

Data Annotation Architecture for Automatic Depression Detection

Senibina Anotasi Data untuk Mengesan Kemurungan Secara Automatik

Yun Yao Chang, Nazlia Omar

*Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia, 43600 Bangi,
Selangor, Malaysia*

**Corresponding author: p104370@siswa.ukm.edu.my*

Received 1 October 2022

Accepted 17 January 2023, Available online 1 June 2023

ABSTRACT

Depression is a mood disorder that causes a person to feel sad, tired and experience a prolonged lack of energy, irritability, and loss of interest in daily activities. Many scholars have contributed in identifying and curbing depression. One of such efforts is the development of a model that can identify and predict depression among Twitter users. However, so far, there is no quality and labeled dataset containing depression from tweet sources. Therefore, the purpose of this study is to propose an architecture that can collect data on social media such as Twitter to detect depression automatically. This study involves text analysis that begins with data scraping, text processing, feature extraction, modeling, evaluation and followed by document corpus analysis using TF-IDF and BOW. The sentiment lexicon derived from two tools, TextBlob and Vader, was used to distinguish the emotions of words. Four machine learning classifiers i.e., Logistic Regression, Decision Tree, Support Vector Machine and K-Nearest Neighbour were used to perform the classification. The final data set management and the use of Logistic Regression produced the expected high accuracy, precision, recall and F1-Score results in predicting depression. For the application, data for Malaysia local COVID-19 tweets was scraped using TWINT. Appropriate hashtags and keywords were used to obtain tweet sentences. The results show that the proposed architecture outperforms the baseline by achieving 92.876% F1-Score through SVM+TFIDF compared to the F-Score obtained by the baseline. This shows that the proposed data annotation architecture has good performance in detecting depression.

Keywords: COVID-19, data annotation, depression, Machine Learning, TWINT

ABSTRAK

Kemurungan adalah gangguan perasaan yang menyebabkan seseorang merasa sedih, keletihan dan kekurangan tenaga yang berpanjangan, mudah marah, dan kehilangan minat dalam aktiviti harian. Banyak kajian lepas telah menyumbang dalam mengenalpasti dan membendung kemurungan. Antara usaha ini ialah pembangunan sebuah model yang dapat mengenalpasti

dan meramal kemurungan di kalangan pengguna Twitter. Namun begitu, setakat ini tiada set data berkualiti dan berlabel yang mengandungi data berkaitan kemurungan dari sumber Twit. Oleh itu, tujuan kajian ini adalah untuk mengusulkan satu seni bina yang dapat mengumpul data di media sosial seperti Twitter bagi mengesan kemurungan secara automatik. Kajian ini melibatkan analisis teks yang bermula dengan pengikisan data, pemprosesan teks, pengekstrakan ciri, pemodelan, penilaian dan diikuti analisis korpus dokumen menggunakan TF-IDF dan BOW. Leksikon sentimen yang berasal dari dua alatan, TextBlob dan Vader, digunakan untuk membezakan emosi kata-kata, serta empat teknik pembelajaran mesin iaitu Regresi Logistik (LR), Pepohon Keputusan (DT), Mesin Vektor Sokongan (SVM) dan K-Jiran Terdekat (KNN) yang digunakan untuk melakukan pengelasan. Pengurusan set data akhir dan penggunaan Regresi Logistik menghasilkan keputusan ketepatan dan Skor-F1 tertinggi yang diharapkan dalam meramal kemurungan. Untuk bahagian aplikasi, data dari twit lokal iaitu Malaysia COVID-19 telah dikikis dengan menggunakan TWINT. Tanda pagar dan kata kunci yang sesuai digunakan untuk mendapatkan ayat Twit Keputusan menunjukkan bahawa seni bina yang dicadangkan mengatasi garis dasar dengan mencapai 92.876% Skor-F1 melalui SVM+TFIDF berbanding Skor-F1 yang diperolehi oleh garis dasar. Ini menunjukkan bahawa seni bina anotasi data yang dicadangkan telah menghasilkan prestasi yang baik dalam mengesan kemurungan.

Kata kunci: COVID-19, Anotasi Data, Kemurungan, Pembelajaran Mesin, TWINT

PENGENALAN

Kesihatan mental merujuk kepada kesejahteraan emosi, psikologi, dan sosial kita. Ia mempengaruhi cara kita berfikir, merasa, dan bertindak. Pertubuhan Kesihatan Sedunia mula memberi tumpuan kepada kesihatan mental sejak tahun 2001. Ini menunjukkan bahawa kesihatan mental sangat penting untuk kesejahteraan keseluruhan individu, masyarakat, dan negara (Organization 1991). Statistik kemurungan di Malaysia telah meningkat sebanyak 29.2% bagi rakyat negara ini yang berusia 16 tahun ke atas berbanding 11.2% yang dicatat pada tahun 2006 (Aris, 2017). Kajian Rosli Hamid (2020) mendapati bahawa pelajar tingkatan empat mengalami kemurungan, kebimbangan dan tekanan pada tahap sederhana. Dengan itu, masyarakat perlu mengambil berat sebelum keadaan mereka menjadi teruk. Selain itu, laporan KKM 2022 menunjukkan sejumlah 631 kes bunuh diri telah berlaku pada tahun 2020, manakala trend tahun 2021 menunjukkan peningkatan ketara sebanyak 81% (1,142) (K. Malaysia 2022). Namun begitu, kekurangan ahli psikiatri profesional juga merupakan salah satu masalah di negara ini. Malaysia memerlukan seramai 3,000 ahli psikiatri, namun negara ini hanya mempunyai 410 ahli psikiatri berdaftar sehingga Julai 2018 (Bernama, 2021). Secara global, hampir 90% individu yang membunuh diri mempunyai masalah mental pada saat sebelum kematian dan 60% daripada mereka mengalami kemurungan (Wan Mohd Rushidi 2018). Abd Malek et al. (2020) menyatakan bahawa individu yang mengalami kemurungan lebih cenderung untuk membunuh diri. Menurut Shaffer et al. (1996), seseorang yang berkecenderungan membunuh diri menderita daripada sekurang-kurangnya satu penyakit, lazimnya kemurungan. Semasa wabak COVID-19, kerajaan Malaysia melaksanakan Perintah Kawalan Pergerakan (PKP) ke atas rakyat bermula 18 Mac 2020, sekaligus mengubah corak dan norma kehidupan orang ramai, yang membawa kepada situasi kemurungan yang semakin ketara (Abdul Aziz, 2020). Oleh itu, penyelidikan bagi mengesan kemurungan peringkat awal perlu dijalankan bagi mengelak salah laku yang tidak diingini. Atas motivasi ini, kajian ini menggunakan pengesanan kemurungan sebagai tujuan penyelidikan utama.

Srividya et al. (2018), Giannakakis et al. (2017) dan Gkotsis et al. (2017) menjalankan penyelidikan mengenai kesihatan mental melalui pembelajaran mesin. Namun begitu, tidak ada satu set data garis asas yang boleh digunakan untuk meramal kemurungan. Martinez (2019) menyatakan bahawa tidak ada set data yang berkualiti untuk melabel perkataan yang mengandungi Twit kemurungan. Walaupun beberapa penyelidik berjaya membangunkan set data yang baik, ianya tidak dikongsi atas dasar 'privasi'. Sekiranya kita hanya mengguna dan mempertimbangkan set data Sentiment140 atau SemEval, emosi sedih tersebut tidak dapat mewakili kemurungan sepenuhnya. Dengan itu, dalam proses latihan pembelajaran mesin, pembersihan set data adalah penting untuk dilakukan. Sekiranya set data yang dimodelkan tidak baik, ia akan menghasilkan dapatan yang juga kurang baik. Oleh itu, untuk mendapatkan hasil ramalan yang baik, data latihan yang mewakili kemurungan sepenuhnya adalah sangat penting.

Di samping itu, walaupun terdapat banyak kajian yang dijalankan menggunakan pembelajaran mesin ke atas kemurungan (Shatte et al. 2019), masih tiada kajian yang lengkap terutamanya dari segi seni bina untuk dijadikan sebagai rujukan atau garis panduan bagi memproses data kemurungan secara automatik. Untuk proses pembinaan set data secara manual, pertama sekali, kita harus menilai emosi ayat demi ayat, untuk mengenalpasti sama ada terdapat situasi kemurungan daripada teks yang dihasilkan oleh seseorang. Contohnya, sekiranya terdapat 100 perkataan, maka 100 penilaian diperlukan. Apabila bilangan perkataan bertambah, ia mengambil masa yang lebih lama. Selain itu, ketiadaan standard yang tetap juga menjadi salah satu isu dalam proses penganotasian secara manual. Individu berbeza mungkin mempunyai cara yang berbeza untuk menilai sama ada ayat itu mengandungi elemen kemurungan. Oleh itu, pakar diperlukan untuk pengesanan data. Walau bagaimanapun, seperti yang disebut dalam bahagian pengenalan, Malaysia tidak mempunyai ramai pakar seperti ini. Proses anotasi data secara manual ini mungkin boleh mengakibatkan kehilangan erti pertimbangan segera kerana data perlu dinilai satu persatu. Oleh itu, jika ada pengesanan model kemurungan secara automatik, ia akan membantu meningkatkan kecekapan mengesan.

Bagi sesetengah sarjana dalam domain lain, kebanyakan kajian menjalankan tinjauan melalui soal selidik. Oleh itu, sekiranya terdapat seni bina yang boleh digunakan dalam pelbagai domain berkaitan pengekstrakan data untuk kemurungan, ia akan digunakan bagi menambah baik penyelidikan mengenai kemurungan dalam bidang yang dikaji. Contohnya, kajian tentang keadaan psikologi masyarakat semasa COVID-19 boleh diperolehi melalui seni bina yang dicadangkan supaya data penyelidikan yang diperolehi lebih relevan.

Tujuan utama kajian ini adalah untuk mengusulkan satu seni bina yang boleh menghasilkan set data berlabel kemurungan secara automatik. Aliran kerja seni bina ini mesti jelas dan boleh dipercayai agar sesuai untuk para sarjana termasuk mereka dari bidang lain untuk melakukannya. Selain itu, hasilnya mesti dapat dikendalikan dan disahkan agar dapat digunakan untuk pengembangan penelitian berkaitan dengan kemurungan, dan hasilnya dapat digunakan bagi meramal perkataan lain. Objektif kajian ini adalah untuk mencadangkan satu aliran kerja seni bina untuk menganotasi set data latihan kemurungan secara automatik.

KAJIAN BERKAITAN

Pengesanan kemurungan

Menurut buku panduan kesihatan mental (Mental Health Handbook 2019), penyelidik mempunyai pemahaman ringkas mengenai gangguan mental biasa termasuk kemurungan, kegelisahan, gangguan bipolar, dan skizofrenia. Kemurungan adalah antara gangguan mental

yang boleh dirawat. Antara 80% dan 90% individu yang mengalami kemurungan sembuh dengan baik setelah menerima rawatan (American Psychiatric Association 2021).

Prapemprosesan

A. Perkataan Biasa Kemurungan (#Tandapagar).

Tanda pagar adalah tag metadata yang diawali oleh simbol #pagar diikuti oleh rentetan aksara alfanumerik, tanpa ruang kosong atau tanda baca (Burgess 2020). Simbol tanda pagar wujud sekitar abad ke-14 (Andersen 2018). Simbol ini kini menjadi ungkapan penting dalam budaya popular, dan umumnya dikaitkan dengan pelbagai dimensi aktiviti di persekitaran media sosial (Berg 2014).

Dalam kajian ini, kata-kata umum kemurungan digunakan di bahagian pengambilan data. Ramos et al. (2016) menemui kaedah insan menyatakan ‘kesedihan’. Kata kuncinya termasuk sedih (sad), kesepian (lonely), dan tidak senang (unhappy), yang mana digunakan dalam pengambilan data seterusnya. Di samping itu, individu yang mengalami kemurungan biasanya menggunakan “saya” (I) lebih daripada “kami” (We).

CAPAIAN DATA

Sumber data kajian ini kebanyakannya diperoleh daripada TWINT dan SemEval. Untuk percubaan ini, hasil yang ditangkap oleh TWINT ditapis sebagai 1 (Positif), manakala ayat dalam SemEval ditapis sebagai 0 (Negatif). Bahagian ini akan diterangkan lebih lanjut dalam bahagian Metodologi.

A. TWINT

TWINT ialah alat pengikis Twitter lanjutan yang ditulis dalam Python yang membolehkan Twit dikikis daripada profil Twitter tanpa menggunakan Twitter API (Haccer 2019). Memandangkan audit untuk menggunakan Twitter API semakin sukar, TWINT menjadi alatan yang lebih senang digunakan untuk mendapatkan maklumat. TWINT boleh digunakan tanpa nama dan tanpa memerlukan pendaftaran, dan persediaan awal boleh diselesaikan dengan lebih cepat (“TWINT: A Twitter information scraping tool” 2019). Di samping itu, cara untuk mendapatkan Set Data Kemurungan melalui TWINT akan dicadangkan dalam kajian ini.

B. SemEval-2019

SemEval ialah satu siri bengkel penyelidikan pemprosesan bahasa tabii antarabangsa. Misinya adalah untuk memajukan analisis semantik terkini dan membantu dalam penciptaan anotasi set data berkualiti tinggi (“SemEval - International Workshop on Semantic Evaluation” n.d.). Acara ini diadakan setahun sekali, di mana SemEval-2018 berlangsung pada tahun 2018. Kajian ini menggunakan data tahun 2018 kerana data yang diekstrak dalam TWINT juga adalah dari tahun 2018.

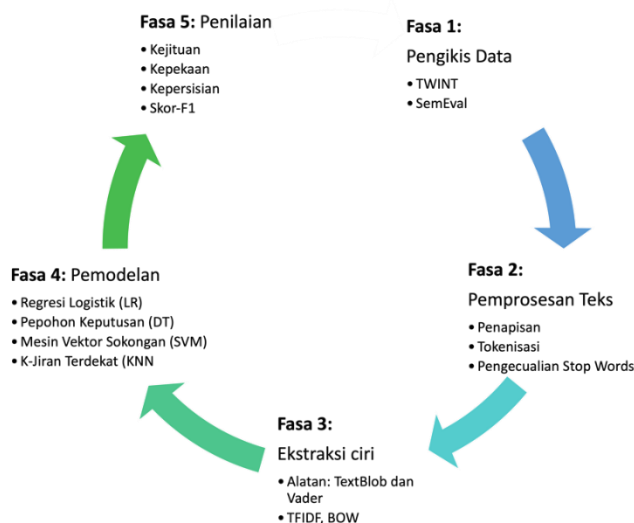
KAJIAN LEPAS

Penyelidikan berkaitan anotasi data telah menarik minat yang amat tinggi dari kalangan penyelidik dalam bidang pengesanan kemurungan. Dalam kajian yang dijalankan oleh Martinez (2019), kaedah set data latihan disediakan dengan menggunakan TWINT untuk mengikis Tweet kemurungan, diikuti pengambilan sampel 8,000 tweet dengan polariti 4 yang positif, dan menggabungkan kedua-dua set data (TWINT dan SemEval). Dalam kajian tersebut,

data rujukan yang digunakan sebagai ‘twit bukan kemurungan’ ialah Sentimen140. Sentimen140 ialah set data yang mengandungi 1,600,000 Twit yang diekstrak dengan menggunakan API Twitter. Twit tersebut telah diberi anotasi 0 = negatif, 2 = neutral, and 4 = positif yang boleh digunakan untuk mengesan sentimen (Kazanova n.d.). Selain itu, beberapa kajian lain seperti Rajaraman et al. (2020) dan Bhargava (2021) menggunakan data Sentimen140. Namun begitu, berdasarkan maklumat yang diperolehi daripada laman web rasmi semasa, Kazanova menyatakan bahawa Sentimen140 (ver2) paling akhir dikemaskini 4 tahun yang lalu. Oleh itu, boleh disimpulkan bahawa keseluruhan Twit yang ada mungkin tidak selaras dengan penggunaan perkataan terkini.

METODOLOGI KAJIAN

Metodologi kajian ini merangkumi lima fasa. Fasa pertama ialah penyediaan data. Bahagian ini ditakrifkan sebagai pengikis data dengan menggunakan SemEval dan data yang diambil dari TWINT. Fasa kedua melibatkan tugas pra-pemprosesan seperti penapisan, tokenisasi, pembuangan kata henti dan sebagainya. Fasa ketiga bertujuan mewakili istilah dalam perwakilan ruang vektor dengan menggunakan TF-IDF dan BoW. Dua alat pengekstrakan ciri iaitu TextBlob dan Vader digunakan dalam fasa ini. Fasa keempat melibatkan klasifikasi oleh empat pengelas iaitu LR, DT, SVM dan KNN. Fasa kelima ialah bahagian untuk penilaian. Fasa ini menyemak sama ada idea yang dicadangkan ditambah baik berbanding dengan penyelidikan semasa. Bahagian ini menggunakan kaedah penilaian yang paling lazim iaitu Keperisian, Kepekaan, Skor-F1, dan Kejituan. Rajah 1 menunjukkan 5 tahap aliran eksperimen ini.



RAJAH 1. Aliran eksperimen

Rajah 2 dan Rajah 3 memperkenalkan cadangan seni bina kajian. Eksperimen terbahagi kepada dua iaitu Eksperimen 1 (Cadangan Senibina) dan Eksperimen 2 (Senibina Garis Dasar). Eksperimen 1 adalah hasil pemprosesan mengikut rangka kerja yang dicadangkan. Eksperimen 2 menggunakan twit asal (tidak diproses) yang dijadikan sebagai penanda aras.

Rajah 2 menunjukkan seni bina Eksperimen 1 serta aliran eksperimen. Keseluruhan seni bina boleh dibahagikan kepada dua bahagian. Twint terutamanya digunakan untuk mendapat twit kemurungan yang berpotensi melalui hashtag tertentu. Selepas itu, Topeng Penapisan digunakan untuk menapis ayat yang tidak diinginkan. Pemprosesan data kemudian dijalankan bagi memenuhi syarat pembelajaran mesin. Bahagian ini mempertimbangkan dua alat sentimen

iaitu Vader dan TextBlob dalam proses pelabelan data. Daripada analisis sentimen, kita dapat melabelkan teks sebagai kemurungan atau bukan kemurungan. Selain data yang diekstrak daripada Twitter, kajian ini juga memperoleh beberapa ayat Twitter daripada SemEval, selain melakukan pelabelan data. Data SemEval mempunyai 4 tag (Kesedihan, Kemarahan, Ketakutan, Kegembiraan), manakala perkataan selepas TextBlob dan Vader dibahagikan kepada 3 sentimen (Positif, Neutral, Negatif). Justeru, terdapat $4 \times 3 = 12$ kombinasi sentimen. Jadual 1 menunjukkan contoh perkataan dalam keadaan sentimen yang berbeza. Menerusi pemerhatian, didapati ayat Kesedihan+Negatif pada dasarnya paling sesuai dengan sentimen ayat Kemurungan. Contohnya, *'this year keeps getting more depressing and disappointing'* (tahun ini semakin menyedihkan dan mengecewakan). Perkataan Kemarahan+Positif, Kemarahan+Neutral, Ketakutan+Positif, Ketakutan+Neutral, Kegembiraan+Positif, Kegembiraan+Neutral jelas lebih positif atau neutral. Oleh itu, perkataan tersebut digunakan sebagai label=0 (perkataan bukan kemurungan). Contohnya, *'Woke up to my neighbors singing Perfect by Ed Sheeran. My mood has been set'* (Terjaga dari tidur setelah terdengar jiran saya menyanyikan lagu Perfect oleh Ed Sheeran. Mood saya telah ditetapkan). *'Everyone wants to win, but swear they never lost at anything. Victory is only sweet when you taste the bitterness of defeat'* (Semua orang mahu menang, tetapi bersumpah mereka tidak pernah kalah dalam apa-apa. Kemenangan hanya manis apabila anda merasai kepahitan kekalahan). *'Whats for breakfast today? I had 2 slices of bread, I toasted them and spread them with butter; strawberry jam. Delicious!'* (Sarapan apa hari ini? Saya bersarapan 2 keping roti yang dibakar dan disapu jem strawberi. Sedap!). Selain itu, ayat gabungan lain agak samar, justeru tidak digunakan untuk penilaian. Ramos et al. (2016) menemui kaedah insan menyatakan 'sedih'. Kata kuncinya termasuk sedih (sad), kesepian (lonely), dan tidak senang (unhappy). Daripada Jadual 1, didapati ayat-ayat yang diperolehi oleh Kesedihan+Negatif mempunyai perkataan 'lost' (hilang arah), 'sad' (kesedihan), 'depressed' (kemurungan), dan 'disappointing' (mengecewakan). Kesemua perkataan ini adalah negative.

JADUAL 1. Contoh perkataan dalam keadaan kombinasi sentimen yang berbeza

Sentimen	Contoh perkataan												
1. Kesedihan	a. Positif												
	<table border="1"> <tbody> <tr> <td>An early night for me love you</td> <td>sadness</td> <td>POSITIVE</td> </tr> <tr> <td>"Let us not become weary in doing good, for at the proper time we will reap a harvest if we do not give up." Galatians 6:9 NIV</td> <td>sadness</td> <td>POSITIVE</td> </tr> <tr> <td>@khizthigh Lol idk really :S khair good luck</td> <td>sadness</td> <td>POSITIVE</td> </tr> <tr> <td>@SonaFanArt Isnt it too early? Thats really interesting 🤔 i wonder what are they filming in kusadasi. It makes me so excited :)</td> <td>sadness</td> <td>POSITIVE</td> </tr> </tbody> </table>	An early night for me love you	sadness	POSITIVE	"Let us not become weary in doing good, for at the proper time we will reap a harvest if we do not give up." Galatians 6:9 NIV	sadness	POSITIVE	@khizthigh Lol idk really :S khair good luck	sadness	POSITIVE	@SonaFanArt Isnt it too early? Thats really interesting 🤔 i wonder what are they filming in kusadasi. It makes me so excited :)	sadness	POSITIVE
An early night for me love you	sadness	POSITIVE											
"Let us not become weary in doing good, for at the proper time we will reap a harvest if we do not give up." Galatians 6:9 NIV	sadness	POSITIVE											
@khizthigh Lol idk really :S khair good luck	sadness	POSITIVE											
@SonaFanArt Isnt it too early? Thats really interesting 🤔 i wonder what are they filming in kusadasi. It makes me so excited :)	sadness	POSITIVE											
	b. Neutral												
	<table border="1"> <tbody> <tr> <td>I graduated yesterday and already had 8 family members asking what job I've got now 🤔 #nightmare</td> <td>sadness</td> <td>NEUTRAL</td> </tr> <tr> <td>@DOBrienAJC What did you do to rile up all the Roberts?</td> <td>sadness</td> <td>NEUTRAL</td> </tr> <tr> <td>#sadness #cry 🤔 dad will never understand his 3rd child</td> <td>sadness</td> <td>NEUTRAL</td> </tr> <tr> <td>@lvankaTrump What's up w the profile pic??</td> <td>sadness</td> <td>NEUTRAL</td> </tr> </tbody> </table>	I graduated yesterday and already had 8 family members asking what job I've got now 🤔 #nightmare	sadness	NEUTRAL	@DOBrienAJC What did you do to rile up all the Roberts?	sadness	NEUTRAL	#sadness #cry 🤔 dad will never understand his 3rd child	sadness	NEUTRAL	@lvankaTrump What's up w the profile pic??	sadness	NEUTRAL
I graduated yesterday and already had 8 family members asking what job I've got now 🤔 #nightmare	sadness	NEUTRAL											
@DOBrienAJC What did you do to rile up all the Roberts?	sadness	NEUTRAL											
#sadness #cry 🤔 dad will never understand his 3rd child	sadness	NEUTRAL											
@lvankaTrump What's up w the profile pic??	sadness	NEUTRAL											
	c. Negatif												
	<table border="1"> <tbody> <tr> <td>With you, I'm lost in the moment. Without you I'm lost in the world. 🤔🤔 #world #moment #lostinthemoment #You</td> <td>sadness</td> <td>NEGATIVE</td> </tr> <tr> <td>ha 🤔 ha 🤔 i just realised that i've never actually been to therapy and i still don't know if i'm depressed or just sad all the time</td> <td>sadness</td> <td>NEGATIVE</td> </tr> <tr> <td>this year keeps getting more depressing and disappointing</td> <td>sadness</td> <td>NEGATIVE</td> </tr> <tr> <td>And all the bullshit in your live ends and nothing will bother you ever again #depression</td> <td>sadness</td> <td>NEGATIVE</td> </tr> </tbody> </table>	With you, I'm lost in the moment. Without you I'm lost in the world. 🤔🤔 #world #moment #lostinthemoment #You	sadness	NEGATIVE	ha 🤔 ha 🤔 i just realised that i've never actually been to therapy and i still don't know if i'm depressed or just sad all the time	sadness	NEGATIVE	this year keeps getting more depressing and disappointing	sadness	NEGATIVE	And all the bullshit in your live ends and nothing will bother you ever again #depression	sadness	NEGATIVE
With you, I'm lost in the moment. Without you I'm lost in the world. 🤔🤔 #world #moment #lostinthemoment #You	sadness	NEGATIVE											
ha 🤔 ha 🤔 i just realised that i've never actually been to therapy and i still don't know if i'm depressed or just sad all the time	sadness	NEGATIVE											
this year keeps getting more depressing and disappointing	sadness	NEGATIVE											
And all the bullshit in your live ends and nothing will bother you ever again #depression	sadness	NEGATIVE											

2. Kemarahan a. Positif

Happy right now I just beaten the second game of crash bandicoot one more to go now didn't get that mad 🤔 at the game lol 🤔	anger	POSITIVE
Everyone want to win,but swear they never lost at anything. Victory is only sweet when you taste the bitterness of defeat.	anger	POSITIVE
im back lol #revenge	anger	POSITIVE
I follow this famous chick on snap and she put 'I'm snapping back' & next snap says 'dudes stop sending your peepes, it's not cute.' 🤔🤔🤔	anger	POSITIVE

b. Neutral

Were starting an @insecurehbo club at my job. Every Monday we must be caught up on required watching. @opjumpstartlb @IssaRae	anger	NEUTRAL
Shakespeare Dictionary: The word 'knotty-pated' means 'block-headed, dull-witted'	anger	NEUTRAL
Coming up on snap chat so add me asap MATEO_BKNY #snapchat	anger	NEUTRAL
I want a friendship group like in Friends 🤔	anger	NEUTRAL

c. Negatif

Humble yourself in the sight of the LORD. If we have died in Christ, then how can we be offended? A dead person cannot feel anything, right?	anger	NEGATIVE
Love implies anger. The man who is angered by nothing cares about nothing. ~Edward Abbey #love	anger	NEGATIVE
Having fun today at Mad museum Stratford upon Avon	anger	NEGATIVE
@Lovelsland That rap reminds me of when the kids were small and they would prepare a performance #cute #crap 🤔	anger	NEGATIVE

3. Ketakutan a. Positif

Absolutely blown away by @BatTheMusical this evening 🤔 Such incredible sets, slick choreography and holy moly what insaaane vocals! #awe	fear	POSITIVE
Every time I fart my dog jumps in fear hahahaha yass	fear	POSITIVE
and CLEARLY all this #breathless coverage isn't covering the REAL #Collusion story - #DNC & #UkraineCollusion !!	fear	POSITIVE
@exceptions Although I have a nice vulva, I choose not to intimidate other women with it.	fear	POSITIVE

b. Neutral

One week on SW and going to gym. Smashed it #weekone #goals #sw	fear	NEUTRAL
#start my YouTube channel techno ayan please follow friends	fear	NEUTRAL
Going over the #script for this weekend! Can't wait to start #filming! #actorslife #ActressLife #actress #horrorfilm #HorrorMovies	fear	NEUTRAL
u can get an oreo shake at burger king	fear	NEUTRAL

c. Negatif

#Past #success or #failure is no #indicator of #future #performance DON'T be #afraid to take a #risk.	fear	NEGATIVE
When you wake up early to workout and then the fitness room smells awful, so you come back to your room to do core...	fear	NEGATIVE
I'll have the first Question for Followers ready in a few. It's directed at all of you, so please don't be shy and respond when you can! 🤔	fear	NEGATIVE
#Sleep is my #drug. My bed is my dealer. My #alarm is the #cops. #School is the #jail. #TeamFollowBack	fear	NEGATIVE

4. Kegembiraan a. Positif

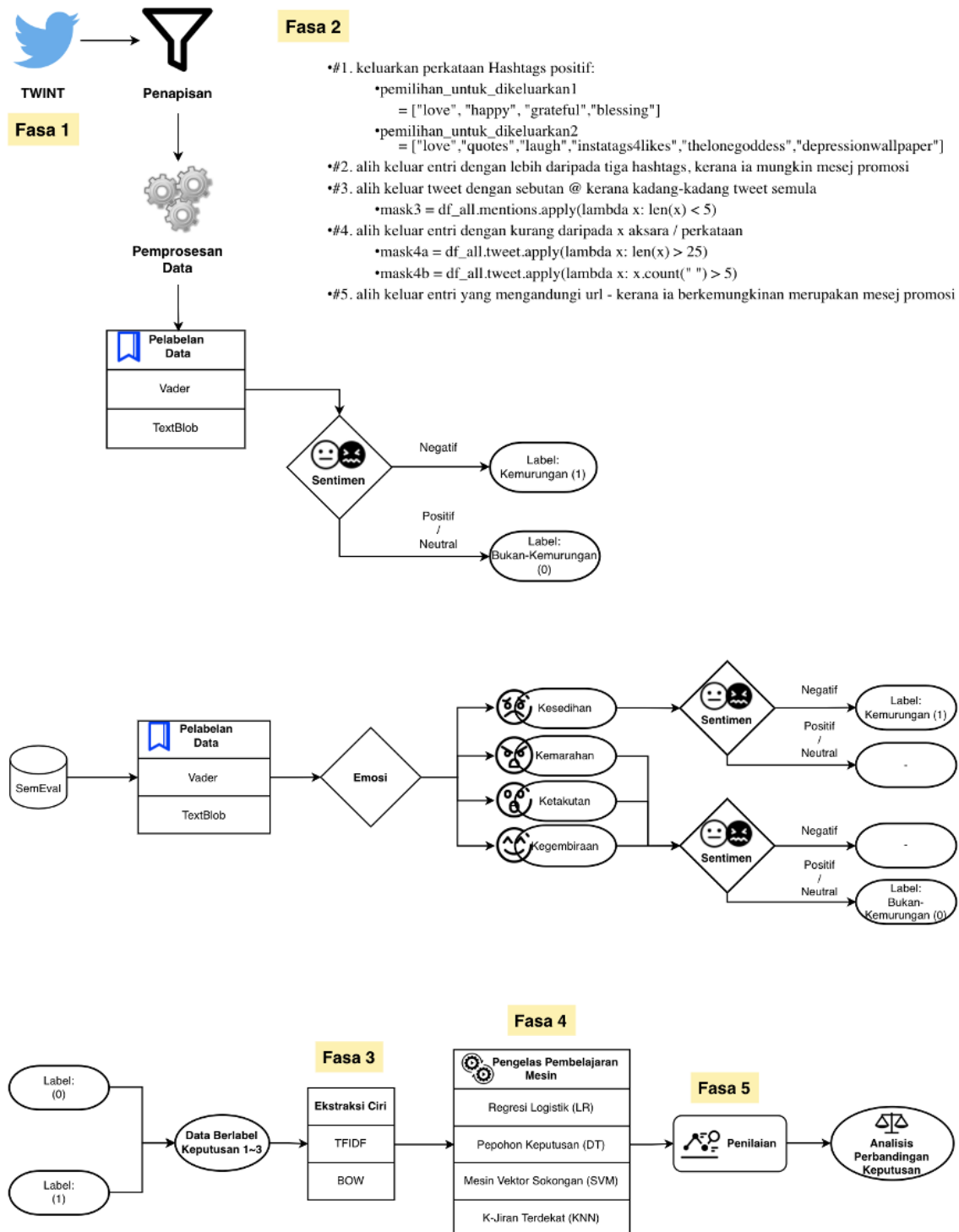
@TommySandhu @bbcasiannetwork tommy you r phenomenal smashed #ajjdin on the mixes #hilarious and #awesome 🤔🤔🤔🤔	joy	POSITIVE
For many years I have despised olives, my thoughts on them have now changed. #delightful	joy	POSITIVE
Woke up to my neighbors singing Perfect by Ed Sheeran. 🤔 My mood has been set. 🤔	joy	POSITIVE
whats for #breakfast today ? i have 2 slices of #bread out i #toast them #spread them with #butter & #strawberry jem and yum #delicious !	joy	POSITIVE

b. Neutral

@TommySandhu @bbcasiannetwork tommy you r phenomenal smashed #ajjdin on the mixes #hilarious and #awesome 🤔🤔🤔🤔	joy	POSITIVE
For many years I have despised olives, my thoughts on them have now changed. #delightful	joy	POSITIVE
Woke up to my neighbors singing Perfect by Ed Sheeran. 🤔 My mood has been set. 🤔	joy	POSITIVE
whats for #breakfast today ? i have 2 slices of #bread out i #toast them #spread them with #butter & #strawberry jem and yum #delicious !	joy	POSITIVE

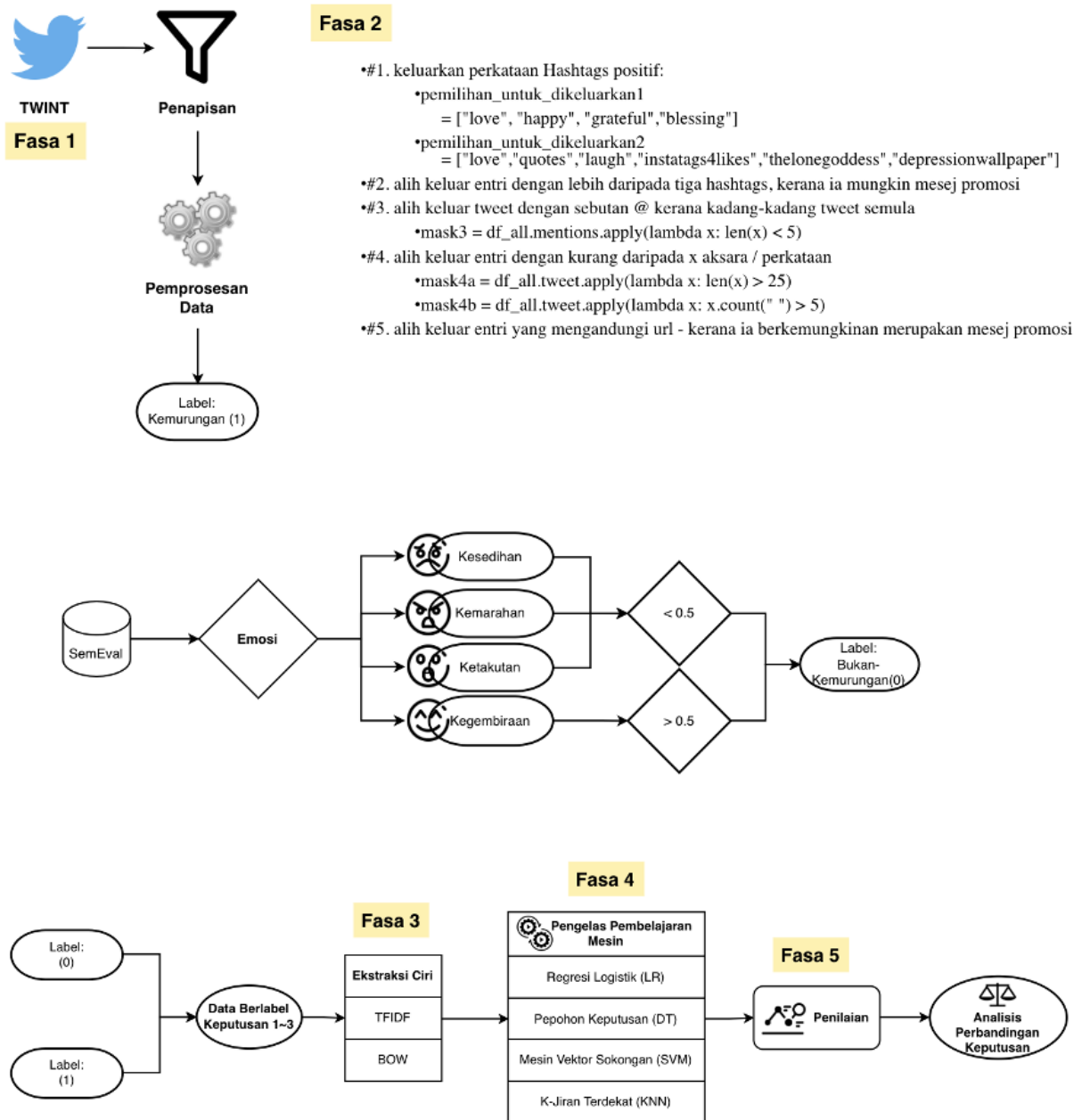
c. Negatif

Whether my glass is half empty or its half full. I'm just grateful I even have a glass and that there's something in it.\n #optimism 🤔	joy	NEGATIVE
As you start and end your day, be thankful for every little thing in your life. You will come to realize how blessed you truly are.	joy	NEGATIVE
So excited for tonight, I actually feel a bit sick 🤔	joy	NEGATIVE
i was so confused but then my inner fujoshi was screaming and like i was trying so hard not to smile or laugh	joy	NEGATIVE



RAJAH 2. Eksperimen 1 (Cadangan senibina)

Rajah 3 menunjukkan seni bina Eksperimen 2. Perbezaannya dengan Eksperimen 1 ialah ketiadaan penilaian sentimen dalam bahagian ini. Pada peringkat pertama, kajian ini menggunakan Twint untuk menangkap tweet kemurungan yang berpotensi melalui tanda pagar tertentu. Selepas menyelesaikan penapisan dan pemprosesan data yang asas, data terus dilabelkan sebagai tweet kemurungan. Bagi tweet tanpa kemurungan, ia diklasifikasikan mengikut empat tag yang digunakan SemEval (Kesedihan, Kemarahan, Ketakutan, Kegembiraan). Ayat dengan Kesedihan+Kemarahan+Ketakutan (polariti kurang daripada 0.5) dan ayat dengan Kegembiraan (polariti lebih daripada 0.5) dilabelkan sebagai ‘Bukan-Kemurungan’.



RAJAH 3. Eksperimen 2 (Senibina garis dasar)

Set data yang digunakan adalah daripada gabungan data TWINT dan SemEval-2018. Data yang diekstrak oleh TWINT diberi anotasi sebagai label=1 (Twit Kemurungan) selepas pemprosesan teks berikutnya. Tag berikut digunakan untuk mengikis: ‘#depressed AND sad, #depressed AND lonely, #depressed AND unhappy’. Dalam latihan pembelajaran mesin, adalah mustahil untuk melatih model hanya dengan mengambil kira label=1. Justeru, kajian ini juga memerlukan beberapa teks Twit bukan kemurungan. Oleh itu, data SemEval-2018 digunakan. Selepas pemprosesan selesai, data yang ditapis diberi anotasi sebagai label=0.

HASIL KAJIAN

Selepas melengkapkan penyediaan data, hasil visualisasi dengan WordCloud digunakan seperti dalam Rajah 4. Dalam bahagian penyelidikan ini, keputusan sentimen berbeza diwakili oleh tiga warna: biru adalah sentimen negatif, merah adalah sentimen positif, dan hijau adalah sentimen neutral. Hasil daripada sentimen negatif (biru) boleh digunakan untuk menggambarkan kata kunci tentang kemurungan. Pada masa yang sama, kata kunci ini juga merupakan kriteria yang digunakan oleh pembelajaran mesin untuk menilai sama ada sesuatu ayat itu menggambarkan kemurungan. Semakin tinggi kekerapan kata kunci ini dalam ayat tersebut, semakin tinggi kebarangkalian untuknya dinilai sebagai kemurungan.



RAJAH 4. WordCloud (Negatif, Neutral, Positif)

A. Perbandingan antara Pengelas

Dari hasil eksperimen, keputusan terbaik ialah SVM+Vader+TFIDF (kejituan: 91.523, kepersisan: 92.632, kepekaan: 93.386, skor-f1: 92.876). Symbol '*' mewakili hasil terbaik. Didapati bahawa Regresi Logistik (LR) dan Pepohon Keputusan (DT) memberi hasil terbaik apabila menggunakan ciri Beg Perkataan (BOW). Manakala Mesin Vektor Sokongan (SVM) dan K-Jiran Terdekat (KNN) memberi hasil terbaik apabila menggunakan ciri Frekuensi Songsang Jangka Kekerapan Dokumen (TFIDF). Jadual 2 hingga Jadual 5 menunjukkan hasil perbandingan antara pengelas.

JADUAL 2. Regresi Logistik (LR)

Regresi Logistik		Kejituan	Kepersisan	Kepekaan	Skor-F1
TextBlob *	BOW *	<u>91.486</u>	<u>93.003</u>	92.197	<u>92.598</u>
	TFIDF	90.484	90.476	93.353	91.892
Vader	BOW	90.581	91.406	92.857	92.126
	TFIDF	90.267	89.109	<u>95.238</u>	92.072

JADUAL 3. Pepohon Keputusan (DT)

Pepohon Keputusan		Kejituan	Kepersisan	Kepekaan	Skor-F1
TextBlob	BOW	85.810	86.761	<u>89.017</u>	87.874
	TFIDF	85.309	87.069	87.572	87.320
Vader *	BOW *	<u>85.871</u>	<u>88.710</u>	87.302	<u>88.000</u>
	TFIDF	84.458	87.399	86.243	86.818

JADUAL 4. Mesin Vektor Sokongan (SVM)

Mesin Vektor Sokongan		Kejituan	Kepersisan	Kepekaan	Skor-F1
TextBlob	BOW	90.150	<u>93.617</u>	89.017	91.259
	TFIDF	91.152	92.464	92.197	92.330
Vader *	BOW	88.697	88.250	<u>93.386</u>	90.746
	TFIDF *	<u>91.523</u>	92.632	93.122	<u>92.876</u>

JADUAL 5. K-Jiran Terdekat (KNN)

K-Jiran Terdekat		Kejituan	Kepersisan	Kepekaan	Skor-F1
TextBlob	BOW	71.619	67.671	97.399	79.858
	TFIDF	78.484	74.058	96.532	83.814
Vader *	BOW	64.521	62.667	<u>99.471</u>	76.892
	TFIDF *	<u>79.435</u>	<u>75.464</u>	96.825	<u>84.820</u>

B. Perbandingan antara Hasil TextBlob dan Vader

Perbincangan seterusnya ialah pengisihan bagi pengelas yang menggunakan TextBlob dan Vader (Rajah 5 dan Rajah 6). Ketepatan adalah tinggi dalam keputusan kedua-duanya. Pada masa yang sama, didapati bahawa Skor-F1 juga tinggi. Oleh itu, bahagian ini menggunakan pengisihan Skor-F1 sebagai piawaian penilaian untuk pengelas. Kombinasi pengisihan untuk hasil TextBlob ialah: LR+BOW -> SVM+TFIDF -> LR+TFIDF -> SVM+BOW -> DT+BOW. Manakala kombinasi pengisihan hasil Vader ialah: SVM+TFIDF -> LR+BOW -> LR+TFIDF -> SVM+BOW -> DT+BOW. Di samping itu, didapati bahawa keputusan BOW adalah yang terbaik dalam pertimbangan TextBlob dan Vader. SVM mendapat hasil yang terbaik dalam TFIDF, manakala DT mendapat hasil terbaik dalam BOW. Selain itu, didapati bahawa KNN tidak muncul dalam keputusan 5 tertinggi TextBlob dan Vader. Oleh itu, ia juga menunjukkan bahawa prestasi KNN secara relatifnya adalah yang paling lemah, justeru tidak sesuai untuk klasifikasi dalam kajian ini.

```
result.sort_values("F1Score", ascending=False).head(5)
```

	Classifier	Methods	Accuracy	Precision	Recall	F1Score	
0	LogisticRegression	BOW	91.485810	93.002915	92.196532	92.597968	
0		SVC	TFIDF	91.151920	92.463768	92.196532	92.329957
0	LogisticRegression	TFIDF	90.484140	90.476190	93.352601	91.891892	
0		SVC	BOW	90.150250	93.617021	89.017341	91.259259
0	DecisionTreeClassifier	BOW	85.809683	86.760563	89.017341	87.874465	

RAJAH 5. Keputusan Skor-F1 dalam TextBlob

```
result.sort_values("F1Score", ascending=False).head(5)
```

	Classifier	Methods	Accuracy	Precision	Recall	F1Score	
0		SVC	TFIDF	91.522763	92.631579	93.121693	92.875989
0	LogisticRegression		BOW	90.580848	91.406250	92.857143	92.125984
0	LogisticRegression		TFIDF	90.266876	89.108911	95.238095	92.071611
0		SVC	BOW	88.697017	88.250000	93.386243	90.745501
0	DecisionTreeClassifier		BOW	85.871272	88.709677	87.301587	88.000000

RAJAH 6. Keputusan Skor-F1 dalam Vader

C. Perbandingan antara Eksperimen 1 (Ex1) dan Eksperimen 2 (Ex2)

Daripada pemerhatian Jadual 6 hingga Jadual 13, didapati bahawa Ex1 (rangka kerja yang dicadangkan) mendapat hasil yang lebih baik daripada Ex2 (Asal). Walaupun terdapat beberapa keputusan kepekaan yang agak rendah, keputusan yang lain hampir cenderung kepada hasil rangka kerja yang dicadangkan. Justeru boleh disimpulkan bahawa keputusan rangka kerja yang dicadangkan adalah lebih baik.

JADUAL 6. Hasil Rangka Kerja Cadangan melalui BOW dengan LR

BOW+LR	Kejituan	Kepersisan	Kepekaan	Skor-F1
Ex2: Garis asas (Eksperimen 2)	82.550	82.429	87.925	85.088
Ex1: Senibina (Eksperimen 1)	<u>91.486</u>	<u>93.003</u>	<u>92.197</u>	<u>92.598</u>

JADUAL 7. Hasil Rangka Kerja Cadangan melalui BOW dengan DT

BOW+DT	Kejituan	Kepersisan	Kepekaan	Skor-F1
Ex2: Garis asas (Eksperimen 2)	76.638	74.502	<u>89.308</u>	81.236
Ex1: Senibina (Eksperimen 1)	<u>85.871</u>	<u>88.710</u>	87.302	<u>88.000</u>

JADUAL 8. Hasil Rangka Kerja Cadangan melalui BOW dengan SVM

BOW+SVM	Kejituan	Kepersisan	Kepekaan	Skor-F1
Ex2: Garis asas (Eksperimen 2)	79.202	77.668	88.805	82.864
Ex1: Senibina (Eksperimen 1)	<u>90.150</u>	<u>93.617</u>	<u>89.017</u>	<u>91.259</u>

JADUAL 9. Hasil Rangka Kerja Cadangan melalui BOW dengan KNN

BOW+KNN	Kejituan	Kepersisan	Kepekaan	Skor-F1
Ex2: Garis asas (Eksperimen 2)	62.678	60.431	<u>98.742</u>	74.976
Ex1: Senibina (Eksperimen 1)	<u>71.619</u>	<u>67.671</u>	97.399	<u>79.858</u>

JADUAL 10. Hasil Rangka Kerja Cadangan melalui TFIDF dengan LR

TFIDF+LR	Kejituan	Kepersisan	Kepekaan	Skor-F1
Ex2: Garis asas (Eksperimen 2)	81.695	82.885	85.283	84.067
Ex1: Senibina (Eksperimen 1)	<u>90.267</u>	<u>89.109</u>	<u>95.238</u>	<u>92.072</u>

JADUAL 11. Hasil Rangka Kerja Cadangan melalui TFIDF dengan DT

TFIDF+DT	Kejituan	Kepersisan	Kepekaan	Skor-F1
Ex2: Garis asas (Eksperimen 2)	73.718	72.468	86.415	78.830
Ex1: Senibina (Eksperimen 1)	<u>85.309</u>	<u>87.069</u>	<u>87.572</u>	<u>87.320</u>

JADUAL 12. Hasil Rangka Kerja Cadangan melalui TFIDF dengan SVM

TFIDF+SVM	Kejituan	Kepersisan	Kepekaan	Skor-F1
Ex2: Garis asas (Eksperimen 2)	84.117	84.129	88.679	86.344
Ex1: Senibina (Eksperimen 1)	<u>91.523</u>	<u>92.632</u>	<u>93.122</u>	<u>92.876</u>

JADUAL 13. Hasil Rangka Kerja Cadangan melalui TFIDF dengan KNN

TFIDF+KNN	Kejituan	Kepersisan	Kepekaan	Skor-F1
Ex2: Garis asas (Eksperimen 2)	60.399	58.911	<u>99.371</u>	73.970
Ex1: Senibina (Eksperimen 1)	<u>79.435</u>	<u>75.464</u>	96.825	<u>84.820</u>

D. Aplikasi COVID-19

Kajian ini seterusnya menggunakan model kemurungan yang dilatih untuk meramalkan ayat COVID-19, dan memerhatikan berapa banyak perkataan COVID-19 yang menggambarkan kemurungan. Data yang ditangkap oleh TWINT diberi anotasi sebagai label=1 (Twit COVID-19) selepas pemprosesan teks berikutnya. Tag berikut digunakan untuk mengikis

‘#kitajagakita, #MCO AND #PKP, #RMCO AND #PKPP, #CMCO AND #PKPB, #FMCO AND #PKP, #EMCO AND #PKPD’. Berikut adalah tiga pengelas terbaik dalam TextBlob dan Vader yang dipilih mengikut keputusan sebelumnya, iaitu TextBlob+LR+BOW, TextBlob+SVM+TFIDF, TextBlob+LR+TFIDF dan Vader+SVM+TFIDF, Vader+LR+BOW, Vader +LR+TFIDF. Rajah 7 menunjukkan keputusan 6 kombinasi pengelasan tersebut. 0 bermakna ayat tersebut tidak mempunyai kemurungan, dan 1 bermakna ayat tersebut mungkin berpotensi kemurungan. Didapati bahawa 6 keputusan pengelas ini tidak semestinya sama.

no	Tweets	tweet_original	TextBlob+LR+BOW	TextBlob+SVM+TFIDF	TextBlob+LR+TFIDF	Vader+SVM+TFIDF	Vader+LR+BOW	Vader+LR+TFIDF
0	besides voicin @Naven, Selva @smallSabri60 Besides voicing out, I think all we can expect is hope for a better #! #kitajagakita		0	0	0	1	0	0
1	to know more! To know more, and to listen to their voices, do check it out. Retweet for my small business, thank you 		0	0	0	0	0	0
2	this pkp is n this pkp 3.0 is making me a Swiftie 🙌 #taylorswift #swiftie #mco #pkp		0	0	0	1	0	0
3	why im creati Why im creating a home gym. #mco #pkp #sop #gym https://t.co/B1rz0xAaq		0	0	0	1	1	1
4	how long mor How long more we should suffer from this pandemic? I don't mean suffering from Covid. I mean suffering from fi		0	0	0	1	1	1
5	how many mc How many more months do you think we'll be under lockdown? 🙌 #PKP #MCO		1	1	0	1	1	1
6	how to cookir HOW TO COOKING HEALTHY FOOD IN HOME AND WHY ? #PKP #MCO #COOKING		0	0	0	1	1	1
7	just let us out Just let us out already - honestly wouldn't make any difference and we would keep our sanity. I've nearly cc		1	0	0	1	1	1
8	pkpd activity PKPD activity 🙌 We love doing Afro Moves. It is so fun. My daughter lost 4kg in 3 days doing this 🙌 Si		0	0	0	1	0	0
9	i believe this s I believe this sit here today. #mco #emco #pkp #pkpd https://t.co/0zcfjz2Xb		1	0	0	1	1	1
10	at this rate th At this rate that we're going, #MCO / #PKP ending at 2030 sounds plausible tho https://t.co/cnByVYN		0	0	0	1	0	0
11	nothing has cl nothing has changed... You all still can't move around. #mco #PKP3 #pkp #LockDown		0	0	0	1	0	0
12	whole countri Whole country under MCO and some places under EMCO ... But today case 7643... 🙌 #Covid19 #MCO #PKP		0	0	0	1	0	0
13	from my bella From my Bella to you #PKP3 #PKP #MCO #MCOCOUNTDOWN https://t.co/cc6aln925y		0	0	0	1	1	1
14	rambut pkp b Rambut PKP. Balah tangah! 🙌 Bring back backstreet boys style #pkp #mco https://t.co/aowCo2NE		0	0	0	1	0	0
15	as per the p ... [as per the Prime Minister's announcement today] 🙌 Current nationwide #MCO #PKP in "total" lockdown		1	0	0	1	1	1
16	bangsar on su Bangsar on Sunday used to be buzzing. It's now a ghost town. If this continues, we won't have to worr		0	0	0	1	1	1
17	congratulatio Congratulations! Automatic extension have been approved 🙌 #PKP #MCO #LOCKDOWN		0	0	0	0	0	0
18	day of move Day 465 of movement control order since 18 March 2020 [Nationwide #MCO #PKP] • 12 May - 7 June		0	0	0	1	1	1
19	honestly just j Honestly just put us in lockdown til august, instead of telling us it's only gonna be for 2 weeks for the past 1.5 ye		1	0	0	1	1	1
35	mco we hav! Mco 3.0 we have to face for another 2 weeks but please give us also a financial aid. #covid19 #mco #pkp		0	0	0	1	0	0

RAJAH 7. Keputusan 6 kombinasi pengelas

Selepas melengkapkan klasifikasi awal pembelajaran mesin, formula ‘countif’ (kaedah excel) digunakan untuk mengira berapa banyak kombinasi secara keseluruhan ayat itu harus diklasifikasikan sebagai kemurungan. Keputusan dibahagikan kepada 6 kategori mengikut tahap yang berbeza (jumlah kombinasi ayat tersebut yang diklasifikasi sebagai kemurungan). Daripada 6 hasil kombinasi yang ditunjukkan dalam Rajah 7, didapati bahawa setiap kombinasi mempunyai pertimbangan yang berbeza dalam menentukan sama ada ayat tersebut menunjukkan kemurungan. 0 bermakna ayat tersebut tidak menunjukkan kemurungan, dan 1 bermakna ayat tersebut mungkin berpotensi menunjukkan kemurungan. Dengan itu, kajian ini menghitung berapa kali label 1 muncul. Semakin tinggi tahapnya, semakin tinggi potensi ayat tersebut menunjukkan kemurungan. Untuk perkataan tahap 4 dan ke atas, perkataan tersebut dikira sebagai ayat yang berpotensi menunjukkan kemurungan. Seterusnya, tahap 0, 1, 2, 3 diabaikan. Hanya ayat tahap 4, 5, dan 6 digunakan untuk menapis ayat kemurungan yang berpotensi. Jadual 14 menunjukkan contoh ayat yang berpotensi menunjukkan kemurungan.

E. Perbandingan dengan Kajian Berkaitan

Untuk menentukan sumbangan kajian ini, bahagian ini melakukan perbandingan terhadap kajian yang berkaitan. Jadual 15 menunjukkan perbandingan keputusan berdasarkan Skor-F1.

Kajian Hassan et al. (2017) menyatakan bahawa penyelidik menggunakan pelbagai alat pembelajaran mesin untuk mengelaskan sentimen. Walau bagaimanapun, tiada pengelas tunggal yang terbaik untuk semua jenis set data. Oleh itu, tiga pengelas yang berbeza telah digabungkan dengan menggunakan pendekatan pengundian. Selain itu, kajian Deshpande dan Rao (2017) menggunakan Emotion AI bagi mengesan kemurungan. Dalam proses ramalan kelas, mesin SVM dan pengelas NB digunakan. Di samping itu, kajian Peng et al. (2019) membina kamus ciri emosi untuk analisis kemurungan mengikut kaedah pengelas CALO, serta mencadangkan model pengecaman kemurungan berdasarkan SVM berbilang kernel. Selain itu, objektif utama kajian Tadesse et al. (2019) adalah untuk memeriksa siaran pengguna Reddit untuk mengesan sebarang faktor yang mungkin mendedahkan sikap kemurungan pengguna dalam talian yang berkaitan. Kajian AlSagri dan Ykhlef (2020) menggunakan kaedah

pembelajaran mesin untuk menganalisis tingkah laku rangkaian dan Twit pengguna. Ia mengenalpasti sama ada pengguna mengalami tekanan berdasarkan ciri yang diekstrak daripada aktiviti pengguna dalam rangkaian. Selain itu, kajian Bhargava (2021) menggunakan data sentimen140 sebagai latihan model. Literatur juga menjelaskan proses mengesan kemurungan yang lengkap. Kaedah pengekstrakan ciri yang digunakan adalah word2vec, manakala bahagian latihan model menggunakan LSTM dan CNN-LSTM.

Berdasarkan keputusan pada Jadual 15, didapati bahawa keputusan kajian ini sedikit bertambah baik secara keseluruhannya.

JADUAL 14. Contoh ayat berpotensi kemurungan

Twits	Tahap
My anxiety just kicked in. #PKP #MCO	4
To be honest, not looking forward to this MCO being lifted in Malaysia. Having a lot of mix feelings about it... 😞 #MCO #PKP #Malaysia	4
if u're not effected in any way during these hard times, it is better to say nothing, anything, to anyone or on social media, no matter your age, race, gender, status #CMCO #PKPB #COVID19	4
Extend loan moratorium please. No money to pay house loan..still trying hard to land a job. Part time job's pay is not enough..need to feed family of five... #CMCO #PKPB	4
Working from home, again. I dislike WFH but I guess we have to get used to the new norm. I'm still gonna ask for a letter from the Human Resource Dept tho so I can go to the office. Memang I hate working from home!! #PKP #MCO	5
When will be the normal life 😞 #mco #pkp #CircuitBreaker	5
This is ME when pkpb sambung. I'm Dead #PKPB #CMCO https://t.co/2lmOynWq1O	5
Real stories of lived realities. It is not just about the disease, but about people. #PKPB #CMCO #CMCOIsATypeOfLockdown #LockdownsAreNotSolutions	5
Just found out architecture firms are going through a very tough time. Salary cuts and deferments, forced unpaid leave ... including the second firm I worked at upon graduation. Hopefully nobody gets retrenched. 😞 #MCO #PKP #jobs	6
I can feel the MCO level stress, uncertainty and anxiety returning. I hate this. #PKPB #CMCO #MuhyiddinOut	6
#CMCO #PKPB I am so worried about all the small businesses...the people who earn a living but have to close again... seriously i sooo wish this was not happening again. 😞 Hang in there ok..i know its gonna be even tougher this time #KitaJagaKita	6
#PKPB #CMCO 2020 is paused again. How do you feel guys? Is this normal or you can still feel the anxiety?	6

JADUAL 15. Perbandingan dengan Kajian Berkaitan

	(Hassan et al. 2017)	(Deshpande & Rao 2017)	(Peng, Hu, & Dang, 2017)	(Tadesse et al. 2019)	(AlSagri & Ykhlef 2020)	(Bhargava 2021)	Hasil Cadangan
Kejituan	91%	83%	83.46%	91%	82.5%	91.35%	91.523
Kepersisan	83%	83.6%	75.56%	93%	73.91%	92%	92.632
Kepekaan	79%	83%	76.69%	90%	85%	91%	93.122
Skor-F1	-	83.29%	76.12%	92%	77.5%	91%	92.876 *

RUMUSAN

Kajian ini berjaya mencapai objektif kajian yang dicadangkan iaitu untuk membangunkan seni bina yang dapat meningkatkan ketepatan pengesanan kemurungan. Pencapaian pertama diwakili oleh cadangan aliran kerja seni bina untuk menyalurkan set data latihan kemurungan secara automatik. Seni bina yang dicadangkan menggunakan penapis dan melalui TextBlob dan Vader (Eksperimen 1) adalah lebih baik daripada garis dasar tanpa penapis dari segi pengesanan kemurungan (Eksperimen 2). Penemuan ini menunjukkan keberkesanan cadangan seni bina dalam Eksperimen 1 untuk mengekstrak kemurungan. Keputusan baik yang dicapai dalam Eksperimen 1 menunjukkan bahawa hubungan semantik telah dikenalpasti dengan betul berbanding garis dasar Eksperimen 2 dengan pencapaian Skor-F1 92.59% dan 92.88% melalui BOW+LR dan SVM+TFIDF berbanding Skor-F1 86.34% yang diperolehi oleh garis dasar. Objektif kedua yang bertujuan menilai seni bina yang terhasil melalui seni bina yang dicadangkan dalam objektif 1 berjaya dicapai dengan menggunakan pengelas RF, DT, SVM, dan KNN. Perbandingan telah dijalankan ke atas garis dasar, di mana seni bina yang dicadangkan menunjukkan hasil yang lebih baik berbanding keputusan garis dasar. Selain itu, LR dan DT mencapai hasil terbaik dalam BOW, manakala SVM dan KNN mencapai hasil terbaik apabila menggunakan TFIDF. Kesimpulannya, seni bina yang dicadangkan telah menunjukkan kepentingannya. Ini membuktikan bahawa struktur seni bina eksperimen memainkan peranan penting dalam pengesanan kemurungan.

RUJUKAN

- Abd Malek, N.S.A., Raop, N.A. & Hassan, M.S. 2020. Peranan Kesihatan Mental Sebagai Moderator Terhadap Kecenderungan Bunuh Diri. *Jurnal Sains Sosial: Malaysian Journal of Social Sciences* 5(1): 87–99.
- Abd Rahman, R., Omar, K., Noah, S.A.M., Danuri, M.S.N.M. & Al-Garadi, M.A. 2020. Application of machine learning methods in mental health detection: a systematic review. *IEEE Access* 8: 183952–183964.
- Abdul Aziz, Abdul Rashid & Mohd Sukor, Nurhafizah & Hamizah, Dan. 2020. Wabak COVID-19: Pengurusan Aspek Kesihatan Mental Semasa Norma Baharu.
- AlSagri, H.S. & Ykhlef, M. 2020. Machine learning-based approach for depression detection in twitter using content and activity features. *IEICE Transactions on Information and Systems* 103(8): 1825–1832.
- Andersen, I. 2018. Everything You Need to Know About Hashtags.
- Aris, A. 4.2 Juta Kemurungan. *Harian Metro*.
- Beckstein, A., Rathakrishnan, B., Hutchings, P.B. & Mohamed, N.H. 2021. The COVID-19 pandemic and mental health in Malaysia: Current treatment and future recommendations. *Malaysian Journal of Public Health Medicine* 21(1): 260–267.

- Berg, J. 2014. The story of the hashtag(#): A practical theological tracing of the hashtag(#) symbol on Twitter. *HTS Teologiese Studies / Theological Studies* 70.
- Bernamea. Stres pembunuh dalam senyap. *Harian Metro*.
- Bhargava, C. 2021. Depression Detection Using Sentiment Analysis of Tweets. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12(11): 5411–5418.
- Boyd, D.M. & Ellison, N.B. 2007. Social network sites: Definition, history, and scholarship. *Journal of Computer-mediated Communication* 13(1): 210–230.
- Bruce, D.F. 2020. Depression Diagnosis.
- Burgess, J. 2020. Friday essay: Twitter and the way of the hashtag.
- Chiong, R., Budhi, G.S. & Dhakal, S. 2021. Combining sentiment lexicons and content-based features for depression detection. *IEEE Intelligent Systems* 36(6): 99–105.
- Deshpande, M. & Rao, V. 2017. Depression detection using emotion artificial intelligence. In *2017 International Conference on Intelligent Sustainable Systems (Iciss)*. pp. 858–862. IEEE.
- DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A. & Lhommet, M. 2014. SimSensei Kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*. pp. 1061–1068.
- Dr. Wan Mohd Rushidi. Bunuh diri, sakit jiwa boleh dicegah. *Berita Harian Online*.
- Eisenstein, J. 2017. Unsupervised learning for lexicon-based classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 31.
- Erfani, E., Samy Helmy Hanna, A. & Boroon, L. 2021. Social network sites use and psychological distress: A systematic review. In *Proceedings of the 54th Hawaii International Conference on System Sciences*. p. 2988.
- Feature Extraction. (n.d.).
- Giannakakis, G., Padiaditis, M., Manousos, D., Kazantzaki, E., Chiarugi, F., Simos, P.G., Marias, K. & Tsiknakis, M. 2017. Stress and anxiety detection using facial cues from videos. *Biomedical Signal Processing and Control* 31: 89–101.
- Gkotsis, G., Oellrich, A., Velupillai, S., Liakata, M., Hubbard, T.J.P., Dobson, R.J.B. & Dutta, R. 2017. Characterisation of mental health conditions in social media using Informed Deep Learning. *Scientific Reports* 7(1): 1–11.
- Go, A., Bhayani, R. & Huang, L. 2009. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford* 1(12): 2009.
- Haccr. 2019. TWINT - twintproject. github.
- Hamdan, H., Bellot, P. & Bechet, F. 2015. lsislif: Feature extraction and label weighting for sentiment analysis in twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. pp. 568–573.
- Hamouda, A., Marei, M. & Rohaim, M. 2011. Building machine learning based senti-word lexicon for sentiment analysis. *Journal of Advances in Information Technology* 2(4): 199–203.
- Hassan, A.U., Hussain, J., Hussain, M., Sadiq, M. & Lee, S. 2017. Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression. In *2017 International Conference on Information and Communication Technology Convergence (ICTC)*. pp. 138–140. IEEE.
- Holmes, E.A., O'Connor, R.C., Perry, V.H., Tracey, I., Wessely, S., Arseneault, L., Ballard, C., Christensen, H., Silver, R.C. & Everall, I. 2020. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *The Lancet Psychiatry* 7(6): 547–560.

- Hutto, C. & Gilbert, E. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International AAAI Conference on Web and Social Media. Vol. 8 pp. 216–225.
- K, A. 2021. DDM — Detecting Depression in Tweets.
- Kannan, S., Karuppusamy, S., Nedunchezian, A., Venkateshan, P., Wang, P., Bojja, N. & Kejariwal, A. 2016. Chapter 3 - Big Data Analytics for Social Media. In Buyya, R., Calheiros, R.N., & Dastjerdi, A.V.B.T.-B.D. (eds.). pp. 63–94. Morgan Kaufmann.
- Kazanova. (n.d.). Sentiment140 dataset with 1.6 million tweets.
- Kolchyna, O., Souza, T.T.P., Treleaven, P. & Aste, T. 2015. Twitter sentiment analysis: Lexicon method, machine learning method and their combination. ArXiv Preprint ArXiv:1507.00955.
- Madasu, A. 2019. A Study of Feature Extraction techniques for Sentiment Analysis. ArXiv Preprint ArXiv:1906.01573.
- Malaysia, K. 2019. Laporan Tahunan Kementerian Kesihatan Malaysia 2019.
- Malaysia, K.K. 2018. Laporan Tahunan Kementerian Kesihatan Malaysia 2018. Centre HI, Editor. Malaysia: Kementerian Kesihatan Malaysia.
- Martinez, V.R. 2019. A machine learning approach for the detection of depression and mental illness in Twitter.
- Malaysia, K.K. 2022. Kenyataan Media Kementerian Kesihatan Malaysia, Sambutan Hari Kesihatan Mental Sedunia 2022. *Centre HI, Editor. Malaysia: Kementerian Kesihatan Malaysia.*
- Mental Health Handbook. 2019. (MMHA), Malaysian Mental Health Association, Malaysian Psychiatric Association, Pfizer (Malaysia) Sdn Bhd.
- Organization, W.H. 1991. Guidelines for the assessment of herbal medicines. Report No. . World Health Organization.
- Peng, Z., Hu, Q. & Dang, J. 2019. Multi-kernel SVM based depression recognition using social media data. *International Journal of Machine Learning and Cybernetics* 10(1): 43–57.
- Rajaraman, P. V, Nath, A., Akshaya, P.R. & Bhujja, C. 2020. Depression Detection of Tweets and A Comparative Test. *International Journal of Engineering Research & Technology* 9.
- Ramos, R.M., Ferrer-Cheng, P.G., Bitsch, J.Á. & Jonas, S.M. 2016. HOW DO I SAY “SAD?” BUILDING A DEPRESSION-LEXICON FOR PSYCHOLOGIST IN A POCKET.
- Rosli Hamid, Muhammad Faizal A. Ghani, Syed Kamaruzaan Syed Ali, Megat Ahmad Kamaludin Megat Daud, Rahma Dewi. 2020. Kemurungan, Kebimbangan dan Tekanan dalam Kalangan Pelajar Tingkatan Empat di Daerah Kota Setar. *Jurnal Kepimpinan Pendidikan*. Oktober 2020, Bil. 7, Isu 4.
- SemEval - International Workshop on Semantic Evaluation. (n.d.).
- Shatte, A.B.R., Hutchinson, D.M. & Teague, S.J. 2019. Machine learning in mental health: a scoping review of methods and applications. *Psychological Medicine* 49(9): 1426–1448.
- Sommar, F. & Wielondek, M. 2015. Combining Lexicon-and Learning-based Approaches for Improved Performance and Convenience in Sentiment Classification.
- Srividya, M., Mohanavalli, S. & Bhalaji, N. 2018. Behavioral modeling for mental health using machine learning algorithms. *Journal of Medical Systems* 42(5): 1–12.
- Swewang. (n.d.). depression detection.
- Tadesse, M.M., Lin, H., Xu, B. & Yang, L. 2019. Detection of depression-related posts in reddit social media forum. *IEEE Access* 7: 44883–44893.
- TextBlob: Simplified Text Processing. (n.d.).
- TWINT: A Twitter information scraping tool. 2019.
- What Is Depression? 2021. American Psychiatric Association.