

ANALISIS SIMPTOM DEPRESI DALAM SET DATA KESIHATAN MENTAL

NUR AIN SOFIA BINTI KHAMIS
DR. LAILATUL QADRI BINTI ZAKARIA

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

ABSTRAK

Depresi atau lebih dikenali sebagai kemurungan adalah merupakan perkara yang serius terhadap kesihatan mental manusia. Depresi menjejaskan kesejahteraan hidup individu kerana boleh menyebabkan individu berasa sedih berpanjangan, keletihan, cepat marah dan hilang minat dalam melakukan aktiviti harian. Isu utama yang berkaitan dengan depresi ini adalah individu yang mengalami depresi mungkin tidak sedar bahawa mereka sedang mengalami simptom depresi. Perkara ini boleh terjadi apabila individu mempunyai kurang pengetahuan tentang simptom-simptom depresi. Oleh itu, projek ini akan membina model pembelajaran mesin untuk menganalisis simptom depresi berdasarkan set data kesihatan mental. Set data yang akan digunakan merupakan set data sumber terbuka berkaitan dengan kesihatan mental yang telah tersedia dan diperolehi daripada aplikasi *Twitter*. Projek ini akan menggunakan teknik pemrosesan bahasa tabii untuk mengekstrak simptom-simptom depresi berdasarkan simptom-simptom terpilih yang terdapat di dalam instrumen Soal Selidik Kesihatan Pesakit-9 (*PHQ-9*). Simptom-simptom tersebut akan digunakan sebagai fitur untuk mengenalpasti teks depresi menggunakan teknik pembelajaran mesin. Antara simptom depresi yang dipilih ialah kurang minat atau keseronokan dalam melakukan sesuatu, berasa tertekan, murung atau putus asa, dan mengalami perubahan selera dan berat badan. Projek ini juga akan turut membangunkan aplikasi berasaskan web ringkas yang bernama *DepressioCare* bagi memudahkan para pengguna mengesan simptom depresi yang terdapat dalam teks dan juga mengenalpasti samaada teks tersebut merupakan depresi atau tidak. Aplikasi *DepressioCare* diharap akan meningkatkan pemahaman terhadap penyakit depresi ini dengan menyediakan informasi tentang simptom-simptom depresi. Secara tidak langsung, model ini juga dapat membantu individu dalam mengenali simptom-simptom depresi pada diri mereka atau orang lain di sekitar mereka berdasarkan teks yang mereka gunakan

PENGENALAN

Isu depresi atau lebih dikenali sebagai kemurungan merupakan salah satu penyakit yang mengakibatkan kesihatan mental terjejas. Menurut Kementerian Kesihatan Malaysia, kemurungan adalah gangguan perasaan seseorang yang menyebabkan seseorang itu merasa

sedih yang berpanjangan, keletihan dan ketiadaan tenaga, cepat marah serta hilang minat dalam aktiviti harian (Kementerian Kesihatan Malaysia 2012). Terdapat pelbagai faktor kemurungan berlaku dan ini termasuklah dengan interaksi kompleks di antara faktor sosial, psikologi dan biologi (*World Health Organisation* t.th.). Kemurungan juga mempunyai pelbagai jenis dan sesetengah daripadanya mungkin berkembang di atas faktor faktor yang tertentu (*National Institute of Mental Health* t.th.).

Individu yang mempunyai kemurungan kebiasaannya akan menunjukkan simptom-simptom seperti kegelisahan, hilang minat dalam melakukan aktiviti yang menyeronokkan dan perasaan sedih atau kosong yang berterusan. Untuk menentukan seseorang individu mempunyai kemurungan adalah dengan cara mendapatkan diagnosis daripada doktor dan juga boleh membuat penilaian sendiri dalam talian untuk permulaan (Kumar 2022). Ujian sendiri yang boleh dilakukan bagi mengenalpasti status kemurungan mereka ialah ujian Soal Selidik Kesihatan Pesakit-9 (*PHQ-9*) dan ujian Skala Kemurungan, Kebimbangan & Tekanan (SKKT) (Kementerian Kesihatan Malaysia t.th.). Keputusan yang diperolehi daripada ujian ini boleh digunakan sebagai panduan untuk mengesan simptom awal dan menggalakkan individu membuat rawatan lanjut dengan merujuk kepada pakar.

Namun begitu, kadar kemurungan di Malaysia telah mencapai tahap yang mengkhawatirkan. Perkara ini berlaku kerana kurangnya kesedaran terhadap kesihatan mental terutamanya depresi. Terdapat segelintir individu tidak dapat mengenal pasti akan simptom-simptom depresi yang ada pada diri sendiri dan individu di sekitar mereka. Hal ini akan menyebabkan individu yang mempunyai depresi mempunyai kesan yang negatif. Individu yang tidak dapat mengidentifikasi simptom-simptom depresi mungkin tidak menyedari mereka memerlukan sokongan mahupun rawatan yang diperlukan.

Oleh itu, projek ini akan membangunkan sistem analisis simptom depresi dalam set data kesihatan mental. Sistem ini dapat menganalisis dan mengekstrak simptom-simptom depresi daripada set data kesihatan mental. Dengan adanya sistem ini, kesedaran terhadap simptom-simptom depresi dapat ditingkatkan. Individu yang menggunakan sistem ini dapat mengenalpasti simptom-simptom awal depresi dalam diri sendiri dan individu di sekeliling mereka. Individu yang menggunakan sistem ini juga dapat mengurangkan stigma sosial terhadap penyakit depresi dengan membantu dalam mengurangkan perspektif yang negatif yang mungkin menghalang individu mengalami depresi daripada mendapatkan bantuan dan rawatan yang diperlukan.

METODOLOGI KAJIAN

Metodologi yang digunakan dalam pembangunan projek ini ialah metodologi Proses Standard Industri Cross untuk Perlombongan Data (*CRISP-DM*). Metodologi Proses Standard Industri Cross untuk Perlombongan Data (*CRISP-DM*) ialah model process yang berfungsi sebagai asas untuk proses sains data (Hotz 2023). Metodologi ini dipilih bagi projek ini adalah kerana metodologi ini adalah bersesuaian dengan projek perlombongan data dan ianya adalah metodologi yang teguh dan terbukti dengan baik.

Fasa pemahaman perniagaan

Fasa ini adalah untuk memahami tujuan utama bagi projek ini iaitu untuk membuat analisis simptom depresi dalam set data kesihatan mental berdasarkan simptom terpilih yang terdapat di dalam instrumen Soal Selidik Kesihatan Pesakit-9 (*PHQ-9*). Di dalam konteks projek ini, pemahaman perniagaan melibatkan keperluan kesedaran tentang simptom depresi dan mengaitkan maklumat tersebut dengan manfaat kepada individu dan masyarakat di sekeliling. Fasa ini juga dapat memastikan bahawa setiap objektif bagi projek ini akan dibangunkan seperti yang telah ditetapkan.

Fasa pemahaman data

Fasa pemahaman data merupakan fasa yang memerlukan pengumpulan maklumat dan pemahaman yang mendalam tentang simptom depresi terpilih yang terdapat di dalam instrumen Soal Selidik Kesihatan Pesakit-9 (*PHQ-9*) dan set data kesihatan mental yang berkaitan dengan depresi yang akan digunakan di dalam projek ini. Pada fasa ini, sorotan susastera juga telah dijalankan kepada kajian tentang pengelasan teks depresi berdasarkan set data depresi yang sedia ada untuk meningkatkan pemahaman yang dapat membantu pembangunan sistem ini.

Fasa penyediaan data

Fasa penyediaan data bagi 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 analisis simptom depresi. Pada fasa ini, pengemaskinian data terhadap leksikon depresi telah dilakukan bagi memastikan setiap simptom depresi yang terpilih meliputi hampir semua istilah yang digunakan yang merujuk kepada simptom tersebut.

Selepas itu, proses memuat naik data pada perisian pengaturcaraan *Google Colab* melibatkan fail json yang mengandungi 10 lajur berkaitan leksikon depresi dan enam fail excel yang mengandungi 15 lajur berkaitan dengan isu kesihatan mental yang telah diperoleh daripada aplikasi *Twitter*. Namun, hanya tiga lajur utama daripada fail json yang akan digunakan iaitu lajur "*signal_1*", lajur "*signal_2*", dan lajur "*signal_5*" yang menunjukkan leksikon depresi bagi simptom pertama, kedua dan kelima daripada instrumen Soal Selidik Kesihatan Pesakit-9 (*PHQ-9*). Manakala 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.

Seterusnya, pra-pemprosesan data akan dilakukan pada data-data yang telah dimuat naik di *Google Colab*. Pra-pemprosesan data yang terlibat ialah membuang karakter bukan *ASCII* seperti petikan dan apostrof keriting, membuang tanda baca seperti apostrof dan tanda seru, membuang pautan dan juga menukar kepada huruf kecil. Selain itu, pembuangan

kata henti terpilih juga turut dilakukan bagi memastikan data yang akan digunakan adalah bersih dan relevan untuk analisis seterusnya. Kata henti yang perlu diambil kira dalam analisis ini adalah seperti perkataan kata ganti diri seperti “saya”, “dia”, “anda” dan juga perkataan yang menunjukkan tentang perasaan seperti “tidak”, “jangan”, dan “down” yang juga menunjukkan kepada depresi. Kelima-lima teknik tersebut telah digunakan bagi memastikan setiap data mempunyai format yang sama untuk memudahkan penggunaan pada peringkat seterusnya iaitu pepadanan rentetan

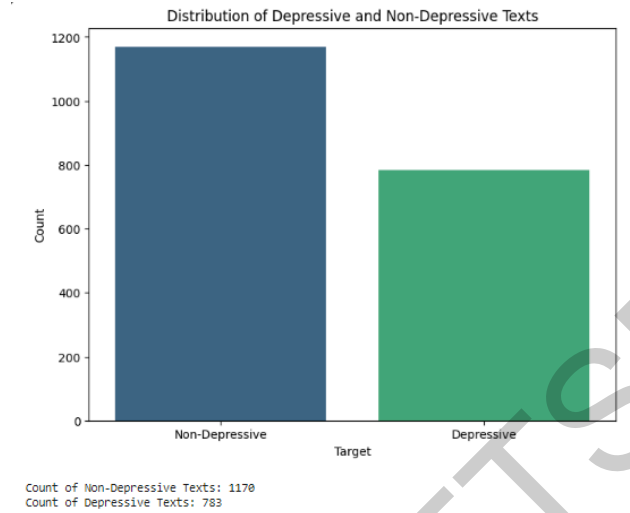
Akhir sekali, fasa penyediaan data diteruskan lagi dengan membuat padanan rentetan bagi kedua-dua fail json dan fail excel untuk mengekstrak maklumat yang penting. Kedua-dua fail tersebut telah digunakan bagi membuat pepadanan rentetan dengan mengekstrak istilah ataupun leksikon depresi bagi simptom pertama, kedua dan kelima di dalam instrumen Soal Selidik Kesihatan Pesakit-9 (PHQ-9) yang terdapat di dalam setiap teks. Hasil pepadanan tersebut disimpan ke dalam rangka data yang baharu bernama “matched_system_df”. Rangka data ini mengandungi beberapa lajur penting yang merangkumi indeks (“Tweet_Index”), teks yang mengandungi tweet yang telah diproses (“Tweet_Processed”), senarai istilah yang sepadan untuk setiap simptom (“Signal_1”, “Signal_2”, “Signal_5”), jumlah kehadiran istilah untuk setiap simptom (“Signal_1_Count”, “Signal_2_Count”, “Signal_5_Count”), penunjuk kewujudan simptom (“Symptom_Presence”), dan nilai sasaran (“Target”) yang menentukan teks tersebut mempunyai depresi ataupun tidak.

Hasil pepadanan ini juga dapat mencipta fitur baharu yang boleh digunakan di dalam model pembelajaran mesin pada peringkat yang seterusnya untuk membantu meningkatkan prestasi model. Selain itu, melalui pepadanan rentetan ini, data-data yang berbeza yang diekstrak seperti data keseluruhan, data sepadan, data rawak, dan data seimbang akan disimpan di dalam rangka pandas baharu untuk digunakan bagi membuat analisis simptom depresi untuk menentukan kaedah dan pendekatan yang mana memberikan prediksi pengelasan pada teks dengan baik dan tepat. Penjelasan tentang data dan fitur yang digunakan adalah seperti yang berikut.

a. Data Analisis Simptom Depresi

Pada fasa ini, analisis simptom depresi akan dilakukan dengan menggunakan data-data yang berbeza. Berikut adalah tentang data yang akan digunakan bagi membuat prediksi pada peringkat seterusnya.

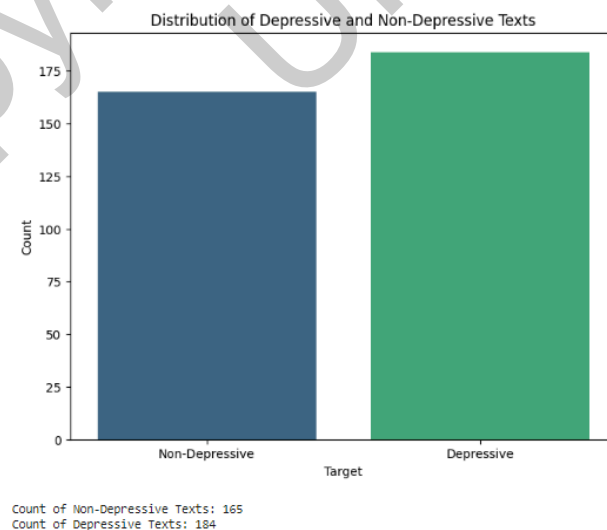
i. Data keseluruhan



Rajah 1 Bilangan Teks Depresi dan Bukan Depresi Bagi Data Keseluruhan

Bagi keseluruhan data, data yang akan digunakan adalah berjumlah 1953 data. Keseluruhan data ini mempunyai 1170 jumlah teks yang menunjukkan bukan depresi manakala 783 jumlah teks yang menunjukkan depresi seperti yang ditunjukkan pada Rajah 1. Data ini merupakan data umum dan ianya bercampur dengan data yang mempunyai simptom yang sepadan dengan simptom yang digunakan.

ii. Data Sepadan

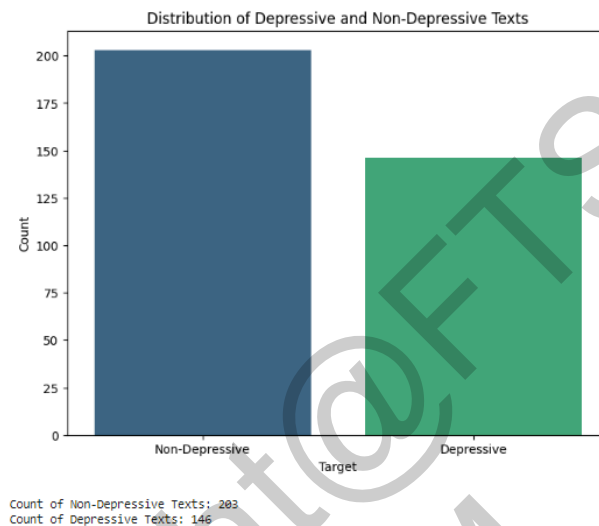


Rajah 2 Bilangan Teks Depresi dan Bukan Depresi Bagi Data Sepadan

Data sepadan ini merupakan data yang diekstrak apabila hanya menjumpai simptom yang digunakan di dalam teks. Simptom yang digunakan adalah hanya simptom pertama, kedua dan kelima yang terdapat dalam instrumen Soal Selidik Kesihatan Pesakit-9 (PHQ-9).

Simptom pertama ialah mempunyai sedikit minat atau keseronokan dalam melakukan sesuatu. Simptom yang kedua ialah berasa tertekan, murung atau putus asa. Simptom yang ketiga adalah mempunyai masalah susah untuk tidur, tetap tidur dan terlalu banyak tidur. Menurut Rajah 2, terdapat 165 jumlah teks yang tidak menunjukkan depresi dan 184 jumlah teks yang menunjukkan depresi.

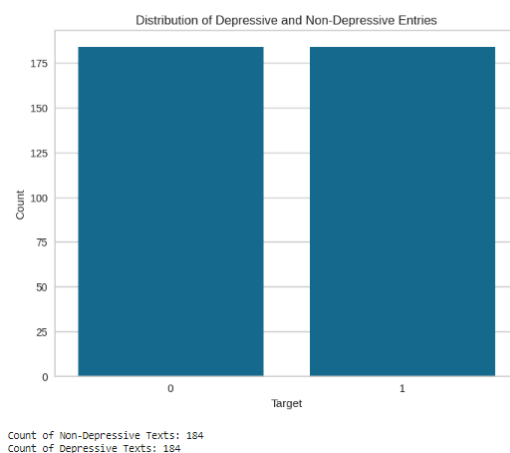
iii. Data Rawak



Rajah 3 Bilangan Teks Depresi dan Bukan Depresi Bagi Data Rawak

Data rawak ialah data yang mempunyai jumlah yang sama dengan jumlah data yang sepadan dengan simptom yang digunakan tetapi ianya mengandungi data yang meliputi semua sama ada sepadan dengan simptom ataupun tidak secara rawak. Menurut Rajah 3, data rawak ini mempunyai jumlah 203 teks yang menunjukkan bukan depresi manakala 146 teks yang menunjukkan depresi.

iv. Data Seimbang



Rajah 4 Bilangan Teks Depresi dan Bukan Depresi Bagi Data Seimbang

Data seimbang adalah data yang mempunyai jumlah yang sama bagi kedua - dua kategori teks. Data seimbang ini juga merangkumi 184 teks yang menunjukkan depresi. Teks yang dikategorikan sebagai depresi ini mempunyai data yang sepadan dengan simptom yang digunakan. Bagi memastikan data ini seimbang, sebanyak 19 jumlah teks yang tidak mempunyai data sepadan dengan simptom yang digunakan ditambah pada data seimbang ini.

b. Fitur Yang Akan Digunakan di Dalam PyCaret

Pada fasa ini, fitur yang akan digunakan adalah untuk membuat prediksi bagi menentukan teks tersebut merupakan depresi atau tidak. Setiap data dan fitur akan diuji data dan fitur yang mana memberikan keputusan terbaik semasa membuat prediksi dengan menggunakan perpustakaan pembelajaran mesin PyCaret. Antara data dan fitur yang akan digunakan adalah seperti berikut.

i. Fitur Teks

```
data_text = data[['Tweet_Processed', 'Target']]
```

	Tweet_Processed	Target
0	mood caused infectious diseases nutritional de...	0
1	unnecessary family drama i feel like moving far ...	1
2	im 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
...
1948	done absolute fuck last 2days	1
1949	i trying 2 make list ppl i twt 2 regularly u w...	1
1950	your 14 year old son looks go new years eve pa...	0
1951	perpetrators admit never fault victim scarred ...	0
1952	i currently accepting new clients online mindf...	0

Rajah 5 Data Mengandungi Fitur Teks

Rajah 5 menunjukkan data yang akan digunakan hanya mengandungi fitur teks sahaja.

ii. Fitur Teks dan Fitur Bilangan Setiap Simptom

```
[ ] # Create a new DataFrame with only the necessary columns include features
data_text_symptoms = data[['Tweet_Processed', 'Target', 'Signal_1_Count', 'Signal_2_Count', 'Signal_5_Count' ]]
```

```
[ ] data_text_symptoms
```

	Tweet_Processed	Target	Signal_1_Count	Signal_2_Count	Signal_5_Count
0	mood caused infectious diseases nutritional de...	0	0	0	1
1	unnecessary family drama i feel like moving far ...	1	0	0	0
2	im going keep banging cos true focus get stop ...	0	0	0	0
3	ugh woke acne covered face morning yyayyy	0	0	0	0
4	portable transparent tobacco rolling machine p...	0	0	0	0
...
1948	done absolute fuck last 2days	1	0	0	0
1949	i trying 2 make list ppl i twt 2 regularly u w...	1	0	0	0
1950	your 14 year old son looks go new years eve pa...	0	0	0	0
1951	perpetrators admit never fault victim scarred ...	0	0	0	0
1952	i currently accepting new clients online mindf...	0	0	0	0

Rajah 6 Data Mengandungi Fitur Teks dan Setiap Bilangan Simptom

Rajah 6 menunjukkan data yang akan digunakan mengandungi fitur teks, dan setiap bilangan simptom yang telah dipadankan dengan simptom yang digunakan sahaja.

iii. Fitur Teks dan Fitur Bilangan Semua Simptom

```
[ ] # Create a new DataFrame with only the necessary columns include features
data_text_symptoms_all= data[['Tweet_Processed', 'Target', 'count_symptoms_all' ]]
```

```
data_text_symptoms_all
```

	Tweet_Processed	Target	count_symptoms_all
0	mood caused infectious diseases nutritional de...	0	1
1	unnecessary family drama i feel like moving far ...	1	0
2	im going keep banging cos true focus get stop ...	0	0
3	ugh woke acne covered face morning yyayyy	0	0
4	portable transparent tobacco rolling machine p...	0	0
...
1948	done absolute fuck last 2days	1	0
1949	i trying 2 make list ppl i twt 2 regularly u w...	1	0
1950	your 14 year old son looks go new years eve pa...	0	0
1951	perpetrators admit never fault victim scarred ...	0	0
1952	i currently accepting new clients online mindf...	0	0

Rajah 7 Data Mengandungi Fitur Teks dan Semua Bilangan Simptom

Rajah 7 menunjukkan data yang akan digunakan hanya mengandungi teks dan bilangan semua simptom yang telah dipadankan dengan simptom yang digunakan. Bilangan semua simptom ini merupakan bilangan setiap simptom yang telah ditambah.

Fasa pemodelan

Fasa pemodelan bagi projek ini akan menggunakan perpustakaan bahasa pengaturcaraan

Python iaitu *PyCaret* untuk meramal prediksi pengelasan teks kepada kategori teks kemurungan dan teks bukan kemurungan. Pembangunan model ini akan melibatkan penggunaan data dan fitur yang berbeza. Semua parameter yang digunakan pada setiap data dan fitur adalah sama iaitu dengan menggunakan kaedah Kekerapan Istilah-Kekerapan Dokumen Terbalik (*TF-IDF*) sebagai kaedah ciri teks. Data yang akan digunakan bagi fasa ini ialah data keseluruhan, data sepadan, data rawak dan data seimbang manakala fitur yang akan digunakan bagi setiap data adalah fitur teks, fitur bilangan setiap simptom, dan fitur bilangan semua simptom. Fasa pemodelan dengan menggunakan *PyCaret* ini akan menentukan algoritma dan model yang terbaik bagi menentukan fitur dan data yang mana yang akan memberikan keputusan ketepatan pada set ujian dan data baharu dengan baik. Tujuan menentukan algoritma dan model terbaik adalah untuk membantu dalam membuat ramalan pengelasan teks depresi atau tidak pada peringkat seterusnya. Rajah 8 menunjukkan kod bagi pembangunan model untuk menentukan algoritma yang terbaik untuk digunakan bagi setiap data dan fitur manakala Jadual 1 menunjukkan keputusan ketepatan algoritma terbaik bagi setiap data dan penggunaan fitur yang berbeza

```
[ ] from pycaret.classification import *

# Initialize the PyCaret setup
# Setting up the classification environment
clf_setup = setup(data=data_text,
                  target='Target',
                  session_id=123,
                  text_features=['Tweet_Processed'],
                  text_features_method = 'tf-idf',
                  train_size = 0.8)

# Compare different models
best_model_text = compare_models()

# Display the best model
print(best_model_text)
```

Rajah 8 Kod untuk Menentukan Algoritma Terbaik

Jadual 1 Jadual Ketepatan Model Terbaik

Data/Fitur	Fitur Teks	Fitur Bilangan Setiap Simptom	Fitur Bilangan Semua Simptom
1. Data Keseluruhan	Algoritma: Regresi Logistik Ketepatan: 0.6825	Algoritma: Regresi Logistik Ketepatan: 0.6812	Algoritma: Regresi Logistik Ketepatan: 0.6818
2. Data Sepadan.	Algoritma: Regresi Logistik Ketepatan: 0.7053	Algoritma: Regresi Logistik Ketepatan: 0.7052	Algoritma: Regresi Logistik Ketepatan: 0.6981
3. Data Rawak.	Algoritma: Pengelas Ridge Ketepatan: 0.6770	Algoritma: SVM - Kernel Linear Ketepatan: 0.6665	Algoritma: SVM - Kernel Linear Ketepatan: 0.6771
4. Data Seimbang	Algoritma: Regresi Logistik Ketepatan: 0.6939	Algoritma: Pengelas Pokok Tambahan Ketepatan: 0.6943	Algoritma: Pengelas Pokok Tambahan Ketepatan: 0.7008

Fasa penilaian

Fasa penilaian akan memfokuskan kepada penilaian model dengan menggunakan model terbaik kepada set ujian dan data baharu. Fasa ini akan menentukan model yang manakah yang akan memberikan keputusan yang terbaik pada set data ujian dan data baharu. Kemudian, model yang memberikan keputusan terbaik akan digunakan bagi mengimplementasi model tersebut kepada aplikasi web ringkas yang bernama *DepressioCare*. Pada fasa penilaian ini juga turut menggunakan data dan fitur yang berbeza yang telah dinyatakan pada sebelum ini. Jadual 2 dan Jadual 3 merupakan rumusan keputusan dalam bentuk jadual bagi ketepatan pada set ujian dan data baharu dengan menggunakan algoritma terbaik semasa pembangunan model dijalankan.

Jadual 2 Jadual Keputusan bagi Ketepatan Pada Set Ujian

Data/Fitur	Fitur Teks	Fitur Bilangan Setiap Simptom	Fitur Bilangan Semua Simptom
1.Data Keseluruhan	0.8111	0.8013	0.8003
2. Data Sepadan.	0.9026	0.8883	0.9054
3. Data Rawak.	0.9140	0.9054	0.9169
4. Data Seimbang	0.9049	0.9375	0.9402

Jadual 3 Jadual Keputusan bagi Ketepatan Pada Data Baharu

Data/Fitur	Fitur Teks	Fitur Bilangan Setiap Simptom	Fitur Bilangan Semua Simptom
1.Data Keseluruhan	0.3750	0.5000	0.6250
2. Data Sepadan.	0.6250	0.5000	0.6250
3. Data Rawak.	0.0000	0.1250	0.0000
4. Data Seimbang	0.3750	0.3750	0.5000

Setelah melakukan ujian ketepatan pada set ujian dan data baharu, didapati bahawa model terbaik yang memberikan nilai ketepatan yang baik dan konsisten ialah algoritma Regresi Logistik dan data sepadan serta fitur teks dan bilangan semua simptom. Oleh itu, pendekatan tersebut akan digunakan bagi mengimplementasikan ke dalam aplikasi berasaskan web *DepressioCare*.

Fasa implementasi

Fasa implementasi melibatkan integrasi model terbaik iaitu algoritma Regresi Logistik dan data sepadan serta fitur teks dan bilangan semua simptom ke dalam aplikasi web ringkas dengan menggunakan *Streamlit* dan *Visual Studio Code*. Aplikasi web *DepressioCare* ini membenarkan pengguna untuk memasukkan input teks berkaitan dengan perasaan mereka kemudian menganalisis teks tersebut dan mengekstrak simptom-simptom terpilih yang terdapat di dalam teks pengguna. Sistem ini juga akan menentukan prediksi teks pengguna mempunyai unsur depresi ataupun tidak dengan menggunakan model terbaik yang telah dilatih. Rajah 9 menunjukkan kod yang digunakan untuk implementasi model terbaik bagi meramal input pengguna.

```

model_path = "best_model_symptoms_all_pipeline.pkl"

# Load the best model
best_model_symptoms_all_new = load_model_from_disk(model_path)

# Make predictions on the new data
predictions_symptoms_all_new = predict_model(best_model_symptoms_all_new,
data=new_data_symptoms_all)

st.write("Here is your result,")
predict_score = predictions_symptoms_all_new['prediction_score'].iloc[0]

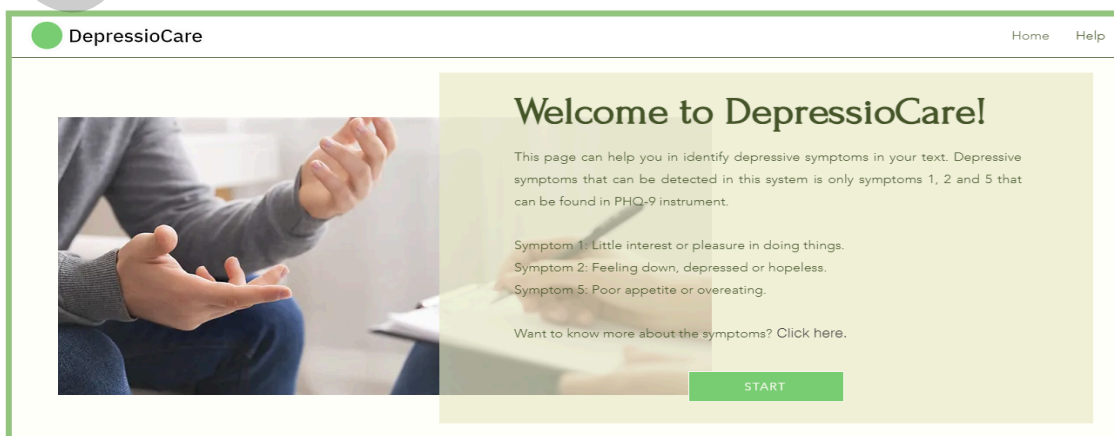
# Print interpretation based on prediction
label = predictions_symptoms_all_new['prediction_label'].iloc[0]
if label == 1:
    st.write(f"Based on prediction score of {predict_score:.2f}, you MAY HAVE depression. Please seek professional help for further information. ")
else:
    st.write(f"Based on prediction score of {predict_score:.2f}, you MAY NOT HAVE depression. Please seek professional help for further information.")

```

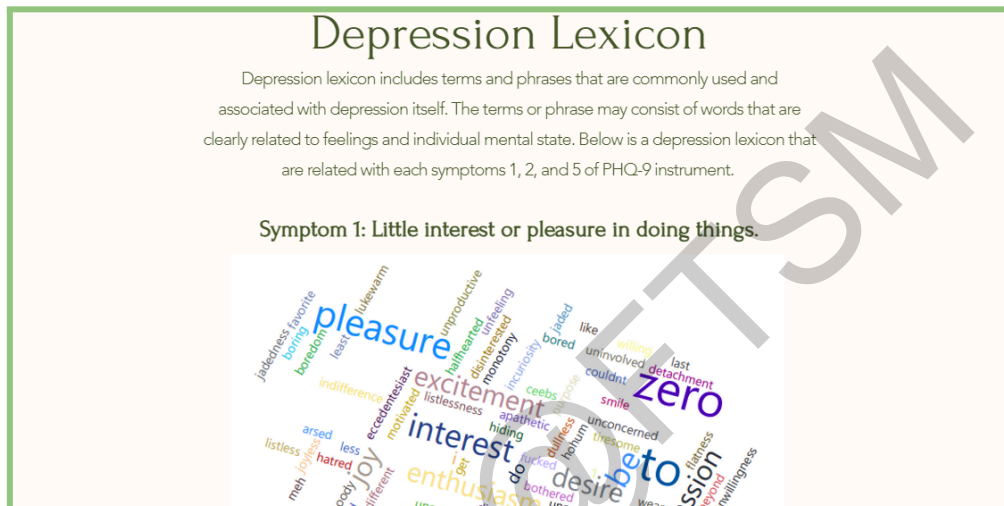
Rajah 9 Kod Implementasi Model Terbaik Ke Dalam Aplikasi Web *Depressiocare*

KEPUTUSAN DAN PERBINCANGAN

Model analisis simptom depresi dalam set data kesihatan mental dan aplikasi web ringkas *DepressioCare* telah berjaya dibangunkan dan semua dokumentasinya telah disiapkan. Semasa proses pembangunan, model analisis ini menggunakan bahasa pengaturcaraan *Python* dan menggunakan perisian pengaturcaraan *Google Colab* untuk membangunkan model dan melatih model dengan menggunakan model yang terbaik. Seterusnya, model terbaik yang telah dilatih akan digunakan pada aplikasi web *DepressioCare* dengan menggunakan rangka kerja sumber terbuka *Streamlit*. *Streamlit* menukarkan kod-kod pembangunan model kepada aplikasi data interaktif dengan menggunakan bahasa pengaturcaraan *Python*. Antara muka interaktif di dalam aplikasi web ringkas *DepressioCare* telah dibangunkan menggunakan *Streamlit* yang membolehkan para pengguna untuk mengisi teks pada ruangan input dan menekan butang untuk mendapatkan output adakah teks pengguna mengandungi unsur-unsur depresi ataupun tidak.

Rajah 10 Halaman Muka bagi Aplikasi Web *DepressioCare*

Rajah 10 menunjukkan halaman muka bagi aplikasi web *DepressioCare*. Laman ini mempunyai butang menu bernama “*Help*” yang membawa pengguna kepada Halaman Panduan Pengguna. Manakala, butang “*Click here*” akan membawa pengguna kepada halaman muka di bawah yang menunjukkan deskripsi tentang leksikon depresi yang digunakan. Butang “*Start*” pula akan membawa pengguna kepada Halaman Utama *DepressioCare* untuk memasukkan input pengguna.



Rajah 11 Halaman Deskripsi Leksikon Depresi bagi Aplikasi Web *DepressioCare*

Rajah 11 menunjukkan halaman deskripsi leksikon depresi bagi aplikasi web *DepressioCare*. Laman ini akan memberikan penerangan tentang leksikon simptom depresi yang digunakan pada model ini. Simptom yang digunakan bagi leksikon depresi ini hanyalah simptom terpilih iaitu simptom pertama, kedua dan kelima yang terdapat di dalam instrumen Soal Selidik Kesihatan Pesakit-9 (*PHQ-9*). Simptom-simptom tersebut adalah mempunyai sedikit minat atau keseronokan dalam melakukan sesuatu, berasa sedih, tertekan, atau putus asa dan mempunyai kurang selera makan atau makan berlebihan.



Rajah 12 Halaman Utama bagi Aplikasi Web *DepressioCare*

Halaman utama bagi aplikasi *DepressioCare* merupakan halaman di mana memerlukan pengguna untuk memasukkan jawapan teks tentang penjelasan perasaan dan luahan atau keadaan mereka sebagai input. Selepas memasukkan input, pengguna boleh meneruskan menggunakan aplikasi dengan menekan butang “*Analyze*”

Based on your text, here are some terms that are related with depressive symptoms:

You have no signs of little interest or pleasure in doing things

You have shown sign of feeling down, depressed or hopeless which is: depression, unhappy, sad

You have no signs of poor appetite or overeating.

You have shown 3 depressive symptoms.

Next

Rajah 13 Halaman Senarai Istilah Berkaitan Dengan Simptom Depresi bagi Aplikasi Web *DepressioCare*

Tindakan menekan butang “*Analyze*” pada halaman utama akan membawa kepada halaman seperti pada Rajah 13. Halaman ini menunjukkan keputusan input jawapan teks tentang penjelasan perasaan pengguna merupakan salah satu simptom depresi yang dinyatakan ataupun tidak. Simptom depresi yang digunakan di dalam sistem ini adalah simptom pertama, kedua dan kelima yang terdapat pada instrumen Soal Selidik Kesihatan Pesakit-9 (*PHQ-9*). Jika input jawapan teks mempunyai istilah yang berkaitan dengan leksikon depresi bagi simptom terpilih tersebut, halaman ini akan memaparkan istilah yang berkaitan dengan depresi bagi setiap simptom yang berkenaan yang telah dinyatakan. Seterusnya, pengguna boleh meneruskan aplikasi dengan menekan butang “*Next*”.

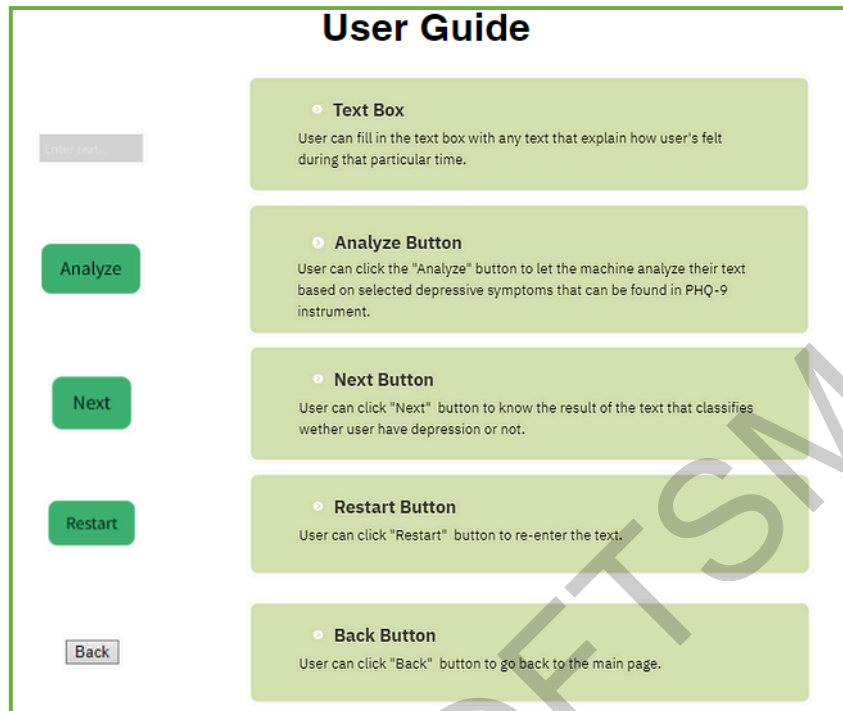
Here is your result,

Based on prediction score of 0.73, you MAY HAVE depression. Please seek professional help for further information.

Restart

Rajah 14 Halaman Keputusan Penentuan Depresi bagi Aplikasi Web *DepressioCare*

Tindakan menekan butang “*Next*” pada halaman sebelumnya akan membawa pengguna kepada halaman paparan keputusan bagi penentuan depresi. Halaman ini akan memberitahu pengguna bahawa teks mereka telah menunjukkan depresi mahupun tidak. Jika keputusannya menunjukkan teks pengguna mempunyai depresi, teks untuk pengguna digalakkan berjumpa dengan pakar untuk maklumat lanjut akan dipaparkan.



Rajah 15 Halaman Panduan Pengguna bagi Aplikasi Web *DepressioCare*

Selepas menekan butang menu bernama "*Help*" pada halaman muka, pengguna akan dibawa ke halaman ini. Halaman yang ditunjukkan di dalam Rajah 15 merupakan halaman bagi 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 *DepressioCare*.

Pengujian Kebolehgunaan

Pengujian kebolehgunaan ialah satu proses yang melibatkan pengujian akhir terhadap sistem yang dilakukan oleh wakil pengguna untuk memastikan model analisis dan aplikasi web *DepressioCare* berfungsi dengan baik dan memenuhi keperluan dan kepuasan pengguna. Kaedah yang digunakan bagi pengujian ini ialah dengan menggunakan format Ujian Penerimaan Pengguna (*UAT*). Ujian ini memfokuskan kepada pengesahan kemesraan pengguna, kefungsian dan prestasi perisian daripada perspektif pengguna. Maklum balas pengujian aplikasi berasaskan web *DepressioCare* telah dikumpulkan melalui borang atas talian yang bernama *Google Form*. Borang tersebut direka untuk pengguna menilai pelbagai aspek terhadap laman web seperti kebolehgunaan dan pengalaman pengguna serta kefungsian sistem. Terdapat sebanyak 14 responden yang telah mengambil bahagian dalam maklum balas ini. Setiap maklum balas diambil kira bagi mengoptimumkan prestasi aplikasi web *DepressioCare*. 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.	Laman web mudah untuk dilayari dan digunakan.			100% (14 responden)
2.	Reka bentuk laman web mudah untuk difahami.		7.1% (1 responden)	92.9% (13 responden)
3.	Maklumat yang terdapat di dalam web memudahkan untuk saya fahami tentang simptom depresi.		7.1% (1 responden)	92.9% (13 responden)
4.	Laman web berfungsi dengan baik.		28.6% (4 responden)	71.4% (10 responden)
5.	Input pengguna boleh dimasukkan dengan baik.		7.1% (1 responden)	92.9% (13 responden)
6.	Model ini boleh mengesan dan mengekstrak simptom depresi dengan baik.		28.6% (4 responden)	71.4% (10 responden)
7.	Model ini boleh meramal input pengguna sama ada input pengguna menunjukkan depresi atau tidak dengan baik.	7.1% (1 responden)	35.7% (5 responden)	57.1% (8 responden)

Berdasarkan maklum balas yang diterima, hampir semua hasil maklum balas amat memuaskan pengguna. Namun, terdapat hanya 1 orang responden sahaja yang berasa neutral terhadap kebolehan model untuk meramal input pengguna sama ada menunjukkan depresi atau tidak dengan baik. Ini terjadi mungkin kerana model gagal meramal input bukan depresi menjadi output sebagai depresi.

Cadangan Penambahbaikan

Selepas menjalankan kajian dan pengujian yang menyeluruh, antara penambahbaikan pada masa hadapan bagi memastikan projek ini dapat memberikan hasil yang lebih baik adalah dengan memastikan data bagi teks depresi dan bukan depresi yang mengandungi terma berkaitan dengan simptom terpilih adalah seimbang. Data yang seimbang dapat mengurangkan bias model terhadap kelas majoriti. Dengan data yang seimbang juga, model akan mempunyai peluang yang sama untuk belajar daripada kedua-dua kelas dan perkara ini dapat mengurangkan bias dan meningkatkan ketepatan prediksi.

Selain itu, penambahbaikan yang seterusnya ialah dengan meningkatkan penggunaan

fitur dengan menyertakan lebih banyak fitur yang relevan. Sebagai contoh, menjadikan semua simptom yang terdapat di dalam instrumen Soal Selidik Kesihatan Pesakit-9 (*PHQ-9*) sebagai fitur dapat membantu model untuk membuat prediksi yang lebih tepat.

KESIMPULAN

Kesimpulannya, aplikasi berasaskan web *DepressioCare* yang telah diterapkan dengan model terbaik iaitu dengan menggunakan algoritma Regresi Logistik, data sepadan dan fitur teks serta fitur bilangan semua simptom telah berjaya disiapkan. Sepanjang proses pembangunan, model analisis dan aplikasi web *DepressioCare* ini telah mencapai objektif utama dengan berkesan seperti yang telah ditetapkan pada peringkat awal pembangunan projek.

Kekuatan Sistem

Bagi projek ini, terdapat beberapa kekuatan projek yang perlu diketengahkan. Antaranya ialah tentang peningkatan senarai istilah yang berkaitan dengan simptom terpilih. Projek ini telah menambah baik senarai istilah yang berkaitan dengan simptom depresi yang terpilih daripada instrumen Soal Selidik Kesihatan Pesakit-9 (*PHQ-9*) iaitu simptom pertama, kedua dan kelima. Menambah baik leksikon depresi ini dapat meningkatkan keupayaan untuk mengenal pasti simptom dengan lebih tepat di dalam teks. Secara tidak langsung, menambah baik leksikon depresi juga dapat meluaskan liputan dalam setiap teks. Selain itu, kekuatan yang seterusnya ialah 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 analisis yang lebih mendalam terhadap kandungan teks pada fasa berikutnya. Di samping itu, kekuatan yang berikutnya ialah model yang dibangunkan dapat membuat prediksi terhadap teks dengan menentukan teks tersebut merupakan depresi ataupun tidak.

Kelemahan Sistem

Terdapat juga kekurangan semasa menjalankan proses pembangunan projek ini. Pada tahap pengujian data baharu, didapati bahawa terdapat beberapa data yang mempunyai sasaran bernilai 0 yang bermaksud teks tersebut tidak mengandungi depresi telah dikelaskan kepada label yang salah iaitu menunjukkan teks tersebut merupakan depresi. Sebagai contoh, teks yang dimasukkan oleh pengguna pada ruangan input ialah "*Today is the happiest day of my life.*". Nilai sasaran ialah 0 dan model telah membuat prediksi pada label ramalan sebagai 1 yang bermaksud teks tersebut merupakan teks depresi. Antara faktor yang menyumbang kepada kejadian ini adalah disebabkan oleh kekangan seperti data yang mempunyai simptom yang sepadan dengan simptom terpilih adalah tidak seimbang. Terdapat jurang perbezaan 19 data yang mempunyai unsur depresi lebih daripada data yang tidak mempunyai unsur depresi. Ketidakeimbangan data di mana teks depresi adalah lebih banyak menyebabkan model prediksi menjadi bias kepada teks depresi. Secara tidak langsung, kekangan ini akan mengurangkan ketepatan bagi teks yang bukan depresi.

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 menyiapkan projek ini dengan jayanya. 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

- Hotz, N. 2023, January 19. What is CRISP-DM?. Data Science Process Alliance. <https://www.datascience-pm.com/crisp-dm-2/> [29 Oktober 2023].
- Kementerian Kesihatan Malaysia (KKM). 2012, February 10. Apaitukemurungan? https://www.moh.gov.my/moh/resources/Penerbitan/Rujukan/NCD/Kesihatan%20Mental/Kemurungan_-_Apa_Yang_Anda_Perlu_Tahu.pdf [1 November 2023].
- Kementerian Kesihatan Malaysia (KKM). t.th. Self-test – Mentari Malaysia official portal. website statistics. <https://mentari.moh.gov.my/malay/self-test/> [29 Oktober 2023].
- Kumar, P. 2022, July 26. Kemurungan 101: Gejala, Punca & Rawatan. Homage Malaysia. <https://www.homage.com.my/ms/kesihatan/kemurungan/> [3 November 2023].
- National Institute of Mental Health. t. th. Depression. National Institute of Mental Health. <https://www.nimh.nih.gov/health/topics/depression> [29 Oktober 2023].
- World Health Organization. t.th. Depression. World Health Organization. https://www.who.int/health-topics/depression#tab=tab_1 [4 November 2023].

Nur Ain Sofia Binti Khamis (A188254)

Dr. Lailatul Qadri Binti Zakaria

Fakulti Teknologi & Sains Maklumat

Universiti Kebangsaan Malaysia