

KAEDAH ANALISIS SENTIMEN DAN PERLOMBONGAN PETUA SEKUTUAN DALAM MENEROKA GEJALA COVID JANGKA PANJANG MELALUI DATA CIAPAN

PRAVINASHWARN A/L RAJA SAGERAN

PROF. DR. AZURALIZA ABU BAKAR

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

ABSTRAK

Penghidap COVID-19 telah dikenal pasti menghadapi kemudaratan kesihatan untuk jangka masa yang panjang setelah dihadapi dengan jangkitan tersebut. Istilah ini dikenali sebagai COVID Jangka Panjang. Hampir 10 hingga 20 peratus penghidap COVID-19 disyaki mengalami COVID Jangka Panjang setelah selepas pulih daripada COVID-19. Media sosial seperti Twitter, Facebook, dan Instagram menjadi medium utama digunakan oleh masyarakat untuk berkongsi pendapat dan perasaan mereka mengenai COVID Jangka Panjang. Namun begitu, Rakyat masih tidak cakna mengenai kesan COVID Jangka Panjang atas sebab gejala-gejala yang dikesan adalah tidak luar biasa. Selain itu, kajian atau penyelidikan yang amat terhad dilakukan atas COVID Jangka Panjang menyebabkan tiada penjelasan yang kukuh terhadap keburukan sindrom COVID Jangka Panjang. Objektif projek ini adalah untuk mengenalpasti dan mengelaskan sentimen terhadap kesan COVID Jangka Panjang menggunakan data ciapan, tema-tema yang timbul berkaitan dengan COVID Jangka Panjang bagi mengenali gejala yang ketara serta meneroka hubungan antara gejala-gejala yang berkaitan dengan COVID Jangka Panjang. Data ciapan didapati daripada pautan Kaggle digunakan untuk mengumpulkan data mengenai pandangan rakyat tentang COVID Jangka Panjang. Data ciapan yang telah didapati dalam Bahasa Inggeris melalui Kaggle dengan kata kunci (long COVID syndrome, long COVID, post-COVID syndrome, postacute sequelae of SARS-CoV-2, long-term COVID, long haulers, dan chronic COVID syndrome) antara May 1, 2020, and Januari 31, 2023. Analisis merangkumi analisis sentimen, dan pemodelan topik dan perlombongan petua sekutuan untuk mengenalpasti dan meneroka topik-topik perbincangan dari semasa ke semasa. Komputeran Bahasa Tabi dan algoritma VADER digunakan untuk menentukan topik data ciapan yang paling dikongsi, mengelaskan sentimen, dan mengenalpasti tema dalam konteks COVID Jangka Panjang. Tiga sentimen utama dijangka dikesan dalam analisis adalah positif, negatif, dan neutral. Hasil analisis sentimen menunjukkan bahawa orang mempunyai tahap negativiti yang lebih tinggi terhadap COVID Jangka Panjang dan dijangkiti oleh pelbagai gejala berkaitan COVID Jangka Panjang. Kajian ini menerangkan tema-tema dominan, hubungan antara gejala COVID Jangka Panjang dan sentimen yang berkaitan dengan krisis kesihatan awam yang sedang berlaku. Kajian ini akan menjadi penting dalam membimbing penyebaran maklumat dan pelaksanaan dasar kerajaan yang diperlukan dalam

menangani sindrom COVID Jangka Panjang di seluruh dunia. Aplikasi web berdasarkan HTML dengan bantuan alat Python Flask akan dibangun untuk membentangkan output model jangkaan serta ketepatan model yang dibina. Project ini adalah sebahagian dari projek perundingan UKM-NIOSH "KAJIAN GEJALA & FAKTOR RISIKO LONG COVID DALAM KALANGAN PEKERJA SEKTOR PEMBUATAN DI MALAYSIA" UKMP-S230424 oleh Prof. Dr. Azuraliza Abu Bakar.

Kata kunci: COVID Jangka Panjang, Twitter, media sosial, analisis sentimen, pemodelan topik, perlombongan petua sekutuan

Pengenalan

Sindrom COVID Jangka Panjang merujuk kepada suatu keadaan di mana seseorang individu secara langsung mengalami gejala-gejala tertentu untuk jangka masa yang agak berterusan atau muncul secara tiba-tiba untuk jangka masa yang panjang setelah pemulihan dari jangkitan asal. Menurut Kementerian Kesihatan Malaysia (2023), bekas pesakit COVID-19 berkemungkinan mengalami gejala-gejala berterusan bagi tempoh 12 minggu atau lebih selepas sembuh. COVID Jangka Panjang adalah keadaan baru yang masih dikaji. Antara gejala-gejala COVID Jangka Panjang yang amat biasa termasuklah badan berasa telalu lemah dan penat, kesesakan nafas, hilang deria bau serta sakit otot (Murugesan 2021).

Menurut kajian dalam talian yang dijalankan dengan pelbagai pesakit yang disyaki dan disahkan COVID oleh Davis (2021) sebanyak 85.9% peserta mengalami kesan berulang, yang sebahagian besar disebabkan oleh senaman, usaha fizikal atau mental, dan tekanan. 86.7% responden yang belum pulih mengalami keletihan pada masa kaji selidik berbanding dengan 44.7% responden yang telah pulih. 45.2% responden meminta dan memerlukan pengurangan jadual kerja mereka berbanding sebelum jangkitan, dan tambahan 22.3% individu tidak bekerja pada masa kaji selidik kerana kesan selepas penyakit. Masalah disfungsi kognitif atau ingatan adalah umum di kalangan semua kumpulan umur (Davis et al. 2021).

Kajian terhadap COVID Jangka Panjang kurang dipergiatkan secara berleluasa kerana kesedaran tentang sindrom jangka panjang sindrom baru mendapat perhatian umum dan gejala-simptom baru-baru ditemui setelah penularan wabak COVID-19. Oleh itu, tiada bukti atau kajian sokongan yang kukuh menyatakan kesan atau gejala-gejala yang jelas. Bagi mengkaji ideologi tentang kesan COVID Jangka Panjang, set data yang dicapai amat jarang yang boleh didasarkannya untuk membuat penyelidikan secara mendalam. Jadi, salah satunya cara yang boleh dijadikan adalah dengan mendapatkan data daripada media sosial. Tujuan media sosial menjadi saluran utama adalah sebab media sosial mudah diperalatkan orang-orang ataupun mana-mana organisasi dalam generasi ini untuk berkongsi idea atau pengalaman mereka secara terbuka. Jadi, sudah pastinya kita boleh mendapatkan maklumat-maklumat terkini mengenai COVID Jangka Panjang ini dengan lebih mudah dan efektif. Di samping itu, pandangan atau kesan mengenai COVID Jangka Panjang ini tidak bersifat mutlak dan mungkin berbeza mengikut pengalaman masing-masing. Oleh itu, keadaan ini membawa kesukaran bagi media sosial untuk digunakan bagi mencapai set data terkini untuk mengkaji sentimen pengguna berkenaan isu COVID Jangka Panjang dengan lebih lanjut.

Setelah meneliti masalah-masalah yang bakal ditimpa bagi menjalankan kajian terhadap gejala COVID Jangka Panjang ini, terdapat beberapa inisiatif yang telah diambil demi mengkaji data COVID Jangka Panjang dengan cara yang lebih efisien. Antara langkah-langkah yang diambil adalah dengan menyaring data ciapan melalui pautan Kaggle. Tujuan data ciapan digunakan melainkan sosial media lain adalah sebab aplikasi Twitter merupakan platform yang terkenal digunakan untuk mendapat informasi berita dan berkongsi pandangan atau pengalaman seseorang dalam bentuk teks dan pra-pemprosesan data teks dilaksanakan bagi memastikan data yang diperoleh bersih dan dapat mengenalpasti sentimen serta istilah penting dalam set data ciapan. Analisis sentimen digiatkan menggunakan algoritma seperti VADER demi mengkaji sentimen yang boleh mendatangkan kepada manusia. Sentimen positif, negatif dan neutral dikelaskan bagi menilai tahap sentimen COVID Jangka Panjang kepada manusia. Hasil kajian ini memberi pengetahuan kepada semua tentang impak COVID Jangka Panjang terhadap manusia secara menyeluruh dan mendalam.

Teknik pemodelan topik dijalankan terhadap set data kajian kita bagi mengelakkan kesan-kesan yang sering diperkatakan apabila COVID Jangka Panjang menjadi topik perbualan rakyat. Dengan ini, ideologi dan persepsi kebanyakan manusia tentang COVID Jangka Panjang dikenal pasti dan penemuan kajian ini dapat membolehkan kesan-kesan atau gejala COVID Jangka Panjang yang ketara dapat dilihat dengan jelas serta mengambil langkah penyelesaian bagi gejala yang membawa kepada keburukan kepada manusia. Perlombongan Petua Kesukutan dilakukan melalui gejala-gejala yang dikenalpasti melalui pemodelan topik untuk meneroka hubungan tersembunyi antara gejala COVID Jangka Panjang.

Keputusan kajian ini juga boleh digunakan oleh jabatan kajian dan penyelidikan sains sebagai dasar untuk membuat ramalan gejala COVID Jangka Panjang terhadap sesebuah individu atau membuat kajian bagi mencari jalan penyelesaian bagi mengurangkan risiko kesan buruk gejala tersebut. Data-data daripada aplikasi telah melalui fasa pembersihan dan diubah sesuai mengikut standard piawaian yang betul bagi memperkayakan data supaya dapat mengelakkan sentimen konteks dengan lebih tepat dan jitu. Hasil kajian ini juga dibentangkan dalam bentuk visual dalam bentuk aplikasi web supaya dapat memudahkan pihak berkenaan boleh mengambil tindakan yang sesuai untuk menyelesaikan masalah gejala COVID Jangka Panjang ini.

METODOLOGI KAJIAN

Metodologi yang digunakan dalam menjalankan kajian analisis sentimen adalah CRISP-DM model. CRISP-DM, yang merujuk kepada Cross-Industry Standard Process for Data Mining, adalah cara yang terbukti dalam industri untuk membimbing usaha perlombongan data. Metodologi ini secara aktif dipergiatkan dalam banyak projek industri kerana ciri-cirinya yang amat fleksibel dan praktikal dipraktikkan dalam kebanyakan isu-isu yang melibatkan kajian atau skop kerja analitik (Schröer et al. 2021). Pembangunan sistem yang terancang boleh dicapai melalui metodologi pembangunan model analitik data CRISP-DM seperti yang ditunjukkan dalam Rajah 1.



Rajah 1 Enam Fasa CRISP-DM
Sumber (HOTZ 2024)

Pemahaman Bisnes

Fasa pertama iaitu pemahaman perniagaan adalah fasa di mana objektif dan matlamat projek perlombongan data ditentukan. Dalam konteks ini, objektif utama perlombongan data mengenai kesan-kesan COVID Jangka Panjang adalah untuk mendapatkan data ciapan pendapat dan pengalaman para pesakit COVID-19 tentang gejala berpanjangan setelah dijangkiti wabak penyakit tersebut. Setelah mendapatkan data ciapan khusus kepada kesan COVID Jangka Panjang, kajian ini berfokus kepada menganalisis sentimen kesan COVID Jangka Panjang bagi membezakan sentimen positif, negatif serta neutral di samping merangka topik pemodelan bagi mengenal pasti gejala yang ketara dengan sentimen negatif yang didapati dalam isu COVID Jangka Panjang tersebut. Hubungan antara gejala yang dikenal pasti juga perlu diterokai setelah melalukan pemodelan topik. VADER, Matplotlib, Apriori dan Latent Dirichlet Allocation (LDA) merupakan antara algoritma yang dicadangkan untuk membantu menjayakan objektif kajian ini. Hasil kajian ini membantu kepada ahli saintis untuk menjalankan kajian lanjutan bagi menangani kesan negatif COVID Jangka Panjang serta membantu rakyat supaya lebih cakna atas pemahaman kesan-kesan COVID Jangka Panjang yang telah diwar-warkan dalam masa ini. Jika data-data yang telah dianalisis dapat berjaya menemukan gejala-gejala ketara dan hubungan antara gejala, kajian tersebut akan dianggarkan berjaya dan dapat diterima oleh pengguna.

Pemahaman Data

Dalam fasa ini, pengumpulan dan penerokaan data akan dimulakan bagi tujuan analisis sentimen kesan COVID Jangka Panjang. Bagi tujuan kes kajian ini, set data telah dicapai melalui Twitter kerana Twitter merupakan wadah utama digunakan oleh hampir seluruh manusia dalam dunia digital bagi berkongsi idea dan persepsi mereka dengan cara yang terbuka. Dengan adanya data ciapan di pautan Kaggle, sudah pastinya kerja pencapaian set data kita menjadi begitu mudah kerana segala maklumat daripada pengalaman orang boleh dijumpai di sana dalam kuantiti yang mendadak. Data ciapan adalah data teks yang amat terkenal dengan penyebaran berita serta perkongsi maklumat secara besar-besaran

berbanding media sosial yang lain seperti Tiktok, Instagram dan Snapchat yang lebih memfokuskan dalam penyebaran gambar dan grafik animasi berkaitan aktiviti harian mereka. Fasa ini juga memerlukan data untuk diterangkan dari segi format, kuantiti dan ciri permukaan lain yang ditemui. Dengan ini, pemeriksaan data amat diperlukan bagi memastikan data yang dikumpul lengkap dan tepat.

text
RT @DrDayaSharma: It's the ultimate in short-term thinking and policy to ignore the health, social and economic costs of #LongCovid.
RT @maosbot: It is hard enough going through #LongCovid *without* the feeling that no-one is learning from you, that your suffering
So a friend of mine who has been double vaccinated and lost his sense of taste has had a partial recovery! Not a fool proof remedy, but as a potential #LongCovid or post-#COVID19 recommendation, I felt it was worth sharing.
This brightened up my weekend to hear about. https://t.co/GemoKlr1dc
@globalnews Our community thanks you for doing such a well-thought-out and written article detailing the cases of many differing #LongCovid patients. Thank you.

Rajah 2 Contoh teks ciapan daripada aplikasi Twitter

Sumber: (Matharaarachchi et al. 2022)

Jadual 1 Contoh Atribut Set Data Ciapan Pesakit COVID Jangka Panjang

Atribut	Keterangan	Format
teks	Data ciapan tentang perkara Covid-19 Jangka Panjang.	String
tarikh	Tarikh ciapan dimuat naik	DateTime

Penyediaan Data

Setelah data difahami dan dicapai, pemprosesan dan penggubalan data perlu dibuat supaya data sesuai untuk tujuan analisis. Fasa ini dikenal sebagai penyediaan data. Data-data yang dicapai perlu melalui pra-pemprosesan data, seperti pembersihan data, penyatuan data, transformasi data dan pengurangan data. Dalam hal ini, kod Python khas telah dihasilkan untuk membersihkan dan pra-pemprosesan data ciapan yang telah disaring dengan mengeluarkan URL, emoji, aksara khas, retweets, nombor, digit, ruang putih, simbol seperti @./, // laman web, dan sebarang maklumat yang boleh dikenal pasti, pautan hiper, simbol pagar serta kata gubahan. Bukan itu sahaja data juga dihantar untuk dinormalisasi menggunakan teknik pembumian (Lemmatization). "Pembumian Kata Wordnet" digunakan untuk mengurangkan kata kepada bentuk asal atau bentuk kamusnya, dan mengenal pasti jenis kata (Part Of Speech) dengan menerapkan analisis morfologi untuk menentukan bentuk asasnya dengan lebih tepat. Pembumian telah digunakan terhadap teks ciapan untuk membenarkan normalisasi teks dijalankan supaya teks mempunyai ciri-ciri subjek dan memiliki kandungan yang paling sesuai untuk analisis sentimen. Perkataan bigram dihasilkan demi mencari gejala-gejala dan vektorisasi matrik TF-IDF digiatkan bagi menapis unsur-unsur yang relevan dalam dokumen. Dengan ini semua teks data telah disusun mengikut struktur yang sama untuk tindakan analisis selanjutnya.

Pemodelan Data

Beralih kepada fasa seterusnya iaitu pemodelan data di mana dalam fasa ini berlaku pemilihan dan penggunaan pelbagai teknik analisis sentimen data untuk mencipta model yang boleh membantu menangani objektif kajian kesan COVID Jangka Panjang ini. Dalam hal ini, beberapa model kajian yang akan ditubuhkan dalam kajian ini untuk mengkaji secara telus data COVID Jangka Panjang. Dengan menggunakan Python sebagai bahasa pengaturcaraan utama, VADER akan dibina bagi membezakan dan mengkaji tahap positifiti dan negativiti data yang secara terus mewakili kesan yang mendatangkan oleh COVID Jangka Panjang. Di samping itu, topik pemodelan mengikut kerelevanan juga dibina menggunakan Latent Dirichlet Allocation (LDA) bagi merumuskan kesan-kesan utama dalam cara yang ringkas dan mudah dibaca dan dianalisis oleh orang lain. Apriori pula digunakan sebagai algoritma dalam proses perlombongan petua sekutuan bagi mencari hubungan antara gejala. Dengan ini, analisis teks dan sentimen dapat dijalankan dengan cara yang bijak.

Penilaian

Setelah selesai membina model, fasa beralih kepada penilaian. Fasa ini membantu kita membuat penilaian sejauh mana keputusan model yang dibangunkan memenuhi kriteria objektif kajian. Dengan ini, segala keputusan yang diperolehi daripada model yang dibangunkan direkodkan dan disimpan. Dalam kes ini, penilaian dilakukan bahawa sejauh manakah COVID Jangka Panjang dapat dibuktikan membawa kesan negatif atau sebaliknya melalui pengelasan sentimen positif, negatif dan neutral. Ujian prestasi setiap algoritma yang dibina dilakukan bagi menilai keberkesanan model dalam setiap analisis yang dilaksanakan. VADER telah dilakukan ujian prestasi dengan menghasilkan laporan klasifikasi dan skor AUC. LDA pula diuji menggunakan skor koheren, kekeliruan, kepelbagaian dan kestabilan topik. Di samping itu, petua sekutuan yang relevan dipilih berdasarkan skor lif, keyakinan dan sokongan yang tinggi. Ulasan projek perlu dilakukan untuk membuat sebuah ringkasan kepada keseluruhan proses dan mengkaji hasil keputusan teks yang telah didapati pada akhir kajian. Selepas itu, keputusan akan dibuat jika keputusan projek sudah memadai. Fasa penilaian adalah kunci dalam analisis sentimen data kerana ia membolehkan pemahaman lebih mendalam tentang bagaimana pengguna Twitter berasa atau mengalami tentang isu atau topik tertentu, yang boleh digunakan untuk pengambilan keputusan dan pemantauan selanjutnya.

Penggunaan

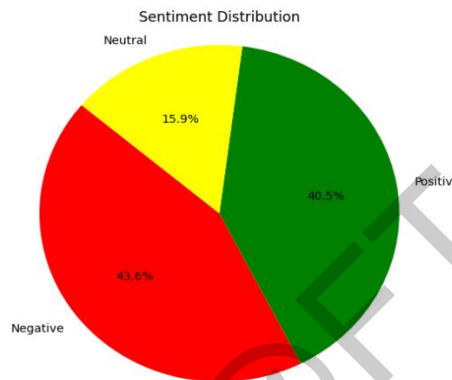
Setelah model yang memuaskan dikenal pasti, ia boleh dikeluarkan sebagai suatu kajian yang lengkap dan sedia untuk proses penerbitan. Dalam fasa ini perkembangan model analisis sentimen, pemodelan topik dan petua sekutuan berjaya dibangunkan mengikut objektif yang dijangkakan. Selain itu, dalam pembentangan akhir, sistem visualisasi laman web dibina menggunakan HTML dengan bantuan Python Flask dibina bagi mengintegrasikan segala keputusan visualisasi model berserta membina model ramalan COVID Jangka Panjang berdasarkan skor keyakinan daripada petua sekutuan dan pemodelan yang telah dianalisis kerana ini akan memudahkan kefahaman dan penggambaran visual yang baik. Akhir sekali, proses dokumentasi perlu dihasilkan untuk merekod prosedur dan idea-idea penting mengenai

cara pengendalian kajian serta pengalaman yang diperoleh semasa menjalankan kajian ini.

KEPUTUSAN DAN PERBINCANGAN

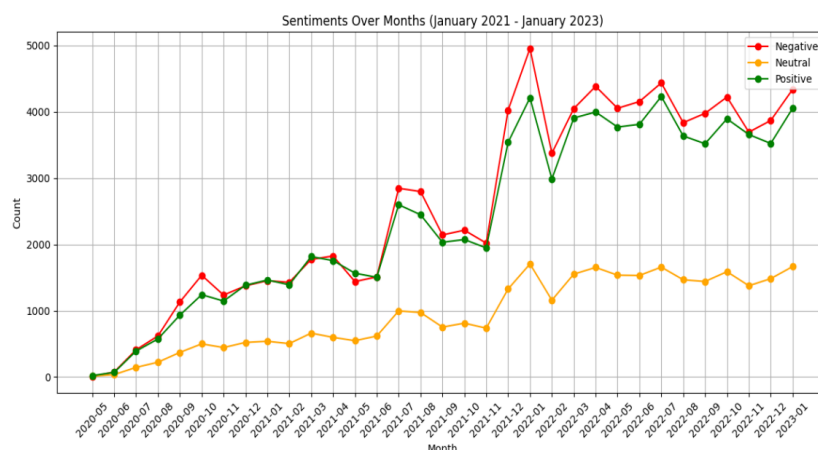
Analisis Sentimen Menggunakan VADER

Set data ciapan yang mengandungi 195476 baris data telah dikelaskan mengikut 3 sentimen utama iaitu positif, negatif dan neutral melalui VADER. Rajah 3 menunjukkan carta pie yang menggambarkan taburan set data yang dibahagikan mengikut sentimen masing-masing.



Rajah 3 Taburan Sentimen Data Ciapan

Dari hasil analisis, 40.5% daripada data ciapan menunjukkan sentimen positif, manakala 43.6% menunjukkan sentimen negatif, dan 15.9% lagi adalah neutral. Hal ini menunjukkan bahawa majoriti data (43.6%) menggambarkan sentimen negatif terhadap gejala COVID Jangka Panjang. Sentimen ini mencerminkan kebimbangan, ketidakpuasan, atau kesan negatif yang dialami oleh individu yang terjejas oleh COVID Jangka Panjang. Sebaliknya, 40.5% data mempunyai sentimen positif, yang boleh menunjukkan adanya pengalaman yang baik atau pemulihan daripada gejala tersebut. Sentimen neutral (15.9%) menunjukkan bahawa sebahagian daripada entri mungkin tidak menunjukkan emosi yang kuat sama ada positif atau negatif. Analisis sentimen ini juga diteruskan dengan mengambil kira taburan data mengikut sentimen menjelang bulan daripada Mei 2020 sehingga Januari 2023 yang merupakan julat tarikh awal dan akhir set data yang dikumpul..



Rajah 5 Graf Garisan Taburan Sentimen Menjelang Bulan

Berdasarkan graf garisan, terdapat peningkatan ketara dalam sentimen positif pada sekitar bulan Mei 2021, yang terus meningkat secara beransur-ansur hingga mencapai puncak pada bulan Januari 2022. Selepas Januari 2022, sentimen positif menurun sedikit tetapi masih kekal pada tahap yang lebih tinggi berbanding tempoh sebelum Mei 2021. Ini mungkin menunjukkan peningkatan dalam pengalaman positif atau laporan pemulihan yang lebih baik daripada gejala COVID-19 dalam kalangan pesakit.

Sentimen negatif menunjukkan fluktuasi yang lebih besar dengan beberapa puncak yang jelas, terutamanya pada sekitar bulan Oktober 2021 dan Januari 2022. Puncak tertinggi sentimen negatif pada Januari 2022 ramai gejala COVID Jangka Panjang yang dikesan pada waktu dalam tempoh selama 2 tahun wabak COVID-19 berleluasa. Selepas Januari 2022, sentimen negatif juga menurun sedikit tetapi masih kekal tinggi, menunjukkan bahawa terdapat masih ramai menghadapi pengalaman negatif dalam mengenai gejala COVID Jangka Panjang setelah bertahun-tahun dunia menghadapi pandemik COVID-19.

Sentimen neutral menunjukkan peningkatan yang lebih stabil dan tidak banyak fluktuasi berbanding sentimen positif dan negatif. Terdapat sedikit peningkatan dalam sentimen neutral dari Mei 2021 hingga Januari 2023, yang mungkin menunjukkan bahawa lebih banyak individu yang melaporkan pengalaman yang tidak terlalu positif atau negatif. Pada bulan Januari 2022, terdapat puncak yang jelas bagi sentimen negatif dan positif, menunjukkan variasi yang besar dalam pengalaman pesakit pada waktu tersebut. Ramai pesakit difahamkan pulih setelah dihadapi gejala COVID Jangka Panjang di samping ramai yang baru dihadapi gejala COVID Jangka Panjang pada jangka masa tersebut. Secara keseluruhan, jumlah sentimen negatif melebihi jumlah sentimen positif sepanjang tempoh masa yang diteliti, menunjukkan bahawa lebih ramai pesakit mungkin mengalami pengalaman negatif dengan gejala COVID Jangka Panjang.

Peningkatan mendadak dalam sentimen positif dan negatif pada akhir tahun 2021 hingga awal tahun 2022 mungkin menunjukkan kejadian atau intervensi khusus, seperti pengenalan vaksinasi COVID-19 yang mungkin berbantu dalam mengelakkan gejala COVID Jangka Panjang atau peningkatan dalam mengesan kes COVID Jangka Panjang yang mempengaruhi pengalaman pesakit. Keberadaan sentimen positif yang hampir seimbang dengan sentimen negatif pada beberapa bulan menunjukkan bahawa walaupun terdapat kesan negatif yang signifikan, terdapat juga laporan pengalaman positif atau pemulihan dalam kalangan pesakit. Laporan pengelasan yang dikeluarkan sebagai keputusan pengujian model VADER juga menggambarkan keberkesanan model VADER dengan lebih lanjut. Jadual 2 menggambarkan nilai setiap metrik yang dikemukakan dalam laporan pengelasan setelah model diuji.

Jadual 2 Laporan Pengelasan VADER

	Kejituan	Dapatan Semula	Skor F1	Sokongan
Positif	0.20	0.58	0.30	27166

Neutral	0.09	0.26	0.13	10509
Negatif	0.89	0.48	0.62	157801
Ketepatan			0.48	195476
Purata Makro	0.39	0.44	0.35	195476
Purata Pemberat	0.75	0.48	0.55	195476

Kejituan purata pemberat adalah 0.75. Ini bermaksud, secara keseluruhan, 75% daripada semua ramalan sentiment diramal oleh VADER adalah tepat. Ketepatan pemberat purata yang tinggi (0.75) menunjukkan bahawa model VADER cukup baik dalam membuat ramalan yang tepat bagi sentimen dan mengelakkan kesalahan dalam pengelasan sentimen. Hal ini penting kerana menunjukkan bahawa kebanyakan sentimen yang diramalkan oleh model adalah benar.

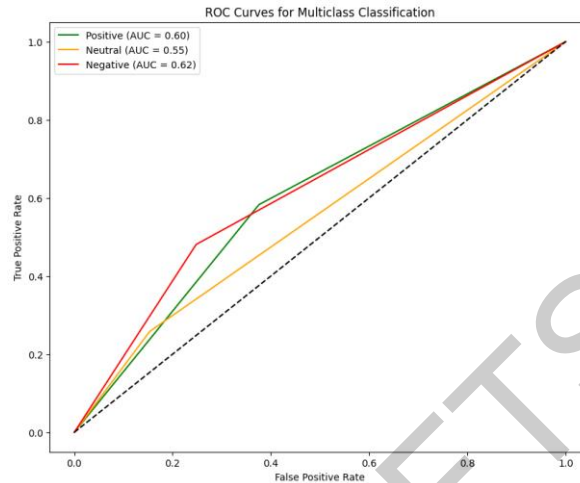
Pemberat purata dapatan semula model VADER adalah sebanyak 0.48. Ini menunjukkan bahawa model hanya dapat mengenalpasti 48% sentimen daripada semua sentimen yang dilabel dengan betul. Purata pemberat dapatan semula yang sederhana (0.48) menunjukkan bahawa model mempunyai kesukaran dalam mengenalpasti semua sentimen dengan tepat, terutama sekali apabila jumlah kes untuk beberapa kategori sentimen adalah jauh berbeza. Namun, ia masih menunjukkan bahawa hampir separuh daripada sentimen yang benar dapat dikenalpasti dengan betul. Purata pemberat skor F1 adalah 0.55. Ini bermaksud, terdapat keseimbangan sederhana antara kejituan dan dapatan semula, memberikan gambaran yang lebih menyeluruh tentang prestasi model.

Purata pemberat skor F1 (0.55) memberikan gambaran seimbang tentang bagaimana model berprestasi dalam mengimbangi kejituan dan dapatan semula. Walaupun model mempunyai dapatan semula yang lebih rendah, kejituan yang tinggi membantu mengekalkan skor F1 yang sederhana. Jumlah sokongan adalah 195476, yang menunjukkan jumlah keseluruhan data yang digunakan dalam laporan ini. Jumlah sokongan yang besar (195476) menunjukkan bahawa model diuji pada set data yang sangat besar, memberikan keyakinan bahawa hasilnya adalah tidak berat sebelah. Hal ini juga menunjukkan bahawa hasil keputusan ini disokong dan diterima kerana diuji pada data yang luas.

Ketepatan keseluruhan merujuk kepada peratusan keseluruhan prediksi yang betul daripada jumlah keseluruhan kes. Ini membawa makna bahawa 48% daripada semua ramalan adalah tepat. Hampir separuh daripada semua ramalan adalah tepat. Walaupun ketepatan dilihat sederhana tetapi jumlah sokongan yang amat besar menyebabkan model diuji adalah masih dipercayai untuk menangkap variasi dalam 92 sentimen COVID Jangka Panjang. Walaupun dapatan semula tidak setinggi kejituan, skor F1 menunjukkan bahawa ada keseimbangan yang baik antara kejituan dan dapatan semula, memberikan gambaran yang adil tentang prestasi model.

Laporan pengelasan VADER menunjukkan bahawa model ini cukup baik dalam mengenalpasti sentimen dalam set data COVID Jangka Panjang. Dengan purata pemberat

kejituan yang tinggi, VADER menunjukkan keupayaan yang kuat dalam membuat ramalan yang benar, sementara skor F1 memberikan keyakinan tambahan bahawa keseimbangan antara kejituan dan dapatan semula adalah memadai. Oleh itu, VADER boleh dianggap sebagai alat yang andal untuk analisis sentimen dalam konteks COVID Jangka Panjang.



Rajah 6 Graf ROC bersama nilai AUC

Graf di Rajah 4.24 menunjukkan ROC dan skor AUC untuk pengelasan berbilang kelas (Positif, Neutral, Negatif) menggunakan model VADER. Skor AUC positif (0.60), menunjukkan bahawa model mempunyai keupayaan pemisahan yang lebih baik sedikit daripada keputusan rawak untuk mengelaskan sentimen positif. Walaupun ini tidak mencerminkan prestasi yang sangat kuat, ia masih menunjukkan bahawa model mampu mengenalpasti lebih daripada separuh ciri positif dalam teks.

Skor AUC (0.55) bagi sentimen neutral pula menunjukkan bahawa model mempunyai keupayaan pemisahan yang agak lemah untuk mengelaskan sentimen neutral. Ini mungkin disebabkan oleh kesukaran dalam membezakan teks neutral daripada teks positif atau negatif, terutamanya dalam konteks COVID Jangka Panjang di mana sentimen sering lebih jelas negatif atau positif.

Skor AUC negatif (0.62) ini adalah yang tertinggi di antara tiga kelas, menunjukkan bahawa model mempunyai keupayaan yang lebih baik untuk mengenalpasti sentimen negatif. Dalam konteks COVID Jangka Panjang, sentimen negatif mungkin lebih mudah dikenalpasti kerana aduan, kebimbangan, dan pengalaman berkaitan gejala negatif sering lebih tidak tersirat. Keputusan AUC amat menguntungkan kajian ini kerana set data sentimen negatif yang bakal digunakan untuk analisis teks selanjutnya bagi meneroka gejala COVID Jangka Panjang.

Semua skor AUC berada di atas 0.5, menunjukkan bahawa VADER mampu membuat diskriminasi yang lebih baik daripada keputusan rawak. Ini menunjukkan bahawa VADER mempunyai asas dalam mengenalpasti sentimen dalam teks walaupun dalam konteks yang mencabar seperti COVID Jangka Panjang.

Pemodelan Topik LDA

Jadual 3 menunjukkan 30 istilah utama yang dikategorikan mengikut kerelevanan 3 topik yang diterokai oleh LDA.

Jadual 3 Keputusan LDA mengikut Topik

Topik LDA	Istilah Kerelevanan
Topik 1	<i>joint_pain, sore_throat, since_march, muscle_pain, chronic_pain, cant_work, health_issue, much_worse, back_pain, side_effect, short_breath, many_person, feel_better, able_work, glandular_fever, sick_march, feel_pain, heart_issue, getting_better, nerve_pain, nerve_damage, much_better, unable_work, cant_even, high_risk, much_pain, nervou_system, cant_sleep, cant_breathe</i>
Topik 2	<i>chest_pain, chronic_fatigue, extreme_fatigue, heart_rate, back_work, mild_case, chronic_illnes, ever_since, go_back, lung_damage, heart_attack, blood_clot, fatigue_pain, immune_system, come_go, heart_lung, constant_pain, symptom_since, health_problem, heart_problem, sick_since, los_taste, back_normal, life_back, many_u, cant_afford, return_work, quality_life, wish_anyone, last_march</i>
Topik 3	<i>mental_health, body_ache, taste_smell, blood_pressure, smell_taste, wear_mask, viral_fatigue, shortnes_breath, sense_smell, breathing_issue, many_symptom, make_feel, came_back, come_back, sore_arm, go_away, cant_walk, hair_los, sense_taste, wearing_mask, cognitive_issue, feel_bad, every_single, brain_damage, since_last, high_blood, getting_worse, rest_life, whole_body, brain_injury</i>

Berdasarkan topik 1, topik ini berpusat kepada gejala-gejala yang berkaitan dengan rasa sakit dan kesukaran fizikal yang berterusan, serta masalah-masalah pernafasan. Topik ini menekankan kepada masalah kesihatan yang teruk dan berpanjangan seperti brain fog (kekeliruan mental), joint pain (sakit sendi), muscle pain (sakit otot), dan chronic pain (sakit kronik). Ia juga mencakupi isu pernafasan seperti short breath (sesak nafas) dan cant breathe (tidak boleh bernafas). Keseluruhannya, topik ini menggambarkan bagaimana ramai individu mengalami penurunan kualiti hidup yang ketara akibat kesan COVID Jangka Panjang, dengan banyak yang melaporkan tidak mampu bekerja atau berfungsi seperti biasa (cant work, unable work).

Topik 2 berfokus kepada masalah jantung dan keletihan kronik serta kesan

berpanjangan yang dirasakan sejak terjangkit COVID. Topik ini menonjolkan gejala seperti chest pain (sakit dada), chronic fatigue (keletihan kronik), heart rate (kadar denyutan jantung), dan lung damage (kerusakan paru-paru). Terdapat juga penekanan pada masalah yang berterusan sejak mula sakit seperti symptom since (gejala sejak) dan sick since (sakit sejak). Banyak individu merasakan sukar untuk kembali kepada keadaan normal sebelum sakit (back normal, life back), dan masalah kesihatan ini turut memberi kesan kepada kualiti hidup mereka (quality life).

Topik 3 pula merangkumi isu kesihatan mental, kehilangan deria rasa dan bau, serta gejala lain yang mempengaruhi keseluruhan tubuh badan. Topik ini merangkumi gejala mental health (kesihatan mental), body ache (sakit badan), taste smell (rasa dan bau), dan blood pressure (tekanan darah). Ia juga menekankan masalah pernafasan seperti shortnes breath (sesak nafas) dan breathing issue (masalah pernafasan). Beberapa individu melaporkan kesan jangka panjang seperti hair loss (kehilangan rambut), cognitive issue (masalah kognitif), dan brain injury (kecederaan otak). Topik ini menggambarkan bagaimana COVID Jangka Panjang memberi kesan menyeluruh kepada fizikal dan mental individu yang terjangkit.

Walaupun terdapat sebahagian istilah yang menimbulkan kekeliruan dalam skop analisis topik LDA, namun ketiga-tiga topik ini menunjukkan betapa luas dan pelbagai gejala COVID Jangka Panjang, yang bukan sahaja merangkumi masalah fizikal tetapi juga mental, serta kesan-kesan berpanjangan yang mengubah kualiti hidup individu yang terkesan secara menyeluruh. Tafsiran ini membantu dalam memahami pelbagai aspek kesan COVID Jangka Panjang dan menyediakan ruang untuk menjadikan sebagai perwakilan beg kata untuk peringkat analisis ARM. Dalam konteks pengujian model LDA, beberapa metrik digunakan untuk menilai kualiti topik yang dihasilkan. Jadual 4 menunjukkan segala keputusan metrik yang digunakan untuk menguji prestasi LDA.

Jadual 4 Skor Prestasi LDA

Metrik Pengujian	Skor
Skor Koheran	0.34380634832191487
Skor Kekeliruan	-93.2951981590944
Skor Kepelbagaian Topik	3.0
Skor Purata Kestabilan Topik	0.4615384615384615

Skor koheran yang lebih tinggi menunjukkan bahawa kata-kata dalam topik tertentu lebih berkaitan dan lebih bermakna, manakala skor yang lebih rendah menunjukkan topik yang kurang koheren atau kurang berkaitan secara semantik. Skor 0.34 menunjukkan bahawa terdapat tahap koherensi yang sederhana dalam topik-topik yang dihasilkan oleh model LDA ini. Ini bermakna topik-topik tersebut agak berkaitan secara semantik, tetapi masih ada ruang untuk penambahbaikan.

Kekeliruan yang lebih rendah menunjukkan model yang lebih baik kerana ia menunjukkan bahawa model tersebut lebih baik dalam meramalkan data yang tidak dilihat.

Kekeliruan yang sangat negatif seperti -93.295 menunjukkan bahawa model ini cukup baik dalam meramalkan data teks yang belum dilihat. Skor kepelbagaian topik 3.0 menunjukkan bahawa model LDA ini telah mengenalpasti tiga topik utama dalam data teks. Ini adalah penemuan yang berguna kerana ia memberikan gambaran tentang berapa banyak kelompok topik yang berbeza yang wujud dalam data tersebut.

Skor kestabilan topik purata yang sederhana (0.46) menunjukkan bahawa topik-topik yang dihasilkan oleh model LDA adalah agak stabil, yang bermaksud bahawa apabila model diterapkan kepada subset data yang berbeza, topik-topik yang dihasilkan adalah agak tekal. Ini menunjukkan bahawa model tersebut mampu mengenalpasti topik yang sama merentasi pelbagai subset data dan tidak bertindih antara kelompok topik dengan kadar yang sederhana.

Secara keseluruhan, hasil LDA ini menunjukkan bahawa model yang digunakan mampu mengenalpasti topik-topik yang agak koheren dengan kekeliruan yang baik dan menghasilkan topik-topik yang stabil. Skor kestabilan topik yang sederhana adalah membolehkan model agak boleh dipercayai dalam mengklasifikasikan teks kepada topik-topik yang tekal.

Algoritma Apriori Untuk Perlombongan Petua Sekutuan

Daripada 48 petua yang dihasilkan kajian telah menjalani penapisan petua di mana hanya 18 petua dikenal pasti boleh dipercayai berdasarkan metrik-metrik yang dikaji. Jadual 4.6 menunjukkan 18 petua ARM yang berjaya diterokai dan dijamin dalam konteks COVID Jangka Panjang melalui perbandingan skor-skor metrik yang diuji. 5 petua utama dalam Jadual 5 dikenal pasti memiliki hubungan yang sangat kuat antara gejala berdasarkan skor keyakinan yang tinggi

Jadual 5 Peraturan ARM yang ditapis

Penyebab	Akibat	Keyakinan	Sokongan	Lif
(brain_fog, loss_smell)	(loss_taste,)	0.528302	0.001850	27.473514
(loss_taste,)	(loss_smell,)	0.697595	0.013414	12.582476
(brain_fog, body_ache)	(cognitive_issue,)	0.528302	0.001850	14.225609
(headache,)	(brain_fog,)	0.493333	0.002445	10.047932
('cognitive_issue',)	('brain_fog',)	0.368327	0.013679	7.501882
('fatigue',)	('body_ache',)	0.26378	0.008855	6.686391
('heart_issue',)	('lung_issue',)	0.156863	0.001586	10.05849
('breathing_issue',)	('brain_fog',)	0.137143	0.001586	2.793247
('chest_pain',)	('fatigue',)	0.112994	0.001322	3.366031
('loss_taste',)	('brain_fog',)	0.109966	0.002115	2.239717
('chest_pain',)	('brain_fog',)	0.101695	0.001189	2.071264
('fatigue',)	('cognitive_issue',)	0.094488	0.003172	2.544288
('fatigue',)	('brain_fog',)	0.09252	0.003106	1.884388
('breathing_issue',)	('fatigue',)	0.091429	0.001057	2.7236
('body_ache',)	('brain_fog',)	0.088777	0.003502	1.808164
('cognitive_issue',)	('body_ache',)	0.085409	0.003172	2.164989
('brain_fog',)	('loss_smell',)	0.071332	0.003502	1.286619

('cognitive_issue')	('brain_fog')	0.368327	0.013679	7.501882
---------------------	---------------	----------	----------	----------

Petua (brain_fog, loss_smell) -> (loss_taste) dipilih kerana memiliki sokongan yang rendah tetapi keyakinan dan lif yang tinggi. Sokongan 0.001850 menunjukkan kejadian yang jarang, tetapi keyakinan 0.53 dan lif 27.47 menunjukkan hubungan yang kuat di antara kehilangan bau, kabut otak, dan kehilangan rasa. Ini menunjukkan bahawa kehadiran kehilangan bau dan kabut otak sering berkait rapat dengan kehilangan rasa.

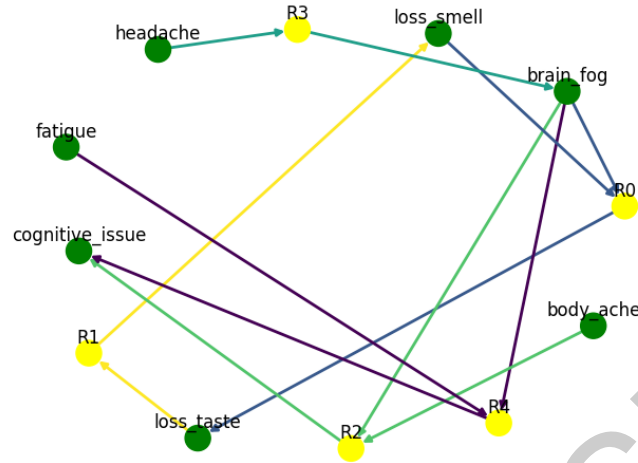
Selain itu, petua (loss_taste) -> (loss_smell) dipilih kerana sokongan yang lebih tinggi (0.0134) dan keyakinan yang tinggi (0.6976). Lif 12.58 menunjukkan hubungan yang signifikan di antara kehilangan rasa dan kehilangan bau, menyarankan kehadiran kehilangan rasa sering berkait dengan kehilangan bau dalam konteks ini. Selain itu, petua (brain_fog, body_ache) -> (cognitive_issue) dipilih dengan sokongan 0.00185, keyakinan 0.5283, dan lif 14.22. Ini menunjukkan bahawa masalah kognitif dan nyeri badan sering berhubungan dengan kabut otak dalam dataset ini.

Petua (headache) -> (brain_fog) dipilih dengan sokongan 0.00245, keyakinan 0.493, dan lif 10.05. Hubungan yang kuat antara sakit kepala dan kabut otak ditonjolkan di sini, menunjukkan sakit kepala sering diikuti dengan gejala kabut otak. Akhirnya, (brain_fog, fatigue) -> (cognitive_issue) dengan sokongan 0.00145, keyakinan 0.468 dan lif 12.60 menunjukkan bahawa sekiranya seseorang dihadapi COVID Jangka Panjang mungkin dihadapi gejala seperti kabut, otak, kepenatan dan masalah kognitif.

Petua-petua ini dipilih kerana mereka menunjukkan kombinasi sokongan yang signifikan, keyakinan tinggi dalam hubungan antara penyebab dan akibat, serta lif yang tinggi menunjukkan hubungan yang kuat dan penting dalam set data COVID Jangka Panjang. Walaupun sokongan dalam semua petua amat rendah, namun jika dibandingkan dengan perwakilan beg kata yang amat besar mengandungi 14997 baris data. Hal ini menyebabkan nilai sokongan diterima dalam menyokong petua kerana kewujudan dalam set data yang besar amat ketara.

Kejituan menunjukkan bahawa hanya 28% dari petua yang ditemui oleh ARM adalah relevan atau tepat dalam konteks analisis yang dilakukan. Ini bermakna sebahagian besar petua yang dihasilkan mungkin tidak relevan atau tidak memberi manfaat yang besar secara praktikal. Dapatan semula mencapai 1.00, menunjukkan kebolehan ARM untuk menemui semua petua yang benar-benar relevan atau tepat dalam set transaksi. Tidak ada petua yang relevan yang terlepas daripada perhatian.

Skor F merupakan purata harmonik antara kejituan dan dapatan semula, memberikan pandangan menyeluruh tentang keberkesanan ARM. Dengan nilai skor F 0.28, walaupun ARM berjaya dalam menemui semua petua yang relevan (dapatan semula yang tinggi), kejituan yang rendah menunjukkan bahawa banyak petua yang dihasilkan mungkin agak relevan. Rajah 7 menggambarkan jaringan ARM melalui gambar anak panah yang mewakili setiap petua yang dikenal pasti setelah penapisan data.



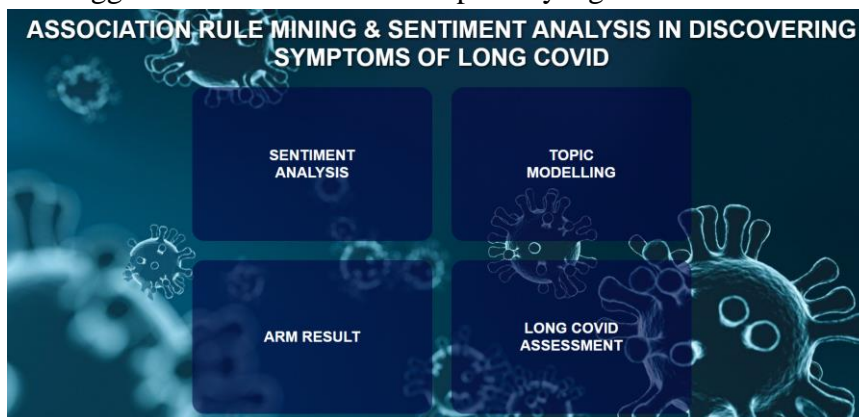
Rajah 7 Visualisasi Anak Panah 5 Petua Utama ARM

Pemrosesan Bahagian Depan (FRONT-END)

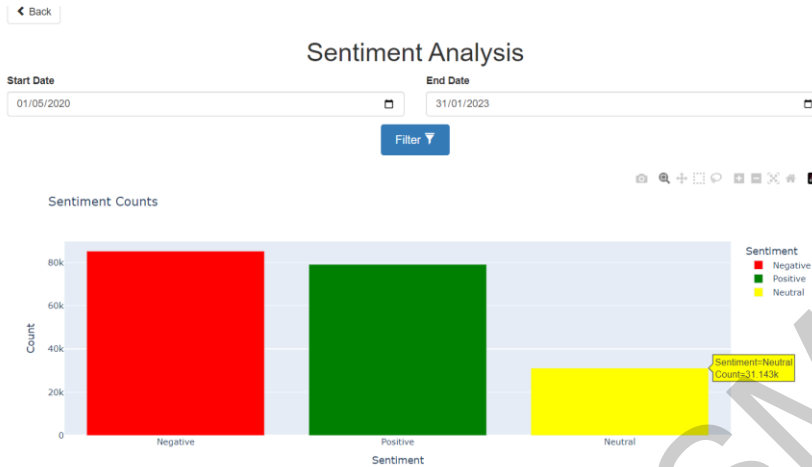
Setelah pemrosesan bahagian belakang berjaya dibina, semua keputusan setiap model di pemproses bahagian belakang perlu dibentangkan di mana sebuah sistem web dibina sebagai ruang visualisasi supaya dapat menjamin mesra pengguna. Sistem ini merangkumi laluan untuk pelbagai fungsi seperti analisis sentimen, pemodelan topik, visualisasi perlombongan petua sekutuan (ARM), dan ruang penilaian kebarangkalian seseorang menghadapi gejala COVID Jangka Panjang berdasarkan skor keyakinan ARM bersama LDA.

Teknologi utama yang digunakan dalam pembangunan sistem ini termasuk Flask, yang digunakan untuk membuat laluan web dan mengendalikan permintaan HTTP (Khalid 2024). Pandas digunakan untuk manipulasi dan analisis data, manakala Plotly digunakan untuk mencipta visualisasi saling tindak. HTML dan CSS digunakan untuk membangunkan antara muka hadapan sistem (Mal et al. 2024).

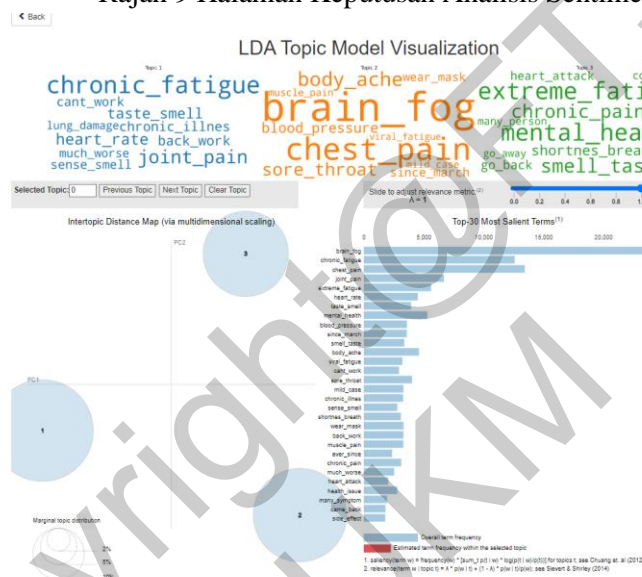
Antara muka pengguna utama mengandungi empat butang yang mengarahkan pengguna ke halaman lain untuk fungsi khusus yang berbeza seperti analisis sentimen, pemodelan topik, visualisasi ARM dan penilaian COVID Jangka Panjang. Rajah 8 hingga Rajah 12 akan menggambarkan visualisasi setiap web yang dibina.



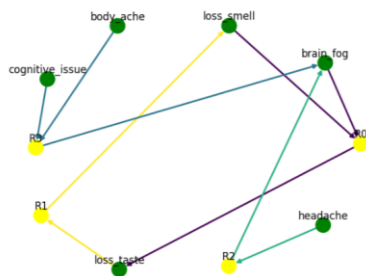
Rajah 8 Halaman Utama Antara Muka



Rajah 9 Halaman Keputusan Analisis Sentimen



Rajah 10 Halaman Keputusan LDA Apriori ARM Visualization



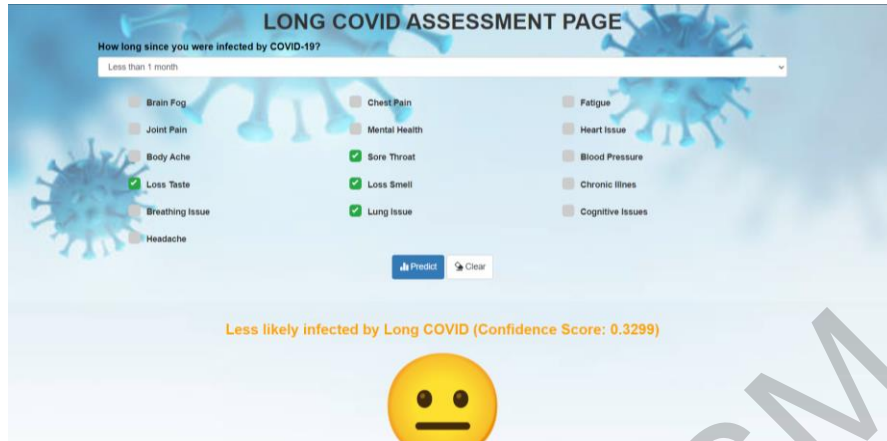
ARM Results Data Table

Below is the table displaying all ARM outputs:

Show 5 entries

Rules	Antecedents	Consequents	Confidence	Support	Lift
1	{'brain_fog', 'loss_smell'}	{'loss_taste'}	0.45	0.0012002400480096	34.96709844559586
2	{'loss_taste'}	{'loss_smell'}	0.5751295336787565	0.0074014802960592	10.317245952847262
3	{'headache'}	{'brain_fog'}	0.42	0.0014002800560112	12.522345924453278
4	{'cognitive_issue', 'body_ache'}	{'brain_fog'}	0.4074074074074074	0.0014669600586784	12.14689639938149
5	{'headache'}	{'body_ache'}	0.36	0.0012002400480096	7.335489130434783

Rajah 11 Halaman Keputusan ARM



Rajah 12 Halaman Penilaian COVID Jangka Panjang

Cadangan Penambahbaikan

Cadangan penambahbaikan untuk projek ini merangkumi beberapa aspek penting yang boleh meningkatkan keberkesanan dan kebolehpercayaan hasil kajian. Pertama, adalah disarankan untuk mengambil set data ciapan yang merangkumi pelbagai bahasa, tidak hanya tertumpu kepada bahasa Inggeris sahaja. Ini akan membolehkan kajian ini lebih mencerminkan kepelbagaian pengguna media sosial di seluruh dunia. Kedua, adalah pembangunan model lain wajar dilakukan sebagai perbandingan prestasi model. Contohnya, penggunaan TextBlob boleh dijadikan sebagai perbandingan terhadap prestasi model VADER yang digunakan. Ini akan memberikan perspektif yang lebih luas tentang kelebihan dan kekurangan setiap model yang digunakan. Ketiga, sistem model yang dibangunkan harus sentiasa diinput dengan data ciapan secara masa nyata untuk memantau gejala COVID Jangka Panjang dari semasa ke semasa. Dengan ini, model akan sentiasa dikemaskini dengan data terkini, menjadikannya lebih relevan dan tepat dalam mengesan perubahan tren gejala COVID Jangka Panjang.

KESIMPULAN

Kajian yang bertajuk Kaedah Analisis Sentimen dan Perlombongan Petua Sekutuan dalam meneroka gejala COVID Jangka Panjang melalui Data Ciapan telah Berjaya dibangunkan dan sistem visualisasi yang telah dibangunkan dapat memenuhi objektif kajian yang telah dirancang dan ditunjukkan pada peringkat pengenalan projek. Objektif kepada kajian ini adalah untuk mencapai set data kesan COVID Jangka Panjang melalui data ciapan daripada pautan Kaggle dan mengamalkan teknik pra-pemprosesan data. Kemudian, data yang dibersihkan dibuat analisis sentimen tentang kesan-kesan COVID Jangka Panjang. Seterusnya, data ciapan digunakan untuk melaksanakan pemodelan topik bagi meneroka gejala dan kesan-kesan COVID Jangka Panjang. Kesan-kesan yang dikaji diperincikan dengan meneroka hubungan antara gejala COVID Jangka Panjang melalui Perlombongan Petua Sekutuan. Dengan ini, semua objektif kajian telah mencapai sepanjang kajian ini dilakukan.

Kekuatan Model

Set data yang didapati mengandungi saiz set data yang cukup besar iaitu sebanyak 195476 baris data ciapan. Selain itu, set data juga amat seimbang dari segi pengelasan sentimen oleh VADER. Setiap model yang dibina telah menjalani ujian prestasi bagi menilai prestasi model di mana segala keputusan ujian melambangkan setiap model yang dibina sesuai digunakan dalam konteks kajian COVID Jangka Panjang. Hasil analisis LDA memberikan tiga topik utama yang berkaitan dengan gejala COVID Jangka Panjang iaitu topik pertama merangkumi rasa sakit dan kesukaran fizikal yang berterusan, serta masalah-masalah pernafasan. Topik kedua pula merangkumi masalah jantung dan keletihan kronik serta kesan berpanjangan yang dirasakan. Topik ketiga mengutarakan isu kesihatan mental, kehilangan deria rasa dan bau, serta gejala lain yang mempengaruhi keseluruhan tubuh badan. ARM juga berjaya meneroka 5 petua gejala yang terkemuka dalam kontek COVID Jangka Panjang. Petua (brain_fog, loss_smell) -> (loss_taste) menunjukkan hubungan yang terkuat antara gejala-gejala COVID Jangka Panjang

Kelemahan Model

Projek ini mempunyai beberapa kelemahan yang perlu diambil kira. Pertama, model yang dibina hanya sesuai digunakan untuk ciapan dalam Bahasa Inggeris sahaja menyebabkan data ciapan yang berlainan bahasa sukar ditafsir oleh model. Selain itu, model algoritma yang digunakan tidak dibandingkan dengan model lain kerana skop kajian sudah ditetapkan dengan matlamat yang tinggi, menyebabkan keterbatasan dalam menilai kelebihan dan kekurangan model yang dibangunkan. Akhir sekali, gejala COVID Jangka Panjang yang dikenal pasti tidak lagi dinilai oleh pakar perubatan untuk kebolehsahan data, menyebabkan kemungkinan ketepatan dan kebolehpercayaan data yang diperolehi diragui.

PENGHARGAAN

Penulis kajian ini ingin ucapkan setinggi-tinggi penghargaan dan jutaan terima kasih kepada Prof. Dr. Azuraliza Abu Bakar, penyelia penulis kajian ini yang telah memberi tunjuk ajar serta bimbingan untuk menyiapkan projek ini dengan jayanya.

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

- Davis, H. E., Assaf, G. S., Mccorkell, L., Wei, H., Low, R. J., Re'em, Y., Redfield, S., Austin, J. P. & Akrami, A. 2021. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *eClinicalMedicine* 38(10)1019.
- Davis, H. E., Mccorkell, L., Vogel, J. M. & Topol, E. J. 2023. Long COVID: major findings, mechanisms and recommendations. *Nature Reviews Microbiology* 21(3): 133-146.
- HOTZ, N. 2024. What is CRISP DM? Data Science Process Alliance <https://www.datascience-pm.com/crisp-dm-2/> [17 July 2024].

- Mal, S., Shrivastava, V., Pandey, A. & Mishra, M. 2024. Web Development Using HTML AND CSS. International Journal of Research Publication and Reviews (5) <https://dl.acm.org/doi/abs/10.5555/1593773>.
- MOH. 2023. SINDROM COVID BERPANJANGAN (LONG COVID SYNDROME) | COVID-19 MALAYSIA. Kementerian Kesihatan Malaysia <https://covid-19.moh.gov.my/semasa-kkm/2023/03/sindrom-covid-berpanjangan-long-covid-syndrome> [7 July 2024].
- Murugesan, M. 2021. Understanding long Covid. News Strait Times, 31-08-2021:
- Matharaarachchi, S., Domaratzki, M., Katz, A. & Muthukumarana, S. 2022. Discovering Long COVID Symptom Patterns: Association Rule Mining and Sentiment Analysis in Social Media Tweets. JMIR Form Res 6(9): e37984.

Pravinashwarn A/L Raja Sageran (A188190)

Prof. Dr. Azuraliza Abu Bakar

Fakulti Teknologi & Sains Maklumat

Universiti Kebangsaan Malaysia