

MENGENALPASTI KEMURUNGAN BERDASARKAN SIMPTOM KEMURUNGAN DALAM TEKS SOSIAL MEDIA

MOHANDASS A/L THIRUVASAGAM
DR. LAILATUL QADRI BINTI ZAKARIA

*Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia, 43600 UKM Bangi,
Selangor Darul Ehsan, Malaysia*

ABSTRAK

Dalam era moden, orang ramai muncul dengan kemurungan sebagai masalah sosial utama. Kemurungan ialah mood rendah yang berlarutan dalam jangka masa yang lama, menjejaskan kehidupan seharian. Ia sering dicetuskan oleh gabungan faktor genetik, psikologi dan persekitaran. Kemurungan adalah masalah sosial yang kritikal di seluruh dunia, tetapi alat untuk mendiagnosis mereka secara objektif adalah kurang dan ia menunjukkan kadar penggunaan perkhidmatan kesihatan mental yang rendah. Berdasarkan kajian menunjukkan bahawa risiko menjadi tertekan boleh ditingkatkan dengan peristiwa kehidupan seperti kemiskinan, kematian orang tersayang, penyakit fizikal atau penderaan terutamanya remaja. Penyelidikan menunjukkan bahawa remaja yang menghabiskan lebih banyak masa di media sosial juga berasa lebih terasing. Tujuan kajian ini adalah untuk mengklasifikasikan ayat yang ditulis oleh pengguna media sosial berdasarkan gejala kemurungan dalam Soal Selidik Kesihatan Pesakit (PHQ-9), menggunakan leksikon emosi untuk menilai secara semula jadi pengguna kemurungan berdasarkan keputusan mereka. Set data yang akan digunakan merupakan set data sumber terbuka berkaitan dengan kesihatan mental yang telah tersedia dan diperoleh daripada aplikasi Twitter. Dalam projek ini menggunakan teknik pembelajaran mesin untuk menentukan simptom depresi berdasarkan simptom- simptom terpilih yang terdapat di dalam instrumen Soal Selidik Kesihatan Pesakit-9 (PHQ-9). Dengan menggunakan leksikon depresi dapat menganalisis dan mengklasifikasikan simptom pesakit kepada positif dan negatif tentang tahap kemurungan mereka yang berkenaan. Di samping itu, ia juga penting untuk mengetahui kekerapan perkataan yang sering muncul dalam gejala biasa pesakit tersebut. Projek ini juga akan turut membangunkan aplikasi web ringkas yang bernama *DepresCare* bagi memudahkan para pengguna mengesan simptom depresi yang terdapat dalam teks. Pengguna aplikasi web *DepresCare* dapat membantu pengguna untuk mengetahui tahap kemurungan dengan mengenali simptom depresi mereka. Kajian ini penting kerana ia menunjukkan kemungkinan menentukan kemurungan dengan menggunakan aplikasi web *DepresCare* berdasarkan data teks.

Kata kunci: PHQ-9, "leksikon", Twitter, *DepresCare*

PENGENALAN

Kemurungan adalah penyakit yang mengancam kesihatan mental orang moden dan diiktiraf sebagai masalah yang perlu diselesaikan, tetapi terdapat kekurangan pemahaman dan persetujuan mengenai rawatan yang betul untuk kemurungan. Kemurungan membawa kepada

kemerosotan dalam fungsi kehidupan seharian, dengan simptom utamanya hilang semangat dan berasa sedih atau tidak gembira. Pertubuhan Kesihatan Sedunia (WHO) menjangkakan kemurungan menjadi penyakit paling membebankan manusia pada 2030, dengan lebih 264 juta orang di dunia yang menghidapnya. WHO juga menyatakan bahawa, secara global, kemurungan adalah punca utama ketidakupayaan dan menyumbang kepada beban penyakit kepada orang ramai (Woody, 2023). Oleh itu, status kesihatan mental orang moden telah muncul sebagai masalah sosial utama dan telah mula dianggap sebagai masalah yang tidak boleh diabaikan lagi.

Menurut penyelidikan menunjukkan bahawa mereka adalah tiga sebab utama untuk penggunaan rendah perkhidmatan kesihatan mental adalah keperluan yang dianggap rendah, halangan struktur, dan halangan sikap. Keperluan yang dirasakan rendah ialah kekurangan kesedaran tentang isu kesihatan mental dan bermakna pesakit itu sendiri berpendapat tiada bantuan diperlukan. Halangan struktur mewakili kebimbangan tentang wang, kekurangan masa, kebolehcapaian, perlindungan insurans, dsb. Halangan sikap termasuk idea bahawa gangguan mental akan membaik dengan sendirinya, prejudis terhadap perkhidmatan kesihatan mental, dan ketidakpercayaan terhadap kesan rawatan, yang mengakibatkan pesakit gagal menggunakan perkhidmatan tersebut. Berbeza dengan penyakit fizikal, pesakit yang menghidap 2 penyakit mental sering tidak memahami tahap penyakit, sering tidak mendapat rawatan kerana motivasi yang rendah, dan sering tidak tahu bahawa masalah boleh diperbaiki dengan menggunakan perkhidmatan kesihatan mental. Secara umum, penyakit mental boleh dirawat sepenuhnya dengan campur tangan awal, tetapi semakin lewat rawatan, semakin serius gangguan itu, jadi kesedaran yang betul dan pengesanan awal penyakit mental adalah langkah penting dan perlu dalam merawat penyakit. Tambahan pula, mengenali penyakit dan mengetahui nama tepatnya meningkatkan kebarangkalian pengesanan awal dan meningkatkan kesan rawatan yang positif.

Salah satu langkah penting dalam merawat kemurungan ialah kesedaran diri yang betul tentang keadaan ini. Diagnosis diri kemurungan membolehkan orang ramai menyemak sendiri tahap kemurungan mereka, dan terdapat banyak ujian diagnosis diri untuk kemurungan. Contoh instrumen diagnosis diri untuk kemurungan termasuk Inventori Kemurungan Beck (BDI), Pusat Skala Kemurungan Kajian Epidemiologi (CES-D), Soal Selidik Kesihatan Pesakit-9 (PHQ-9), dan Skala Kemurungan Geriatrik (GDS). Terdapat pelbagai jadual diagnosis diri untuk kemurungan, tetapi sukar bagi orang kurang upaya mental untuk mengenal pasti keadaan mereka melalui kaedah ini atas sebab yang sama seperti kadar penggunaan rendah perkhidmatan kesihatan mental. Oleh itu, alternatif boleh menjadi sistem yang secara automatik (tanpa penglibatan pesakit tertentu) mengenal pasti tahap kemurungan dalam pesakit tersebut.

METODOLOGI KAJIAN

I Fasa Pengumpulan Data

Fasa pengumpulan data merupakan langkah pertama yang perlu dilakukan dalam kajian analisis emosi ini. Dalam kajian ini, data iaitu teks akan diambil daripada pengguna media sosial yang mempunyai simptom masalah kemurungan berdasarkan emosi yang ditunjukkan melalui teks. Set data yang digunakan bagi projek ini dikumpul melalui laman sosial iaitu Twitter. Semua data tersebut akan disimpan dalam 6 format .csv untuk kemudahan penggunaan pada fasa-fasa seterusnya. Data yang dikumpulkan adalah data yang mentah dan mempunyai jumlah yang besar.

II Fasa Prapemprosesan

Prapemprosesan teks ialah langkah penting dalam pemprosesan bahasa semula jadi (NLP) yang melibatkan pembersihan dan mengubah data teks tidak berstruktur untuk menyediakannya untuk analisis. Ia termasuk tokenisasi, penyingkiran kata henti sebelum menjalankan analisis emosi. Fasa ini penting kerana ia boleh memberikan kualiti hasil kajian yang tepat. Kemudian, untuk pembersihan data, tokenisasi ialah proses menggantikan data sensitif dengan simbol pengenalan unik yang mengekalkan semua maklumat penting tentang data tanpa menjejaskan keselamatannya. Akhir sekali, data yang telah dikumpulkan di fasa ini akan disimpan untuk digunakan di fasa seterusnya.

III Fasa Pengekstrakan Ciri

Dalam kajian ini fasa pengekstrakan ciri menggunakan analisis emosi. Fasa pengekstrakan teks ialah proses mencari dan mengekstrak butiran atau ciri yang berkaitan daripada teks data yang boleh digunakan untuk mengesan emosi dan mengaitkan emosi. Dalam fasa ini terdiri daripada model Bag of words adalah salah satu teknik untuk mengekstrak data. Ia dipanggil (BOW) kerana sebarang maklumat tentang susunan atau struktur perkataan dalam data tidak diambil kira dan akan melihat data yang diketahui sahaja (Purva,2023). Selain itu, Frekuensi Istilah-Frekuensi Dokumen Terbalik (TF-IDF). Ia mengukur betapa pentingnya istilah dalam dokumen berbanding koleksi dokumen (iaitu, relatif kepada korpus). Di samping itu, Leksikon ialah koleksi perkataan atau frasa yang dikaitkan dengan keadaan emosi iaitu gembira, sedih atau kecewa. Model ini dapat menentukan polariti tanggapan atau pendapat terhadap masalah kemurungan.

IV Fasa pengelasan

Fasa pengelasan merupakan fasa yang penting dalam mesin algoritma pembelajaran. Teknik pengesanan emosi Pendekatan berasaskan Leksikon Pendekatan ialah pendekatan carian berasaskan kata kunci yang mencari kata kunci emosi yang diberikan kepada beberapa keadaan psikologi (Rabeya et al. 2017). Penggunaan algoritma emosi analisis data akan diterapkan untuk menganalisis dan mengekstrak simptom-simptom depresi daripada dataset.

V Fasa Penilaian

Fasa ini akan memaparkan keputusan analisis emosi berdasarkan emosi yang telah dianalisis

tergolong dalam emosi leksikon. Seterusnya, dapat mengambil tindakan awal apabila dapat mengenal pasti kemurungan terhadap seseorang yang menghadapi masalah tersebut berdasarkan data yang telah dikaji. Akhir sekali, hasil keputusan akan divisualisasikan dalam bentuk graf supaya mudah difahami dan digunakan sebagai rujukkan.

KEPUTUSAN DAN PERBINCANGAN

Pembangunan Model

Projek ini telah melibatkan beberapa langkah seperti kemaskini data, muat naik data, pra-pemprosesan data dan pengekstrakan maklumat penting untuk dijadikan sebagai data baharu yang akan menghasilkan keputusan model yang terbaik bagi tujuan klasifikasi simptom depresi. Pada fasa ini, pengemaskinian data terhadap leksikon depresi telah dilakukan bagi memastikan setiap simptom depresi yang terpilih mengandungi semua istilah yang digunakan untuk memastikan leksikon merujuk kepada simptom tersebut.

Seterusnya, proses memuat naik data pada perisian pengaturcaraan *Google Colab* melibatkan enam fail excel yang mengandungi 15 lajur berkaitan dengan data pengguna berkenaan kesihatan mental yang telah diperolehi daripada aplikasi *Twitter* dan fail json yang mengandungi 10 lajur berkaitan leksikon depresi. Namun, hanya tiga lajur utama daripada fail json yang akan digunakan iaitu lajur "*signal_3*", lajur "*signal_4*", dan lajur "*signal_8*" yang menunjukkan leksikon depresi bagi simptom ketiga, keempat dan kelapan daripada instrumen Soal Selidik Kesihatan Pesakit-9 (PHQ-9). Bagi fail excel, hanya dua lajur utama sahaja yang akan digunakan bagi projek ini, iaitu lajur "*tweet_processed*" dan lajur "*Target*". Lajur "*tweet_processed*" mengandungi teks yang telah dikumpul daripada aplikasi *Twitter* berkenaan isu depresi manakala lajur "*Target*" merupakan label teks tersebut menunjukkan depresi ataupun tidak. Bagi lajur ini, label "0" menunjukkan teks itu tidak mengandungi unsur depresi manakala label "1" menunjukkan teks itu mengandungi unsur depresi.

Selain itu, pra-pemprosesan data telah dilakukan bagi memastikan data tersebut meningkatkan produktiviti keseluruhan dan membolehkan maklumat berkualiti tinggi dalam membuat keputusan. Antara pembersihan data yang dilakukan adalah membuang tanda baca seperti apostrof dan tanda seru, membuang teks berulang, membuang nombor dan juga menukar kepada huruf kecil. Pembuangan kata henti terpilih juga turut dilakukan bagi memastikan data yang akan digunakan adalah bersih dan relevan untuk klasifikasi. Pembuangan baris yang tidak mempunyai data dengan klasifikasi simptom depresi supaya data tersebut tidak menyumbang kepada model pembelajaran mesin bagi melakukan prediksi. Teknik- Teknik tersebut digunakan untuk memastikan setiap data mempunyai format yang sama untuk memudahkan penggunaan pada peringkat seterusnya iaitu pepadanan rentetan (*String Matching*).

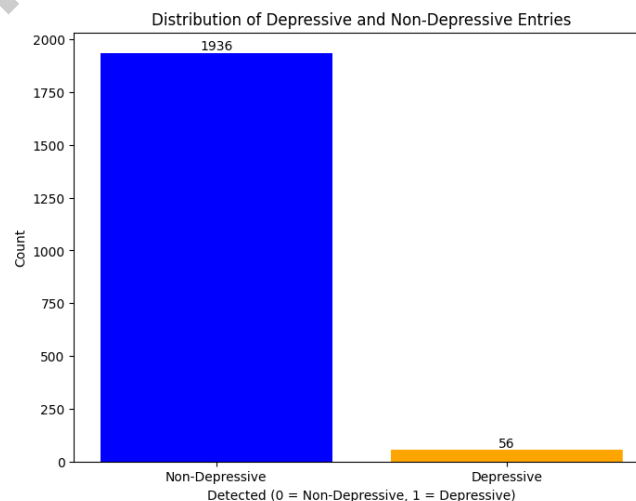
Akhir Sekali, Pemadanan rentetan ini telah melibatkan dua rangka data pandas iaitu rangka data “mergedata” dan “combine_lexicon”. Pada rangka data “mergedata”, lajur “tweet_processed” akan digunakan untuk membuat pemadanan rentetan. Pada rangka data “combine_lexicon” pula, hanya lajur “Trouble falling or staying asleep, or sleeping too much?”, “Feeling tired or having little energy?” dan “Moving or speaking so slowly that other people could have noticed.?” digunakan untuk membuat pemadanan rentetan. Kedua-dua rangka data ini digunakan untuk membuat pemadanan dengan mengekstrak istilah ataupun leksikon depresi bagi simptom ketiga, keempat dan kelapan di dalam instrumen Soal Selidik Kesihatan Pesakit-9 (PHQ-9) yang terdapat di dalam setiap teks. Daripada hasil pemadanan tersebut disimpan ke dalam rangka data yang baharu bernama “all_values_df”. Rangka data ini mengandungi beberapa lajur penting iaitu teks yang mengandungi tweet yang telah diproses (“tweet_processed”), senarai istilah yang sepadan untuk setiap simptom (“PHQ-3 (auto)”, “PHQ-4 (auto)”, “PHQ-8 (auto)”), jumlah kehadiran istilah untuk setiap simptom (“PHQ-3_Count”, “PHQ-4_Count”, “PHQ-8_Count”), penunjuk kewujudan simptom (“Detected”).

Hasil pemadanan ini dapat mencipta fitur baharu yang boleh digunakan di dalam model pembelajaran mesin pada peringkat yang seterusnya untuk membantu meningkatkan prestasi model.

Klasifikasi Data Simptom Depresi

Bagi mengklasifikasikan data simptom depresi akan dilakukan dengan menggunakan data-data yang berbeza iaitu data keseluruhan dan data seimbang. Berikut adalah data yang akan digunakan bagi membuat prediksi pada peringkat seterusnya.

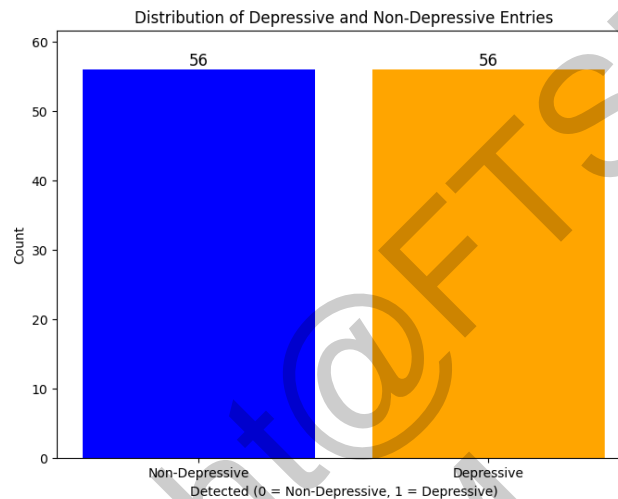
i. Data Keseluruhan



Rajah 1 Data Keseluruhan

Data Keseluruhan merupakan data keseluruhan yang akan digunakan sebanyak 1992 data. Rajah 1 menunjukkan keseluruhan data yang terdiri daripada 1936 jumlah teks yang menunjukkan bukan depresi manakala 56 jumlah teks yang menunjukkan depresi. Data ini merupakan data umum dan bercampur dengan data yang mempunyai simptom yang sepadan dengan keseluruhan simptom yang digunakan.

ii. Data Seimbang



Rajah 2 Data Seimbang

Data seimbang, merupakan data yang mempunyai jumlah yang sama bagi kedua - dua kategori teks. Data seimbang ini juga merangkumi 56 teks yang menunjukkan depresi dan tidak depresi seperti Rajah 2. Teks yang dikategorikan sebagai depresi ini mempunyai data yang sepadan dengan simptom yang digunakan. Bagi memastikan data ini sepadan, perlu mengambil data rawak yang tidak mempunyai depresi untuk menjadikan data itu seimbang bagi menjalankan prediksi seterusnya.

Menentukan Fitur Dengan Menggunakan Pycaret

i. Fitur Keseluruhan Simptom yang Dikesan (*Target*)

```
# Create a new DataFrame with only necessary columns
data = mergedata[['tweet_processed', 'Detected']]
data
```

	tweet_processed	Detected
0	mood caused infectious diseases nutritional de...	0
1	unnecessary family drama feel like moving far aw...	0
2	i'm going keep banging cos true focus get stop...	0
3	ugh woke acne covered face morning yyayyy	0
4	portable transparent tobacco rolling machine p...	0
...
2004	done absolute fuck last days	0
2005	trying make list ppl twt regularly u wanna b t...	0
2006	year old son looks go new year's eve party stu...	0
2007	perpetrators admit never fault victim scarred ...	0
2008	currently accepting new clients online mindful...	0

Rajah 3 Fitur Keseluruhan Simptom yang Dikesan (*Target*)

Berdasarkan Rajah 3 menunjukkan data teks yang telah diproses akan digunakan dan keseluruhan simptom yang sepadan dengan dengan simptom untuk membuat klasifikasi.

ii. Fitur PHQ-3_Count

```
# Create a new DataFrame with only necessary columns
data = mergedata[['tweet_processed', 'PHQ-3_Count']]
data
```

	tweet_processed	PHQ-3_Count
0	mood caused infectious diseases nutritional de...	0
1	unnecessary family drama feel like moving far aw...	0
2	i'm going keep banging cos true focus get stop...	0
3	ugh woke acne covered face morning yyayyy	0
4	portable transparent tobacco rolling machine p...	0
...
2004	done absolute fuck last days	0
2005	trying make list ppl twt regularly u wanna b t...	0
2006	year old son looks go new year's eve party stu...	0
2007	perpetrators admit never fault victim scarred ...	0
2008	currently accepting new clients online mindful...	0

Rajah 4 Fitur *PHQ-3_Count*

Berdasarkan Rajah 4 menunjukkan data teks yang telah diproses akan digunakan dan simptom dari *PHQ-3_Count* yang sepadan dengan dengan simptom untuk membuat klasifikasi.

iii. Fitur PHQ-4_Count

```
# Create a new DataFrame with only necessary columns
data = mergedata[['tweet_processed', 'PHQ-4_Count']]
data
```

	tweet_processed	PHQ-4_Count
0	mood caused infectious diseases nutritional de...	0
1	unnecessary family drama feel like moving far aw...	0
2	i'm going keep banging cos true focus get stop...	0
3	ugh woke acne covered face morning yyayyy	0
4	portable transparent tobacco rolling machine p...	0
...
2004	done absolute fuck last days	0
2005	trying make list ppl twt regularly u wanna b t...	0
2006	year old son looks go new year's eve party stu...	0
2007	perpetrators admit never fault victim scarred ...	0
2008	currently accepting new clients online mindful...	0

Rajah 5 Fitur *PHQ-4_Count*

Berdasarkan Rajah 5 menunjukkan data teks yang telah diproses akan digunakan dan simptom dari *PHQ-4_Count* yang sepadan dengan dengan simptom untuk membuat klasifikasi.

iv. Fitur *PHQ-8_Count*

```
# Create a new DataFrame with only necessary columns
data = mergedata[['tweet_processed', 'PHQ-8_Count']]
data
```

	tweet_processed	PHQ-8_Count
0	mood caused infectious diseases nutritional de...	0
1	unnecessary family drama feel like moving far aw...	0
2	i'm going keep banging cos true focus get stop...	0
3	ugh woke acne covered face morning yyayyy	0
4	portable transparent tobacco rolling machine p...	0
...
2004	done absolute fuck last days	0
2005	trying make list ppl twt regularly u wanna b t...	0
2006	year old son looks go new year's eve party stu...	0
2007	perpetrators admit never fault victim scarred ...	0
2008	currently accepting new clients online mindful...	0

Rajah 6 Fitur *PHQ-8_Count*

Berdasarkan Rajah 6 menunjukkan data teks yang telah diproses akan digunakan dan simptom dari *PHQ-8_Count* yang sepadan dengan dengan simptom untuk membuat klasifikasi.

Klasifikasi Data Dengan Menggunakan Pycaret

PyCaret ialah perpustakaan pembelajaran mesin kod rendah sumber terbuka dalam Python yang mengautomatiskan proses latihan dan menggunakan model pembelajaran mesin. Ia direka

bentuk supaya mudah digunakan dan boleh mempercepatkan pembangunan model pembelajaran mesin dengan ketara. *PyCaret* mempunyai modul pengelasan yang digunakan untuk mengelaskan elemen kepada Kumpulan dengan tujuan untuk meramalkan label kelas kategori (*PyCaret Binary Classification*). Dalam fasa ini, *PyCaret* akan digunakan untuk meramalkan prediksi pengelasan teks kepada kategori teks kemurungan dan teks bukan kemurungan.

i. **Menentukan Model dengan menggunakan Data berbeza**

Dalam Kajian ini tiga model yang terkenal iaitu model Regresi Logistik, Pohon Keputusan (*Decision Tree*) dan K-Jiran Terdekat akan digunakan untuk membangunkan model.

Jadual 1 Keputusan Data Keseluruhan

Metrik Penilaian	Ketepatan	Kejituan	Dapatan Semula	F-Skor
Regresi Logistik	0.9910	1.0000	0.7150	0.8067
Pohon Keputusan	0.9716	0.2000	0.0750	0.1067
K-Jiran Terdekat	1.0000	1.0000	1.0000	1.0000

Jadual 2 Keputusan Data Seimbang

Metrik Penilaian	Ketepatan	Kejituan	Dapatan Semula	F-Skor
Regresi Logistik	0.5643	0.5679	0.7167	0.6208
Pohon Keputusan	0.7054	0.6833	0.5333	0.5800
K-Jiran Terdekat	0.5214	0.4967	0.4917	0.4746

Setelah membuat perbandingan dengan tiga model yang terkenal iaitu model Regresi Logistik, Decision Tree dan K-Jiran Terdekat, Regresi Logistik merupakan model yang sesuai digunakan kerana ia berfungsi dengan baik apabila hubungan antara ciri input dan output adalah lebih kurang linear. Dalam konteks klasifikasi teks, ciri (seperti skor TF-IDF atau pembedaan perkataan) selalunya mempunyai hubungan linear dengan label sasaran.

Keputusan Pemodelan Terbaik

Berikut merupakan rumusan keputusan dalam bentuk jadual bagi ketepatan pada set ujian

Jadual 3 Keputusan ketepatan pada set ujian

Data	Fitur Teks dan Keseluruhan Simptom yang Dikesan (<i>Target</i>)	Fitur Teks dan Fitur PHQ-3_Count	Fitur Teks dan Fitur PHQ-4_Count	Fitur Teks dan Fitur PHQ-8_Count

Data Keseluruhan	0.9725	0.9960	0.9900	0.9854
Data Seimbang	0.8868	0.9286	0.8214	0.7411

Bedasarkan keputusan yang diperoleh ,didapati bahawa model terbaik yang memberikan nilai ketepatan yang baik dan konsisten ialah algoritma Regresi Logistik dan data keseluruhan serta Keseluruhan Simptom yang Dikesan (*Target*). Oleh itu, pendekatan tersebut akan digunakan bagi mengimplementasikan ke dalam aplikasi berasaskan web *DepresCare*.

Papan Pemuka

Pembinaan web akan dijalankan dengan menggunakan pembelajaran mesin “*Streamlit*”. Penggunaan *Streamlit* dijalankan khususnya untuk membangunkan papan pemuka (Dashboard) iaitu antara muka interaktif yang membolehkan pengguna meneroka dan mengklasifikasi teks yang bersesuaian dengan kajian ini iaitu untuk Mengenalpasti Kemurungan Berdasarkan Simptom Kemurungan Dalam Teks Sosial Media. Papan Pemuka yang telah direka membolehkan pengguna memuat naik teks emosi dan perasaan pada keadaan semasa , mengklasifikasikan teks pengguna dan memaparkan keputusan berdasarkan teks. Oleh itu, bagi projek ini, model Regresi Logistik telah dipilih sebagai model yang dapat meramal ramalan yang tepat.



Rajah 7 Menunjukkan Halaman Utama bagi Aplikasi Web *DepresCare*

Rajah 7 menunjukkan halaman utama bagi projek Mengenalpasti Kemurungan Berdasarkan Simptom Kemurungan Dalam Teks Sosial Media yang memberikan sedikit gambaran awal tentang projek ini kepada pengguna baru. Laman ini mempunyai butang menu bernama “*Classify Depression*” akan membawa pengguna kepada Halaman *Classify Depression* untuk memasukkan input pengguna bagi mengklasifikasikan teks tersebut mempunyai depresi atau tidak. Selain itu, butang “*Contact*” yang memberikan informasi untuk mendapatkan bantuan professional sekiranya pengguna mempunyai “*Depression*”. Butang “*User Manual*” yang membawa pengguna kepada Halaman Panduan Pengguna.

Symptoms are used to detect depression based on PHQ-9 are as below:

	Trouble falling or staying asleep, or sleeping too much?	Feeling tired or having little energy?	Moving or speaking slowly
0	active at night	bushed	adagio
1	all nighter	debilitate	agitated
2	all nighter	did nothing	angry
3	awake	dog tired	annoyed
4	bad sleep	don't have the strength	annoying
5	be up late	drained	antsy
6	can't fall back asleep	droopy eyes	anxiety
7	can't get out of bed	enervated	anxiety disord
8	can't get up	enervation	anxious
9	can't sleep	ennui	babbling

Click the button below to check your Depression based on PHQ-9

CLICK

Rajah 8 Menunjukkan Simptom Depresi yang digunakan di Aplikasi Web *DepresCare*

Rajah 8 menunjukkan Simptom Depresi yang digunakan di web *DepresCare*. Simptom depresi yang digunakan iaitu masalah jatuh atau terus tidur, atau terlalu banyak tidur, rasa letih atau kurang tenaga dan bergerak atau bercakap dengan perlahan sehingga orang lain dapat perasan. Pengguna perlu menekan "CLICK" untuk pergi ke halaman *Classify Depression*.

Welcome to DepresCare

Text classifier for users based on emotions and feelings during that situations for detect depression

Enter your text here

Feeling increasingly frustrated, I was both angry and annoyed at the constant setbacks. With a sigh, I admitted to myself, "I am weak," unable to handle the pressure any longer.

CLASSIFY

Welcome to DepresCare

Text classifier for users based on emotions and feelings during that situations for detect depression

Enter your text here

I am so happy

CLASSIFY

Rajah 9 Menunjukkan Halaman *Classify Depression* bagi Aplikasi Web *DepresCare*

Rajah 9 menunjukkan halaman “*Classify Depression*” yang mudah digunakan dan difahami oleh pengguna. Pengguna hanya perlu memasukkan teks input ke dalam ruangan yang disediakan tentang penjelasan perasaan atau emosi mereka. Setelah itu, pengguna telah memasukan teks perlu menekan butang “CLASSIFY” untuk mulakan proses klasifikasi teks pengguna.

Based on your text, here are the detected depression symptoms:

Detected Signals for Trouble falling or staying asleep (PHQ-3) : []

Detected Signals for Feeling tired (PHQ-4) : ['i am weak']

Detected Signals for Moving or speaking so slowly (PHQ-8) : ['angry', 'annoyed']

Prediction: Depressed

Based on the result, you MAY HAVE DEPRESSION. Please click NEXT for further information to seek for professional help .

Please click NEXT for further information or RETURN to classify another text.

RETURN NEXT

No depression signals detected in the text: i am so happy.

Based on the depression signals of Trouble falling or staying asleep (PHQ-3), Feeling tired (PHQ-4) and Moving or speaking so slowly (PHQ-8) symptoms.

Prediction: Not Depressed

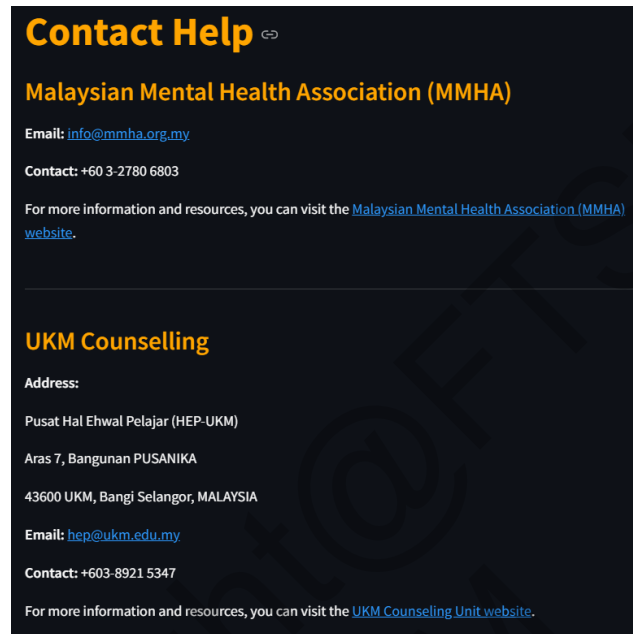
Please click RETURN to classify another text.

RETURN

Rajah 10 Halaman Senarai Istilah Berkaitan Dengan Simptom Depresi bagi Aplikasi Web *DepresCare*

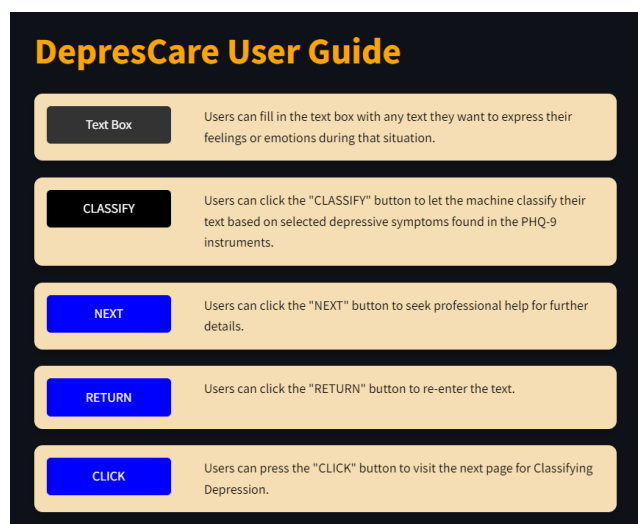
Setelah pengguna menekan “CLASSIFY” akan papar halaman proses simptom depresi. Halaman ini menunjukkan keputusan input jawapan teks tentang penjelasan perasaan atau emosi pengguna yang merupakan salah satu simptom depresi yang dinyatakan ataupun tidak seperti Rajah 10. Simptom depresi yang digunakan di dalam sistem ini adalah simptom ketiga, keempat dan kelapan yang terdapat pada instrumen Soal Selidik Kesihatan Pesakit-9. Jika input jawapan teks mempunyai istilah yang berkaitan dengan leksikon depresi bagi simptom tersebut, halaman ini akan memaparkan istilah yang berkaitan dengan depresi bagi setiap simptom yang berkenaan yang telah dinyatakan. Seterusnya, halaman ini akan memberitahu

pengguna bahawa mereka telah mempunyai depresi ataupun tidak. Jika keputusannya menunjukkan pengguna mempunyai beberapa simptom depresi, dalam aplikasi web *DepresCare* akan menunjukkan teks untuk pengguna untuk mengambil tindakan dengan mengalakkan berjumpa dengan pihak kaunseling atau pakar untuk maklumat lanjut. Pengguna boleh meneruskan aplikasi dengan menekan butang “NEXT” untuk mendapat maklumat lanjut atau menekan butang “RETURN” untuk kembali mengklasifikasi teks.



Rajah 11 Halaman *Contact Help* bagi Aplikasi Web *DepresCare*

Rajah 11 di bawah menunjukkan halaman “*Contact Help*” setelah pengguna menekan butang “NEXT” di halaman “Senarai Istilah dan Keputusan Penentuan Depresi” yang memberikan informasi untuk mendapatkan bantuan professional sekiranya pengguna mempunyai “Depression”.



Rajah 12 Halaman Panduan Pengguna bagi Aplikasi Web *DepresCare*

Rajah 12 di bawah menunjukkan halaman panduan penggunaan kepada pengguna. Setiap panduan telah dinyatakan dengan jelas bagi memudahkan pengguna untuk berinteraksi dan bernavigasi merentas semua halaman yang terdapat di dalam aplikasi web *DepresCare*.

Pengujian Kebolehgunaan

Pengujian kebolehgunaan ialah satu proses yang melibatkan pengujian akhir terhadap sistem yang dilakukan oleh wakil pengguna untuk memastikan model klasifikasi dan aplikasi web *DepresCare* berfungsi dengan baik dan memenuhi keperluan dan kepuasan pengguna. Pengujian aplikasi berasaskan aplikasi web *DepresCare* ini dilakukan dengan menggunakan format Ujian Penerimaan Pengguna (UAT). Ujian ini memfokuskan kepada pengesahan kemesraan pengguna, kefungsiannya dan prestasi perisian daripada perspektif pengguna. Pengujian aplikasi web *DepresCare* telah dikumpulkan melalui borang atas talian yang bernama *Google Form*. Borang tersebut direka untuk pengguna menilai pelbagai aspek terhadap aplikasi web seperti kebolehgunaan aplikasi web dan maklum balas daripada penggunaan sistem *DepresCare*. Terdapat sebanyak 13 responden yang telah mengambil bahagian dalam maklum balas ini. Setiap maklum balas diambil kira untuk menilai tahap prestasi aplikasi web *DepresCare*. Jadual 4 menunjukkan tafsiran skala bagi setiap soalan yang perlu dijawab oleh responden manakala Jadual 5 menunjukkan skala majoriti yang diterima bagi setiap soalan.

Jadual 4 Tafsiran Skala

Skala	Tafsiran
1.00	Sangat Tidak Setuju
2.00	Tidak Setuju
3.00	Neutral
4.00	Setuju
5.00	Sangat Setuju

Jadual 5 Skala Majoriti bagi Setiap Soalan

Bil	Soalan	3	4	5
1.	Saya berpendapat aplikasi web ini mudah digunakan.		30.8% (4 responden)	69.2% (9 responden)

2.	Saya berpendapat bahawa saya TIDAK memerlukan sokongan teknikal untuk menggunakan aplikasi web ini.		46.2% (6 responden)	53.8% (7 responden)
3.	Saya dapati reka bentuk antara muka aplikasi web ini mudah difahami.		30.8% (4 responden)	69.2% (9 responden)
4.	Saya dapati semua reka bentuk antara muka di dalam aplikasi web ini relevan dengan fungsinya.		30.8% (4 responden)	69.2% (9 responden)
5.	Saya mempunyai pengetahuan tentang <i>Patient Health Questionnaire (PHQ-9)</i> ?	23.1%(3 responden)	46.2% (6 responden)	30.8% (4 responden)
6.	Adakah teks data dapat mengenal pasti simptom-simptom depresi yang ada dalam PHQ-9 ?		38.5% (5 responden)	61.5% (8 responden)
7.	Adakah aplikasi web ini mengesan dan mengekstrak simptom depresi dengan baik apabila mendapat keputusan.		30.8% (4 responden)	69.2% (9 responden)
8.	Adakah aplikasi web ini memberikan ramalan yang betul sama ada input pengguna menunjukkan depresi atau tidak dengan baik.		7.7%(1 responden)	92.3%(12 responden)
9.	Secara keseluruhan, saya berpuas hati menggunakan aplikasi web <i>DepresCare</i> .		46.2%(6 responden)	53.8%(7 responden)

Berdasarkan maklum balas yang diperoleh, hampir semua maklum balas daripada pengguna memberikan keputusan yang baik. Namun, terdapat hanya 1 orang responden sahaja yang berasa neutral terhadap pengetahuan tentang *Patient Health Questionnaire (PHQ-9)*. Hal ini kerana PHQ-9 merupakan alat yang sangat efisien untuk menentukan depresi dengan pelbagai simptom dan mungkin tidak difahami dengan baik oleh pengguna.

Cadangan Penambahbaikan

Terdapat beberapa cadangan penambahbaikan yang diperlukan pada masa hadapan untuk memastikan projek ini dapat memberikan hasil yang lebih baik. Antaranya, menggunakan teknik seperti pensampelan berlebihan, pensampelan terkurang atau penjanaan data sintetik untuk mengimbangi set data. Selain itu, menilai secara teliti dengan menggunakan metrik penilaian komprehensif untuk memahami model tersebut supaya dapat memberikan keputusan yang tepat. Akhir sekali, meningkatkan penggunaan fitur bagi semua simptom yang terdapat di dalam instrumen Soal Selidik Kesihatan Pesakit-9 (PHQ-9) untuk diuji supaya dapat membantu model untuk membuat prediksi yang lebih tepat.

KESIMPULAN

Kesimpulannya, kajian ini berjaya mencapai objektif yang telah ditetapkan dengan mengumpulkan dan mengklasifikasikan teks data untuk menentukan depresi ataupun tidak dengan menggunakan simptom yang dipilih. Selain itu, aplikasi web *DepresCare* yang telah diterapkan dengan model terbaik iaitu dengan menggunakan algoritma Regresi Logistik, data keseluruhan dan fitur teks serta fitur keseluruhan simptom yang dikesan telah berjaya disiapkan. Akhir sekali, kajian ini dapat memberikan manfaat kepada pengguna untuk menentukan depresi atau tidak berdasarkan perasaan dan emosi pengguna di aplikasi web *DepresCare*.

Kekuatan Sistem

Bagi projek ini, terdapat beberapa kekuatan projek yang menyumbang kepada kejayaan kajian ini. Antaranya ialah peningkatan senarai istilah leksikon yang berkenaan dengan simptom terpilih. Projek ini telah menambah baik senarai istilah leksikon yang berkaitan dengan simptom depresi yang terpilih daripada instrumen Soal Selidik Kesihatan Pesakit-9 (PHQ-9) iaitu simptom ketiga, keempat dan kelapan. Hasil penambahan leksikon depresi ini dapat meningkatkan keupayaan untuk mengenal pasti simptom dengan lebih tepat di dalam data teks. Selain itu, model yang dibangunkan dapat mengenalpasti dan mengekstrak senarai istilah yang berkaitan dengan setiap simptom depresi yang terpilih. Dengan keupayaan mengenalpasti dan mengekstrak istilah-istilah tersebut, ia dapat memudahkan bagi membuat klasifikasi yang lebih mendalam terhadap kandungan teks data.

Kelemahan Sistem

Seterusnya, terdapat sedikit kekurangan semasa menjalankan pendekatan ini. Terdapat beberapa data yang mempunyai sasaran bernilai 1 yang bermaksud teks tersebut mengandungi depresi. Namun, apabila mengimplementasikan pendekatan terbaik yang dipilih, prediksi pada label ramalan adalah bernilai 0 yang menunjukkan teks tersebut tidak mengandungi unsur depresi. Sebagai contoh, teks yang dimasukkan oleh pengguna pada ruangan input ialah "feeling lazy pack want go home yet.". Oleh itu, data teks menjadi bias dan tidak dapat menunjukkan ketepatan yang tepat.

PENGHARGAAN

Penulis kajian ini ingin ucapkan setinggi-tinggi penghargaan dan jutaan terima kasih kepada Dr. Lailatul Qadri Binti Zakaria selaku penyelia projek tahun akhir saya yang telah banyak memberi tunjuk ajar dan membimbing saya sepanjang melakukan projek ini. Terima kasih juga yang tidak terhingga kepada beliau kerana sentiasa bersabar dan tabah dalam membimbing dan mengajar saya sepanjang penyediaan projek.

Penulis kajian ini juga ingin mengucapkan terima kasih kepada semua pihak yang membantu secara langsung mahupun tidak langsung dalam menyempurnakan projek ini. Segala bantuan

yang telah dihulurkan amatlah dihargai kerana tanpa bantuan mereka, projek ini tidak dapat dilaksanakan dengan baik. Semoga tuhan merahmati dan memberikan balasan yang terbaik.

RUJUKAN

- Gillis, A. S. 2022, March 14. What is user acceptance testing (UAT)? - searchsoftwarequality. Software Quality.
<https://www.techtarget.com/searchsoftwarequality/definition/user-acceptance-testing-UAT>
- Martin, M. (2023, December 30). *Functional vs Non Functional Requirements*. Guru99.
<https://www.guru99.com/functional-vs-non-functional-requirements.html>
- Md. Tazmim Hossain, Md. Arafat Rahman Talukder, Nusrat Jahan.(2021). Depression prognosis using natural language processing and machine learning from social media status. *Journal of Electrical and Computer Engineering (IJECE)*.
<https://ijece.iaescore.com/index.php/IJECE/article/view/25789/15662>
- Minsu Park, Chiyong Cha, Meeyoung Cha.(2012). Depressive Moods of Users Portrayed in Twitter.
https://nyuscholars.nyu.edu/ws/portalfiles/portal/134720119/depressive_moods_kdd.pdf
- Mior Amir.(2019,July 10).Reka Bentuk Antara Muka.
<https://prezi.com/p/fxehee1vkotw/reka-bentuk-antara-muka-pengguna/>
- Mukesh Chaudhary. (2020, April 24). TF-IDF Vectorizer scikit-learn.
<https://medium.com/@cmukesh8688/tf-idf-vectorizer-scikit-learn-dbc0244a911a>

Mohandass A/L Thiruvassagam (A189202)
Dr. Lailatul Qadri Binti Zakaria
Fakulti Teknologi & Sains Maklumat
Universiti Kebangsaan Malaysia