Asia-Pacific Journal of Information Technology and Multimedia 14(2) 2025: 438 – 452
https://doi.org/10.17576/apjitm-2025-1402-06

# Sentiment Analysis on Indonesian COVID-19 Pandemic based on 2020 Twitter Data

# Analisis Sentimen Tentang Pandemik COVID-19 Indonesia Berasaskan Data Twitter 2020

*Suhaila Zainudin*[*]*, Nur Aliah Ahmad Basri, dan Lailatul Qadri Zakaria, Fadilla 'Atyka Nor Rashid*

*Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor*

[*]*Corresponding author: suhaila.zainudin@ukm.edu.my*

*Received 18 September 2024*
*Accepted 24 July 2025, Available online 15 December 2025*

## ABSTRACT

From the end of 2019 until 2022, the COVID-19 pandemic spread globally, and most countries implemented movement control measures or curfews to contain its spread. At the same time, citizens from various parts of the world have used social media as a platform for sharing issues, problems, and emotional expressions in their own mother tongue. A high amount of Twitter data was collected during this era, and this leads to the motivation of the study, which is to utilize Twitter data sources from various languages , since the majority of existing studies focus on the English language. This is largely due to the abundance of available resources in English. Another issue is the difficulty of performing manual sentiment classification on social media data from application providers. Social media data is vast in size and requires a feature extraction process before further processing. This study uses COVID-19 Indonesian Twitter data collected between May and July 2020. This study has pre-processed the dataset and also translated the Indonesian tweets into English. The dataset is then analyzed using TextBlob to determine the polarity label for each tweet. Then, to preserve important information, feature extraction approaches such as Bag-of-Words (BOW) and Term Frequency-Inverse Document Frequency (TF-IDF) are used. For the classification task, this study uses Naïve Bayes (NB), Decision Tree (DT), and Support Vector Machine (SVM) techniques to predict sentiment classification. Results, including Accuracy, Precision, Recall, and F-Score, were used to evaluate the performance. The results of the study show that BOW produces features that improve the performance of the Decision Tree (DT) up to an accuracy of 89%.

Keywords**:** *Analisis Sentimen; Pembelajaran Mesin: Pengekstrakan Fitur; BOW; TF-IDF*

## ABSTRAK

Bermula dari hujung 2019 sehingga 2022, pandemik COVID-19 telah merebak secara global dan kebanyakan negara melaksanakan perintah berkurung bagi mengawal penularan pandemik. Pada masa yang sama, warga dari berbagai pelusuk dunia telah menggunakan media sosial seperti Twitter atau kini dikenali sebagai X sebagai tapak perkongsian berbagai isu, masalah dan luahan emosi menggunakan bahasa ibunda tersendiri. Terdapat banyak data twit yang terkumpul semasa era perintah berkurung dan ini memotivasi kajian ini untuk memanfaatkan sumber data twit dari pelbagai bahasa memandangkan majoriti kajian sedia ada memfokus kepada bahasa Inggeris. Ini disebabkan kebanyakan sumber penyelidikan sedia ada menggunakan bahasa Inggeris. Satu lagi isu adalah sukar melakukan pengelasan sentimen secara manual ke atas data media sosial secara langsung dari penyedia aplikasi. Lazimnya, data media sosial bersaiz amat besar dan perlu melalui proses pengekstrakan fitur sebelum diproses menggunakan pembelajaran mesin. Kajian ini menggunakan data COVID-19 Indonesia Twit yang dikumpul antara Mei dan Julai 2020. Set data ini kemudian dipraproses dan kajian ini turut menterjemah twit bahasa Indonesia ke twit bahasa Inggeris. Set data kemudian diproses menggunakan Textblob bagi menentukan label polariti bagi setiap twit. Kemudian bagi memelihara maklumat penting, pendekatan pengekstakan fitur seperti Bag of Words (BOW) dan Term Frequency-Inverse Document Frequency (TF-IDF). Untuk tugas pengelasan, kajian ini menggunakan Teknik Naïve Bayes (NB), Pohon Keputusan (DT), dan Mesin Sokongan Vektor (SVM) bagi meramal pengelasan sentimen. Hasil Ketepatan, Kejituan, Dapatan Semula, dan Skor-F1 digunakan bagi menilai hasil. Hasil kajian adalah BOW menghasilkan fitur yang meningkatkan prestasi Pohon Keputusan (DT) sehingga mencapai ketepatan 89%.


Kata kunci**:** *Analisis Sentimen; Pembelajaran Mesin: Pengekstrakan Fitur; BOW; TF-IDF*

## PENGENALAN

Pandemik COVID-19 telah menular secara global bermula dari hujung 2019 sehingga 2022 (Kementerian Kesihatan Malaysia, 2023). Apabila pandemik COVID-19 mula berleluasa serata dunia, beberapa negara telah mengambil inisiatif melaksanakan perintah kawalan pergerakan bagi mengurangkan jangkitan pandemik COVID-19 terhadap masyarakat. Malaysia turut melaksanakan perintah kawalan pergerakan dan dibahagi kepada beberapa fasa; antaranya Perintah Kawalan Pergerakan(PKP), Perintah Kawalan Pergerakan Bersyarat(PKPB) dan Perintah Kawalan Pergerakan Pemulihan (PKPP) (Ismail et al., 2020).

Indonesia mula dilanda pandemik COVID-19 dari Februari 2020 (Astuti & Mahardhika, 2020). Pandemik ini telah memberi kesan kepada penurunan kadar aktiviti ekonomi dan memberi impak kepada pelbagai sektor seperti perniagaan, pelaburan, dan institusi kewangan. (Kennedy, 2020) telah mengkaji kesan sistem perintah berkurung berdasarkan strategi pengurusan perubahan yang dilaksanakan oleh pemimpin yang mempunyai kuasa penuh dan menggunakan teori ekonomi yang maju. Oleh kerana perkerjaan di Indonesia lebih tertumpu di sektor tidak formal, jika perintah berkurung dilaksanakan, kesannya terhadap rakyat adalah lebih besar berbanding dengan negara lain kerana majoriti aktiviti ekonomi rakyat Indonesia bergantung kepada pendapatan harian (Kennedy, 2020). Natijah ke atas ekonomi setempat memberi impak sangat besar dan kesan ini tidak dapat diramalkan memandangkan kadar penyebaran pandemik COVID-19 sangat cepat dan ketiadaan vaksin untuk mengawal penularan pandemik.

Rentetan itu, pada pertengahan bulan April 2020, Presiden Indonesia mengumumkan larangan balik kampung atau ‘mudik’ untuk sambutan Aidil Fitri (Gorbiano, 2020). Ini bagi membendung

penularan COVID-19 ke lain-lain daerah di Indonesia. Semua kaedah perjalanan pulang ke kampung halaman dengan menaiki pengangkutan awam seperti bas, penerbangan komersial, pengangkutan laut serta kereta api diharamkan sehingga bulan Jun 2020. Lantaran itu, rakyat Indonesia mengambil peluang menggunakan platform media sosial seperti Twitter atau Facebook sebagai tempat luahan emosi dan pandangan mereka masing-masing terhadap perlaksanaan perintah ini.

Selain Indonesia, India juga melaksanakan sistem perintah berkurung. Perintah berkurung sepenuhnya di India melaksanakan pertama kali pada 24 Mac 2020 dan berlangsung selama 21 hari bagi mengurangkan penyebaran pandemik COVID-19. Semua individu dihadkan dari berkunjung ke rumah individu lain dan mereka yang keluar untuk perkhidmatan penting perlu mematuhi piawai jarak sosial yang ketat (Kaur & Ranjan, 2020). Pergerakan manusia yang dihadkan menyebabkan mereka menghabiskan banyak masa di rumah atau tempat tinggal. Ini mewujudkan senario yang ideal untuk menyatakan luahan, pengalaman, dan pandangan mereka di laman rangkaian sosial seperti Twitter.

Data media sosial yang terkumpul ini sesuai untuk diproses menggunakan kaedah analisis sentiment (Madhoushi et al, 2019) (Madhoushi et al., 2023) (Ezuana Sukawai & Nazlia Omar, 2020). Analisis sentimen ialah aplikasi pemprosesan bahasa semula jadi bertujuan mengenal pasti ekspresi berasaskan pendapat peribadi seseorang (Li & Hovy, 2017). Kajian ini bertujuan untuk menganalisis sentimen menggunakan data Twitter Indonesia yang direkodkan dari Mei hingga Julai 2020. Persembahan kajian ini terbahagi kepada beberapa seksyen iaitu seksyen 1 merangkumi pengenalan kepada kajian, diikuti dengan seksyen 2 iaitu latar belakang kajian. Seterusnya diikuti dengan seksyen 3 iaitu metodologi dan seksyen 4 yang membincangkan Hasil Kajian. Seksyen 5 mengakhiri kertas ini dengan membincangkan kesimpulan kajian.

## ANALISIS SENTIMEN MENGGUNAKAN BAHASA INDONESIA

Beberapa kajian lepas yang memfokus kepada analisis sentimen menggunakan bahasa Indonesia telah dilaksana oleh beberapa penyelidik seperti (Afdhalul Ihsan, 2021) (Kalanjati et al., 2024) (Wicaksana et al., 2022) (Zulfa & Winarko, 2017) (Utami). Kajian Zulfa & Winarko (2017) mengkaji sentimen twit Indonesia menggunakan kaedah Rangkaian Kepercayaan Dalam (Deep Belief Network) untuk membina model klasifikasi. Mereka melaporkan bahawa aplikasi kaedah Rangkaian Kepercayaan Dalam dengan kaedah BOW sebagai pengekstrak fitur tidak mendapat ketepatan yang optimum berbanding dengan kaedah Naive Bayes dan SVM dengan menggunakan BOW sebagai pengekstrak fitur.

Manakala, Kalanjati et al., (2024) melaporkan bahawa sentimen dan pendapat mengenai COVID-19 dan vaksinasi COVID-19 di Twitter berbahasa Indonesia jarang dilaporkan dalam satu kajian komprehensif. Oleh itu, kajian ini menganalisis berita dan fakta palsu, dan penglibatan Twitter untuk memahami persepsi dan kepercayaan orang ramai yang menentukan kesedaran tentang kesihatan awam. Kalanjati et al., (2024) mendapati bahawa memahami sentimen dan pendapat orang ramai boleh membantu penggubal dasar merancang strategi terbaik untuk menangani pandemik. Ini terbukti dengan sentimen positif dan pendapat berasaskan fakta tentang COVID-19, dan vaksinasi COVID-19 yang telah dikaji.

Wicaksana et al., (2022) menyatakan media sosial ialah sumber data yang sangat besar dan mempunyai perubahan yang berterusan dan sangat singkat, jangka masa perubahannya juga mempunyai perkaitan yang besar untuk pelbagai bidang. Tujuan kajian mereka adalah menggunakan analisis sentimen untuk menganalisis status twitter yang membincangkan topik harian untuk menentukan sikap pendapat daripada twit yang disiarkan mengenai topik yang

popular dalam talian. Kajian ini menggunakan kaedah Naive Bayes Classifier.

Utami telah membina sistem analisis sentimen ulasan terhadap kosmetik Indonesia menggunakan klasifikasi Naïve Bayes (NB). Kajian ini menyimpulkan bahawan analisis sentimen untuk membuat keputusan dari ulasan produk kosmetik menggunakan dengan Pengelas Naïve Bayes Classifier menghasilkan ketepatan setinggi 80%. Sistem ini menghasilkan ketepatan lebih baik apabila terdapat emotikon di dalam ulasan. Penggunaan stopword dan penapisan pada tahap pra-pemprosesan teks preprocessing dapat mengurangkan perkataan dan mengurangkan beban komputasi PengelasNaïve Bayes Classifier, sehingga hanya kata penting yang akan dihitung dan langsung dapat meningkatkan ketepatan.

Merdeka Belajar Kampus Merdeka (MBKM) ialah dasar yang diperkenal oleh Menteri Pendidikan dan Kebudayaan Indonesia untuk meningkatkan keupayaan, kreativiti dan inovasi pelajar kolej (Suhud et al., 2023). Program Kampus Mengajar merupakan satu program MKBM yang memberi peluang kepada peserta dalam menyelesaikan masalah, pembangunan strategik, pembelajaran yang berkesan, inovatif dan menyeronokkan. Bagi mengenalpasti pendapat umum terhadap pelaksanaan Kampus Mengajar, kajian ini menjalankan analisis sentimen menggunakan set data Twitter Seribu lima ratus yang mengandungi Kampus Mengajar sebagai kata kunci. Kajian menggunakan prapemprosesan sebelum mengklasifikasikan set data dalam Naive Bayes. Kaedah SMOTE dijalankan untuk mengatasi data yang tidak seimbang. Hasil kajian menunjukkan tahap ketepatan dalam menentukan kategori ialah 77.45% dan purata mikro ialah 77.45% dalam menentukan sentimen dan mempunyai tahap ketepatan 81.46% dan ingatan semula sebanyak 77.45%.

Kajian Utami melaksana analisis berasaskan statistik dan pembelajaran mesin pada 40,000 twit, yang dikumpulkan dalam dua rangka masa yang berbeza. Twit dikumpulkan dari tapak Twitter antara 3/07/2020 hingga 11/07/2020 dan 01/08/2020 hingga 06/08/2020. Pelbagai perpustakaan berasaskan Python digunakan untuk pemerolehan data, prapemprosesan data dan proses analisis data. Sebagai fasa prapemprosesan data pada mulanya ayat dibersihkan. Kemudian dengan mengira kekutuban dan ukuran subjektiviti tweet dikategorikan kepada tiga kumpulan (negatif, neutral, dan positif}). Selepas itu, dalam fasa kemudian dengan menggunakan skema pengekstrakan ciri frekuensi-songsang dokumen Jangka (TF-IDF) dengan bantuan teknik uni-gram, bi-gram dan tri-gram ciri yang berbeza diekstrak untuk menyediakan set data untuk disuap ke dalam model ramalan. 70% daripada set data digunakan untuk melatih pengelas Gaussian Naïve Bayes (G-NB), Bernoulli's Naïve Bayes (B-NB), Random forest (RF) dan Mesin vektor Sokongan (SVM) untuk menjana model ramalan yang berbeza. Akhirnya, 30% daripada data diuji pada model pembelajaran tersebut. Keputusan eksperimen menunjukkan bahawa model RF dan B-NB berprestasi lebih baik daripada dua model pengelas yang lain. Kajian eksperimen juga menunjukkan bahawa SVM mempunyai kos yang lebih tinggi.

Semasa tempoh kawalan pergerakan berkaitan pandemik global COVID-19, pengguna menyatakan kebimbangan mereka tentang krisis melalui rangkaian sosial. Kaur & Ranjan, (2020) menganalisis tindak balas orang ramai terhadap sekatan 21 hari pencegahan COVID-19 di India dari 25 Mac 2020 hingga 14 April 2020. Masyarakat umum India secara aktif bertindak balas terhadap hashtag Twitter berkaitan kawalan pergerakan secara positif dan negatif. Set data twit sepanjang tempoh kawalan pergerakan dianalisis setiap hari untuk mengukur respons orang ramai. Kaur & Ranjan (2020) mengenal pasti perkataan dan pasangan perkataan yang paling kerap digunakan sebagai fokus perbualan. Hasil kajian turut mengenalpasti perkataan negatif dan positif yang paling tinggi digunakan.

Khan et al., (2021) melaksanakan analisis sentimen menggunakan data twit dari Amerika

Syarikat menggunakan pembelajaran mesin dan pendekatan analisis leksikon. Kajian ini membuktikan bahawa TF-IDF boleh meningkatkan prestasi model pembelajaran mesin yang diselia dan gradient bossting machine atau mesin penggalak kecerunan mengatasi kaedah lain dan mencapai ketepatan tinggi 96% apabila dipasangkan dengan ciri TF-IDF. Analisis ini dilakukan untuk menganalisis bagaimana keadaan semasa pandemic COVID 19 dikendalikan oleh rakyat Amerika Syarikat.

Masalah utama dalam kajian analisa sentimen adalah majoriti penerbitan kajian hanya menggunakan ulasan bahasa Inggeris. Ini disebabkan oleh kekurangan sumber dalam bahasa lain (Liaqat MI, 2022). Terdapat kekurangan dari segi sumber korpora yang lengkap dalam bahasa selain Bahasa Inggeris (Ghallab et al., 2020). Sumber korpora ini penting untuk melatih model pembelajaran mesin.

Masalah seterusnya berkait dengan data media sosial yang terlalu besar untuk diproses, maka perlunya menggunakan kaedah pengekstrakan fitur yang sesuai. Pengekstrakan fitur adalah proses mengekstrak subset fitur yang informatif dari set data untuk meningkatkan ketepatan klasifikasi dan sangat penting untuk mengenal pasti fitur berkaitan dengan betul dalam teks (Kalaivani et al., 2020; Suhaidi, 2021). Data dalam jumlah besar yang terdapat di laman rangkaian sosial di mana manusia mengekspresi emosi, idea, dan sudut pandang mereka. Ini telah mendorong penyelidkan dalam kajian analisis sentimen (Kalaivani et al., 2020). Analisis sentimen memberi tumpuan kepada sikap yang menunjukkan atau menyimpulkan perasaan positif atau negatif.

Teknik pengekstrakan fitur diperlukan untuk menukar teks menjadi matriks (atau vektor) ciri. Teknik yang boleh sesuai untuk mengekstrak fitur daripada data teks seperti BOW atau TF-IDF (Suhaidi, 2021). BoW berupaya untuk memproses teks pendek dengan berkesan. BOW sangat sesuai untuk analisis sentimen twit atau mengklasifikasikan artikel berita pendek (Lee et al., 2023). Walau bagaimanapun, dengan membuang susunan perkataan dan tatabahasa, BoW kehilangan maklumat semantik dalam struktur teks (Abubakar & Umar, 2022). Untuk menangani batasan ini, beberapa strategi boleh digunakan seperti prapemprosesan seperti tokenisasi, penyingkiran perkataan henti dan pemangkasan/lemmatisasi boleh meningkatkan prestasi BoW dengan ketara (Abubakar & Umar, 2022).

TF-IDF mengambil kira kekerapan perkataan dalam dokumen (TF) dan sebaran perkataan di dalam korpus (IDF). Perkataan yang kerap muncul dengan dalam dokumen dianggap lebih penting untuk kandungannya (Qi, 2020). IDF mengeluarkan perkataan yang kerap muncul berlaku dalam banyak dokumen, mempromosikan (Gupta, 2020). Harsha Kadam (2020) memanfaatkannya untuk klasifikasi e-mel berbilang label, manakala Haryadi et al. (2022) menggunakannya untuk mengklasifikasikan keberkesanan ubat berdasarkan ulasan pesakit. TF-IDF mempunyai kekangan dalam menganalisis set data yang besar kerana TF-IDF menggunakan sumber komputeran yang tinggi (Gupta, 2020). Selain itu, menala parameter TF-IDF untuk prestasi optimum bergantung pada tugas tertentu dan ciri data (Gellerman, 2021). Tambahan pula, mengendalikan kata henti berkesan adalah penting, kerana kata henti dengan kekerapan tinggi boleh memberi kesan ke atas pemberat dan mengimpak ketepatan (Haque et al., 2023).

Keberkesanan pengekstrakan fitur juga akan menghasilkan prestasi model klasifikasi yang lebih tepat. Prestasi model analisis sentimen dapat dibangunkan dengan kaedah menggabungkan pendekatan leksikon dan klasifikasi menggunakan kaedah pembelajaran mesin yang terbukti dalam kajian sebelumnya. Algoritma seperti model Naïve Bayes (NB), Decision Tree (DT), dan Support Vector Machine (SVM) (Rosly et al., 2023) merupakan algoritma yang paling kerap digunakan dan berpotensi tinggi dalam klasifikasi teks dan analisis sentimen.

Singh et al., (2023) telah melakukan analisis sentimen terhadap maklum balas pengguna yang diterima di laman web dengan menggunakan kaedah pembelajaran mesin dan model pembelajaran mendalam (BERT). Proses analisis sentimen menggunakan algoritma untuk menilai nada maklum balas, menunjukkan sama ada positif, negatif atau neutral. Maklumat ini kemudiannya digunakan untuk meningkatkan kualiti perkhidmatan atau produk yang disediakan oleh organisasi. Kajian ini mendapati kaedah pembelajaran mendalam adalah lebih tepat dalam menentukan sentimen maklum balas yang diterima di laman web. Keputusan yang diperoleh daripada analisis sentimen membantu penambahbaikan laman web untuk memenuhi keperluan pengguna.

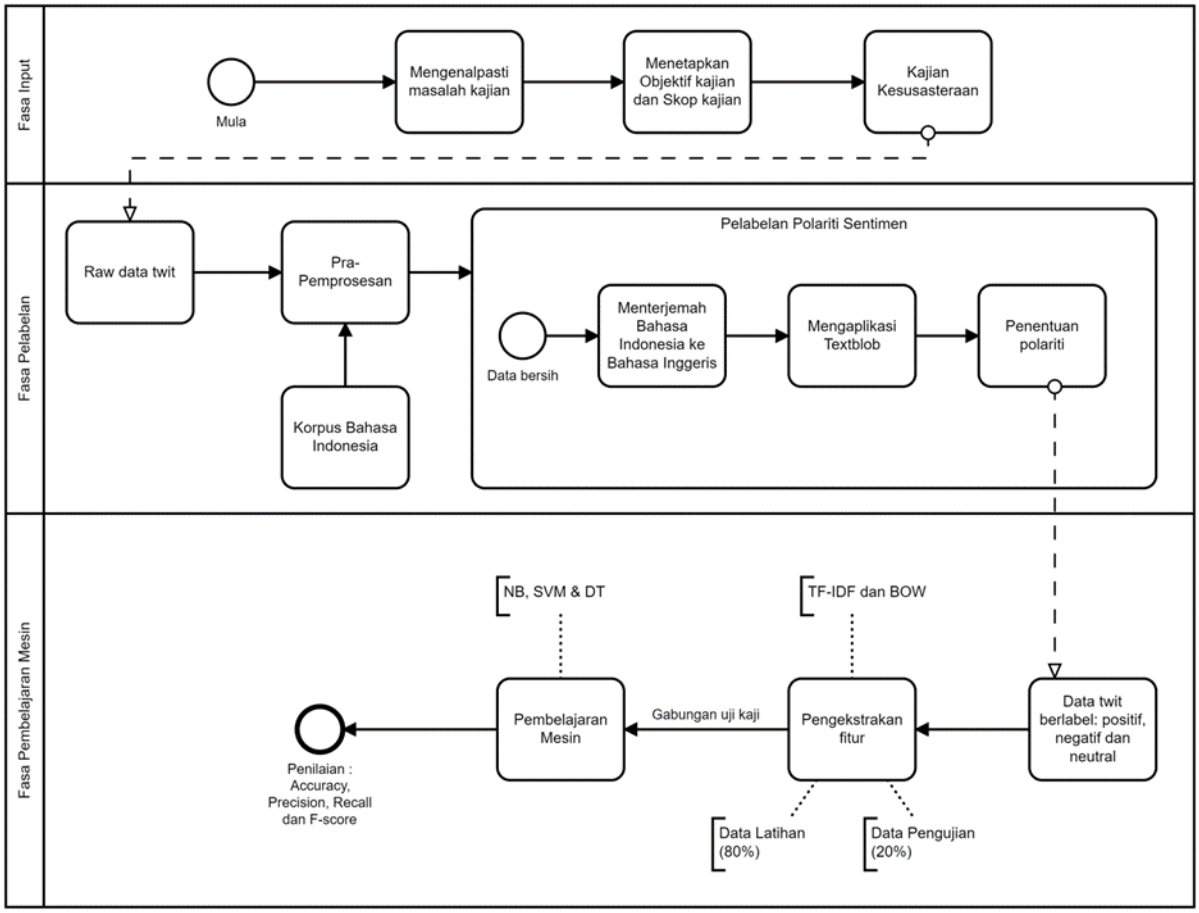
Naïve Bayes (NB) adalah kumpulan algoritma berasaskan kebarangkalian untuk klasifikasi analisis sentiment (Suhud et al., 2023). NB memberikan kebarangkalian bahawa perkataan atau frasa yang diberikan adalah positif atau negative (Wicaksana et al., 2022). Kajian Sutabri et al. (2018) menggunakan kaedah klasifikasi Naïve Bayes untuk menganalisis data testimoni dalam talian dari laman web e-perjalanan terkemuka. Kaedah klasifikasi Naïve Bayes telah dibangunkan bagi mengira nilai kebarangkalian setiap perkataan dan menyediakan penilaian untuk setiap kelas. Naïve Bayes adalah salah satu teknik klasifikasi berdasarkan Teori Bayes dan andaian bagi setiap ramalan adalah tidak bersandar. Di samping itu, pengelas Naïve Bayes membuat andaian bahawa setiap ciri atau ciri dalam kelas tidak berkaitan dengan yang lain.

Selain daripada SVM atau Naive Bayes, Decision Tree (DT) turut digunakan dalam teknik analisis sentimen menggunakan pembelajaran mesin. DT digunakan untuk mengira potensi kejayaan siri keputusan yang berbeza yang dibuat untuk mencapai matlamat tertentu. DT adalah sejenis pemodelan ramalan yang membantu dalam pemetaan beberapa pilihan atau penyelesaian kepada hasil tertentu (Rosly et al., 2023). DT adalah di dalam kategori pembelajaran mesin terselia. DT yang merupakan teknik klasifikasi yang mudah dan digunakan secara meluas. DT adalah teknik pembelajaran yang terselia yang mempunyai pemboleh ubah sasaran yang telah ditetapkan dan sering digunakan dalam masalah klasifikasi (Ghallab et al., 2020).

Rosly et al., (2023) meneroka pelbagai strategi pemilihan ciri (penapis, pembalut dan terbenam) pada data pengajian pasca siswazah daripada universiti awam. Set data ini tidak seimbang, dan kebanyakan kelas adalah pelajar yang menamatkan pengajian mereka. Kelas minoriti adalah untuk pelajar yang belum menamatkan pengajian mereka. Kajian ini menggunakan kaedah pensampelan berlebihan yang dinamakan SMOTE pada data tidak seimbang dalam latihan untuk membolehkan algoritma klasifikasi belajar daripada set data seimbang. Lima algoritma yang diselia, Pokok Keputusan, Hutan Rawak, Naive Bayes, Perceptron Berbilang Lapisan dan Regresi Logistik, telah dinilai untuk model ramalan. Hasilnya menunjukkan bahawa Random Forest ialah algoritma pengelasan terbaik untuk set data ini. Setelah membincangkan latarbelakang kajian, seksyen seterusnya akan memperihalkan metodologi kajian yang telah dilaksanakan

## METODOLOGI

Kajian ini bertujuan untuk melakukan proses analisis sentimen terhadap data daripada media sosial (Twitter) berkaitan dengan COVID-19 dan arahan pemerintahan negara (Pemerintah) sebagai usaha untuk menangani masalah sebelum ini. Metodologi kajian terbahagi kepada tiga (3) fasa iaitu Fasa Input, Fasa Perlabelan dan Fasa Output (Rajah 1). Set data yang digunakan di dalam kajian mengandungi 52959 data berstruktur dengan 10 atribut (Jadual 1). Set data format CSV boleh dimuat turun dari laman web Kaggle https://www.kaggle.com/dionisiusdh/COVID19-indonesian-twitter-sentiment?select=COVID-sentiment.csv. Set data ini mengandungi ulasan Twit pengguna Indonesia yang mengandungi kata kunci "Corona and Pemerintah" atau "COVID dan Pemerintah" dari Mei hingga Julai 2020.




RAJAH 1. Metodologi Kajian

JADUAL 1 : Nama Lajur Medan dan perincian di dalam Data Twitter COVID-19 Indonesia

| No. | Lajur | Jenis data | Penerangan |
|---|---|---|---|
| 1 | date | object | Tarikh twit mengikut: Tahun, Bulan dan Hari. Contoh 2020-05-07 |
| 2 | time | object | Masa twit mengikut sistem 24 jam: Jam , minit dan saat. Contoh 23:57:30 |
| 3 | user_id | int64 | Id bagi setiap pengguna twit. Contoh 1258425982907637761 |
| 4 | username | object | Nama yang digunakan oleh pengguna twit |
| 5 | tweet | object | Teks twit pengguna |
| 6 | mentions | object | Twit yang mengandungi nama pengguna lain |
| 7 | replies_count | int64 | Jumlah kekerapan pengguna membalas twit |
| 8 | retweets_count | int64 | Jumlah kekerapan pengguna retweets |
| 9 | likes_count | int64 | Jumlah kekerapan pungguna lain suka twit tersebut |
| 10 | hashtags | object | "#" yang ditulis di depan topik tertentu agar pengguna lain boleh mencari topik yang sama yang ditulis oleh orang lain juga. Contoh #dudukdirumah |

## FASA INPUT

Fasa input bertujuan untuk mengenalpasti isu-isu utama kajian iaitu analisis sentimen dari media sosial. Untuk menyelami lebih mendalam tentang isu yang difokuskan, analisis terhadap ulasan kesusasteraan telah dilaksanakan. Bahan dari buku, artikel, blog, laman web atau laman sosial media adalah antara sumber maklumat dan pengetahuan mengenai topik atau isu. Selain itu juga, dalam ulasan sastera ini dapat menambah fahaman mengenai masalah berkaitan analisis sentimen, menambah pengetahuan mengenai teknik-teknik penyelidikan yang sesuai dalam membangunkan model yang baik.

## FASA PERLABELAN

Aktiviti pertama dalam fasa ini adalah penyingkiran twit berulang dan menyingkirkan rekod yang tidak mempunyai twit atau twit kosong. Medan utama untuk dianalisis adalah medan yang mengandungi twit yang ditulis pengguna Bagi setiap twit berulang, hanya satu rekod unik akan dikekalkan. Seterusnya diikuti dengan menyingkirkan perkataan yang tidak penting dengan berpandukan korpus koleksi slang bahasa Indonesia daripada https://www.kaggle.com/datasets/oswinrh/indonesian-stoplist. Korpus koleksi slang bahasa Indonesia digunakan serta menggunakan Stop words Indonesia untuk membuan bahasa Indonesia seperti “sih”, “nya”, “iya”, “nih”, “biar”, “tau”, “kayak”, “banget” dan lain-lain. Tujuan utama proses penyingkiran kata henti ini adalah mengurangi jumlah kata dalam sebuah dokumen yang akan mempengaruhi proses analisis seterusnya.

Oleh kerana data twit dalam kajian ini tidak mengandungi bahasa Inggeris sebaliknya twit ini mengandungi bahasa Indonesia, maka dalam proses ini menggunakan perpustakaan Python Sastrawi (Robbani, 2016) (Mas Diyasa et al., 2021) dalam proses stemming bahasa Indonesia. Python Sastrawi merupakan perpustakaan yang dapat mengubah kata Indonesia menjadi bentuk dasarnya.

Selepas menyingkirkan twit berulang dalam praprosesan, proses seterusnya menyingkirkan karakter yang tidak terlibat dalam kajian ini. Apabila melibatkan set data teks mentah, terdapat perkataan dan karakter yang tidak diperlukan dalam teks kajian ini seperti ruang kosong, penamat baris, simbol tanda baca dan digit, serta pemegang twit seperti “@EmilDardak” dan URL.

Menurut Mas Diyasa et al. (2021), teks dalam Bahasa Indonesia tidak boleh diklasifikasi menggunakan perpustakaan Texblob memandangkan Textblob menerima input dalam bahasa Inggeris. Oleh itu, dalam kajian ini melaksanakan penterjemahan set data dari bahasa Indonesia ke Bahasa Inggeris. Penterjemahan menggunakan berbagai sumber antaranya Bing Microsoft, Google Translate dan Kamus Dwibahasa Oxford Fajar. Oleh kerana dalam data twit ini bukan sahaja mengandungi Bahasa Indonesia sahaja bahkan ada twit campuran bahasa. Oleh itu, penterjemahan ini dilakukan bagi mempiawaikan keseluruhan teks twit bagi pelaksanaan proses label sentimen setelah selesai menterjemahkan twit ke Bahasa Inggeris.

Setelah selesai praprosesan serta terjemahan ke Bahasa Inggeris, aktiviti selanjutnya mengaplikasikan perpustakaan Texblob bagi tujuan mendapatkan skor polariti dan subjektiviti serta melabelkan mengikut sentimen. Perpustakaan Textblob digunakan untuk mengklasifikasikan tiga jenis sentimen iaitu positif, negatif, dan neutral (Mas Diyasa et al., 2021). Untuk mendapatkan klasifikasi tersebut, textblob melakukan pengiraan yang menghasilkan nilai polariti. Menurut Bose et al. (2020) skor bagi polariti jatuh di bawah lingkungan [-1 hingga 1] manakala subjektif jatuh di bawah lingkungan [0 hingga 1]. Berpandukan Bose et al. (2020), skor

dalam Jadual 2 dapat membantu dalam melabelkan polariti.

JADUAL 2. Skor lingkungan untuk polariti

| No | Skor | Sentimen |
|---|---|---|
| 1 | 0.0 hingga 1.0 | Positif |
| 2 | 0.0 | Neutral |
| 3 | -1.0 hingga 0.0 | Negatif |

## FASA PEMBELAJARAN MESIN

Analisis sentimen terdiri daripada komen, pendapat, atau ulasan yang diklasifikasi mengikut kategori seperti positif, negatif dan neutral. Analisis Sentimen terlibat dalam penyelesaian masalah dalam lapangan pembelajaran mesin. Kebanyakan kajian menggunakan satu set data berlabel dikenali sebagai set data latihan yang dilabelkan secara manual atau dilabel menggunakan teknik-teknik leksikon seperti Textblob.

Matriks kekeliruan merupakan cara menyampaikan prestasi klasifikasi yang digunakan untuk mencari ketepatan model. Teknik pelaporan Matriks Kekeliruan akan digunakan bagi mengetahui prestasi model terbaik selepas fasa latihan dilakukan. Nisbah digunakan dalam kajian ini ditetapkan kepada 80:20 sebagaimana yang dilaksana oleh (Khan et al., 2021). Jadual 3 merupakan hasil matriks kekeliruan untk gabungan pengekstrakan fitur dan pembelajaran mesin.

JADUAL 3. Matriks Kekeliruan

| Model | Ketepatan | Kelas | Precision | Recall | F-score | Support |
|---|---|---|---|---|---|---|
| SVM | 0.85 | Negatif | 0.87 | 0.62 | 0.73 | 1334 |
| | | Neutral | 0.80 | 0.98 | 0.88 | 3079 |
| | | Positif | 0.92 | 0.82 | 0.87 | 2958 |
| NB | 0.75 | Negatif | 0.61 | 0.60 | 0.61 | 1334 |
| | | Neutral | 0.84 | 0.71 | 0.77 | 3079 |
| | | Positif | 0.72 | 0.85 | 0.78 | 2958 |
| DT | 0.89 | Negatif | 0.80 | 0.74 | 0.77 | 1334 |
| | | Neutral | 0.92 | 0.96 | 0.94 | 3079 |
| | | Positif | 0.90 | 0.88 | 0.89 | 2958 |

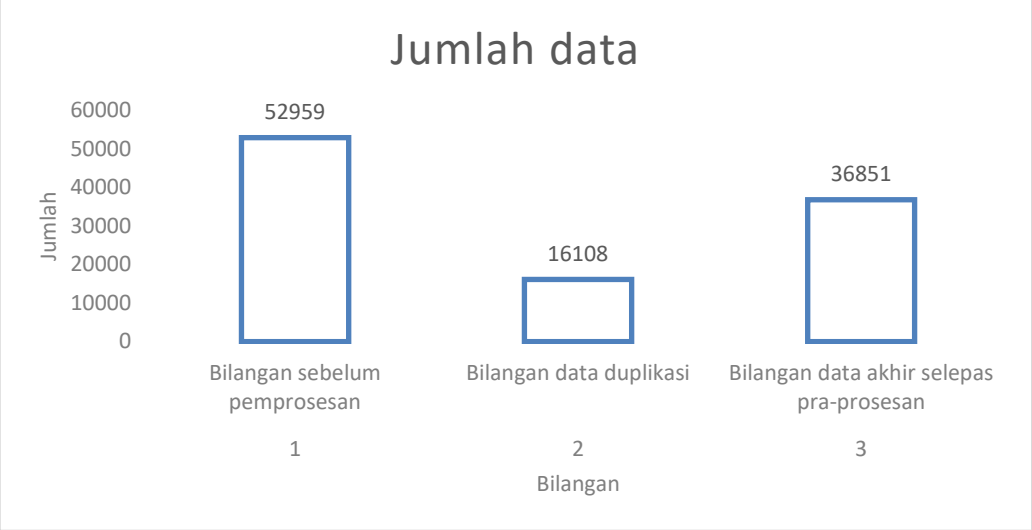
## HASIL KAJIAN

Dalam kajian analisis deskriptif dilakukan ke atas data Twitter COVID-19 yang telah dibersihkan bagi tujuan menganalisis trend dan membuat penilaian metrik. Kaedah ini menggunakan Python untuk menganalisis dan membentangkan data dalam bentuk taburan dan graf. Analisis deskriptif ini dapat divisualizasi data Twitter yang dikaji dalam format yang mudah difahami termasuk ciri-ciri data dan ringkasan berkenaan data tersebut. Di bawah terdapat koleksi analisis deskriptif yang dilakukan melalui penerokaan data.

## TWIT SEBELUM DAN SELEPAS PRAPROSESAN

Berdasarkan Rajah 2, bilangan data mentah yang terdapat dalam pangkalan data selepas proses dimuat turun data dari lama web Kaggle.com adalah sebanyak 52959 jumlah data sebelum prapemprosesan. Data mentah yang tinggal adalah 36851 sebaik sahaja prapemprosesan selesai dijalankan. Jumlah data yang tinggal selepas prapemprosesan menunjukkan bahawa terdapat bunyi atau noise dalam data mentah telah dikurangkan dengan kajian pembersihan seperti

penyingkiran duplikasi data.



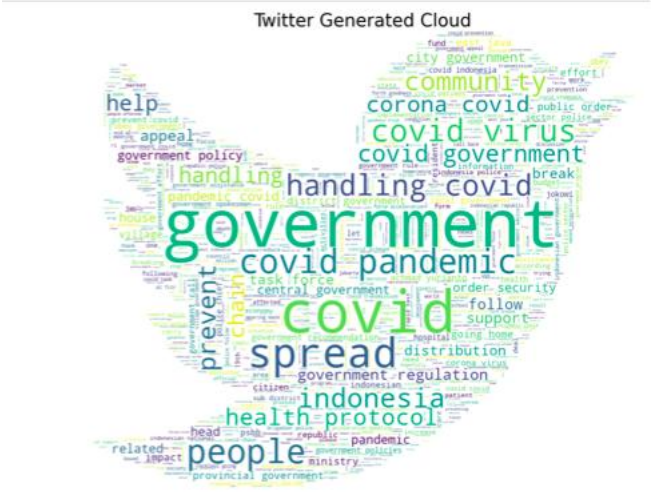

RAJAH 2. Jumlah Bilangan DataTwit

DATA TWIT KEKERAPAN PERKATAAN (WORD CLOUD)

Rajah 3 menunjukkan perkataan twit popular yang digunakan mengenai COVID-19 dalam teks Bahasa Indonesia sebelum dipraproses. Di mana perkataan "perintah" adalah perkataan paling popular digunakan mengenai COVID-19. Perkataan ini turut digambarkan dalam Word Cloud. Hasil daripada terjemahan twit ke Bahasa Inggeris, Rajah 3 menganalisa menggunakan aplikasi Word Cloud. Di mana perkataan "*government*" adalah perkataan paling popular dan kerap dibahas di dalam media sosial berkait COVID-19.

Bahasa Indonesia


Bahasa Inggeris




RAJAH 3. Word Cloud Ulasan Twit

KEKERAPAN TWIT BAGI KATA KUNCI (HASHTAG)

Rajah 4 menunjukkan senarai 10 tempat pertama kata kunci dengan jumlah tertinggi di dalam medan data terkumpul setelah data dibersihkan. Hasil daripada analisi ini, boleh dilihat bahawa

kekerapa pengguna twit menunjukan [] atau tiada array erti kata lain unsur indeks sifar. Ini menunjukkan bahawa sebanyak 26609 daripada jumlah keseluruhan pengguna sahaja tidak menggunakan kata kunci apabila bertwit. Jumlah kata kunci ini dikira dalam medan data setelah melakukan pembersihan data.

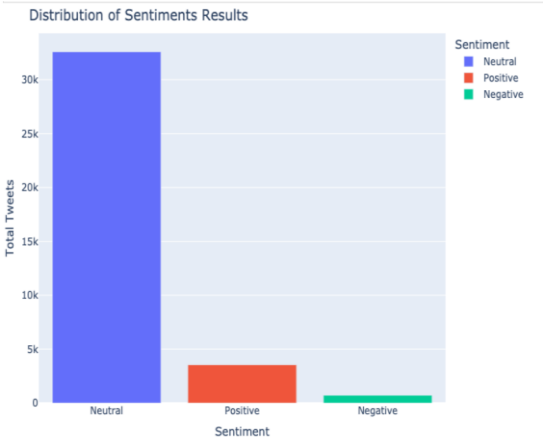

RAJAH 4. Bilangan kekerapan kata kunci dalam data twit

PERBANDINGAN POLARITI SENTIMEN INDONESIA DAN INGGERIS

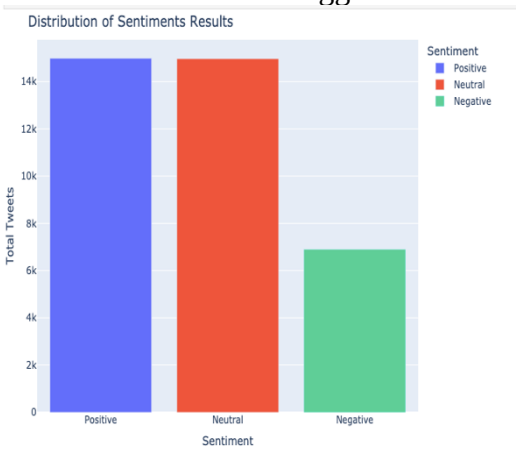
Rajah 5 adalah hasil data twit bahasa Inggeris setelah analisis menggunakan TextBlob menunjukkan bahawa kelas sentimen dengan jumlah tertinggi sebanyak 14987 kategori positif. Diikuti dengan neutral sebanyak 14967 manakala kelas sentimen mempunya jumlah paling rendah ialah negatif dengan sebanyak 6897.

Makala twit bahasa Indonesia menunjukkan bahawa polariti sentimen bagi set data twit dengan jumlah tertinggi sebanyak 32592 kategori neutral. Diikuti dengan positif sebanyak 3547 manakala polariti sentimen negatif hanya berbaki sebanyak 712 yang turut juga menggunakan TextBlob. Ini menunjukkan bahawa, pengguna Textblob ke atas Bahasa Indonesia tidak dapat dikelas dengan begitu tepat. Akibatnya, ulasan penggguna twit mengenai terhadap COVID-19 dan pemerintah mendapat hanya maklum balas neutral.

Twit Bahasa Indonesia




Twit Bahasa Inggeris




RAJAH 5. Perbandingan Polariti Sentimen

## PERBANDINGAN PEMBELAJARAN MODEL

Jadual 4 merujuk kepada laporan perbandingan keputusan antara model pembelajaran mesin, Ketepatan berdasarkan pengekstrakan fitur. Skor Ketepatan bagi gabungan SVM dan BOW adalah 84%, manakala bagi SVM dan TF-IDF adalah 88%. Skor Ketepatan bagi gabungan NB dan BOW mendapat 75% dan gabungan NB dan TF-IDF ialah 69%. Model DT mendapat skor Ketepatan sebanyak 86% dalam pendekatan TF-IDF manakala dalam pendekatan BOW mencapai sebanyak 89%.

JADUAL 4. Keputusan Pembelajaran Mesin

| Model | Ketepatan | Pengkstrakan fitur |
|---|---|---|
| SVM | 0.88 | TF-IDF |
| | 0.85 | BOW |
| NB | 0.69 | TF-IDF |
| | 0.75 | BOW |
| DT | 0.86 | TF-IDF |
| | 0.89 | BOW |

Hasil daripada pengujian menggunakan model dan pengekstrakan fitur (Jadual 4), dapat diperhatikan bahawa model DT dengan pendekatan BOW memberikan ketepatan yang tertinggi berbanding model yang lain. Jika diperhatikan dengan lebih mendalam, model DT dan SVM menghampiri sama tinggi ketepatannya. Ini bermakna model SVM turut mempunyai ketepatan ramalan kelas sentimen yang baik setanding model pembelajaran mesin yang lain.

## KESIMPULAN

Hasil kajian menunjukkan model DT dengan pendekatan BOW memberi bacaan ketepatan yang tertinggi berbanding model yang lain. Menunjukkan dalam eksperimen ini, BOW lebih baik berbanding TF-IDF. Kerana dalam pemprosesan bahasa semulajadi dan pengambilan maklumat, BOW adalah teknik mudah untuk mengekstrak ciri daripada teks atau data yang dipermudahkan. BOW mengira bilangan kali perkataan muncul dalam teks dan mencipta vektor ciri yang mengandungi bilangan kali setiap perkataan unik muncul. BOW juga digunakan terutamanya untuk memperluaskan vokal kata semua perkataan unik dan melatih model pembelajaran berdasarkan kekerapan mereka. Akhirnya kaedah berasaskan DT boleh menghasilkan ketepatan yang stabil pada beberapa topik bagi data mengenai COVID-19 Indonesia.

Penemuan dari kata kunci yang popular membantu memberikan pemahaman yang lebih baik mengenai apa yang diulaskan. Model pembelajaran mesin DT dapat dilihat bahawa analisis sentimen dalam membuat keputusan dan pengurusan ketika pandemik COVID-19.Data mentah untuk kajian ini meggunakan perkataan seperti campuran bahasa Inggeris serta bahasa pasar. Perkataan ini akan menyebabkan mendapat nilai ketepatan tidak tepat serta akan ada kesilapan ketika dalam praprosesan. Menggunakan praprosesan serta teknik data klasifikasi mengikut polariti sentimen dan penerokaan trend perbincangan pengguna dapat dijadikan rujukan untuk penyelidikan dan menyiasatan terhadap setip pandang dan ulasan pengguna terhadap pemerintah. Dari hasil eksperimen, teknik pengektrakan fitur TF-IDF dan BOW membantu peningkatan klasifikasi sentimen berdasarkan perbandingan model yang telah dilaksanakan. Klasifikasi Decision Tree merupakan algoritma klasifikasi sentimen twit terbaik ke atas set data twit COVID-19 berbanding Naïve Bayes dan Support Vector Machine.

Sumbangan kajian ini boleh dimanfaat untuk kajian masa hadapan ialah menyediakan data berlabel untuk penyelidikan analisis sentimen COVID-19 dan twit. Kajian ini turut membantu kerajaan semasa dengan menganalisa data media sosial berkait COVID-19 dan meramal sama ada pelaksanaan sesuatu perintah drastik wajar dilaksanakan. Contoh, adakah perlu melaksanakan Perintah Kawalan Pergerakan semula menjelang musim perayaan. Kajian ini berjaya membangunkan teknik terkini bagi meningkatkan prestasi pengelasan sentimen melalu kaedah penggabungan leksikon dan pembelajaran mesin.

Walau bagaimanapun, masih terdapat banyak ruang untuk penambahbaikan dan pengembangan dalam bidang penyelidikan ini. Antaranya set data asal merupakan data twit ulasan dalam bahasa Indonesia. Tetapi dalam kajian ini memfokuskan kepada set data twit yang ditulis dalam bahasa Inggeris. Kajian lanjutan menggunakan set data ulasan pengguna bukan sahaja bahasa Indonesia juga bahasa Malaysia boleh dijadikan sebagai penyelidikan pada masa akan datang. Selain daripada polariti, satu lagi cabang analisis sentimen adalah untuk mengkategorikan emosi twit kepada kepada meramal perasaan gembira, sedih, takut, atau marah. Kajian ini berpotensi ditambah baik dengan menggunakan Deep Learning untuk menganalisa sentimen semasa peningkatan kes COVID-19.

## PENGHARGAAN

Kajian ini merakamkan ucapan terima kasih kepada Universiti Kebangsaan Malaysia dan Kementerian Pengajian Tinggi atas sokongan di dalam kajian ini menerusi geran penyelidikan Fundamental Research Grant Scheme dengan kod FRGS/1/2022/ICT02/UKM/02/7.

## RUJUKAN

Abubakar, H. D., & Umar, M. 2022. Sentiment Classification: Review of Text Vectorization Methods: Bag of Words, Tf-Idf, Word2vec and Doc2vec. *SLU Journal of Science and Technology*.

Afdhalul Ihsan, E. R. 2021. Optimization Of K-Nearest Neighbour to Categorize Indonesian's News Articles. *Asia-Pacific Journal of Information Technology and Multimedia*, *10*(1), 43-51. https://doi.org/dx.doi.org/10.17576/apjitm-2021-1001-04.

Astuti, P., & Mahardhika, A. 2020. COVID-19: How does it impact to the Indonesian economy? *Jurnal Inovasi Ekonomi*, *5*. https://doi.org/10.22219/jiko.v5i3.11751.

Bose, R., Aithal, S., & Roy, S. 2020. Sentiment Analysis on the Basis of Tweeter Comments of Application of Drugs by Customary Language Toolkit and TextBlob Opinions of Distinct Countries. *International Journal of Emerging Trends in Engineering Research*, *8*, 3684-3696. https://doi.org/10.30534/ijeter/2020/129872020.

Ezuana Sukawai, N. O. 2020. Corpus Development for Malay Sentiment Analysis Using Semi Supervised Approach. *Asia-Pacific Journal of Information Technology and Multimedia*, *9*(1), 94-109. https://doi.org/dx.doi.org/10.17576/apjitm-2020-0901-08.

Gellerman, O. B. a. P. (2021). *Multi-Class Text Classification of Naturalistic Driving Log Annotations* Chalmers University of Technology]. https://odr.chalmers.se/bitstream/20.500.12380/302526/1/GellermanBrask_2021.pdf.

Ghallab, A., Mohsen, A., & Ali, Y. 2020. Arabic Sentiment Analysis: A Systematic Literature Review. *Applied Computational Intelligence and Soft Computing*, *2020*(1), 7403128. https://doi.org/https://doi.org/10.1155/2020/7403128.

Gorbiano, M. I. 2020. BREAKING: Jokowi bans 'mudik' as Ramadan approaches. *The Jakarta Post*. https://www.thejakartapost.com/news/2020/04/21/breaking-jokowi-bans-mudik-

as-ramadan-approaches.html.

Gupta, H. 2020. *Text classification of reviews using machine learning models, review summarization and building content-based hotel recommender system* Trinity College Dublin]. https://publications.scss.tcd.ie/theses/diss/2020/TCD-SCSS-DISSERTATION-2020-066.pdf.

Haque, R., Pareek, P. K., Islam, M. B., Aziz, F. I., Amarth, S. D., & Khushbu, K. G. 2023. Improving Drug Review Categorization Using Sentiment Analysis and Machine Learning. 2023 International Conference on Data Science and Network Security (ICDSNS).

Harsha Kadam, S. a. P., K. 2020. *Text analysis for email multi label classification* Chalmers University of Technology and University of Gothenburg].

Haryadi, D., Atmaja, D. M. U., Hakim, A. R., & Witanti, W. 2022. Classification of Drug Effectiveness Based on Patient's Condition Using Text Mining With K-Nearest Neighbor. 2022 International Conference on ICT for Smart Society (ICISS).

Ismail, S., Ariffin, H. S. a., & Abdullah, Z. H. 2020. Media Sosial Semasa Pandemik. In: Universiti Sains Islam Malaysia.

Kalaivani, K. S., Uma, S., & Kanimozhiselvi, C. S. 2020. A Review on Feature Extraction Techniques for Sentiment Classification. 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC).

Kalanjati, V., Hasanatuludhhiyah, N., d'Arqom, A., Arsyi, D., Marchianti, A., Muhammad, A., & Purwitasari, D. 2024) Sentiment analysis of Indonesian tweets on COVID-19 and COVID-19 vaccinations [version 4; peer review: 1 approved, 2 approved with reservations]. *F1000Research*, *12*(1007). https://doi.org/10.12688/f1000research.130610.4.

Kaur, R., & Ranjan, S. 2020. Sentiment Analysis of 21 daysCOVID-19 Indian lockdown tweets.

Kennedy, P. S. J. a. S., Timothy Wisnu Harya P. and Tampubolon, Emma and Fakhriansyah, Muhammad. 2020. ANALISIS STRATEGI LOCKDOWN ATAU PEMBATASAN SOSIAL DALAM MENGHAMBAT PENYEBARAN COVID-19. *Jurnal Riset Manajemen, 9*, 48-64. https://ejournal.upi.edu/index.php/image/article/view/24189.

Khan, R., Rustam, F., Kanwal, K., Mehmood, A., & Choi, G. S. 2021. US Based COVID-19 Tweets Sentiment Analysis Using TextBlob and Supervised Machine Learning Algorithms. 2021 International Conference on Artificial Intelligence (ICAI).

Lee, E., Rustam, F., Shahzad, H.-F., Washington, P.-B., Ishaq, A., & Ashraf, I. (2023). Drug Usage Safety from Drug Reviews with Hybrid Machine Learning Approach. *Computer Systems Science and Engineering*, *45*(3), 3053--3077. http://www.techscience.com/csse/v45n3/50726.

Li, J., & Hovy, E. 2017. Reflections on Sentiment/Opinion Analysis. In E. Cambria, D. Das, S. Bandyopadhyay, & A. Feraco (Eds.), *A Practical Guide to Sentiment Analysis* (pp. 41-59). Springer International Publishing. https://doi.org/10.1007/978-3-319-55394-8_3

Liaqat MI, A. H. M., Shoaib M, Khurshid SK, Shamseldin MA. 2022. Sentiment analysis techniques, challenges, and opportunities: Urdu language-based analytical study. *PeerJ Computer Science, 8*(e1032). https://doi.org/10.7717/peerj-cs.1032.

Madhoushi, Z., Hamdan, A. R., & Zainudin, S. 2023. Semi-Supervised Model for Aspect Sentiment Detection. *Information*, *14*(5), 293. https://www.mdpi.com/2078-2489/14/5/293.

Malaysia, K. K. (2023). *Soalan Lazim (FAQ) Rasmi Berkaitan COVID-19* Retrieved 19 Mac 2024 from https://COVID-19.moh.gov.my/faqsop/faq-COVID-19-kkm.

Mas Diyasa, I. G. S., Marini Mandenni, N. M. I., Fachrurrozi, M. I., Pradika, S. I., Nur Manab, K. R., & Sasmita, N. R. 2021. Twitter Sentiment Analysis as an Evaluation and Service Base On Python Textblob. *IOP Conference Series: Materials Science and Engineering*, *1125*(1), 012034. https://doi.org/10.1088/1757-899X/1125/1/012034.

Qi, Z. 2020. The Text Classification of Theft Crime Based on TF-IDF and XGBoost Model. 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA).

Robbani, H. A. 2016. *Sastrawi 1.0.1*. https://pypi.org/project/Sastrawi/.

Rosly, M. A. B. M., Zainudin, S., & Kassim, J. M. 2023. Analysing Imbalanced Dataset for Postgraduate Student Dropout Using Predictive Analytics. 2023 International Conference on Electrical Engineering and Informatics (ICEEI),

Singh, A., Srivastava, H., Aman, M., & Dubey, G. 2023. Sentiment Analysis on User Feedback of a Social Media Platform. 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS),

Suhaidi, M., Kadir, R. A., & Tiun, S. 2021. A Review Of Feature Extraction Methods On Machine Learning. *Journal of Information System and Technology Management,, 6*(22), 51-59. https://doi.org/10.35631/JISTM.622005.

Suhud, R., Febriandirza, A., Permatasari, I., & Ramadan, F. 2023. Recognizing Public Satisfaction Toward Kampus Mengajar Program with Naive Bayes. 2023 International Conference on Computer, Control, Informatics and its Applications (IC3INA).

Sutabri, T., Suryatno, A., Setiadi, D., & Negara, E. S. 2018. Improving Naïve Bayes in Sentiment Analysis For Hotel Industry in Indonesia. 2018 Third International Conference on Informatics and Computing (ICIC).

Utami, P. D. *Analisis Sentimen Review Kosmetik Bahasa Indonesia Menggunakan Algoritma Naïve Bayes Classifier.* Politeknik Negeri Jakarta. https://osf.io/3pjdy/download.

Wicaksana, C. A., Fatkhurrokhman, M., Pratama, H. P., Tryawan, R., Alimuddin, & Febriani, R. 2022. Twitter Sentiment Analysis in Indonesian Language using Naive Bayes Classification Method. 2022 International Conference on Informatics Electrical and Electronics (ICIEE).

Zohreh Madhoushi, A. R. H., Zainudin, S. 2019. Aspect-Based Sentiment Analysis Methods in Recent Years. *Asia-Pacific Journal of Information Technology and Multimedia*, *7*(2), 79-96. https://doi.org/dx.doi.org/10.17576/apjitm-2019-0801-07.

Zulfa, I., & Winarko, E. 2017. Sentimen Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems); Vol 11, No 2 (2017): JulyDO - 10.22146/ijccs.24716.* https://jurnal.ugm.ac.id/ijccs/article/view/24716.