ANALISIS SENTIMEN SISTEM EMERIT UNIVERSITI KEBANGSAAN MALAYSIA

NUR AMIRA FARISHA BINTI MOHD YUSRI

PROF. MADYA DR. MASNIZAH MOHD

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

ABSTRAK

Kajian Analisis Sentimen Sistem Emerit Universiti Kebangsaan Malaysia (UKM) bertujuan untuk meneliti persepsi dan impak Sistem Emerit yang baharu diperkenalkan di UKM pada Sesi 2022/2023. Sistem ini direka untuk mengukur kelayakan pelajar dalam mendapatkan penempatan kolej bagi sesi akademik seterusnya, dengan mengambil kira pelbagai faktor seperti pendapatan isi rumah, badan beruniform, i-STAR, dan kecemerlangan akademik. Kajian ini menggunakan teknik Pemprosesan Bahasa Tabii (PBT) untuk menganalisis sentimen yang dinyatakan dalam pelbagai data teks, termasuk kiriman media sosial dan komen-komen terbuka dari saluran Telegram "UKM Confessions". Data ini dikumpulkan daripada pelajar, pihak kolej, atau pentadbir secara awanama untuk mendapatkan pandangan menyeluruh mengenai Sistem Emerit. Analisis kajian memfokuskan kepada penerimaan dan persepsi komuniti UKM terhadap sistem ini, termasuk faktor-faktor yang mempengaruhi sentimen seperti kriteria ganjaran, ketelusan sistem, dan kesannya terhadap motivasi pelajar dan pencapaian akademik. Penemuan dari projek ini memberikan pandangan yang lebih mendalam tentang kelebihan atau kekurangan Sistem Emerit, serta mengemukakan cadangan untuk penambahbaikan. Kesimpulan daripada kajian ini dapat memberi panduan kepada pihak universiti dalam meningkatkan keadilan, penerimaan, dan keberkesanan sistem ganjaran berasaskan merit, sekaligus turut memanfaatkan komuniti UKM secara keseluruhan.

Kata kunci: Analisis Sentimen, Sistem Emerit, Pemprosesan Bahasa Tabii, Kolej UKM, UKM Confessions, i-STAR

PENGENALAN

Analisis sentimen yang juga dikenali sebagai perlombongan pendapat, merupakan bidang pemprosesan bahasa tabii yang memberi tumpuan kepada sentimen atau pendapat yang dinyatakan dalam data teks yang telah ditentukan dan dikeluarkan. Terma sentimen membawa maksud pendapat, pandangan, atau sikap, terutamanya yang berasaskan emosi dan bukannya alasan. Matlamat utama analisis sentimen adalah untuk memahami maklumat subjektif yang disampaikan didalam teks sama ada ia bersifat positif, negatif, atau neutral. Analisis sentimen

mempunyai pelbagai aplikasi, termasuk pemantauan media sosial, analisis maklum balas kaji selidik, ulasan produk, dan penyelidikan pasaran.

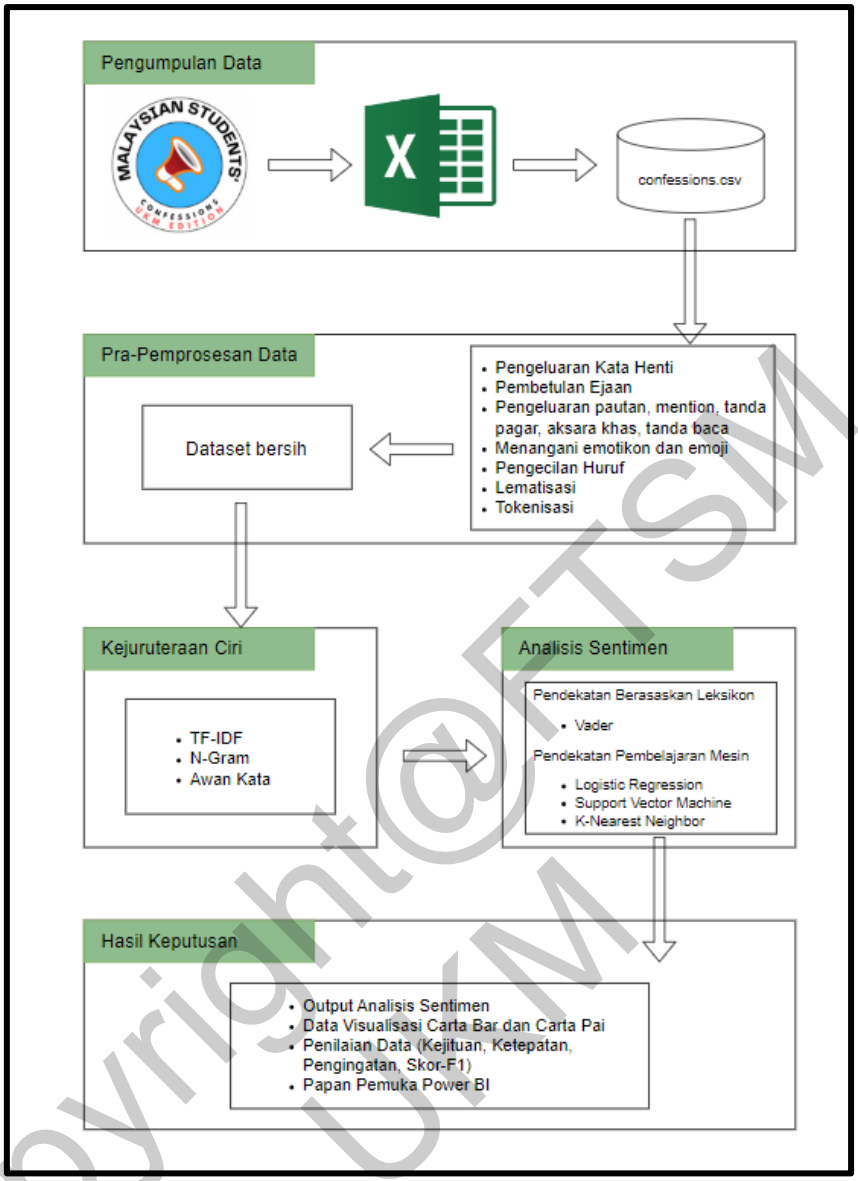
Bagi projek tahun akhir ini, analisis sentimen dikaji dalam domain Sistem Emerit UKM. Dalam era komunikasi digital, pelbagai platform atas talian seperti media sosial dan ulasan kaji selidik menghasilkan aliran data teks yang sangat besar. Memahami sentimen dalam teks ini adalah kritikal bagi perniagaan, penyelidik, dan pembuat dasar, dengan cabaran utama terletak pada pembangunan sistem analisis sentimen yang dapat mengelaskan sentimen teks secara tepat sebagai positif, negatif, atau neutral. Sistem tersebut perlu mengendalikan pelbagai jenis teks termasuk bahasa tidak formal dan singkatan yang digunakan dalam media sosial seperti di saluran Telegram "UKM Confessions." Selain itu, ia mesti mampu menganalisis teks dalam pelbagai bahasa termasuk Bahasa Malaysia, bahasa pengantar utama di platform UKM. Sentimen juga berbeza berdasarkan konteks, jadi sistem perlu melampaui padanan kata kunci dan mempertimbangkan nada keseluruhan teks untuk analisis yang lebih tepat.

Objektif utama projek ini adalah untuk memperoleh pandangan tentang emosi dan kebimbangan pelajar melalui maklum balas, ulasan, atau komen berkaitan Sistem Emerit UKM, sambil mengenal pasti faktor-faktor yang mempengaruhi sentimen tersebut dengan berkesan. Selain itu, projek ini bertujuan untuk mengumpulkan hasil analisis sentimen bagi terus meningkatkan Sistem Emerit UKM, memastikan organisasi memenuhi keperluan yang berkembang dan kekal dikemaskini. Tambahan lagi, projek ini juga membolehkan organisasi pengurusan kolej memahami perspektif pengguna berdasarkan data yang diperoleh daripada spekulasi, menekankan kepentingan analisis sentimen dalam proses membuat keputusan.

METODOLOGI KAJIAN

Terdapat empat fasa yang telah digunakan sepanjang projek ini berjalan. Fasa pertama adalah penyataan masalah. Seterusnya, fasa kedua adalah pengumpulan data dan pra-pemprosesan data. Pengumpulan data telah dijalankan dengan mengambil teks sentimen dari saluran Telegram “UKM Confessions” dengan membuat carian kata kunci ‘merit’ untuk menapis segala kandungan yang ada di dalam laman tersebut. Sejak sesi lalu, sebanyak 1000 ‘confession’ dengan kata kunci ‘emerit’ dan ‘merit’ telah didapati di saluran Telegram “UKM Confessions” menunjukkan bahawa para pelajar UKM sentiasa memberikan pendapat dan ulasan mereka terhadap sistem Emerit yang baharu ini. Manakala, pra-pemprosesan data pula merupakan langkah dimana teks yang telah dikumpul perlu dibersihkan dan dipra-proses untuk menghilangkan simbol, tanda baca, nombor dan aksara khas yang tidak relevan. Kemudian, fasa yang ketiga adalah peringkat pembangunan iaitu fasa kejuruteraan data dan fasa analisis sentimen. Akhir sekali ialah fasa pengujian.

Rajah aliran data adalah gambaran yang menunjukkan aliran informasi dalam system yang menunjukkan langkah-langkah dalam proses olahan data *confessions*, mulai dari penerimaan data hingga hasil analisis sentimen. Rajah 1 menunjukkan aliran data yang membantu dalam mengenalpasti langkah-langkah yang diambil oleh sistem dan gambaran hubungan antara entiti dalam sistem.



RAJAH 1 Rajah Aliran Data

Fasa penyataan masalah

Di peringkat penyataan masalah, domain dan topik kajian telah dikenalpasti dengan lebih terperinci untuk menentukan faktor utama penyataan masalah tersebut. Langkah ini adalah penting untuk memastikan penyataan masalah, cadangan penyelesaian, objektif kajian, skop, dan kekangan kajian adalah relevan dan dapat memberikan dasar yang kuat untuk langkah-langkah seterusnya.

Terdapat rubrik dalam sistem pemarkahan emerit di mana rubrik ini menilai prestasi pelajar dalam kolej kediaman berdasarkan beberapa empat bahagian: A-Penglibatan dalam badan beruniform (30%), B-Aktiviti pelajar (30%), C-Kecemerlangan (20%), dan D-Prestasi akademik serta status pendapatan isi rumah (20%). Rubrik ini menyediakan kerangka komprehensif untuk menilai prestasi pelajar berdasarkan jawatan, pencapaian, prestasi akademik, dan disiplin. Perincian setiap bahagian dan markah adalah seperti di Jadual 1.

JADUAL 1 Rubrik Pemarkahan Emerit Kolej Kediaman

BIL.	%	BAHAGIAN DAN SUB-BAHAGIAN		MARKAH
A	30	**Persatuan/Kelab/Badan Beruniform**		
		Jawatan	Presiden/Pengerusi/Yang Dipertua/ Ketua Eksekutif	10
			Naib Presiden/Naib Pengerusi/Naib Yang Dipertua/Timbalan Ketua Eksekutif	7
			Setiausaha/Bendahari	6
			Penolong Setiausaha/ Penolong Bendahari	5
			Exco	3
			Ahli Jawatankuasa	2
			Ahli biasa	1
		Badan Beruniform	Pasukan Latihan Pegawai Simpanan Tentera Datar (PALAPES) Sukarelawan Polis Siswa Siswi (SUKSIS) Bulan Sabit Merah Malaysia (BSMM) Kumpulan Latihan Kelanasiswa Malaysia (KLKM) Siswa Siswi Pertahanan Awam (SISPA)	30
B	30	**Aktiviti Pelajar**		
		Jawatan	Ketua Pengarah Projek	15
			Penceramah/Ahli Panel/Pembentang	10
			Timbalan Ketua Pengarah Projek	7
			Setiausaha/Bendahari	6
			Penolong Setiausaha/ Penolong Bendahari/ Exco/ Felo/ Pemantau/ Fasilitator	5
			Ahli Jawatankuasa	3
			Ahli Projek	2
			Peserta	1
		Peringkat	Antarabangsa	20
			Kebangsaan	10
			Negeri	8
			Daerah	6
			Universiti	4
			Fakulti/Kolej	2
			Persatuan/Kelab	1
C	20	**Kecermerlangan**		
		Kecermerlangan	Johan/Tempat Pertama/Emas	10
			Naib Johan/Tempat Kedua/Perak	7
			Tempat Ketiga/Gangsa	4
			Saguhati	2
			Lain-lain (Nyatakan)	1
		Peringkat Perwakilan	Antarabangsa	10
			Kebangsaan	7
			Negeri	4
			Daerah/Universiti Awam/Swasta (Luar)/Universiti (Dalam UKM)	2
			Fakulti/Kolej/Persatuan/Kelab	1
		Peringkat Kecermerlangan	Negara	15
			Negeri	7
			Universiti	4
			Fakulti/Institut/Pusat/Kolej/Sendiri	1
D	20	**Akademik & B40**		
		Akademik	≥ 3.67	20
			≥ 3.00	15
			≥ 2.67	10
			< 2.67	5
			0.00	0
		B40	Ya	30
			Tidak	0
		Demerit	Min 5%	

Fasa pra-pemprosesan

Proses pra-pemprosesan adalah proses pembersihan yang dijalankan bagi mengekstrak kata kunci dan simbol. Bahan sumber yang diperolehi daripada saluran Telegram perlu melalui pemprosesan teks sebelum ke fasa analisis bagi tujuan pembersihan data. Berikut merupakan beberapa langkah pra-pemprosesan data:

I. Pengecilan Huruf dalam Teks

Proses ini menukarkan semua teks kepada huruf kecil bagi memastikan keseragaman dan konsistensi dalam analisis seperti di Jadual 2. Ini mencegah model daripada mengendalikan teks versi huruf besar dan versi huruf kecil dari perkataan yang sama sebagai berbeza.

JADUAL 2 Pengecilan Huruf dalam Teks

Sebelum	Selepas
Harap2 sistem emerit kolej ni ditambah baik, selain drpd ada pointer utk B40 ada juga pointer utk tahap jauh alamat rumah student dgn UKM	**harap2** sistem emerit kolej ni ditambah baik, selain drpd ada pointer utk **b40** ada juga pointer utk tahap jauh alamat rumah student dgn **ukm**

II. Mengeluarkan Kata Henti

Mengeluarkan kata henti yang tidak membawa makna seperti “adalah”, “dan”, “itu” dan sebagainya untuk mengurangkan kebisingan dalam data dan membantu meningkatkan kecekapan proses analisis sentimen sperti di Jadual 3.

JADUAL 3 Mengeluarkan Kata Henti

Sebelum	Selepas
harap2 sistem emerit kolej ni ditambah baik **selain drpd** ada pointer utk b40 ada juga pointer **utk** tahap jauh alamat rumah student **dgn** ukm	sistem emerit kolej ni ditambah baik ada pointer utk b40 ada juga pointer tahap jauh alamat rumah student ukm

III. Pencantasan atau Lematisasi (Stemming or Lemmatization)

Mengurangkan perkataan kepada bentuk asas atau kata akar mereka yang membantu mengurangkan dimensi data dan menangkap makna asas perkataan. Pemotongan melibatkan penghapusan imbuhan awalan atau akhiran dari perkataan, manakala lematitasi melibatkan pengurangan perkataan kepada bentuk asas atau kata akar menggunakan analisis kosa kata dan morfologi seperti di Jadual 4.

JADUAL 4 Pencantasan atau Lematisasi

Sebelum	Selepas
sistem emerit kolej **ni di**tambah baik **ada** pointer b40 **ada juga** pointer tahap jauh alamat rumah student ukm **so yg mana asal** sabah sarawak kelantan perlis johor **etc ni ada la**	sistem emerit kolej tambah baik pointer b40 pointer tahap jauh alamat rumah student ukm sabah sarawak kelantan perlis johor

IV. Tokenisasi

Langkah yang penting dalam pra-pemprosesan data dengan memecahkan teks kepada satu-satu perkataan atau token supaya penganalisan sentimen yang berkaitan dapat dilaksanakan dengan setiap perkataan dalam teks, membantu membina representasi bermakna bagi teks input seperti di Jadual 5.

JADUAL 5 Tokenisasi

Sebelum	Selepas
sistem emerit kolej tambah baik pointer b40 pointer tahap jauh alamat rumah student ukm sabah sarawak kelantan perlis johor	[“sistem”, “emerit”, “kolej”, “tambah”, “baik”, “pointer”, “b40”, “pointer”, “tahap”, “jauh”, “alamat”, “rumah”, “student”, “ukm”, “sabah”, “sarawak”, “kelantan”, “perlis”, “johor”

V. Menangani Emotikon dan Emoji

Emotikon dan emoji boleh membawa maklumat sentimen. Bergantung pada konteks, mereka mungkin perlu ditukarkan kepada teks atau dikeluarkan, bergantung kepada keperluan khusus analisis.

VI. Pemeriksaan dan Pembetulan Ejaan

Membetulkan kesilapan ejaan dalam teks adalah langkah yang penting kerana kesilapan ejaan boleh mempengaruhi ketepatan analisis sentimen, jadi adalah baik untuk menanganinya semasa pra-pemprosesan seperti di Jadual 6.

JADUAL 6 Pemeriksaan dan Pembetulan Ejaan

Sebelum	Selepas
[“sistem”, “emerit”, “kolej”, “tambah”, “baik”, “pointer”, “b40”, “pointer”, “tahap”, “jauh”, “alamat”, “rumah”, “student”, “ukm”, “sabah”, “sarawak”, “kelantan”, “perlis”, “johor”	[“sistem”, “emerit”, “kolej”, “tambah”, “baik”, “pointer”, “b40”, “pointer”, “tahap”, “jauh”, “alamat”, “rumah”, “student”, “ukm”, “sabah”, “sarawak”, “kelantan”, “perlis”, “johor”]

Fasa kejuruteraan ciri

Kejuruteraan ciri telah dilakukan dengan mengekstrak ciri-ciri tambahan daripada data seperti mengenal pasti sentimen komen dan faktor sentimen tersebut. Visualisasi seperti word cloud dan carta bar juga dilakukan untuk menunjukkan taburan sentimen, perkataan atau frasa yang sering disebut.

Fasa analisis sentimen

Analisis sentimen dijalankan dimana teknik Pemprosesan Bahasa Tabii telah digunakan. Alat atau perpustakaan analisis sentimen seperti VADER telah dipilih untuk melatih set data yang diperoleh. Set data telah dikategorikan kepada sentimen positif, sentimen negatif atau sentimen neutral. Seterusnya, model pembelajaran mesin seperti Logistic Regression, Support Vector Machine dan K-Nearest Neighbors dibangunkan untuk meramalkan sentimen, kepuasan dan cadangan pembaikan daripada pengguna.

I. Pemprosesan Bahasa Tabii

Pemprosesan bahasa tabii (NLP) dalam analisis sentimen melibatkan penggunaan teknik dan algoritma untuk memahami, mengekstrak, dan menilai sentimen atau emosi yang terkandung dalam teks manusia. Beberapa pendekatan dan konsep dalam NLP yang digunakan dalam analisis sentimen termasuk:

A. N-grams

N-gram merujuk kepada urutan 'n' item berturutan, biasanya perkataan, dalam suatu teks. Dalam konteks analisis sentimen, n-gram digunakan sebagai ciri untuk menangkap konteks linguistik tempatan perkataan dalam suatu teks. Dengan mempertimbangkan urutan perkataan daripada perkataan tunggal, n-gram membantu model analisis sentimen memahami hubungan dan nuansa antara perkataan. Penggunaan n-gram dalam analisis sentimen membolehkan model menangkap bukan sahaja perkataan yang membawa sentimen tetapi juga maklumat kontekstual dan hubungan antara perkataan. Ini boleh berguna untuk memahami sentimen yang dinyatakan dalam frasa, idiom, atau gabungan perkataan khusus yang membawa sentimen.

B. Pencantasan (Stemming)

Pencantasan dalam konteks analisis sentimen biasanya merujuk kepada proses mengurangkan saiz atau kompleksiti model dengan mengeluarkan ciri-ciri, parameter, atau struktur yang tidak relevan. Ini dilakukan untuk meningkatkan kecekapan, kelajuan, dan kadangkala kebolehterjemahan model analisis sentimen. Teknik pemangkasan boleh diterapkan kepada pelbagai komponen model analisis sentimen, termasuk pemilihan ciri, senibina model, atau set data.

C. Penghuraian (Parsing)

Parsing dalam konteks analisis sentimen adalah proses mengolah teks menjadi struktur yang lebih mudah dianalisis. Parsing melibatkan pemecahan teks menjadi frasa atau perkataan sederhana, yang kemudian dianalisis untuk mengidentifikasi sentimen atau emosi yang terkait dengan frasa tersebut. Parsing merupakan langkah penting dalam analisis sentimen kerana ia membolehkan sistem untuk mengolah teks yang kompleks dan berstruktur menjadi format yang lebih mudah dianalisis, sehingga membolehkan analisis sentimen lebih efektif dan tepat.

II. Pendekatan Pembelajaran Mesin

Pendekatan pembelajaran mesin melibatkan penggunaan algoritma untuk memberikan sistem keupayaan untuk belajar dari data dan membuat keputusan atau ramalan tanpa program yang secara eksplisit ditulis untuk tugas tersebut. Terdapat beberapa pendekatan dalam pembelajaran mesin, dan di bawah ini adalah beberapa yang umum digunakan:

A. Logistic Regression

Logistic Regression adalah pendekatan pembelajaran mesin yang digunakan untuk pengelasan. Ia memodelkan hubungan antara satu atau lebih ciri bebas (input) dan pemboleh ubah bersandar binari (output), yang biasanya dikategorikan sebagai 0 atau 1. Model ini menentukan sempadan keputusan yang memisahkan kelas-kelas berdasarkan kebarangkalian yang dianggarkan. Logistic Regression sering digunakan dalam masalah pengelasan seperti pengesanan penipuan, diagnosis penyakit, dan analisis sentimen.

B. *Support Vector Machine* (SVM)

Support Vector Machine (SVM) merupakan kaedah pembelajaran berpenyelia yang digunakan untuk tugas pengelasan dan regresi. Dalam konteks analisis sentimen, SVM adalah algoritma pembelajaran pengelasan yang bertujuan untuk mencari *hyperplane* terbaik yang memisahkan data ke dalam kelas-kelas yang berbeza. Beberapa penelitian telah menggunakan SVM untuk analisis sentimen pada teks Twitter. Hasil penelitian ini menunjukkan bahawa SVM dapat meningkatkan kberecekapan dalam analisis sentimen ketika fungsi kernel yang tepat dipilih.

C. *K-Nearest Neighbors* (KNN)

K-Nearest Neighbors (KNN) adalah algoritma pembelajaran mesin yang juga boleh digunakan dalam tugas analisis sentimen. KNN adalah algoritma pembelajaran berpenyelia, tanpa parameter yang mengelaskan suatu titik data berdasarkan kelas majoriti dari k tetangga terdekatnya dalam ruang ciri. Dalam konteks analisis sentimen, ini bermakna sentimen bagi sepotong teks ditentukan oleh sentimen tetangganya yang terdekat dalam ruang ciri. Dalam analisis sentimen, KNN boleh menjadi pilihan yang sesuai untuk senario tertentu, terutamanya apabila berurusan dengan set data yang kecil relatifnya dan apabila sempadan keputusan tidak terlalu kompleks.

III. Pendekatan Berasaskan Leksikon

Pendekatan berasas leksikon dalam analisis sentimen menggunakan leksikon sentimen yang telah disiapkan sebelumnya untuk memberi skor pada dokumen dengan menggabungkan skor sentimen dari semua kata dalam dokumen tersebut. Pendekatan ini dapat digunakan untuk menentukan polariti atau sentimen untuk pengelasan teks ke dalam kategori negatif, neutral, dan positif.

A. Pendekatan Berasaskan Kamus

Pendekatan berasaskan kamus atau leksikon dalam analisis sentimen melibatkan penggunaan senarai perkataan atau frasa yang telah ditetapkan bersama dengan skor atau label sentimen yang berkaitan. Pendekatan ini bergantung pada kamus atau leksikon yang mengkategorikan perkataan ke dalam sentimen positif, negatif, atau neutral. Sentimen suatu teks ditentukan dengan menggabungkan skor sentimen perkataan atau frasa yang terdapat dalam teks tersebut. Pendekatan ini sesuai untuk situasi di mana sentimen teks adalah mudah difahami dan leksikon merangkumi kosa kata yang relevan.

B. Pendekatan Berasaskan Korpus

Pendekatan berdasarkan korpus dalam analisis sentimen melibatkan penciptaan kamus atau leksikon sentimen berdasarkan koleksi teks yang besar (korpus). Berbanding dengan penyusunan kamus secara manual, pendekatan ini menggunakan kaedah statistik dan algoritma untuk menentukan polariti sentimen bagi perkataan berdasarkan kejadian dan konteks mereka dalam korpus. Kaedah ini memanfaatkan pengedaran semantik perkataan dalam set data besar untuk menyimpulkan sentimen.

Perbandingan Kaedah dan Pendekatan Analisis Sentimen

JADUAL 7 Perbandingan Kaedah dan Pendekatan Analisis Sentimen

Bil.	Kajian	Huraian Kajian	Set data/ Atribut	Kaedah
1	Khalifa Chekima, Rayner Alfred; *Sentiment Analysis of Malay Social Media Text*; 2020	Kajian ini menganalisis sentimen dalam teks media sosial Bahasa Melayu, mengatasi cabaran Bahasa Rojak, Bahasa SMS, emotikon, dan peralihan valens. Penyelidikan menghasilkan RojakLex, sebuah leksikon baru yang meningkatkan ketepatan analisis berbanding kaedah garis dasar.	Set Data - 12,240 komen dikumpulkan dari Facebook dan Twitter berfokuskan teks bahasa Melayu dari pelbagai domain seperti ulasan filem, politik, dan produk.	Pembelajaran Mesin: - *Support Vector Machine* - Rangkaian Neural - *Naïve Bayes* - *Maximum Entropy* Pendekatan Berasaskan Leksikon: - RojakLex
2	Ezuana Sukawai, Nazlia Omar; Pembangunan Korpus Bagi Analisis Sentimen Dalam Bahasa Melayu Secara Separa Selia; 2020	Kajian ini membina korpus analisis sentimen Bahasa Melayu menggunakan data Twitter, menggabungkan teknik leksikon dan pembelajaran mesin. Leksikon sentimen dan emotikon digunakan untuk melabel data latihan awal. Proses ini diikuti oleh pengayaan data latihan dengan contoh baharu menggunakan algoritma Multinomial Naïve Bayes, yang efektif untuk pengelasan.	Set Data - Dikumpulkan dari Twitter yang berfokus kepada domain politik di Malaysia dan menggunakan bahasa Melayu. - Menggunakan RapidMiner Atribut - Bilangan ulasan - Polariti ulasan	Pembelajaran Mesin: - *Naïve Bayes* - *Multinomial Naïve Bayes* - *Support Vector Machine* - *Logistic Regression* - *Expectation Maximization*
3	Tashvini Sri A/P Nagendran, Prof. Dr. Mohd Juzaiddin Ab Aziz; Analisis Sentimen Yang Berfokuskan Emoji; 2022	Kajian ini menganalisis sentimen teks media sosial yang mengandungi emoji, dengan fokus pada peranan penting emoji dalam menggambarkan perasaan atau pendapat. Algoritma yang dikembangkan mengatasi kesukaran mengenalpasti sentimen yang melibatkan emoji menggunakan teknik analisis sentimen tradisional.	Set Data - Dikumpulkan dari Twitter yang berfokuskan kepada ayat yang mengandungi emoji.	Pendekatan Mesin: - Naïve Bayes Pendekatan Leksikon: - *VADER* - *Sentimen Intensity Analyzer* - *NLTK*
4	Wong Wai Jian, Sabrina Binti Tiun; Analisis Sentimen Masa Nyata Tweet Bahasa Melayu Terhadap Vaksin; 2022	Kajian ini menganalisis sentimen masa nyata terhadap tweet Bahasa Melayu yang berkaitan dengan vaksin. Ia menggunakan pembelajaran mesin untuk mengemaskini analisis sentimen secara masa nyata dan memahami pandangan orang ramai di Malaysia terhadap vaksin melalui tweet.	Set Data - Dikumpulkan dari Twitter menggunakan Twitter API yang berfokuskan tweet bahasa Melayu berkaitan dengan domain vaksin. Atribut - Teks tweet - Label Sentimen	Pembelajaran Mesin: - *Support Vector Machine* - *Logistic Regression* - *Naïve Bayes* - *K-Nearest Neighbors*
5	Ian Ho, Hui-Ngo Goh, Yi-Fei Tan; *Preprocessing Impact on Sentiment Analysis Performance on Malay Social Media Text*; 2022	Kajian ini menyelidik kesan prapemprosesan terhadap pengelasan sentimen yang diselia untuk kandungan media sosial bahasa Melayu termasuk menilai kesan penormalan dan penyingkiran kata henti, membandingkan senarai kata henti umum berbanding domain khusus, dan menilai pengaruh kaedah vektorisasi ciri dan saiz set data.	Set Data - Set Data 1: Dikumpulkan dari Malaya Dataset yang mengandungi 2 juta tweets bahasa Melayu dari Twitter. - Set Data 2: Dikumpulkan dari set data Maklum Balas Filem Pang & Lee	Pembelajaran Mesin: - *Logistic Regression* - *Random Forest Classifiers*

Fasa pengujian

Peringkat pengujian telah menyertakan butiran mengenai penilaian ke atas prestasi dan ujian rentas, untuk menilai sejauh mana model analisis sentimen ini berkualiti dan efektif. Hasil penilaian digunakan untuk mengesahkan kesahihan dan kualiti hasil analisis sentimen serta memberikan rekomendasi untuk perbaikan jika diperlukan, menjadikan metodologi ini penting dalam menilai kesahihan analisis sentimen yang dihasilkan.

KEPUTUSAN DAN PERBINCANGAN

Pengumpulan Data dan Pra-pemprosesan Data

I. Pengumpulan Data

Di fasa pengumpulan data pengumpulan data telah dijalankan dengan mengambil teks sentimen dari saluran Telegram "*UKM Confessions*" dengan menggunakan pustaka TelegramClient. Diantara perkara yang diperlukan untuk mengambil data tersebut adalah:

i. Pautan saluran Telegram "UKM Confession"
ii. Kata kunci 'emerit' atau 'merit'
iii. Tempoh tarikh pengambilan data adalah dari Mac 2023 sehingga terkini

Di antara perkara yang diekstrak adalah:

i. Tarikh mesej 'confession'
ii. Mesej 'confession'
iii. Jumlah lihatan mesej 'confession'

Rajah 2 merupakan hasil output pengestrakan set data dari saluran Telegram "*UKM Confession*".

	Date	Message	Views
1	2024-07-04 14:32:15	merit 45 ada peluang ke nak dapat kolej😔 tak l...	732
2	2024-07-04 14:31:18	hai semua pewaris watan, untuk korang yang ada...	772
3	2024-07-04 09:14:09	hai semua pewaris watan, untuk korang yang ada...	1730
4	2024-07-03 23:03:33	Actually yang ckp tipu merit aktiviti 40 betul...	1761
5	2024-07-03 23:02:09	Mana yang risau merit rendahlah apalah tak dap...	1626
...	...	...	...
1095	2023-03-04 11:10:24	#UKMC5924\nNasihat kepada semua MEP kolej, kal...	849
1096	2023-03-04 01:28:56	#UKMC5924\nNasihat kepada semua MEP kolej, kal...	873
1097	2023-03-04 01:27:41	#UKMC5920\nmcm mana nak check merit?\nhttps://...	738
1098	2023-03-03 03:28:24	#UKMC5889\nape maksud ni aaaaaaaa aku kumpul m...	788
1099	2023-03-01 08:35:11	#UKMC5854\nDah byk menyertai aktiviti ii kolej...	780

1099 rows × 3 columns

RAJAH 2 Hasil Output Pengumpulan Data

II. Pra-pemprosesan Data

Di fasa pra-pemprosesan data, diantara langkah-langkah yang telah dilaksanakan adalah:

a. Mengembangkan singkatan dan pembetulan ejaan. Sebagai contoh: 'takde' : 'tidak ada', 'byk' : 'banyak', 'taktau' : 'tidak tahu'
b. Penyingkiran pautan, mention dan tanda pagar. Sebagai contoh: https://forms.gle/1MdkkeXUV3u4HdFTA, @amirul, #MariSertai
c. Penyingkiran kata henti dan aksara khas dan tanda baca tidak relevan.
d. Lematisasi dan tokenisasi.

Rajah 3 merupakan hasil output pra-pemprosesan set data yang tidak bersih.

	Original_Teks	Cleaned_Teks
1	merit 45 ada peluang ke nak dapat kolej😔 tak l...	merit 45 peluang kolej tak larat alik
2	hai semua pewaris watan, untuk korang yang ada...	pewaris watan terfikir join kelab suggestkan j...
3	hai semua pewaris watan, untuk korang yang ada...	pewaris watan terfikir join kelab suggestkan j...
4	Actually yang ckp tipu merit aktiviti 40 betul...	actually tipu merit aktiviti 40 betul senanye ...
5	Mana yang risau merit rendahlah apalah tak dap...	risau merit rendahlah apalah tak kolej risau m...
...	...	...
872	Knp ehh dlm emerit kolej tak ada merit b40? Ca...	ehh emerit kolej tak merit b40 contact suruh u...
873	koranggg, aq nak survey markah merit semua org...	koranggg aq survey merit normal tak reaction s...
874	ukm x de club running ek. Sambil sambil boleh ...	ukm x de club running ek sambil sambil boleh m...
875	Korang merit ni dorg kira mcm mana ehh? Aku ad...	merit dorg kira ehh join dua pertandingan tuli...
876	Kalau merit bawah 5 memang tak dapat kolej ke ...	merit 5 tak kolej

876 rows × 2 columns

RAJAH 3 Hasil Output Pra-Pemprosesan Data

Kejuruteraan Ciri

I. TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) adalah teknik yang digunakan dalam pemprosesan bahasa semula jadi (NLP) dan pemulihan maklumat untuk menilai kepentingan suatu perkataan dalam dokumen relatif kepada sekumpulan dokumen (korpus). Dalam konteks unigram, setiap ciri adalah satu perkataan tunggal. Rajah 4 menunjukkan hasil output TF-TIDF yang menunjukkan kekerapan perkataan yang ada dalam set data yang telah diperoleh:

	Word	TF-IDF Score
0	merit	68.570588
1	kolej	57.187859
2	tak	52.037103
3	program	36.166997
4	masuk	33.544049
...	...	...
95	time	4.937317
96	bayar	4.934616
97	renew	4.923156
98	nama	4.904057
99	guy	4.890926

100 rows × 2 columns

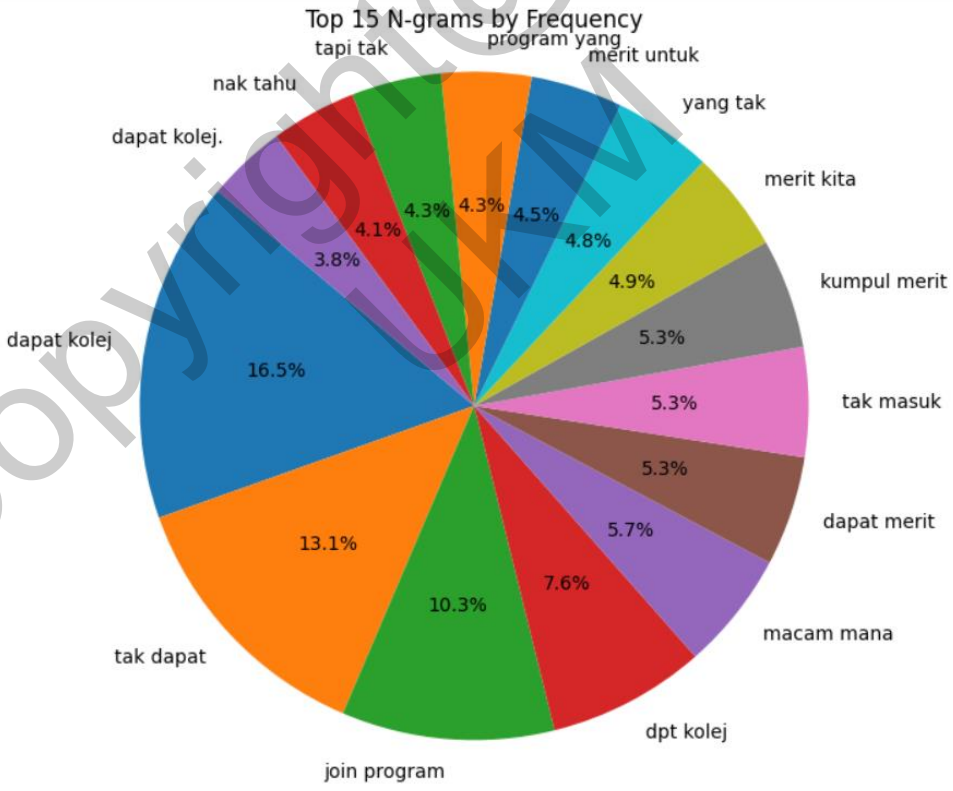
RAJAH 4 Hasil Output TF-IDF

Output tersebut menunjukkan 100 perkataan teratas berdasarkan skor TF-IDF (Term Frequency-Inverse Document Frequency). Beberapa contoh perkataan yang mempunyai skor tertinggi adalah “merit”, “kolej”, “tak”, “program” dan “masuk”. Selain itu, beberapa contoh yang mempunyai skor yang lebih rendah adalah “time”, “bayar”, “renew”, “nama” dan “guy”. Perkataan-perkataan ini diurutkan berdasarkan skor TF-IDF mereka, yang menunjukkan betapa pentingnya setiap perkataan dalam dokumen tersebut. Perkataan dengan skor TF-IDF yang tinggi seperti "merit" dan "kolej" lebih penting dan lebih kerap muncul dalam dokumen yang dinilai, berbanding dengan perkataan dengan skor lebih rendah seperti "bayar" dan "time".

II. N-Grams

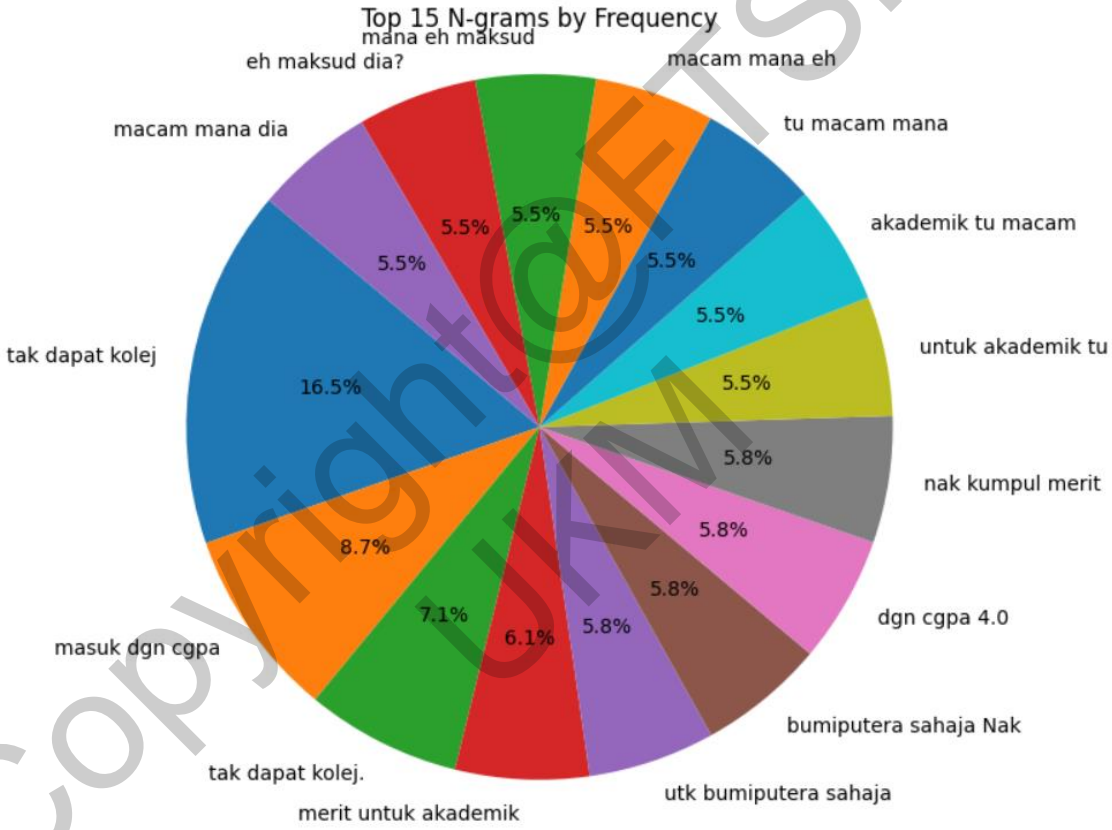
Penjanaan N-gram dimulakan dengan memecahkan teks kepada n-gram berdasarkan nilai "n" yang dipilih. Ini dilakukan dengan meluncurkan tetingkap sepanjang teks untuk menangkap n-gram. Contohnya, untuk teks "Saya tidak dapat kolej", bigram akan menjadi ["Saya tidak", "tidak dapat", "dapat kolej"]. Rajah 5 dan Rajah 6 menunjukkan hasil output bagi Bigram dan Trigram dengan mengubah bilangan n kepada n=2 dan n=3.

Rajah 5 menunjukkan carta pai ini dengan 15 bigrams paling kerap muncul berdasarkan frekuensi dalam data yang dianalisis. N-gram "dapat kolej" adalah yang paling dominan dengan frekuensi 16.5%, diikuti oleh "tak dapat" dengan 13.1%, dan "join program" dengan 10.3%. N-grams lain yang signifikan termasuk "macam mana" (7.6%), "dapat merit" (5.7%), "tak masuk" (5.3%), "kumpul merit" (5.3%), dan "merit kita" (4.9%). N-grams yang muncul kurang kerap, tetapi masih ketara, termasuk "yang tak" (4.8%), "program yang" (4.5%), "merit untuk" (4.3%), "tapi tak" (4.3%), "nak tahu" (4.1%), dan "dapat kolej." (3.8%). Rajah ini menunjukkan tema utama dalam data, dengan fokus pada isu-isu berkaitan kolej dan program, serta kebimbangan mengenai merit dan kemasukan.



RAJAH 5 Carta Pai Bigram

Rajah 6 menunjukkan carta pai ini dengan 15 trigrams paling kerap muncul berdasarkan frekuensi dalam data yang dianalisis. N-gram "tak dapat kolej" adalah yang paling dominan dengan frekuensi 16.5%, diikuti oleh "masuk dgn cgpa" dengan 8.7%, dan "tak dapat kolej." dengan 7.1%. N-grams lain yang signifikan termasuk "merit untuk akademik" (6.1%), "bumiputera sahaja Nak" (5.8%), "dgn cgpa 4.0" (5.8%), "nak kumpul merit" (5.8%), dan "untuk akademik tu" (5.5%). N-grams yang muncul kurang kerap tetapi masih ketara termasuk "akademik tu macam" (5.5%), "tu macam mana" (5.5%), "macam mana eh" (5.5%), "mana eh maksud" (5.5%), "eh maksud dia?" (5.5%), dan "macam mana dia" (5.5%). Rajah ini menunjukkan tema utama dalam data, dengan fokus pada isu-isu berkaitan kemasukan ke kolej, CGPA, merit akademik, dan keutamaan kepada bumiputera.



RAJAH 6 Carta Pai Trigram

III. Awan Kata

Awan kata (*word cloud*) adalah visualisasi grafik yang digunakan untuk mewakili frekuensi atau kepentingan perkataan dalam teks dengan ukuran perkataan yang berbeza. Semakin besar perkataan dalam awan kata, semakin sering perkataan itu muncul dalam teks atau semakin penting perkataan itu berdasarkan metrik tertentu seperti TF-IDF. Dalam analisis sentimen, awan kata digunakan untuk mendapatkan gambaran keseluruhan tentang perkataan yang dominan dalam teks dan membantu dalam mengenal pasti pola dan tema utama. Rajah 7 menunjukkan awan kata bagi set data setelah menyingkirkan kata henti yang tidak relevan.

RAJAH 7 Awan Kata Set Data

Analisis Data

I. Pendekatan Berasaskan Leksikon

VADER (*Valence Aware Dictionary and sEntiment Reasoner*) adalah salah satu algoritma analisis sentimen berasas leksikon yang digunakan untuk menilai sentimen teks dalam projek ini. Rajah 8 menunjukkan hasil output teks yang telah dipra-proses dan dikelaskan kepada sentimen positif, neutral atau negatif:

	Original_Teks	Cleaned_Teks
1	merit 45 ada peluang ke nak dapat kolej😔 tak l...	merit 45 peluang kolej tak larat alik
2	hai semua pewaris watan, untuk korang yang ada...	pewaris watan terfikir join kelab suggestkan j...
3	hai semua pewaris watan, untuk korang yang ada...	pewaris watan terfikir join kelab suggestkan j...
4	Actually yang ckp tipu merit aktiviti 40 betul...	actually tipu merit aktiviti 40 betul senanye ...
5	Mana yang risau merit rendahlah apalah tak dap...	risau merit rendahlah apalah tak kolej risau m...
...	...	...
872	Knp ehh dlm emerit kolej tak ada merit b40? Ca...	ehh emerit kolej tak merit b40 contact suruh u...
873	koranggg, aq nak survey markah merit semua org...	koranggg aq survey merit normal tak reaction s...
874	ukm x de club running ek. Sambil sambil boleh ...	ukm x de club running ek sambil sambil boleh m...
875	Korang merit ni dorg kira mcm mana ehh? Aku ad...	merit dorg kira ehh join dua pertandingan tuli...
876	Kalau merit bawah 5 memang tak dapat kolej ke ...	merit 5 tak kolej

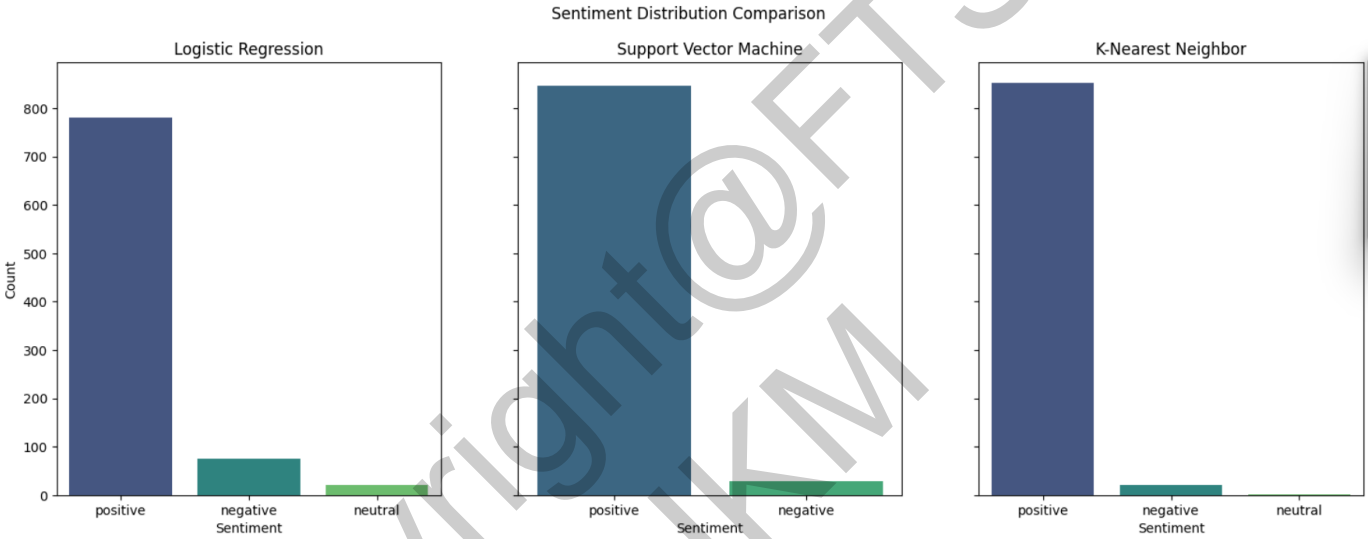
876 rows × 2 columns

RAJAH 8 Hasil Output Analisis Sentimen Menggunakan VADER

Peratus pengagihan sentimen dalam dataset yang dianalisis mendapati bahawa majoriti sentimen dalam dataset adalah positif, dengan 83.6% daripada teks yang dianalisis menunjukkan sentimen positif. Sentimen negatif membentuk 12.2% daripada dataset, manakala sentimen neutral adalah yang paling sedikit dengan 4.1%. Ini menunjukkan bahawa teks yang dianalisis cenderung mempunyai nada positif secara keseluruhan.

II. Pendekatan Pembelajaran Mesin

Pendekatan pembelajaran mesin dengan menggunakan Logistic Regression Classifier, Support Vector Machine dan K-Nearest Neighbor untuk mengelaskan sentimen dalam projek ini. Berikut merupakan perbandingan pengagihan sentimen kepada sentimen positif, negatif atau neutral bagi ketiga-tiga pendekatan pembelajaran mesin dalam bentuk carta bar mengikut bilangan:

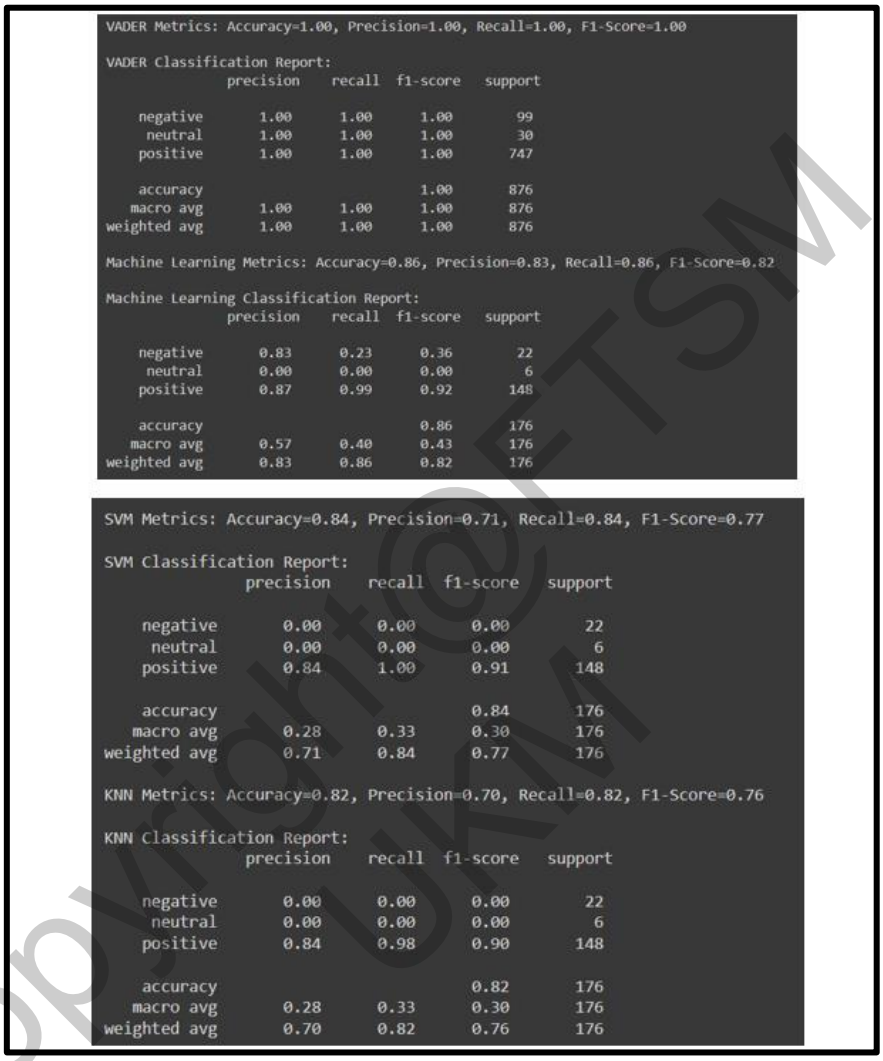


RAJAH 9 Hasil Output Analisis Sentimen Menggunakan Logistic Regression

Rajah tersebut menunjukkan perbandingan pengagihan sentimen menggunakan tiga model pembelajaran mesin: Logistic Regression, Support Vector Machine, dan K-Nearest Neighbor. Pada semua model, sentimen positif adalah yang paling dominan dengan jumlah yang jauh lebih tinggi berbanding sentimen negatif dan neutral. Logistic Regression dan K-Nearest Neighbour menunjukkan corak yang serupa dengan kiraan sentimen positif yang lebih daripada 700, sementara Support Vector Machine juga menunjukkan kiraan yang tinggi untuk sentimen positif tetapi sedikit lebih rendah daripada dua model yang lain. Sentimen negatif dan neutral dalam ketiga-tiga model adalah jauh lebih rendah berbanding sentimen positif, dengan kiraan sentimen negatif lebih tinggi sedikit daripada neutral dalam setiap model. Ini menunjukkan kecenderungan data untuk lebih banyak sentimen positif berbanding negatif atau neutral.

III. Penilaian Data

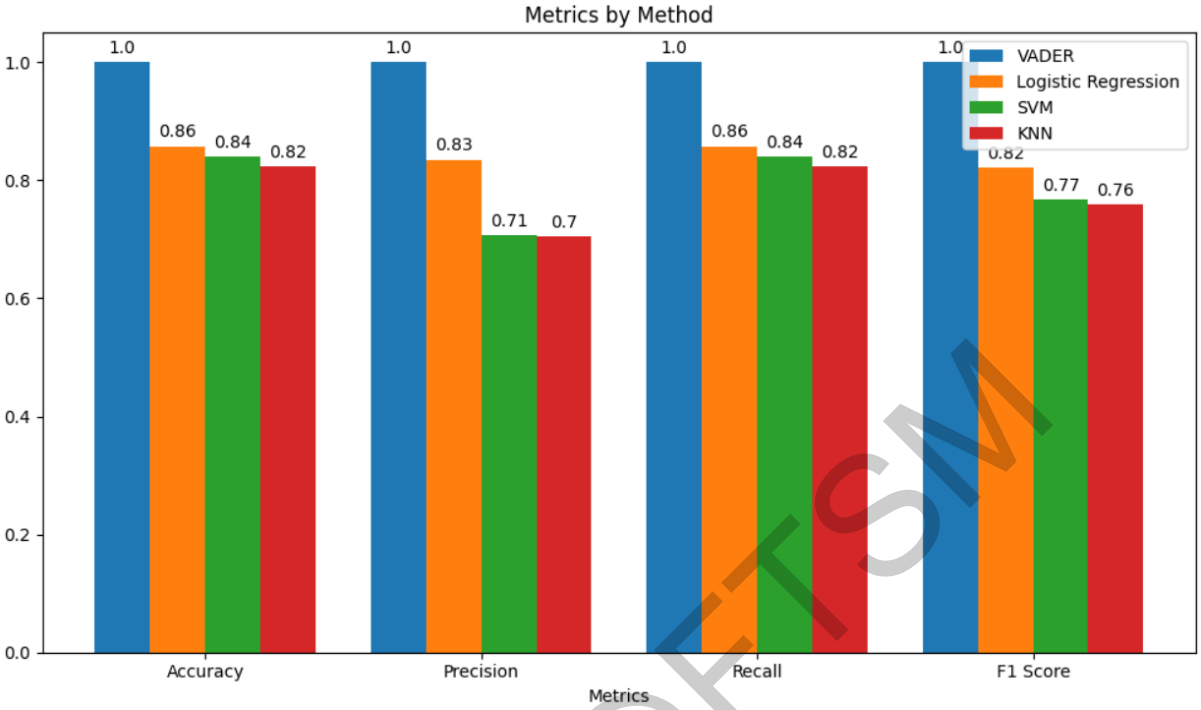
Rajah 10 menunjukkan hasil output penilaian data bagi kedua-dua pendekatan berasaskan leksikon (VADER) dan pendekatan pembelajaran mesin (Logistic Regression) dalam bentuk kejituan, ketepatan, ingatan dan skor-F1:



```
VADER Metrics: Accuracy=1.00, Precision=1.00, Recall=1.00, F1-Score=1.00

VADER Classification Report:
              precision    recall  f1-score   support

    negative       1.00      1.00      1.00        99
     neutral       1.00      1.00      1.00        30
    positive       1.00      1.00      1.00       747

    accuracy                           1.00       876
   macro avg       1.00      1.00      1.00       876
weighted avg       1.00      1.00      1.00       876

Machine Learning Metrics: Accuracy=0.86, Precision=0.83, Recall=0.86, F1-Score=0.82

Machine Learning Classification Report:
              precision    recall  f1-score   support

    negative       0.83      0.23      0.36        22
     neutral       0.00      0.00      0.00         6
    positive       0.87      0.99      0.92       148

    accuracy                           0.86       176
   macro avg       0.57      0.40      0.43       176
weighted avg       0.83      0.86      0.82       176

SVM Metrics: Accuracy=0.84, Precision=0.71, Recall=0.84, F1-Score=0.77

SVM Classification Report:
              precision    recall  f1-score   support

    negative       0.00      0.00      0.00        22
     neutral       0.00      0.00      0.00         6
    positive       0.84      1.00      0.91       148

    accuracy                           0.84       176
   macro avg       0.28      0.33      0.30       176
weighted avg       0.71      0.84      0.77       176

KNN Metrics: Accuracy=0.82, Precision=0.70, Recall=0.82, F1-Score=0.76

KNN Classification Report:
              precision    recall  f1-score   support

    negative       0.00      0.00      0.00        22
     neutral       0.00      0.00      0.00         6
    positive       0.84      0.98      0.90       148

    accuracy                           0.82       176
   macro avg       0.28      0.33      0.30       176
weighted avg       0.70      0.82      0.76       176
```

RAJAH 10 Hasil Output Penilaian Data

Rajah tersebut membandingkan prestasi empat kaedah pengelasan - VADER, Logistic Regression, Support Vector Machine, dan K-Nearest Neighbor - berdasarkan empat metrik: Ketepatan (Accuracy), Ketepatan (Precision), Ingatan (Recall), dan Skor F1 (F1 Score). VADER mencapai skor sempurna (1.0) untuk semua metrik, menunjukkan prestasi yang terbaik. Logistic Regression, dan Support Vector Machine mempunyai prestasi yang serupa, dengan ketepatan, ingatan, dan skor F1 masing-masing sekitar 0.84 hingga 0.86, manakala ketepatan untuk Logistic Regression sedikit lebih tinggi pada 0.83 berbanding 0.71 untuk Support Vector Machine. K-Nearest Neighbor mempunyai prestasi terendah antara empat kaedah dengan nilai metrik di antara 0.70 hingga 0.82. Ini menunjukkan bahawa VADER adalah yang paling cekap dalam pengelasan sentimen dalam dataset yang diuji, diikuti oleh Logistic Regression, Support Vector Machine, dan K-Nearest Neighbor.



RAJAH 11 Carta Bar Penilaian Data

Cadangan Penambahbaikan

Selepas menjalankan kajian yang menyeluruh, cadangan untuk menambahbaik analisis sentimen ini pada masa hadapan adalah dengan membuat pengelasan kategori selain daripada pengelasan sentimen. Pengelasan kategori boleh dilakukan mengikut kategori pembahagian markah Emerit iaitu bahagian Persatuan/Kelab/Badan Beruniform, Aktiviti Pelajar, Kecemerlangan, dan Akademik & B40. Selain itu, untuk meningkatkan lagi ketepatan dan kefahaman analisis sentimen, pendekatan pembelajaran mendalam seperti Long Short-Term Memory (LSTM) dan Convolutional Neural Networks (CNN) boleh diterapkan. Model-model ini mampu memahami konteks dan hubungan antara perkataan dengan lebih baik, serta mengatasi beberapa batasan yang ada pada pendekatan tradisional. Akhir sekali, projek ini hanya memaparkan visualisasi data hasil output dalam bentuk carta bar dan carta pai. Dengan menghasilkan papan pemuka yang berinformasi dengan grafik yang tinggi, lebih banyak analisis yang bermakna dapat ditunjukkan. Power BI boleh digunakan sebagai platform papan pemuka yang sesuai, yang boleh menghubungkan antara pemboleh ubah sentimen yang telah dijana dengan aliran masa sepanjang kajian.

KESIMPULAN

Cadangan utama daripada kajian ini adalah untuk meningkatkan komunikasi dan memberikan penjelasan yang lebih terperinci mengenai kriteria penilaian dalam Sistem Emerit. Pihak universiti perlu memastikan semua maklumat berkaitan Sistem Emerit disampaikan dengan jelas kepada semua pelajar melalui saluran komunikasi yang berkesan seperti e-mel, laman web, dan media sosial, serta mengadakan sesi taklimat dan bengkel untuk memberi penerangan secara langsung. Selain itu, dokumentasi yang jelas dan terperinci mengenai

kriteria penilaian perlu disediakan dan mudah diakses oleh semua pelajar dan staf universiti. Penjelasan terperinci mengenai setiap kriteria yang digunakan dalam penilaian, termasuk contoh-contoh penilaian, akan membantu pelajar memahami bagaimana markah diberikan berdasarkan kriteria yang ditetapkan. Dengan melaksanakan cadangan-cadangan ini, pihak universiti dapat meningkatkan kepercayaan dan keadilan dalam kalangan pelajar terhadap sistem ini, yang seterusnya akan meningkatkan penerimaan dan keberkesanan sistem ganjaran berasaskan merit. Ini akan memberi manfaat yang besar kepada komuniti UKM secara keseluruhan dengan meningkatkan motivasi pelajar untuk mencapai kecemerlangan dalam akademik dan aktiviti kokurikulum. Secara keseluruhan, penemuan dan cadangan dari kajian ini diharapkan dapat membantu pihak universiti dalam meningkatkan keadilan, penerimaan, dan keberkesanan Sistem Emerit, sekaligus memberi manfaat kepada komuniti UKM secara keseluruhan.

Kelemahan Sistem

Terdapat beberapa kekangan dalam membuat projek ini, diataranya ialah kekurangan responden yang berkaitan dengan projek ini kerana sistem Emerit UKM baru sahaja dilaksanakan pada sesi 2022/2023 yang lepas. Maka, hanya terdapat satu kelompok yang boleh memberi ulasan mereka terhadap kajian ini. Selain itu, garis masa juga merupakan salah satu kekangan untuk menyiapkan projek ini dengan sempurna kerana isu sistem Emerit hanya timbul mengikut musim iaitu semasa tamat sesi akademik (keputusan skor Emerit dikeluarkan) dan juga semasa awal sesi akademik (keputusan kelayakan kolej kediaman dikeluarkan) bagi pelajar-pelajar UKM. Akhirnya, kekangan dari segi ulasan yang diperoleh mengandungi ayat yang bercampur bahasa (bahasa malaysia dan bahasa inggeris), bahasa singkatan dan juga kata-kata berbentuk sindiran menjadi cabaran dalam menganalisis sentimen dan komen pengguna kerana alat pemprosesan teks dalam bahasa malaysia yang sesuai dan berkualiti adalah sangat terhad.

PENGHARGAAN

Syukur Alhamdulillah dan setinggi-tinggi kesyukuran dipanjatkan kehadrat ilahi kerana dengan izin kurnianya dapat saya menyempurnakan tugasan projek tahun akhir (T4086) ini dengan jayanya. Saya juga ingin mengucapkan ribuan terima kasih kepada semua pihak yang tidak putus-putus dalam usaha membantu menyempurnakan tugasan ini terutamanya kepada penyelia saya Prof. Madya Dr. Masnizah Binti Mohd atas budi bicara beliau dalam memberi tunjuk ajar sepanjang masa tugasan ini dijalankan. Selain itu, saya turut berterima kasih kepada rakan-rakan seperjuangan saya kerana telah banyak menghulurkan bantuan dan kerjasama bagi merealisasikan usaha menyempurnakan tugasan ini dengan jayanya.

Ucapan ini juga ditujukan kepada semua pihak yang telah terlibat dalam menjayakan tugasan ini sama ada secara langsung atau tidak langsung. Segala bantuan yang telah mereka hulurkan amatlah saya hargai kerana tanpa bantuan dan sokongan mereka semua tugasan ini mungkin tidak dapat dilaksanakan dengan baik.

RUJUKAN

Admin. (2023, July 18). UKM lancar Sistem Emerit. Berita UKM. https://www.ukm.my/beritaukm/ukm-lancar-sistem-Emerit/

Adminlp2m. (2022, February 21). *Analisis Sentimen (Sentiment Analysis) : Definisi, Tipe dan Cara Kerjanya*. Lembaga Penelitian Dan Pengabdian Masyarakat. https://lp2m.uma.ac.id/2022/02/21/analisis-sentimen-sentiment-analysis-definisi-tipe-dan-cara-kerjanya/

Apa itu Analisis Sentimen? - Penjelasan tentang Analisis Sentimen - AWS. (n.d.). Amazon Web Services, Inc. https://aws.amazon.com/id/what-is/sentiment-analysis/

Apakah carta aliran? (n.d.). Dropbox. https://experience.dropbox.com/ms-my/resources/flowcharts

Apakah itu Rajah Konteks: Definisi dan Garis Panduan untuk Dicipta. (2022, April 19). https://www.mindonmap.com/ms/blog/context-diagram/

Aspect-level sentiment analysis based on static pruning of syntactic trees. (2023, May 26). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/10177040

Bab 4 DFD design. (n.d.). Google Docs. https://docs.google.com/presentation/d/1J-J-dI-JKcKgIq5XN9yAnLrdYFMzKEzCOManafF31c4/htmlpresent

Barney, N. (2023, December 21). *sentiment analysis (opinion mining)*. Business Analytics. https://www.techtarget.com/searchbusinessanalytics/definition/opinion-mining-sentiment-mining

Butt, M. M. (2022, June 16). Sentiment Analysis of Twitter's US Airlines Data using KNN Classification. *Medium*. https://towardsdatascience.com/sentiment-analysis-of-twitters-us-airlines-data-using-knn-classification-91c7da987e13

Chulan, & Chulan. (2022, September 24). UKM perkenal Sistem Emerit. My TV ONLINE. https://my-tv.online/2022/09/24/ukm-perkenal-sistem-Emerit/

Damarta, R., Hidayat, A., & Abdullah, A. S. (2021). The application of k-nearest neighbors classifier for sentiment analysis of PT PLN (Persero) twitter account service quality. *Journal of Physics. Conference Series, 1722*(1), 012002. https://doi.org/10.1088/1742-6596/1722/1/012002

Editor. (2023, September 11). Sentiment Analysis: types, tools, and use cases. *AltexSoft*. https://www.altexsoft.com/blog/sentiment-analysis-types-tools-and-use-cases/

Educative. (n.d.). *How to use SVM for sentiment analysis*. https://www.educative.io/answers/how-to-use-svm-for-sentiment-analysis

Liu, B. (2020). The problem of sentiment analysis. In *Cambridge University Press eBooks* (pp. 18–54). https://doi.org/10.1017/9781108639286.003

Liu, B. (2012). The problem of sentiment analysis. In *Synthesis lectures on human language technologies* (pp. 9–22). https://doi.org/10.1007/978-3-031-02145-9_2

Mangipudiprashanth. (2020, April 18). *Twitter Sentiment Analysis using ML & NLP*. https://www.kaggle.com/code/mangipudiprashanth/twitter-sentiment-analysis-using-ml-nlp

Himma, F. (2022, December 19). *Analisis Sentimen adalah: Pengertian, Contoh, Tipe*. https://majoo.id/solusi/detail/analisis-sentimen-adalah

Instagram. (n.d.). https://www.instagram.com/p/CiwM7MIveK8/

Hamid, A. A. (2022, October 28). 'Mana Kolej Kami?, ini penjelasan UKM. *Berita Harian*. https://www.bharian.com.my/berita/nasional/2022/10/1017765/mana-kolej-kami-ini-penjelasan-ukm

Hampir 13,000 penginapan disediakan kepada pelajar sesi akademik 2022/2023 - UKM. (n.d.). www.astroawani.com. Retrieved October 28, 2022, from

https://www.astroawani.com/berita-malaysia/hampir-13000-penginapan-disediakan-kepada-pelajar-sesi-akademik-20222023-ukm-388310

Han, K., Chien, W., Chiu, C., & Cheng, Y. (2020). Application of Support Vector Machine (SVM) in the sentiment analysis of Twitter DataSet. *Applied Sciences*, *10*(3), 1125. https://doi.org/10.3390/app10031125

Hick, H., Bajzek, M., & Faustmann, C. (2019). Definition of a system model for model-based development. *SN Applied Sciences/SN Applied Sciences*, *1*(9). https://doi.org/10.1007/s42452-019-1069-0

Kannan, S., Karuppusamy, S., Nedunchezhian, A., Venkateshan, P., Wang, P., Bojja, N., & Kejariwal, A. (2016). Big data analytics for social media. In *Elsevier eBooks* (pp. 63–94). https://doi.org/10.1016/b978-0-12-805394-2.00003-9

Lexicon-based sentiment analysis: What it is & how to conduct one | KNIME. (n.d.). KNIME. https://www.knime.com/blog/lexicon-based-sentiment-analysis

Onan, A., Korukoğlu, S., & Bulut, H. (2017). A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification. *Information Processing & Management*, *53*(4), 814–833. https://doi.org/10.1016/j.ipm.2017.02.008

Perisian dan perkakasan yang digunakan. (n.d.). https://smksy.tripod.com/Perisian.htm

Phienthrakul, T., Kijsirikul, B., Takamura, H., & Okumura, M. (2009). Sentiment Classification with Support Vector Machines and Multiple Kernel Functions. In *Lecture notes in computer science* (pp. 583–592). https://doi.org/10.1007/978-3-642-10684-2_65

Riochr. (n.d.). *Analisis-Sentimen-ID/kamus/rootword.txt at master · riochr17/Analisis-Sentimen-ID*. GitHub. https://github.com/riochr17/Analisis-Sentimen-ID/blob/master/kamus/rootword.txt

Semmlow, J. (2012). Linear systems analysis in the time domain. In *Elsevier eBooks* (pp. 261–316). https://doi.org/10.1016/b978-0-12-384982-3.00007-9

Sentiment Analysis Approach based on N-Gram and KNN classifier. (2018, December 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/8703350

Sukawai, E., & Omar, N. (2020). Corpus Development for Malay Sentiment Analysis using semi Supervised approach. *Asia-Pacific Journal of Information Technology and Multimedia*, *09*(01), 94–109. https://doi.org/10.17576/apjitm-2020-0901-08

Summing three lexicon based approach methods for sentiment analysis? (n.d.). Data Science Stack Exchange. https://datascience.stackexchange.com/questions/83991/summing-three-lexicon-based-approach-methods-for-sentiment-analysis

System Requirements Specification (SRS) - Kamus Projek NIISE. (n.d.). https://ppk.moha.gov.my/niise/mediawiki/index.php?title=System_Requirements_Specification_%28SRS%29

Venugopalan, M., & Gupta, D. (2020). An Unsupervised Hierarchical Rule Based Model for Aspect Term Extraction Augmented with Pruning Strategies. *Procedia Computer Science*, *171*, 22–31. https://doi.org/10.1016/j.procs.2020.04.303

Nur Amira Farisha Binti Mohd Yusri (A186976)
Prof. Madya Dr. Masnizah Mohd
Fakulti Teknologi & Sains Maklumat
Universiti Kebangsaan Malaysia