

ANALISIS SENTIMEN TERHADAP KESEDARAN MASALAH KESIHATAN MENTAL BAGI TWITTER PERTUBUHAN BUKAN KERAJAAN

FATIN HAZIQAH MOHAMAD ZAINI
ROHIZAH ABD. RAHMAN

Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia

ABSTRAK

Statistik dari Kementerian Kesihatan Malaysia menunjukkan terdapat peningkatan bilangan masalah kesihatan mental di kalangan rakyat di Malaysia. Namun demikian, masyarakat Malaysia sering memandang remeh terhadap permasalahan ini. Tujuan utama projek ini dijalankan adalah untuk melihat sejauh mana kesedaran masyarakat Malaysia terhadap masalah kesihatan mental ini. Projek ini akan melihat kepada sentimen analisis bagi kesedaran masalah kesihatan mental di kalangan rakyat di Malaysia dengan menggunakan data yang diambil dari aplikasi Twitter bagi Pertubuhan Bukan Kerajaan (NGO) berkaitan kesihatan mental di Malaysia. NGO sering menyebarkan maklumat seperti statistik kesihatan, punca-punca dan juga cara mengatasi masalah ini di Aplikasi Twitter mereka. Walau bagaimanapun, pengumpulan data dari aplikasi API Twitter memerlukan keizinan dan permohonan dari pihak Twitter walaupun data adalah secara umum. Pembangunan projek ini akan dilaksanakan menggunakan metod secara eksperimen di mana analisis sentimen merupakan salah satu cara yang sesuai untuk mengkaji kesedaran masyarakat Malaysia terhadap masalah kesihatan mental. Antara eksperimen yang akan dijalankan adalah seperti pengumpulan data, pembersihan data, pengekstrakan ciri melalui Bags-of-Words (BOW) dan N-grams serta implementasi beberapa teknik pembelajaran mesin. Eksperimen yang dijalankan akan dilarikan menggunakan bahasa pengaturcaraan Python dan juga perisian R-Studio.

1 PENGENALAN

Analisis Sentimen merupakan proses untuk mengesan emosi pengguna bagi data berbentuk teks diantara ianya positif, negatif ataupun neutral mengenai sesuatu topik. Ia juga merujuk kepada penggunaan Pemprosesan Bahasa Tabii, Analisis Teks, Linguistik Pengkomputeran dan Biometrik untuk mengenalpasti secara sistematik. Data akan dibaca, diekstrak dan pemprosesan data akan dilakukan. Setelah itu, data akhir akan dihasilkan. Hasil daripada data akhir ini, barulah keputusan dapat dibuat sama ada data tersebut adalah Positif, Negatif ataupun Neutral.

Terdapat dua kaedah yang biasa digunakan iaitu Pendekatan Berdasarkan Peraturan dan Analisis Sentimen Automatik. Kaedah pertama adalah Pendekatan Berdasarkan Peraturan iaitu bermaksud kaedah ini berdasarkan algoritma dengan maksud pendapat yang diberikan dengan jelas untuk dikenalpasti termasuk mengenalpasti polariti, subjektiviti atau pendapat kepada sesuatu topik itu. Selain itu, Pendekatan Berdasarkan Peraturan ini merangkumi Pemprosesan Bahasa Tabii yang asas seperti Tokenisasi, Pencantasan, Penandaan POS (*Part-of-speech*) dan Analisis Leksikon. Teknik yang digunakan dalam Pendekatan Berdasarkan Peraturan ini adalah seperti dua senarai perkataan iaitu senarai perkataan positif dan senarai perkataan negatif. Seterusnya, algoritma akan melalui teks tersebut, dan mengenalpasti perkataan yang hampir dengan kriterianya. Dan akhir sekali, algoritma akan mengira berapa banyak perkataan yang hampir dengan teks tersebut. Jika terdapat banyak perkataan positif, maka teks tersebut akan dianggap mempunyai polariti positif.

Kaedah yang kedua adalah, Analisis Sentimen Automatik. Kaedah ini akan mengkaji sebuah teks tersebut dengan lebih mendalam kerana ia menggunakan Pembelajaran Mesin untuk mencari makna kepada mesej tersebut. Sehubungan dengan itu, Pembelajaran Mesin Diselia juga termasuk ke dalam kaedah ini bagi tujuan algoritma klasifikasi. Secara amnya, Analisis Sentimen merupakan salah satu contoh yang bagus untuk digunakan klasifikasi secara maksimum. Selain itu, Pembelajaran Mesin Tidak Diselia juga digunakan untuk mencari data. Oleh itu, ketepatan Sentimen Analisis ini meningkat dan boleh memproses maklumat yang lebih banyak tanpa berlaku apa-apa masalah. Selain itu juga, terdapat beberapa jenis algoritma klasifikasi lain yang sering digunakan seperti '*Linear Regression*', '*Naïve Bayes*' dan '*Support Vector Machines*'.

2 PENYATAAN MASALAH

Kementerian Kesihatan Malaysia menyatakan hampir setengah juta rakyat Malaysia mengalami masalah kesihatan mental (Suraya Roslan, 2020). Pertubuhan Kesihatan Sedunia (*World Health Organization* -WHO) telah mendefinisikan maksud kesihatan mental sebagai kebolehan mengekalkan hubungan yang harmoni dengan orang lain, kebolehan mengambil bahagian dalam aktiviti kemasyarakatan dan kebolehan untuk menyumbang kepada masyarakat. Selain itu, pihak WHO juga menganggarkan seramai 12% penduduk dunia terjejas akibat penyakit ini (World Health Organization 2018). Tidak kurang statistik Malaysia yang mana telah ramai masyarakat Malaysia mengalami masalah kesihatan mental seperti tekanan, kemurungan, keresahan dan cubaan membunuh diri. Kesedaran mengenai kesihatan mental oleh masyarakat Malaysia sering menjadi tanda tanya, adakah masyarakat Malaysia kini peka akan maklumat yang telah diberikan atau pun tidak.

Menurut statistik pada tahun 2019, jumlah keseluruhan penghidap penyakit mental di Malaysia ialah sebanyak 11.2% iaitu ia bersamaan sekitar 3.3 juta dari keseluruhan 30 juta rakyat Malaysia. Sehubungan dengan itu, jantina yang mengalami lebih ramai pesakit mental adalah kalangan golongan wanita dengan kiraan 55% berbanding golongan lelaki iaitu 45%. Justeru, kadar peratusan bagi cubaan untuk membunuh diri adalah sebanyak 6.4% dan kadar tertinggi pula melibatkan remaja dan peringkat awal remaja (berumur 16 hingga 24 tahun) iaitu 11% (Muhamad Firdaus Rahim 2019).

3 OBJEKTIF KAJIAN

Objektif bagi projek yang dijalankan ini adalah seperti berikut:

- i. Mengenalpasti kesedaran masyarakat Malaysia terhadap maklumat Kesihatan Mental melalui data di aplikasi Twitter NGO Kesihatan Mental.
- ii. Menjalankan proses Analisis Sentimen bagi melihat data yang telah diperolehi.
- iii. Menjalankan analisis deskriptif bagi melihat kesedaran awal masyarakat Malaysia terhadap maklumat Kesihatan Mental.

4 METOD KAJIAN

Metodologi yang akan digunakan bagi projek ini adalah “*Cross Industry Standard Process for Data Mining*” ataupun dikenali sebagai “CRISP-DM”. “CRISP-DM” ini merupakan satu model proses yang mempunyai 6 fasa yang merangkumi kitaran hidup Data Sains. Metodologi ini sering kali digunakan bagi menyusun dan melaksanakan projek yang berkaitan dengan Data Sains. Terdapat 6 fasa di dalam metodologi “CRISP-DM” (Kuliahkomputer 2018) iaitu :



Rajah 1 Carta Alir bagi model “CRISP-DM”

4.1 Fasa Perancangan

Berdasarkan masalah yang telah dinyatakan, perancangan bagi projek ini akan mengkaji kesedaran rakyat Malaysia di laman sosial Twitter Pertubuhan Bukan Kerajaan (Non-Governmental Organization-NGO) menggunakan analisis sentimen. Ini kerana, Analisis Sentimen sering digunakan dalam kawalan media sosial kerana ia membenarkan untuk mendapat pandangan orang ramai terhadap sesuatu topik yang terperinci. Sehubungan dengan itu, pihak NGO telah banyak memberikan maklumat berhubung dengan masalah kesihatan mental melalui medium aplikasi Twitter. Analisis Sentimen merupakan salah satu cara yang terbaik untuk menyelesaikan masalah ini. Melalui analisis sentimen, kita dapat mengetahui sejauh manakah

rakyat Malaysia mengambil perhatian mengenai masalah kesihatan mental. Data akan diekstrak dari aplikasi Twitter bagi NGO kesihatan mental di Malaysia dan seterusnya beberapa pemrosesan data akan dilaksanakan.

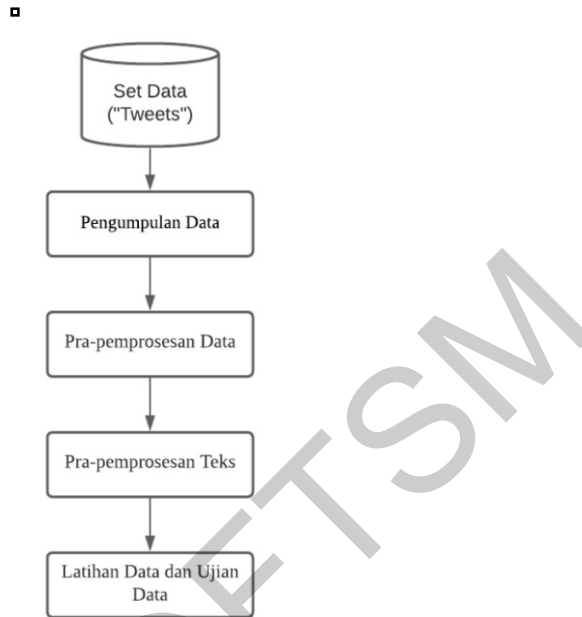
Terdapat empat akaun Twitter NGO kesihatan mental di Malaysia yang akan digunakan iaitu *Mental Illness Awareness and Support Association* (MIASA), MindaKami, SOLS Health dan Relate Malaysia. Kesemua NGO ini mempunyai matlamat yang sama iaitu untuk membantu mereka yang mempunyai kesihatan mental dan menggunakan platform Twitter sebagai tempat untuk menyebarkan maklumat mengenai kesihatan mental . Justeru, data yang dikumpulkan merupakan pada 1 April 2020 sehingga 1 April 2021, di mana ia merupakan fasa Perintah Kawalan Pergerakan di Malaysia. Terdapat beberapa berita yang telah dikenalpasti, bahawa masalah kesihatan mental meningkat pada fasa Perintah Kawalan Pergerakan ini. Jadi data tersebut amatlah sesuai untuk digunakan dalam kajian ini.

4.2 Fasa Analisis

Selain daripada melaksanakan analisis sentimen, terdapat beberapa analisis yang dilakukan seperti dari segi peratusan kadar interaksi, analisis perkataan sentimen dan analisis emosi yang bersesuaian dengan data yang dikumpulkan. Hal ini kerana, kesedaran masyarakat juga mestilah diambil kira dari pandangan yang lain untuk melihat corak kesedaran masyarakat terhadap masalah kesihatan mental ini.

4.3 Fasa Reka Bentuk

Pada fasa ini, reka bentuk bagi setiap analisis yang dikaji telah dikenalpasti. Setiap daripada analisis ini mempunyai kaedah yang berlainan untuk memastikan hasil keputusan dapat dicapai. Hasil daripada kesemua proses reka bentuk ini, telah dinyatakan pada ruangan Hasil Kajian. Bagi analisis sentimen, bahasa Python telah digunakan untuk mengumpul data, membersihkan data dan menganalisis data. Berikut merupakan carta alir bagi proses analisis sentimen yang dilakukan.



Rajah 2 Proses Analisis Sentimen

Pada proses permulaan bagi fasa ini, data telah dikumpulkan dengan menggunakan Twitter API. Setelah dikumpulkan, pra-pemrosesan data telah dilakukan bagi membuang beberapa data yang tidak diperlukan. Setelah itu, data teks tersebut diproses untuk membuang beberapa elemen yang dapat mengganggu proses analisis dijalankan seperti menukarkan kepada huruf kecil, membuang nombor, membuang tanda baca dan sebagainya. Setelah mendapat data yang bersih, data tersebut telah digunakan untuk beberapa proses analisis. Analisis yang pertama merupakan analisis yang utama iaitu analisis sentimen. Pendekatan berdasarkan Leksikon Vader telah diaplikasikan bagi mengkalsifikasikan ciapan kepada positif atau negatif. Dalam Leksikon Vader ini, nilai kompaun bagi setiap telah dikira mengikut nilai perkataan positif dan negatif. Setelah itu, nilai tersebut akan menentukan nilai kompaun bagi keseluruhan satu ciapan tersebut.

Analisis yang seterusnya adalah mengira peratusan kadar interaksi. Proses ini bertujuan untuk mengira kadar interaksi pengikut terhadap ciapan yang telah dimuat naik daripada jumlah “*Retweets*” dan “*Likes*” yang merupakan tanda kesedaran terhadap ciapan tersebut. Rumus bagi mengira peratusan kadar interaksi adalah seperti berikut:

$$\text{PERATUSAN KADAR INTERAKSI} = \frac{\text{JUMLAH "RETWEET" DAN "LIKES"}}{\text{JUMLAH PENGIKUT}} \times 100$$

Selain itu, data yang telah dibersihkan juga telah digunakan untuk analisis klasifikasi perkataan sentimen dengan menggunakan *WordCloud*. *WordCloud* merupakan salah satu kaedah untuk memvisualisasikan perkataan yang mempunyai frekuensi terbanyak bagi suatu data NGO. Melalui proses ini, dapat dilihat perkataan positif dan negatif yang sering digunakan dalam ciapan yang telah dimuat naik. Bagi proses ini juga, *WordCloud* telah diaplikasikan dengan menggunakan Python.

Analisis yang terakhir adalah analisis klasifikasi dengan menggunakan perisian RStudio. Dalam proses ini, pakej '*Syuzhet*' telah diaplikasikan bagi mengenalpasti emosi akan sesuatu teks tersebut. Terdapat lapan emosi yang telah diklasifikasi bagi ciapan setiap NGO seperti :

- i- kemarahan (*anger*)
- ii- jangkauan (*anticipation*)
- iii- jijik (*disgust*)
- iv- ketakutan (*fear*)
- v- kegembiraan (*joy*)
- vi- kesedihan (*sadness*)
- vii- kejutan (*surprise*)
- viii- kepercayaan (*trust*)

4.4 Fasa Pengujian

Pada fasa ini, hasil daripada analisis sentimen telah diuji dengan menggunakan beberapa klasifikasi bagi mengenalpasti ketepatan klasifikasi iaitu klasifikasi *Naïve Bayes* , klasifikasi *Support Vector Machine* dan klasifikasi *Neural Network*. Proses pengujian ini telah menggunakan bahasa Python dan pakej library yang digunakan adalah *scikit-learn*. Terdapat beberapa elemen klasifikasi yang telah diuji seperti :

- i- *R-Squared score*
- ii- *Precision*
- iii- *Recall*
- iv- *F1-score*
- v- *Support*
- vi- *Cross validation result*

Dengan menggunakan ketiga-tiga klasifikasi tersebut, satu laporan klasifikasi telah dihasilkan daripada hasil pengujian tersebut. Laporan klasifikasi adalah seperti Jadual 1 berikut:

Jadual 1 menunjukkan Laporan Klasifikasi

Klasifikasi/Elemen Klasifikasi	NAÏVE BAYES	SUPPORT VECTOR MACHINE	NEURAL NETWORK
<i>R-Squared Score</i>	80.41%	86.88%	86.60%
<i>precision (positif)</i>	0.90	0.87	0.89
<i>precision (negatif)</i>	0.56	0.79	0.76
<i>recall (positif)</i>	0.84	0.96	0.95
<i>recall (negatif)</i>	0.68	0.52	0.59
<i>f1-score (positif)</i>	0.87	0.91	0.92
<i>f1-score (negatif)</i>	0.61	0.63	0.67
<i>support (positif)</i>	150	150	150
<i>support (negatif)</i>	44	44	44
<i>cross validation result</i>	74.44%	84.41%	84.20%

5 HASIL KAJIAN

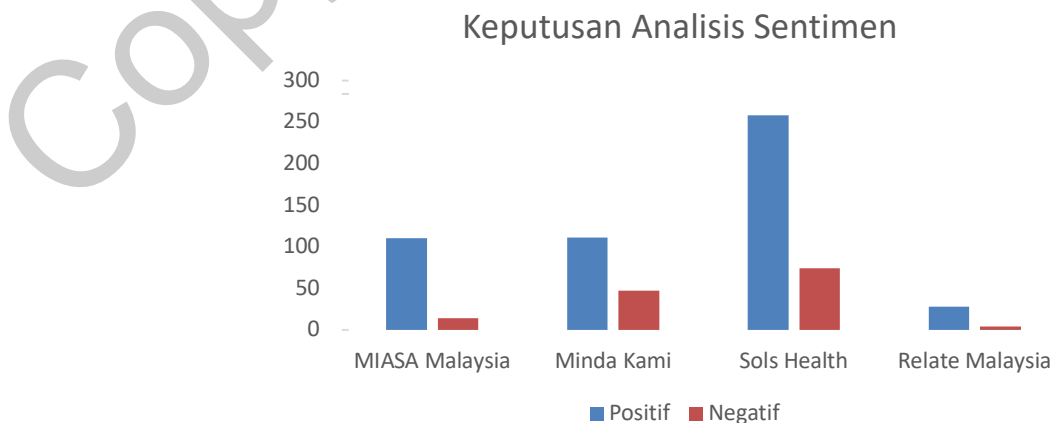
Terdapat beberapa hasil kajian berdasarkan kesemua analisis yang telah dilakukan. Antaranya seperti analisis sentimen, analisis peratusan kadar interaksi, klasifikasi perkataan sentimen WordCloud dan klasifikasi emosi.

5.1 Analisis Sentimen

Berdasarkan Jadual 2 di bawah, jumlah ciapan yang mempunyai sentimen positif dan sentimen negatif yang telah dikenalpasti hasil daripada jumlah ciapan yang telah dikumpulkan mengikut NGO.

Jadual 2 menunjukkan Keputusan Analisis Sentimen

NAMA NGO	JUMLAH CIAPAN	JUMLAH CIAPAN POSITIF	JUMLAH CIAPAN NEGATIF
MIASA MALAYSIA	124	110	14
MINDA KAMI	158	111	47
SOLS HEALTH	332	258	74
RELATE MALAYSIA	32	28	4



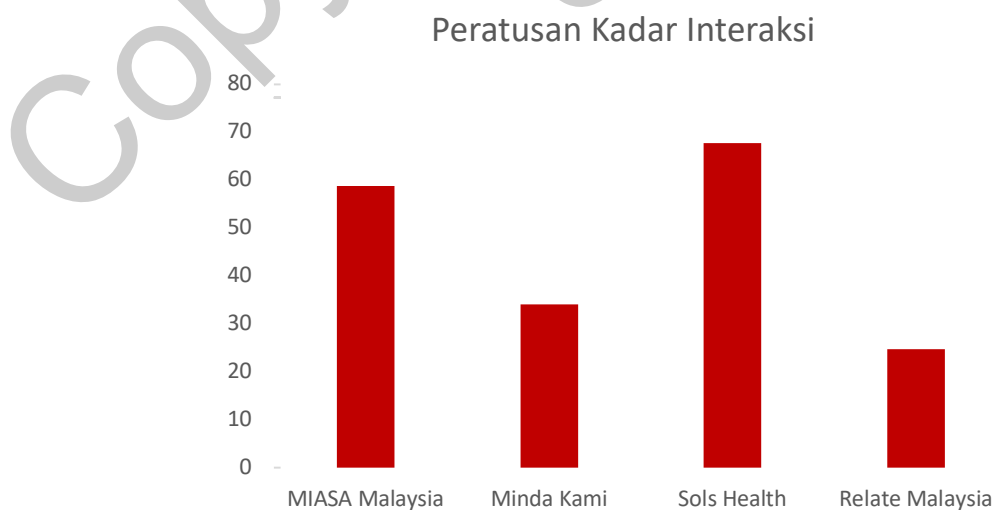
Rajah 3 Keputusan Analisis Sentimen

5.2 Analisis Peratusan Kadar Interaksi

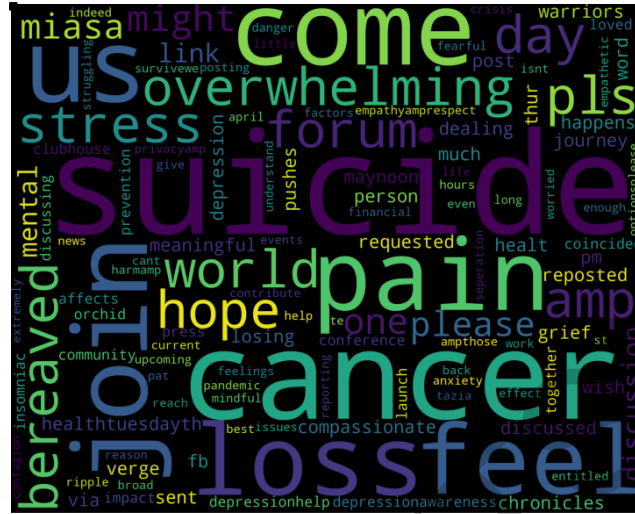
Jadual 3 dan Rajah 4 menerangkan peratusan kadar interaksi bagi setiap NGO yang terlibat. Peratusan kadar interaksi yang tertinggi dalam kalangan NGO terlibat adalah Sols Health dengan kadar peratusan 67.78%. Walau bagaimanapun, dengan perbezaan sebanyak 9.01% iaitu 58.77% telah meletakkan MIASA Malaysia pada tempat kedua tertinggi bagi peratusan kadar interaksi ini. Selain itu, Minda Kami telah mendapati peratusan kadar sebanyak 34.09% dan akhir sekali, peratusan kadar terendah adalah Relate Malaysia dengan 24.83%. Hasil daripada keputusan ini menunjukkan bahawa Sols Health mempunyai lebih banyak interaksi yang menunjukkan rakyat sedar akan kesihatan mental melalui ciapan yang telah dimuat naik di laman akaun Twitter mereka.

Jadual 3 menunjukkan Analisis Peratusan Kadar Interaksi

Nama NGO	Jumlah “Retweets” Dan “Likes”	Jumlah Pengikut	Peratusan Kadar Interaksi (%)
MIASA MALAYSIA	1732	2947	58.77
MINDAKAMI	6238	18 300	34.09
SOLS HEALTH	2559	3300	67.78
RELATE MALAYSIA	403	1636	24.63



Rajah 4 Peratusan Kadar Interaksi

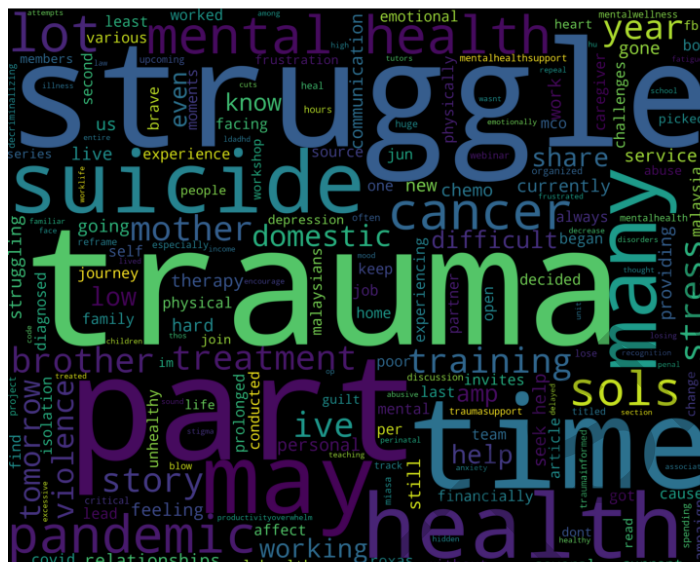


Rajah 6 WordCloud Perkataan Negatif bagi @MIASAMalaysia

ii- **MindaKami**

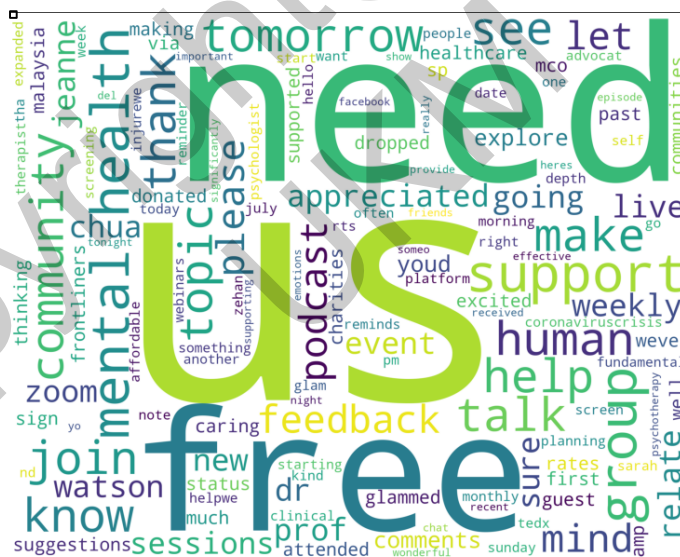


Rajah 7 WordCloud Perkataan Positif bagi @mindakami

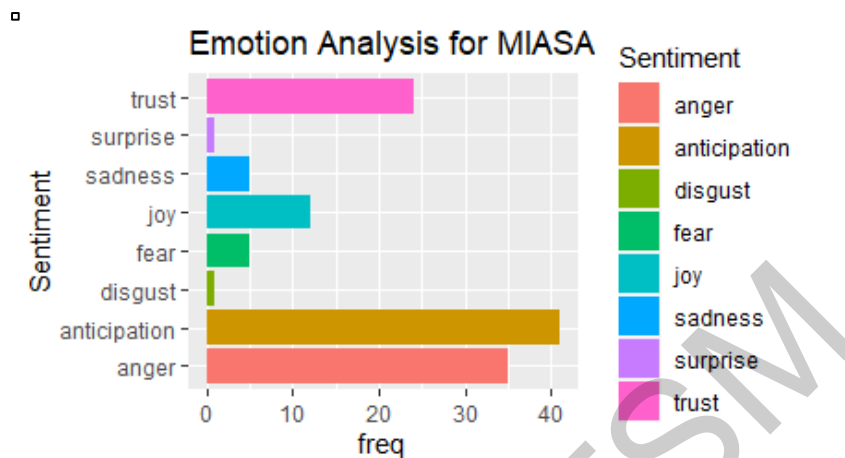


Rajah 10 *WordCloud* Perkataan Negatif bagi @solshealth

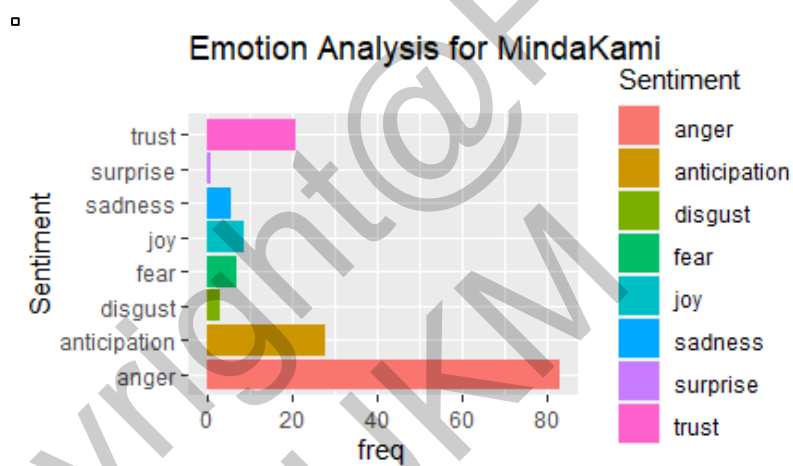
v. Relate Malaysia



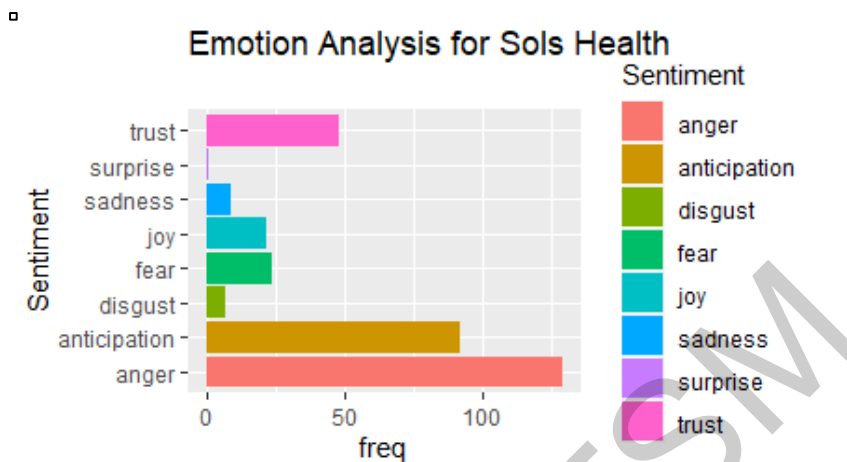
Rajah 11 *WordCloud* Perkataan Positif bagi @relatemalaysia



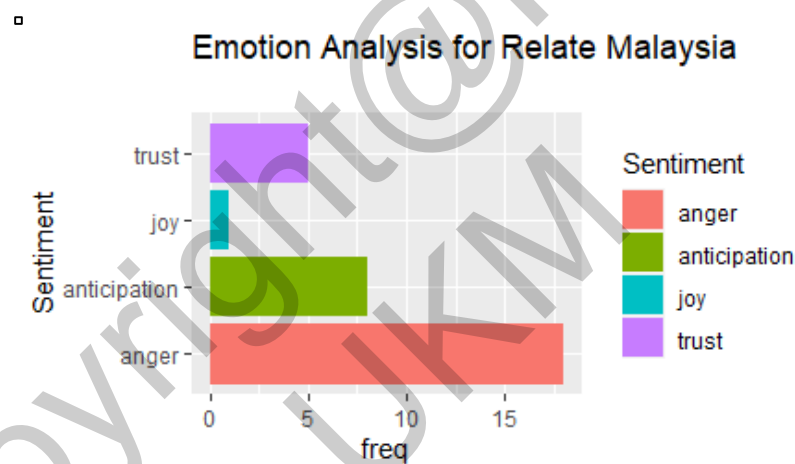
Rajah 13 Klasifikasi Emosi bagi MIASA Malaysia



Rajah 14 Klasifikasi Emosi bagi Minda Kami



Rajah 15 Klasifikasi Emosi bagi Sols Health



Rajah 16 Klasifikasi Emosi bagi Relate Malaysia

6 KESIMPULAN

Analisis sentimen terhadap kesedaran masalah kesihatan mental bagi Twitter pertubuhan bukan kerajaan ini dapat membantu NGO dalam mengenalpasti sejauh manakah kesedaran rakyat Malaysia terhadap masalah kesihatan mental yang semakin lama, semakin meningkat sejak Perintah Kawalan Pergerakan semasa pandemik COVID-19 menular di Malaysia pada 1 April 2020 sehingga 1 April 2021. Hasil daripada analisis sentimen ini juga dapat membantu NGO untuk menambahbaikkan dari segi kandungan ciapan yang telah dimuat naik di aplikasi Twitter bagi disebarkan kepada pengikut-pengikut mereka sama ada ianya kepada ciapan yang positif atau ciapan negatif. Selain daripada analisis sentimen, terdapat juga analisis lain menggunakan data yang sama telah dilakukan dalam projek ini yang dapat memberikan gambaran kesedaran dari sudut yang lain. Contohnya seperti analisis peratusan kadar interaksi untuk mengira kadar interaksi pengikut di Twitter NGO tersebut terhadap ciapan yang telah dimuat naik. Selain itu, klasifikasi emosi dapat mengenalpasti emosi bagi setiap ciapan tersebut.

Sehubungan dengan itu, penggunaan bahasa seperti Python dan RStudio yang digunakan amatlah bersesuaian dengan projek ini kerana ianya mudah dan cekap untuk digunakan. Walaubagaimanapun, masih terdapat lagi kekangan yang telah dikenalpasti didalam projek ini dan memerlukan masa yang lebih untuk mendapatkan hasil analisis yang lebih tepat.

7 RUJUKAN

Kuliahkomputer. 2018, CRISP-DM. <http://www.kuliahkomputer.com/2018/07/crisp-dm.html>

Muhamad Firdaus Rahim. 2019, 7 Statistik Berkenaan Kesihatan Mental Di Malaysia. <https://hellodoktor.com/gaya-hidup-sihat/tips-kesihatan/7-statistik-berkenaan-kesihatan-mental-di-malaysia/#gref>

Suraya Roslan. 2020, Hampir 0.5 juta warga Malaysia kemurungan. , <https://www.hmetro.com.my/utama/2020/10/629299/hampir-05-juta-warga-malaysia-kemurungan>

World Health Organization. 2018, Mental health: Strengthening our response. (n.d.). <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response>

Copyright@FTSM
UKM