

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







www.dosm.gov.my



www.mycensus.gov.my







2016 - 2030



PRESENTATION OUTLINE





1. INTRODUCTION





01

THE VITAL OF ONLINE NEWS

Ŧ



.0

Social media and electronic media play a vital role as a main reference for publics to gather information.



The rapid development of electronic media provides organizations with the opportunity to measure perceptions more quickly and transparently by implementing mediums such as a dedicated portal for monitoring online media reporting.



Agarwal et al. (2016) identified that news mining based on sentiment analysis could be analysed and alkhwarizmi test can be use to classified news statements with positive or negative sentiments that could help the government in decision making process.

04

In order to monitoring online economic news is in line with official statistical released and at par with national policy, DOSM has embarked on a data project initiative known as Statistics Big Data Analytics or STATSBDA.

Public Maturity Assessment on Official Statistics (PMAOS) was developed to provide DOSM with holistic and comprehensive insight of perception developed by media.



MEDIA INTELLIGENCE – STATSBDA INITIATIVE



mmcloud.ebdesk.com			Media Intelligence my topic 1 n	ny topic 2 resume map edit	H DISCOVER ANALYZE ACTION
Media Intelligence	username password	SIGN IN	DOSM 2	Population 1	GDP 1
eBdesk			KES sindiket melibatkan 47 rakyat Malaysia ditipu ditawar pekerjaan sebanyak US\$1,500 (RM6,000) sebulan di Kemboja mencerminkan keterdesakan rakyat untuk 3 months ago, Sarawak Voice	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	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
It tells what you dor	n't know Media Monit	oring	Labour Force 2 Labour Force 2 KES sindiket melibatkan 47 rakyat Malaysia ditipu ditawar pekerjaan sebanyak US\$1,500 (RM6,000) sebulan di Kemboja mencerminkan keterdesakan rakyat untuk 3 months ago, Astro Awani	Salary & wages 4 Image: Solar Sola	CPI-BPHPP 1 Image: CPI-BPHPP 1
Media Campaign	Media Analy	sis	DOSM Industry	HIES O 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



MEDIA INTELLIGENCE ANALYSIS



Media Intelligence my topic 1 m	I DI	DISCOVER ANALYZE		
DOSM 3	Population 4	GDP	Select Topic	
			Dashboard	
			Timeline	
			Top Issues	
An NGO today proposed a congrate division or agoncy to	An NGO today proposed a constrate division or agoncy to	Internativ nal Eit	Top Persons	
handle senior citizen affairs, saying their number is	handle senior citizen affairs, saying their number is	gross do nestic	Influencers tio	
14 hours ago, The Sun Daily	14 hours ago, The Sun Daily	1 month igo, T	Media Share	
Labour Force	Salary & wages 0	CPI-E PHP	Maps	
SULTAN Y			Sentiments	
			Ontology	
			Reporting	
Menteri Besar Johor Datuk Dr Sahruddin Jamal hari ini menasihati para graduan masa kini supaya tidak hanya	Young people, who are considered prime contributors to	Malaysia ; inflat	Trending _{Ise}	
mempertaruhkan kelulusan akademik semata-mata untuk	economic growth, are spending way beyond their means, and this has brought many of them to the brink of	1.5% in / ugust Ratings. The m	People Connection	









ANALYSIS ON MEDIA SHARE AND SENTIMENT (NEWS HEADLINES)



Media - Population1 | index | sentiments | Media - Population1 | index | sentiments ACTION H DISCOVE ANALYZE ANALYZE ACTION News Adv News Т W Μ Q Y Adv Т W М Q Υ Mentioned in News Related Influencers 2019 Sentiment Analysis News Title The Edge Markets MORE The Edge Markets EDGE EDCE 150 news Malaysia on track to achieve 150 news LATEST 7.5 Sustainable... Positive MY Stock 118 Dec 28th. 2018 - The Edge Markets MY Stock 118 Malaysia is on the right track in reaching the 96 news 96 news 111 Neutral Sustainable Development Goals (SDGs) The Sun Daily 94 news Indicators 2030, said the Department of The Sun Dailv **isun**da Statistics Malaysia. In a statement today, it said the governments 🏅 🦾 🥈 🎑 94 news Negative commitment to the sustainable development agenda is transpired through the Mid-Term Review of the 11th Malaysia Plan (11th MP) The Malay Mail Online The Malay Mail Online 0 Sentiment for 2019 presented on Oct 18. In this regard, the commitment to Agenda Jan Feb Mar Apr May Sep Dec Jun Jul Aug Oct Nov 69 news 69 news 2030 for Sustainable Development has been aligned with the Statements MORE MORE **A** Positives MORE Neutrals strategies and initiatives of the 11th Mala. MORE Bernama Δ R Bernama 50 news 50 news SDGs indicators cover 17 goals, 169 targets and Malaysia on track to achieve Sustaina... Total supply and use in Malaysia up 5 November inflation growth slows mor... 244 indicators to measure the achievement of Dec 28th, 2018 - The Edge Markets Dec 26th, 2018 · The Edge Markets Dec 20th, 2018 - The Edge Markets MSN Malaysia MSN Malaysia universal development which comprise social, msn 50 news Ringgit volatility drops on global trad.. Cover Story: Missing funds will have I ... 50 news Malaysia's economy likely to grow in ... economic and Dec 12th, 2018 · The Edge Markets Dec 24th. 2018 · The Edge Markets Sep 13th, 2018 - The Edge Markets Mohd Uzir Mahidin 18 - The Edge Markets News 2019 The Borneo Post The Borneo Post Dec 28th, 2018 - The Edge Markets FDI figures based on approved invest... Framework for sustainable retirement . Concern over rising inequality in Sabah 43 news E. E 20 43 news Dec 10th, 2018 - The Edge Markets Sep 6th, 2018 · The Edge Markets Dec 23rd, 2018 · The Edge Markets "Out of the 244 indicators, 118 indicators (48%) ree Malavsia Todav are Available, 55 indicators (23%) Partially Malaysia's economic growth to ease i... Malaysia Nov CPI up 0.2% on-year Inflation to rise to 1% in July, says RAM Free Malaysia Today 26 news 15 Nov 22nd, 2018 · The Edge Markets Dec 19th, 2018 · The Edge Markets Aug 23rd, 2018 · The Edge Markets news Available, 36 indicators (15%) Partially Available 26 news Need Further Malaysia Digest CPI eases to 0.5% in Q3 The Week Ahead: November CPI, cent... The Week Ahead: Keeping an eye on ... MD Mohd Uzir Mahidin 18 · The Edge Markets Malaysia Digest Nov 16th, 2018 - The Edge Markets Dec 17th, 2018 · The Edge Markets Aug 20th, 2018 · The Edge Markets 25 news Dec 28th, 2018 - The Edge Markets 25 news The Week Ahead: Post-budget rally, in... Malaysia October IPI up 4.2% on-year Cover Story: Sizeable land income for ... New Strait Times Nov 5th, 2018 · The Edge Markets Dec 12th, 2018 · The Edge Markets Aug 16th, 2018 - The Edge Markets New Strait Times 5 "SDGs with a theme Leaving No One Behind 25 news focuses on the development of 5Ps namely 25 news My Say: Increase wages, not retireme... Malaysia Oct manufacturing sales up ... Malaysia's IPI grows at a slower rate o... 71 Media Prev Next 71 Media Prev Next Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

f 💿 😏 @StatsMalaysia @MyCensus2020

ysia @MyCensus2020 🚳 www.dosm.gov.my www.mycensus.gov.my #MyCensus2020 #PastikanAndaDIBANCI



ANALYSIS ON MEDIA SHARE AND SENTIMENT (INFLUENCERS)













PROBLEMS 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;



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



Current portal are based on subscription with no customization basis.

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





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

2. METHODOLOGY





SCOPE OF RESEACRH











3. LITERATURE REVIEW





LITERATURE REVIEW



Author: Ordenes el Author: al. (2014) Author: Nanli et Author: al. (2012) Zamahsyari and **Title: Analyzing** Nurwidyantoro Customer Title: Sentiment (2016)**Experience Analysis:** A **Feedback Using** Literature **Title: Sentiment Text Mining: A Review Analysis of** Linguistics-based **Economic News** Findings: An Approach. in Bahasa analysis of Indonesia Using sentiments or Findings: Text **Majority Vote** mining is the opinion aiming to Classifier process of analyzing investigate, textual information in analyse, extract Findings: Online subjective texts an attempt to news has the discover structure that involve user ability to influence and implicit opinion, public readers on preferences and the opinion of an meanings "hidden" sentiments. within text. article.

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.

f 💿 😏 @StatsMalaysia | @MyCensus2020



LITERATURE REVIEW











4. DATA COLLECTION





News

DATA COLLECTION – TEXT DATA TYPE

ALC: 10-14





Keyword By	Query V	
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	1
Media Tag	OR T A	bb

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
1	Bernama	Bernama	50	130	20	Other
		Malay Mail	69			
		Mtimes	11			
2	Borneo	Borneo	43	51	I	
		Sabahtime	8			
3	Business Time	Business Time	18	18	Ι	
4	Chronicle	Chronicle	19	19	Ι	
5	Daily n Digital	Daily Express	9	14	Ī	
		Digital News Asia	5			
6	Focus	Focus	18	18	Ι	
7	Free M	Free M	26	26	Ī	
8	M Digest	M Digest	25	25	Ι	
9	M Insight	M Insight	23	23	Ī	
10	MSN	MSN	50	50	I	
11	MyStock 118	MyStock 118	96	96	Ι	
12	NST	NST	25	25	Ι	
13	Others	Cilisos	6	38		
	Mainstream	Dap	6			
		Din Merican	5			
		kini	14			
		M MustKnow	1			
		M Today	4			
		People's Parliament	1			
		Umno	1			
14	Singapore	Btimes SG	6	32	I	
		S BizReview	4			
		Yahoo S	22			
15	The Edge	The Edge	149	149	Ι	
16	The Star	The Star	12	12	Ι	
17	The Sun Daily	The Sun Daily	92	92	Ι	
18	Vietnam	V Plus	5	8	I	
		VietNam Net	3			
19	Yahoo M	Yahoo M	20	20		

	Media	Total News	Grand Total
rs	ABC	1	86
	ABS CBN	1	
	Aliran	3	
	Aninetto	1	
	Asia One	3	
	Astro Awani	1	
	Bangkok Post	1	
	Biz Insider	1	
	Biz Today	1	
	Blogs Wsj	1	
	Çhannel	6	
	CoverBureau	4	
	Express	1	
	Ipoh Echo	2	
	Jakarta Post	1	
	Khmertimeskh	1	
	KTemocKonsiders	1	
	Lim Sian See	2	
	LKS Blog	1	
	M Reserve	9	
	Mole	10	
	My Sinchew	6	
	News Xin	6	
	Reuters	3	
	Says M	4	
	Srwk Voice	1	
	Stop	1	
	Strait Times	6	
	Taiping	1	
	Tech In Asia	1	
	Today Online	4	
	Xinmsn	1	
	TOTAL NEWS		932

f 💿 🔰 🖸 @StatsMalaysia @MyCensus2020

www.dosm.gov.my www.mycensus.gov.my #MyCensus2020 #PastikanAndaDIBANCI

5. STATISTICAL ANALYSIS





01

02

03

04

05



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

upload data as corpus

start pre-processing

build-in

format vector

create document-term matrix







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





The main two methods of sentiment analysis are lexiconbased 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 lexiconbased 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:





<pre>> get_sentiments("afinn")</pre>				
# A tibble: 2,476 x 2				
word score				
<chr> <int></int></chr>				
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				

02

> 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_sentiment	s("nrc")
# A tibble: 13,	901 x 2
word	sentiment
<chr></chr>	<chr></chr>
l 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 13,8	91 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

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





RESULT – FREQUENCY OF WORDS







RESULT – CLUSTER ANALYSIS









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



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



Folds

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

Mogntin

Desitiv

	negati	ve	
lag	misuse lack ging UNEM	falling ployed	
negati	_{ve} opp	osition	
concerns is	ssues C	crude #	all slower
problem issu	e deo	cline	poor
	gros	S fell c	oncerned
	ading	a st	boost benefit
trump worth	good e	njoy S P	improve
afford impro	oved s	support	won
encourage Sav	ings in	portant	ease
ontribution	positiv	ve ^{sup}	ported

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

Folder

Focus

Fo

Th



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

negative
inequality declining concerns slow debt issue cancer age is lack decline panerisk negative fall cikk negative fall cikk negative slow debt issue cancer weak risks benefit POVERTY works free top good stable upbeat support contribution effective commitment POSITIVE
positive

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

negative challenging negative difficult slower problem weak limited issue decline poverty lack debt inequality risks hard crude gross risk fall S concern affordable loss fell issu improved savings top good afford significant effective work led support supported strong important leading to skilled positive recovery improve stronger contribution robust positive



RESULT – SENTIMENT ANALYSIS





f 🖸 🎔 🖸 @StatsMalaysia @MyCensus2020

www.dosm.gov.my www.mycensus.gov.my #MyCensus2020 #PastikanAndaDIBANCI







negative challenging negative difficult slower problem weak limited issue poverty decline lack inequality poor risks burden hard fall gross crude S concern Ð loss affordable fell ŝ improvement improved S savings top good mature afford significant work support led effective strong supported important leading boost positive skilled improve recovery stronger contribution robust positive





positive





boost

Contribution to sentiment

5

20

15

10



RESULT - CONCLUSION







RESULT - CONCLUSION



compared

population market





H DISCOVER

Penang Forum: Penang Forum would like to respond to Penang Chief Minister Chow Kon Yeow's statement that the Penang population, as projected

ANALYZE

ACTION

Y

Adv

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 ...



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 ...



Table 4.6:

The Frequency of Key Words Used by Each Folder

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. RESEARCH LIMITATION











8. THE SIGNIFICANT OF RESEARCH







Strengthening the Role as a THRUST DEPARTMENT OF STATISTICS MALAYSIA

Producer of Statistics STRATEGIC

STRATEGY 1

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

Programme Strengthening Statistical Communication

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

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

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



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

TRANSFORMATION

PLAN

2015 - 2020



THE SIGNIFICANT OF RESEARCH





9. RESEARCH FINDINGS AND IMPLICATIONS TO NATIONAL POLICY





THE SIGNIFICANT OF THE RESEARCH



11th MALAYSIA PLAN **Chapter 9: Transforming public service to** 100 productivity Focusing on the outcome of the diversity of skill and expertise in new areas such as data science and economics **PUBLIC SECTOR BIG DATA ANALYTICS Government Data Optimization and Transformation Services (GDOTS)** Initiative with the purpose of improving data driven Government service delivery system

PUBLIC SECTOR ICT STRATEGIC PLAN 2016-2020

Thrust 2: Data-driven Government

- Manage and realise the value of data in a holistic and efficient manner
- Strengthen cross agency data sharing

NATIONAL POLICY ON INDUSTRY 4.0

Technologies advancement and convergence

enabling richer insights through big data analyti

NATIONAL TRANSFORMATION 2050

advancement in computational, big data analytics and intelligence, smart applications, advanced robotics and 3D printing.

10. CONCLUSION AND DISCUSSION



- Current project were based on machine learning method (maximum entropy) with fee for subscription;
- 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	









ANALISIS TEKS DARI BERITA ATAS TALIAN: BERITA PENGANGGURAN DI MALAYSIA PADA TAHUN 2019



Berdasarkan kajian: pelaporan media pada tahun 2019:

- 40% berita pengangguran jenis Frictional dengan isu utama adalah keperluan pihak berkaitan untuk mengkaji kesesuaian subjek pembelajaran di universiti berbanding kemahiran yang diperlukan oleh pemain industri.
- 35% berita jenis **Structural** dengan isu utama adalah kemahiran terkini yang diperlukan untuk persediaan Revolusi Industri 4.0 dan Data Raya;
- 25% berita jenis Cyclical dengan isu utama yang dilaporkan adalah berita pemberhentian pekerja dari media Utusan dan inisiatif SOCSO di dalam membangunkan Sistem EIS (*Employment Insurance System*) di dalam membantu pekerja-pekerja yang diberhentikan kerja



THANK YOU

MENGHITUNG MALAYSIA

7.7.2020 - 30.9.2020 (e-Census)

7.10.2020 - 24.10.2020 [Face To Face]

