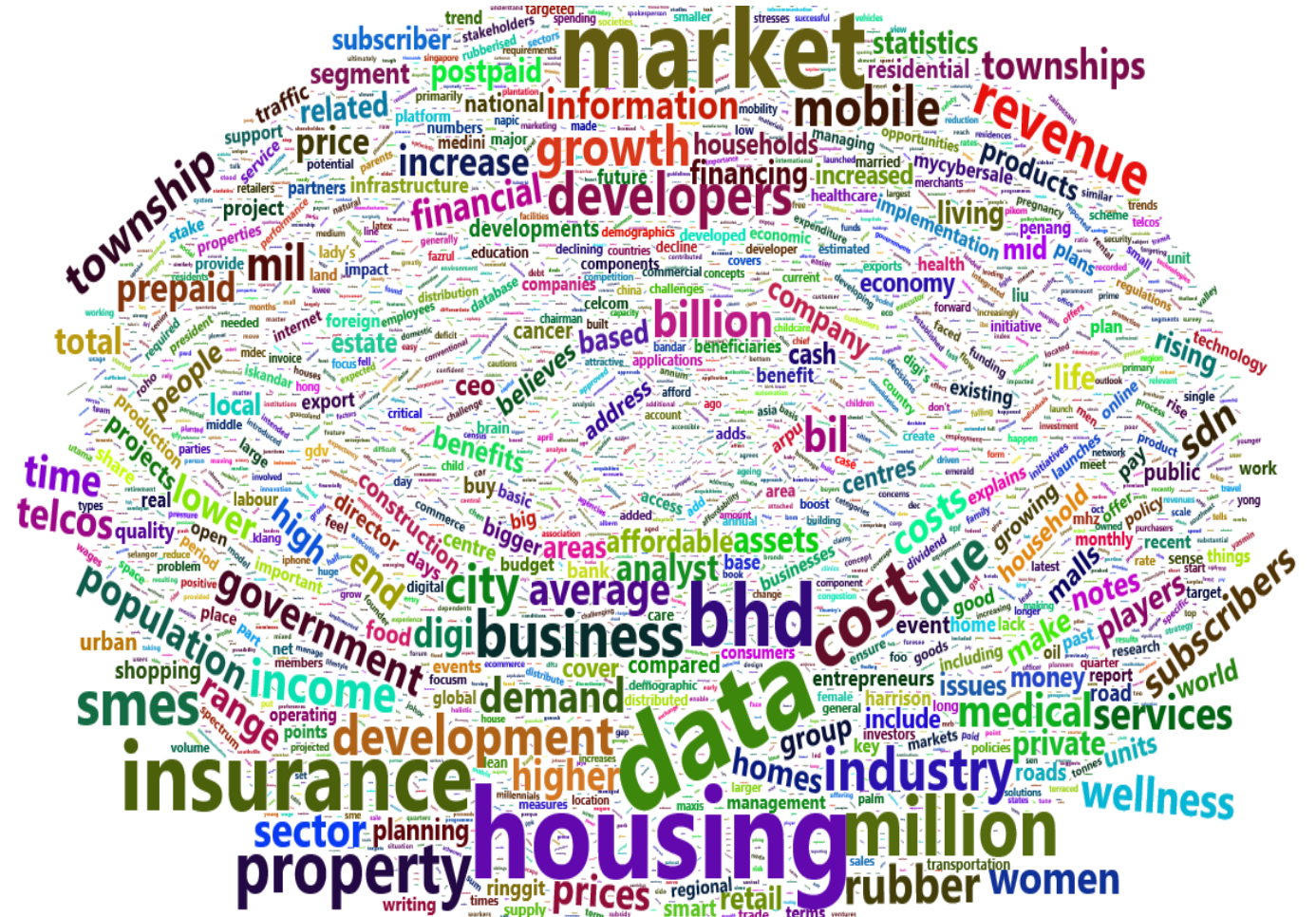
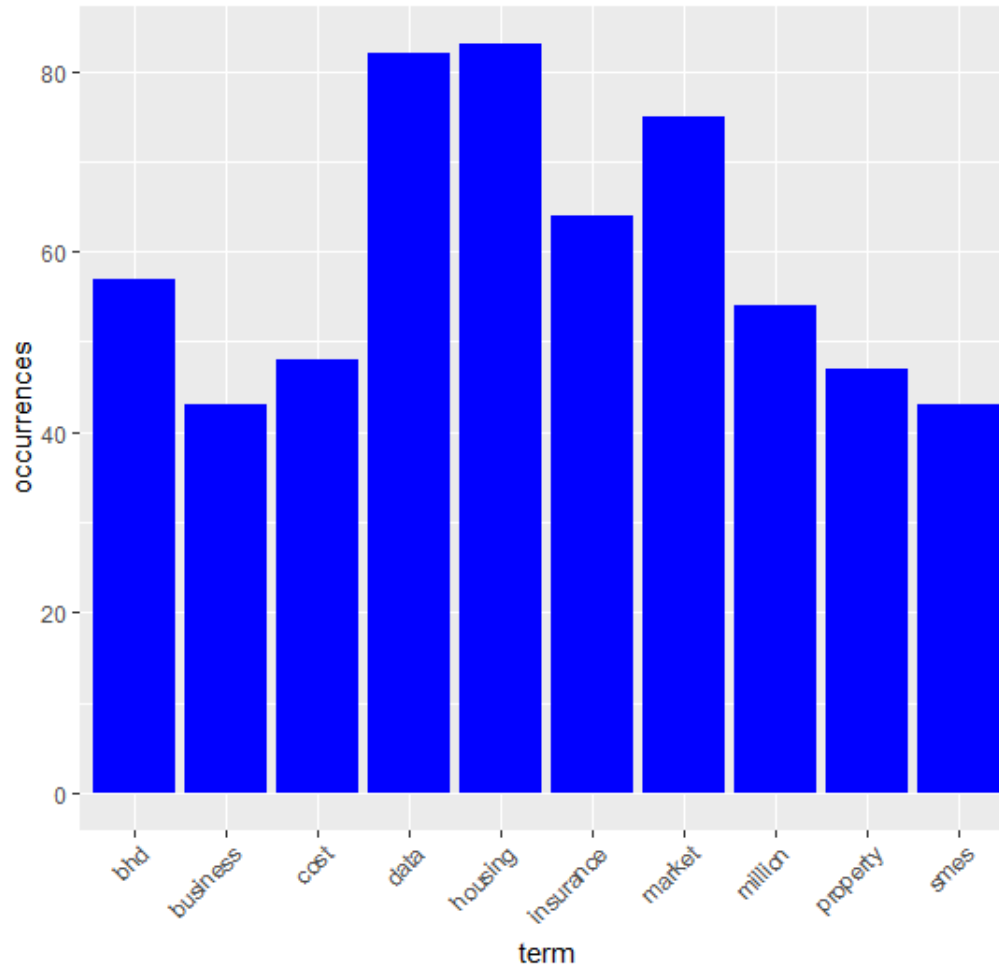
01

02

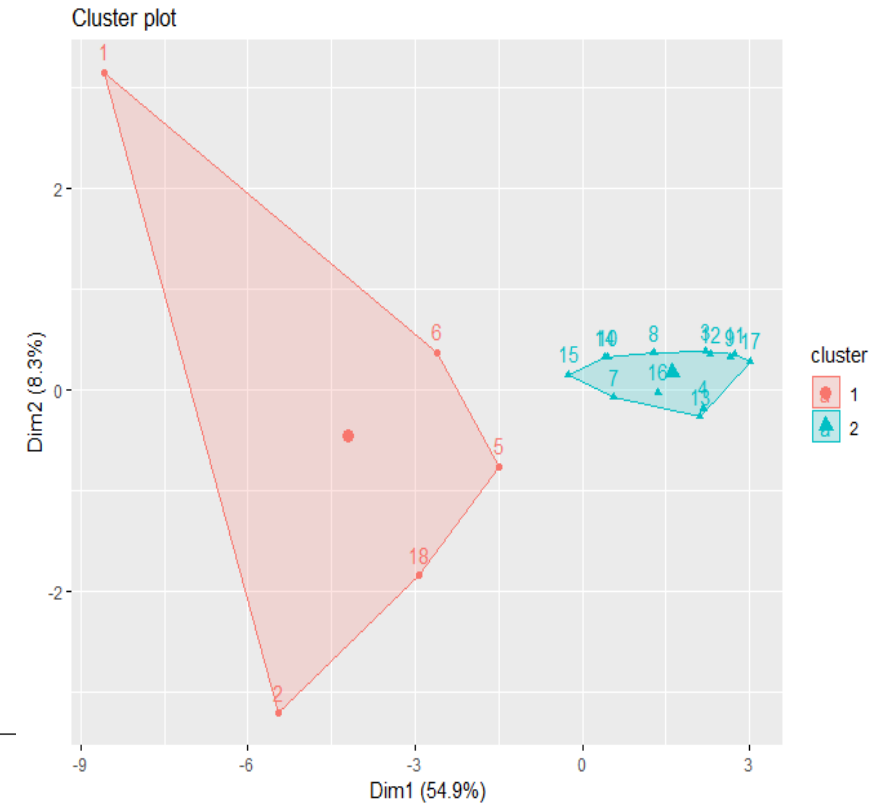
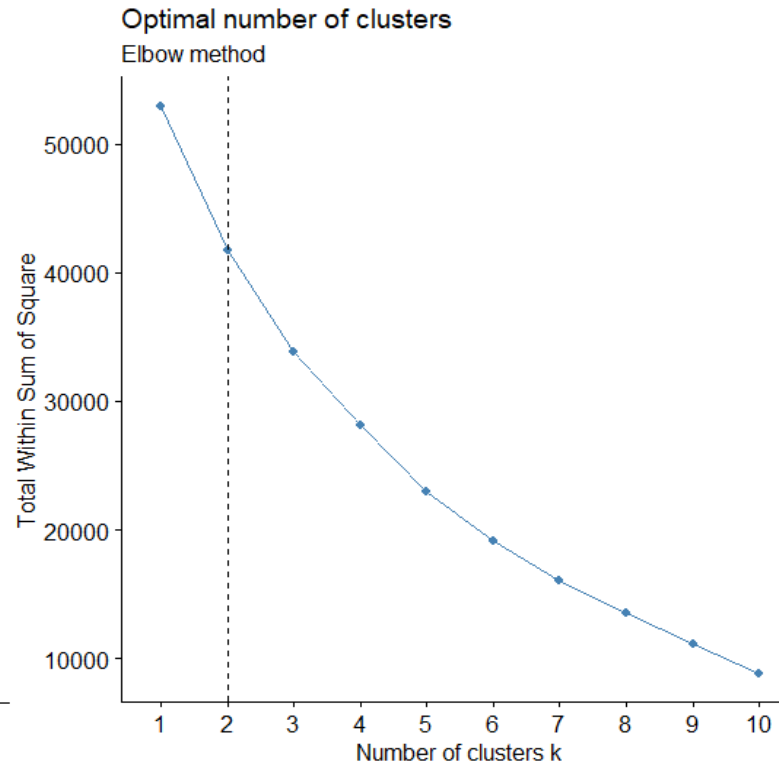
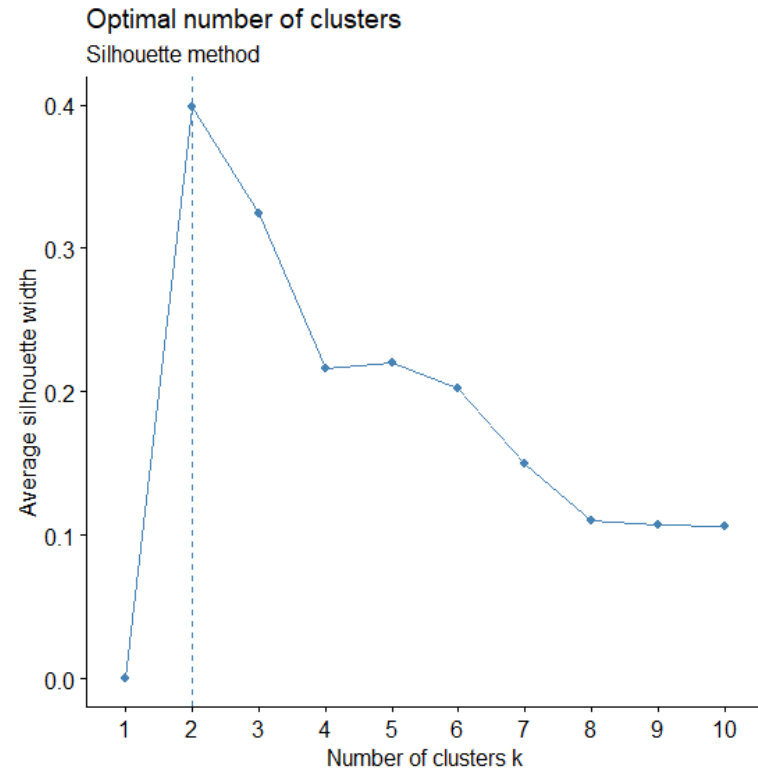
Bing and NRC have more negative than positive words, but the ratio of negative to positive words is higher in the Bing lexicon than the NRC lexicon. This will contribute to the effect in the plot.

6. Result - Frequency of Words

Chart on Frequency of Terms



6. Result – Cluster Analysis



$$\text{Idf}(\text{term}) = \ln \frac{\text{ndocuments}}{\text{ndocuments containing term}}$$

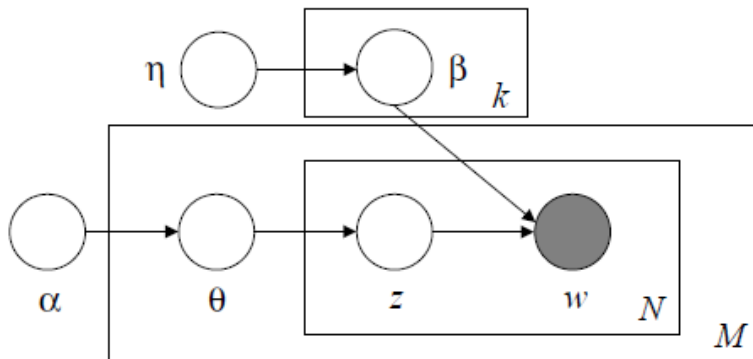
6. Result - Latent Dirichlet Allocation (LDA)

LDA

LDA is a method for finding the best topic in a group of words from a collection of documents.

LDA

LDA is a generative probabilistic model for collections of discrete data such as text corpora and a three-level hierarchical Bayesian model.



Graphical Model Representation of the Smoothed LDA Model.

Document	Cluster	Probabilities	Document	Cluster	Probabilities
1	1	0.866	10	1	0.690
2	2	0.865	11	1	0.696
3	1	0.639	12	1	0.669
4	2	0.790	13	2	0.777
5	2	0.857	14	2	0.722
6	1	0.885	15	1	0.850
7	2	0.852	16	2	0.676
8	1	0.747	17	1	0.695
9	1	0.737	18	2	0.885

6. Result - Topic Model (The Edge)

No.	Topic 1 (Economy)	Topic 2 (Social)	Topic 3 (General)
1	growth	market	income
2	bhd	property	tax
3	billion	retirement	labour
4	prices	housing	workers
5	exports	price	government
6	statistics	financial	higher
7	manufacturing	affordable	economic
8	data	time	low
9	trade	population	foreign
10	expected	development	high
11	index	high	cost
12	products	house	household
13	inflation	age	wage
14	million	klang	living
15	global	land	median
16	oil	work	average
17	compared	developers	economy
18	sector	government	report
19	week	public	people
20	gdp	term	wages
21	due	long	lower
22	quarter	units	monthly
23	bank	people	state
24	rate	adds	rate
25	sales	good	jobs
26	increase	working	gst
27	research	home	revenue
28	export	prices	growth
29	increased	million	households
30	production	demand	country

#top 30 terms in each topic

6. Result -Sentiment Analysis

Folder	Positive	Negative
Bernama	leading	gross
	work	decline
	improved	opposition
	good	crude
	support	issues
	strong	issue
	savings	unemployed
	top	negative
	free	tell
	worth	concerns



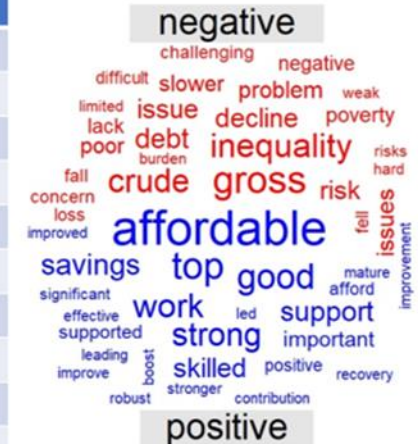
Folder	Positive	Negative
Focus	affordable	issues
	benefits	cancer
	smart	lack
	good	problem
	benefit	decline
	work	critical
	support	declining
	important	stresses
	lean	fell
	boost	gross



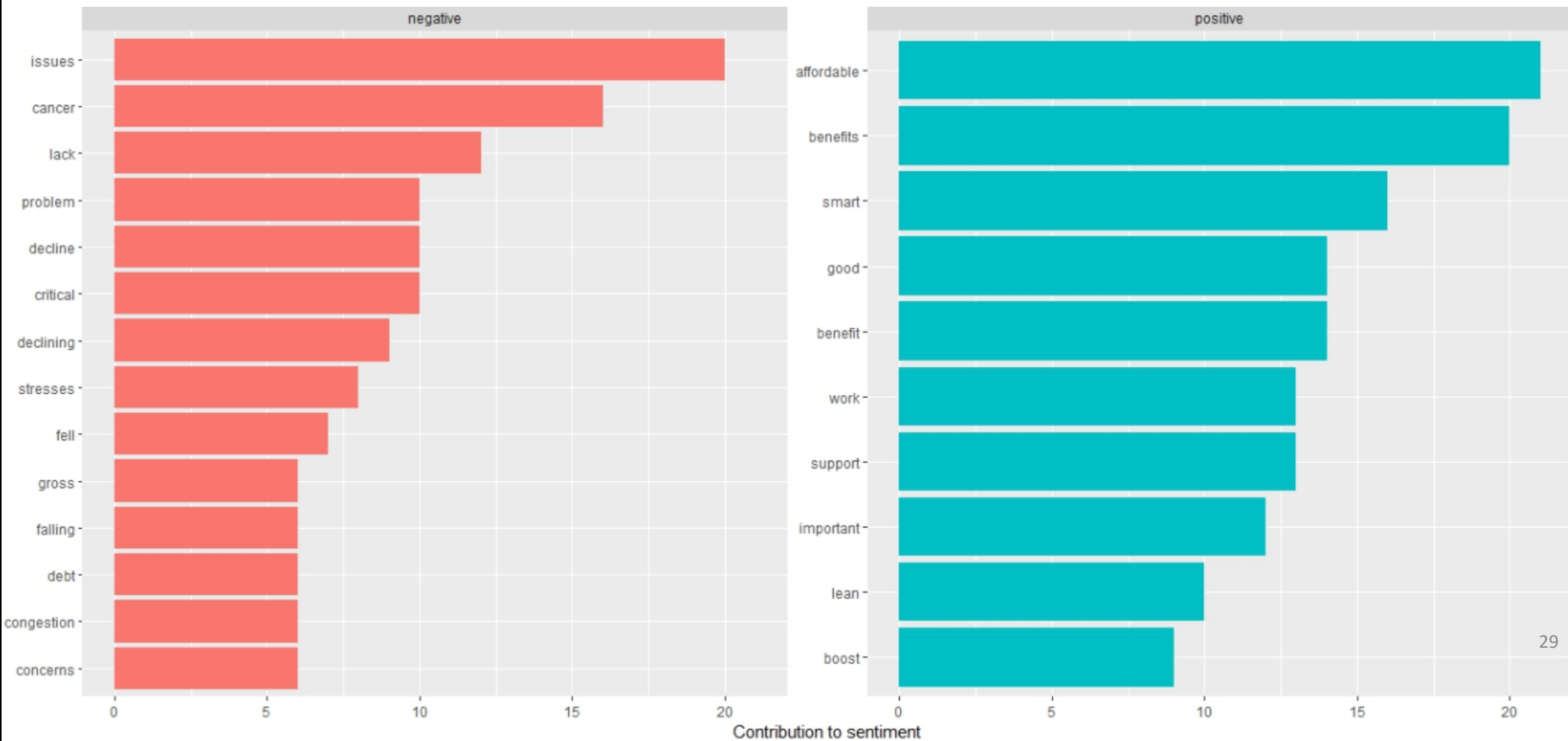
Folder	Positive	Negative
Borneo	positive	poverty
	top	negative
	tender	poor
	strong	lack
	support	issue
	stable	decline
	work	risk
	free	gross
	benefit	debt
	improve	crude



Folder	Positive	Negative
The Edge	affordable	gross
	top	inequality
	good	crude
	work	debt
	strong	decline
	savings	issue
	support	risk
	skilled	problem
	important	issues
	supported	slower

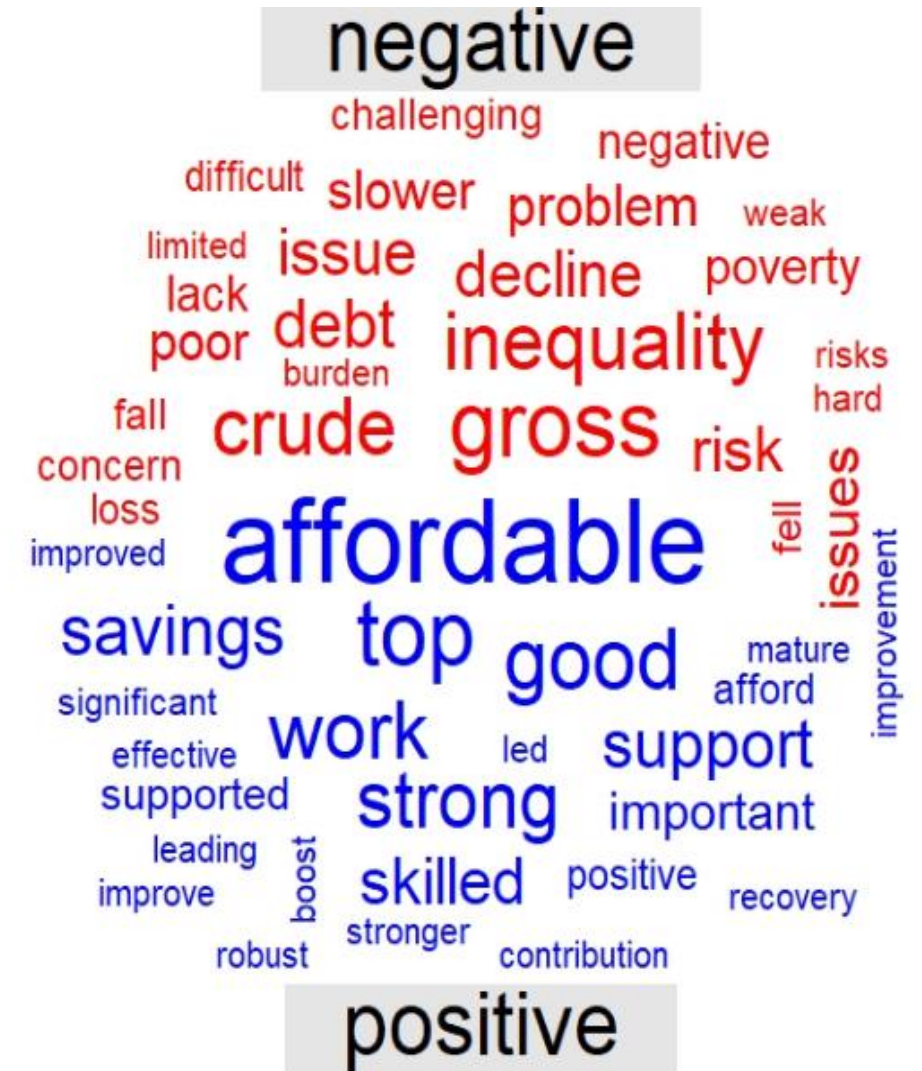
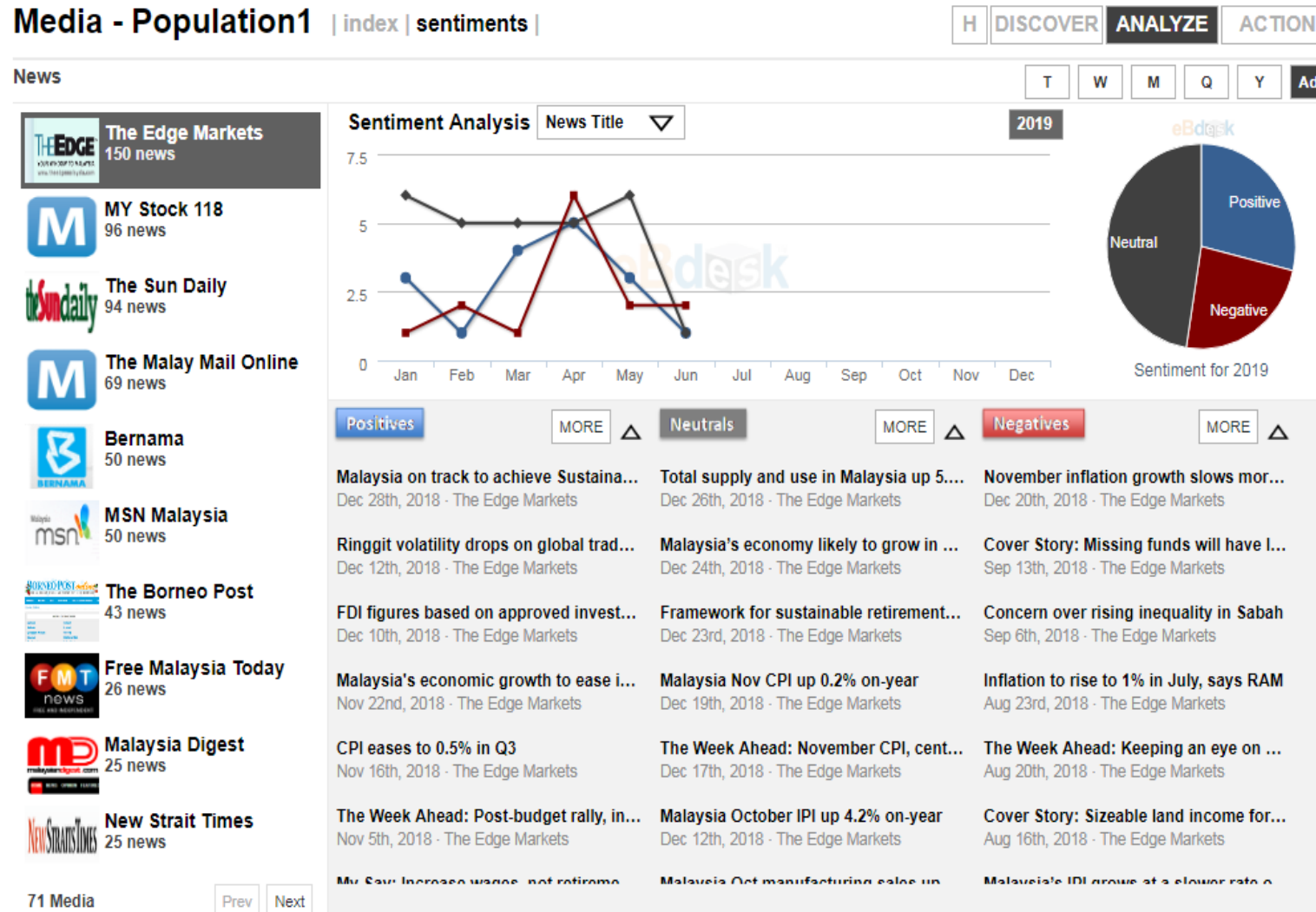


6. Result -Sentiment Analysis:
most common positive & negative words (Focus)



7. Conclusion

Analysis Comparison with STATSBDA (PMAOS and RTOS)



7. Conclusion

Analysis Comparison with STATS BDA (PMAOS and RTOS)

Media - Population1

| index | sentiments |

H DISCOVER ANALYZE ACTION

News

The Edge Markets
150 news

MY Stock 118
96 news

The Sun Daily
94 news

The Malay Mail Online
69 news

Bernama
50 news

MSN Malaysia
50 news

The Borneo Post
43 news

Free Malaysia Today
26 news

Malaysia Digest
25 news

New Strait Times
25 news

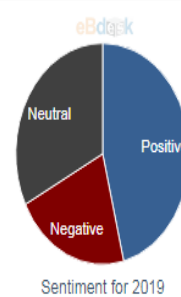
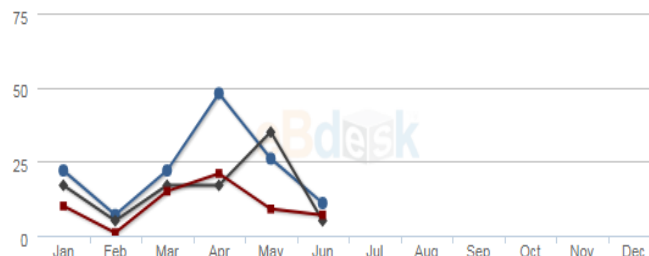
71 Media

Prev Next

Sentiment Analysis

Statements

2019



Positives

MORE

Neutrals

MORE

Negatives

MORE

Mohd Uzir Mahidin
Dec 28th, 2018 - The Edge Markets
SDGs indicators cover 17 goals, 169 targets and 244 indicators to measure the achievement of universal development which comprise social, economic and environmental aspects.

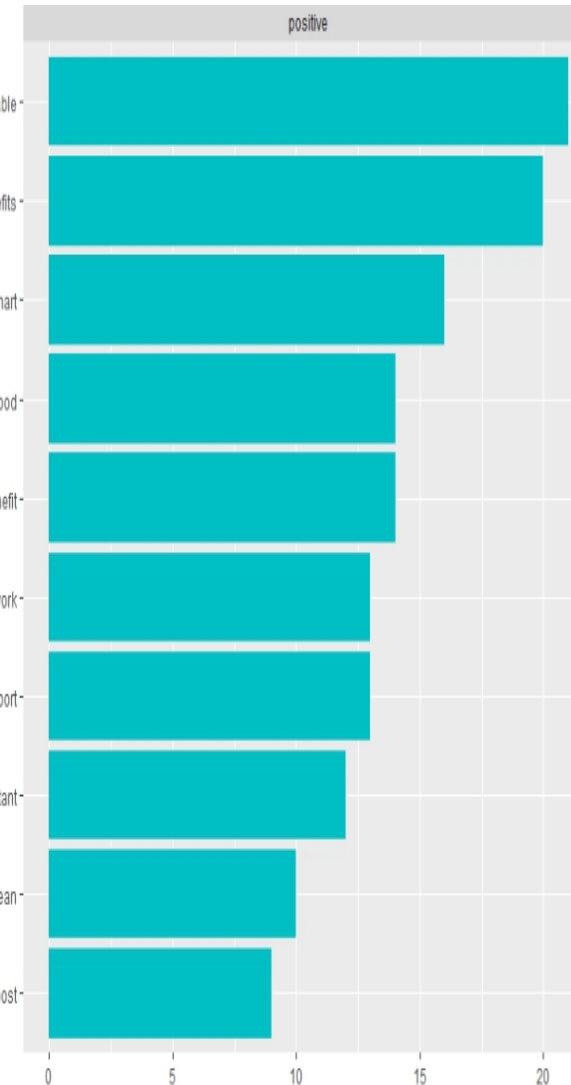
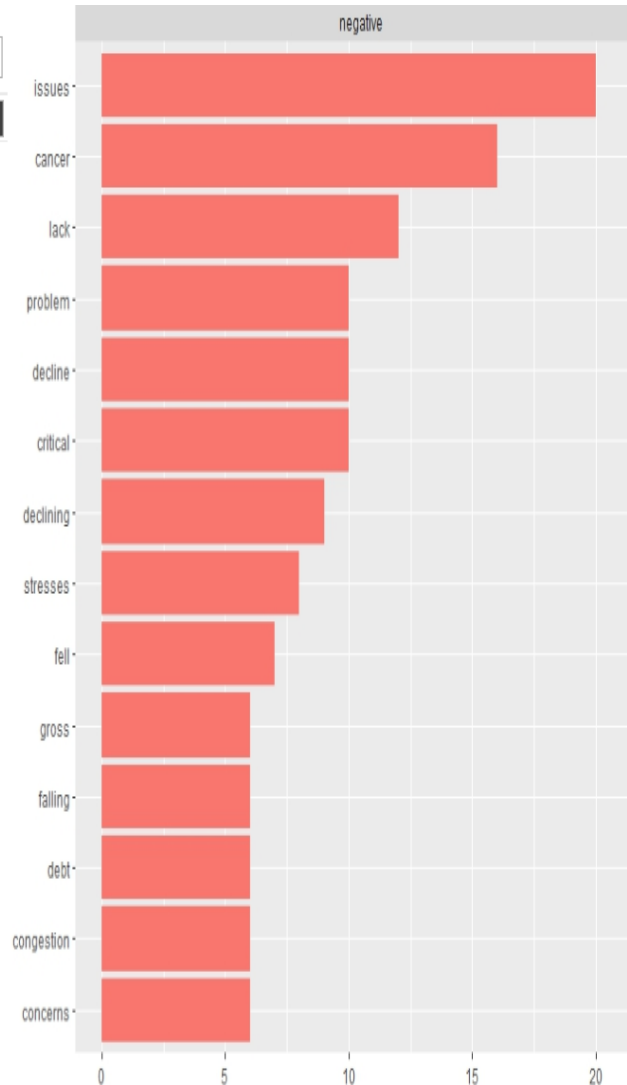
Mohd Uzir Mahidin
Dec 12th, 2018 - The Edge Markets
the total number of employees engaged in the manufacturing sector in October 2018 was 1.076 million persons, an increase of 2.2% from 1.053 million persons in October 2017.

Mohd Uzir Mahidin
Dec 28th, 2018 - The Edge Markets
"Out of the 244 indicators, 118 indicators (48%) are Available, 55 indicators (23%) Partially Available, 36 indicators (15%) Partially Available Need Further Development, 23..."

Mohd Uzir Mahidin
Dec 28th, 2018 - The Edge Markets
"SDGs with a theme Leaving No One Behind focuses on the development of 5Ps namely People, Planet, Prosperity, Peace and Partnership. Thus, the availability of these SDGs..."

Lim Guan Eng
Dec 10th, 2018 - The Edge Markets
FDIs in the manufacturing sector from May to September this year hit RM35 billion compared to RM7.3 billion in the same period of 2017, up 379%.

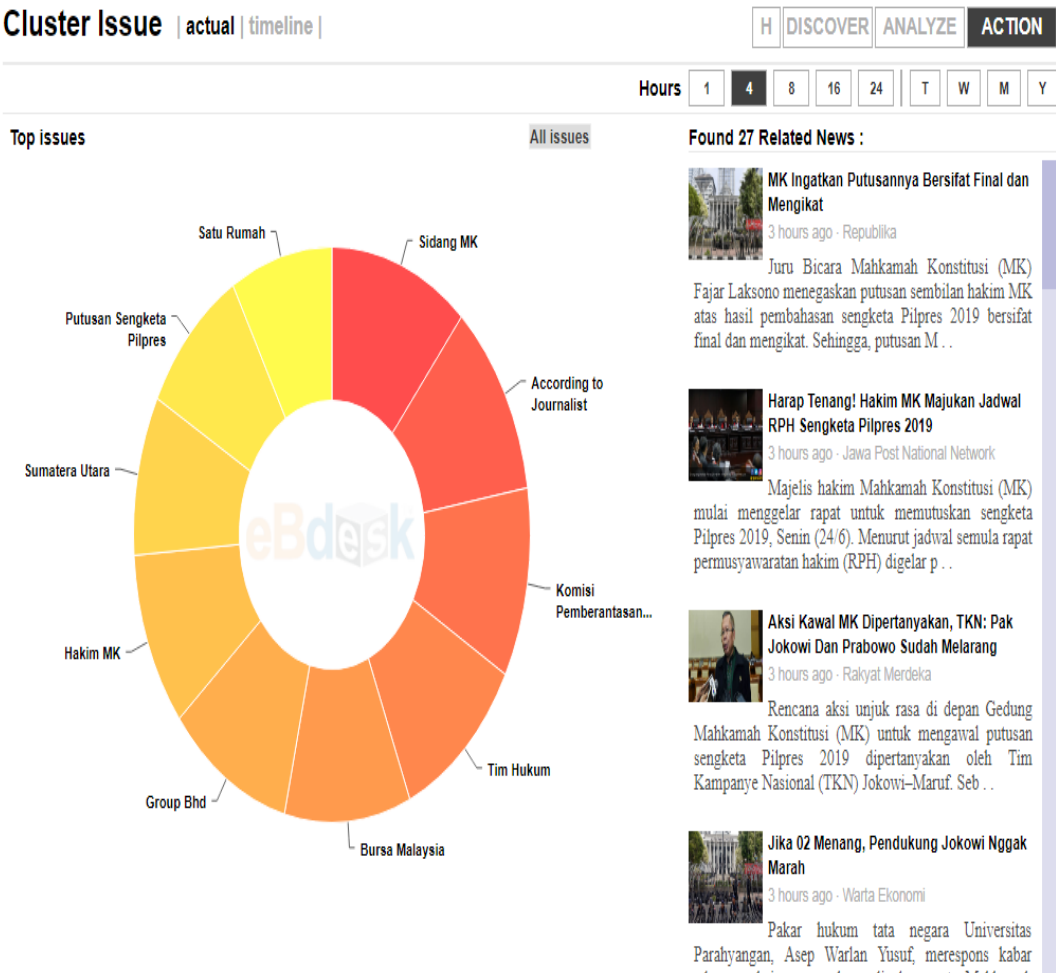
Mohd Uzir Mahidin
Dec 12th, 2018 - The Edge Markets
the growth registered in sales value in October 2018 was due to the increase in transport equipment and other manufactured products (13.3%), electrical and electronics products ...



Contribution to sentiment

7. Conclusion

Analysis Comparison with STATSBD (PMAOS and RTOS)

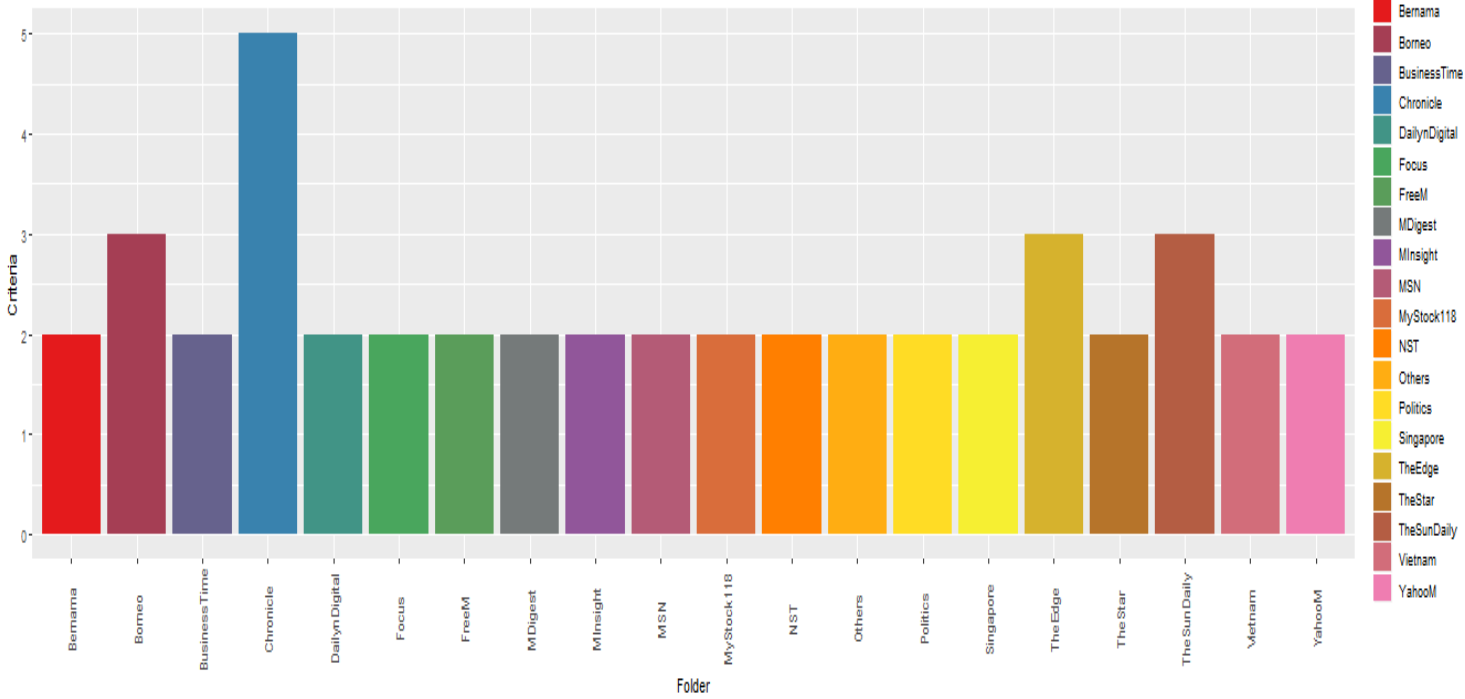


Folder Population

Select Input:

Cluster

Bar chart for Cluster Population.



Analysis Comparison with STATSBD A (PMAOS and RTOS)

Top Issues - Population1

| actual | history | comparison |

News | Twitter | Facebook

T W M Q Y Adv

Top issues

All issues

Issue	Color
Senior Citizens	Yellow
Know about...	Red
Household Income	Orange-red
Chinese Vote...	Orange
Cost of Living	Light orange
Malaysia s...	Pale yellow
Cover Story	Very light yellow
Penang Population	White
New AG...	White
Statistics Dept...	White
eBdesk	White

Found 39 Related News :

Who is plucking Penang population figures from the air

Nov 27th, 2018 - Anilnetto

This statement has just been released by Penang Forum: Penang Forum would like to respond to Penang Chief Minister Chow Kon Yeow's statement that the Penang population, as projected...

Who is plucking Penang population figures from the air

Nov 27th, 2018 - Aliran

Penang Forum would like to respond to Penang Chief Minister Chow Kon Yeow's statement that the Penang population, as projected by the project delivery partner for the RM46bn transport...

Cover Story: Factors of locality must not be overlooked

Aug 23rd, 2018 - The Edge Markets

Affordable housing has been talked about for many years now but ask anyone about it and it is likely he or she will not be too clear about what that really means. What actually is an ...

Empowering senior citizens to be independent, live with dignity

Aug 2nd, 2018 - The Borneo Post

Malaysia's senior citizen demography is trending upwards, with the percentage of older persons nearing almost 20 per cent of the population by 2040, according to estimates by ...



No.	Key Word	Frequency
1	growth	19
2	statistics	15
3	billion	14
4	economic	14
5	million	14
6	government	13
7	economy	12
8	income	12
9	people	11
10	data	10

7. Conclusion

- ❖ Current project were based on machine learning method (maximum entropy);
- ❖ Based on this capstone project result, the sentiment analysis can be used for determine the degree of “happiness” of the community towards the official statistics media released and the key words are related to statistics theme in DOSM; and
- ❖ Text analytics can be used as tools for planning any improvements in producing statistical products and services that meet user’s expectations.

Positive	Negative
affordable	issues
benefits	cancer
smart	lack
benefit	critical
good	decline
support	problem
work	declining
important	stresses
lean	fell
boost	concerns
afford	congestion
positive	debt
attractive	falling
free	gross
easier	challenging
easy	difficult
fast	poor
integrated	risk
lead	
successful	
top	

