



TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN OPINI MASYARAKAT TERHADAP TEKANAN MENTAL PADA MEDIA SOSIAL TWITTER

Dhira Atika¹, Styawati², Ahmad Ari Aldino³

Sistem Informasi^{1,2,3}, Fakultas Teknik dan Ilmu Komputer Universitas Teknokrat Indonesia
JL.ZA. Pagar Alam No. 9-11, Labuhan Ratu, Kec. Kedaton, BandarLampung^{1,2,3}

Email: dhiraatika185@gmail.com¹, styawati@teknokrat.ac.id², aldino@teknokrat.ac.id³

Received: (24 Januari 2022) Accepted: (30 Desember 2022) Published: (30 Desember 2022)

Abstract

Twitter is a medium of information available to all Internet users. Indonesia has 19.5 million Twitter users from 500 million worldwide and continues to grow from time to time. The use of Twitter by the government uses dalma to carry out large social programs (PSBB) related to the travel, school and business sectors. In psychology, mental stress can be caused by several factors, which can be biological, genetic or environmental. The mental pressure that is felt starts from worrying about health, hoax information and mental pressure related to income and income. Twitter's tweets about mental stress generated a lot of public opinion. Comments on Twitter social media users' tweets will be used as research data. Based on tweets made by Twitter social media users, this study will seek public opinion about the mental stress of Twitter users in Indonesia. One of the objectives of this study is to determine the level of mental stress of Twitter users in Indonesia with 2 classifications, namely being exposed to mental stress and not being exposed to mental stress. To facilitate the process of classifying Twitter tweet data, a sentiment analysis is needed. The technique for classifying the sentiment analysis of an algorithm in this study uses the Support Vector Machine (SVM) algorithm with a linear kernel, which is combined with the TF-IDF extraction feature. The validity test applied in this study used a confusion matrix. The use of TF-IDF feature extraction and the Support Vector Machine (SVM) method is able to perform an accuracy value of 99.34%, meaning that the value is classified as good or classified as good.

Keywords: *Sentiment Analysis, TF-IDF, Support Vector Machine (SVM).*

Abstrak

Twitter adalah media informasi yang tersedia untuk semua pengguna Internet. Indonesia memiliki 19,5 juta pengguna Twitter dari 500 juta di seluruh dunia dan terus bertambah dari waktu ke waktu. Penggunaan Twitter oleh pemerintah dipergunakan dalma melakukan kebijakan pembatasan sosial berskala besar (PSBB) berkaitan dengan sektor perjalanan, sekolah dan usaha. Dalam ilmu psikologi tekanan mental dapat terjadi disebabkan oleh beberapa faktor bisa dari faktor biologis, genetik atau lingkungan. Tekanan mental yang dirasakan mulai dari khawatir terhadap kesehatan, informasi hoax dan tekanan mental yang terkait dengan penghasilan dan pendapatan. tweet Twitter tentang tekanan mental menghasilkan banyak pendapat masyarakat. Komentar tweet pengguna media sosial Twitter akan dapat dijadikan data penelitian. Berdasarkan tweet yang dilakukan pengguna media sosial Twitter pada penelitian ini akan mencari opini masyarakat tentang tekanan mental pengguna Twitter di Indonesia. Salah satu tujuan dari penelitian ini adalah mengetahui tingkat tekanan mental masyarakat pengguna Twitter di Indonesia dengan 2 klasifikasi yaitu terkena tekanan mental dan tidak terkena tekanan mental. Untuk memudahkan proses pengklasifikasian Data tweet Twitter dibutuhkan suatu sentimen analisis. Teknik untuk melakukan klasifikasi pada analisis sentimen diperlukan suatu algoritma dalam penelitian ini menggunakan algoritma Support Vector Machine (SVM) dengan kernel linier, yang dikombinasikan dengan fitur ekstraksi TF-IDF. Uji validitas yang diterapkan pada penelitian ini menggunakan matrik konfusi. Penggunaan ekstraksi fitur TF-IDF dan

metode Support Vector Machine (SVM) mampu melakukan nilai accuracy sebesar 99,34 % artinya bahwa nilai tersebut good classification atau klasifikasi baik.

Kata Kunci: Analisis Sentimen, *TF-IDF*, *Support Vector Machine (SVM)*.

To cite this article:

Dhira Atika, Styawati, Ahmad Ari Aldino. (2022). TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN OPINI MASYARAKAT TERHADAP TEKANAN MENTAL PADA MEDIA SOSIAL TWITTER. Jurnal Teknologi dan Sistem Informasi, Vol (1), 86-97.

1. Pendahuluan

Pada masa *Pandemic Covid-19* di Indonesia mendeteksi kasus pertama pada awal bulan Maret tahun 2020. Pada bulan Mei 2021, sekitar 1,7 juta kasus dan \pm 46 ribu kasus kematian dicatatkan. Pemerintah melakukan kebijakan Pembatasan Sosial Berskala Besar (PSBB) berkaitan dengan sektor perjalanan, sekolah dan usaha. Pandemi ini berdampak juga terhadap enam aspek utama yaitu (1) Kemiskinan, (2) Pembelajaran, (3) Kesehatan, (4) Kesehatan Mental, (5) Gizi dan (6) Akses ke layanan umum (Unicef, 2021).

Pemerintah Indonesia melakukan tindapanan merespon kecemasan masyarakat Indonesia terkait dampak dari Pandemi Covid-19, Pemerintah Indonesia melakukan perluasan ruang fiscal dalam menghadapi keadaan darurat dan mendukung pemulihan sosial dan ekonomi Indonesia. Kebijakan yang dilakukan adalah (1) Paket Stimulus fiscal yang dianggarkan tahun 2020 mencapai Rp695 triliun (USD\$49,6 miliar) atau 4,3% dari PDB, (2) Melakukan program perlindungan sosial dalam bentuk bantuan tunai atau nontunai untuk memenuhi kebutuhan sehari-hari, (3) Melakukan pembelajaran jarak jauh atau daring dan (4) Melakukan upaya pencegahan, pemetaan dan upaya sosialisasi terhadap kesehatan mental masyarakat (Unicef, 2021).

Berdasarkan hasil survei Center for Human Capital Development (CHCD) menyatakan bahwa 80% responden mengalami gejala tekanan mental mulai dari tingkat ringan hingga berat selama pandemik berlangsung (Arriani et al., 2020). Dalam ilmu Psikologi tekanan mental dapat terjadi disebabkan oleh beberapa faktor bisa dari faktor biologis, genetik atau lingkungan. Tekanan mental yang dirasakan mulai dari khawatir terhadap kesehatan, informasi hoax dan tekanan mental yang terkait dengan penghasilan dan pendapatan. Menurut data dari Badan Pusat Statistik, Produk domestik bruto (PDB) Indonesia yang mengalami penurunan sekitar 2,1% pada tahun 2020, hingga mengakibatkan hilangnya pekerjaan dalam skala nasional dan angka pengangguran meningkat 7,07 % pada tahun 2020 (BPS RI., 2020).

Menurut Kementerian Komunikasi dan Informatika Republik Indoensia penggunaan media sosial di Indonesia mencapai peringkat ketiga di Asia.

Dari 245 juta penduduk Indonesia sekitar 55 juta pengguna media sosial. Dibeberapa kelompok tingkat masyarakat yang paham penggunaan media sosial menyalurkan keluh kesahnya akan tekanan mental yang dialaminya dengan menuliskan di media sosial agar mendapatkan sebuah dukungan dari beberapa pengguna media sosial lainnya. Media yang sering dipakai masyarakat Indonesia adalah Instagram, Facebook, WhatsApp dan Twitter (Aditya et al., 2020).

Twitter merupakan media sosial yang dibuat pada tahun 2006 dan sampai saat ini telah digunakan lebih dari 217 juta pengguna diseluruh dunia per kuartal-IV 2021 berdasarkan laporan keuangan untuk kuartal-IV (Q4) tahun fiskal 2021 (Riyanto, 2021). Menurut Kementerian Komunikasi dan Informatika Republik Indoensia tercatat ada sekitar \pm 19,5 juta pengguna media sosial Twitter (Kominfo, 2013). Twitter memiliki salah satu fitur yaitu tweet. Tweet merupakan tulisan yang dapat dibuat oleh pengguna dengan panjang karakter maksimal 280 karakter dalam satu kali tweet dan dapat membagikan kepada pengguna lainnya. Tweet dapat berupa teks, foto atau video. Twitter sendiri selalu up to date terhadap topik yang sedang banyak dibicarakan oleh penggunanya. Informasi yang terdapat di tweet berupa opini yang bersifat positif dan negative (Yan et al., 2019).

Dari satu topik pembicaraan di Twitter tentang tekanan mental menghasilkan banyak pendapat masyarakat. Komentar tweet pengguna media sosial Twitter akan dapat dijadikan data penelitian. Berdasarkan tweet yang dilakukan pengguna media sosial Twitter pada penelitian ini akan mencari opini masyarakat tentang tekanan mental pengguna Twitter di Indonesia. Salah satu tujuan dari penelitian ini adalah mengetahui tingkat tekanan mental masyarakat pengguna Twitter di Indonesia dengan 2 klasifikasi yaitu terkena tekanan mental dan tidak terkena tekanan mental. Untuk menyelesaikan masalah yang sudah dijabarkan diatas maka peneliti melakukan studi literatur pada penelitian sebelumnya yang dilakukan oleh (Yan et al., 2019) melakukan penelitian yang berjudul "Analisis Sentimen Komentar Netizen Twitter Terhadap Kesehatan Mental Masyarakat Indonesia yang menghasilkan analisis sentiment Kesehatan mental adalah negative dengan klasifikasi negatif sebesar 50,8% dengan menggunakan metode Naïve Bayes".

Dalam penelitian ini, berfokus pada tweet atau opini masyarakat terhadap tekanan mental dengan analisis sentimen. Penelitian ini bertujuan untuk mengetahui sentiment masyarakat terhadap tekanan mental pada pengguna media sosial Twitter menggunakan metode klasifikasi Support Vector Machine (SVM). SVM adalah metode klasifikasi jenis supervised karena ketika proses training diperlukan target pembelajaran tertentu. SVM merupakan algoritma yang bekerja menggunakan pemetaan nonlinier untuk mengubah data penelitian ke dimensi yang lebih tinggi (Handayanto & Herlawati, 2020), sebelum dilakukan klasifikasi menggunakan SVM, data harus dilakukan pembobotan (Styawati & Mustofa, 2019). Metode Term Frequency – Inverse Document Frequency (TF-IDF) adalah suatu proses untuk melakukan perubahan data dari data teks ke dalam data numeric untuk melakukan proses pembobotan pada tiap kata dan untuk mencari nilai kata pada setiap dokumen tweet yang dihasilkan (Aditya et al., 2020). Dari penjelasan diatas, maka penelitian ini berjudul “Term Frequency – Inverse Document Frequency Support Vector Machine untuk Analisis Sentimen Opini Masyarakat terhadap Tekanan Mental pada Media Sosial Twitter”.

2. Tinjauan Pustaka

Pada bab ini memuat penelitian terdahulu yang berguna sebagai bahan literatur penulis untuk melakukan penelitian. Berikut adalah penelitian yang sudah dilakukan sebelumnya sebagai literatur Pustaka.

1. Penelitian oleh (Aulia & Amelia, 2020) yang berjudul “Analisis Sentimen Twitter Pada Isu Mental Health Dengan Algoritma Klasifikasi Naive Bayes” penelitian ini mengangkat masalah “Penelitian ini bertujuan menerapkan algoritma naïve bayes dalam proses klasifikasi data tweets terkait isu kesehatan mental. bertujuan mencari tahu bagaimana stigma yang berkembang di masyarakat mengenai isu kesehatan mental. Pencarian opini dilakukan dengan mencari data tweets dari Twitter menggunakan Twitterscraper. Data diklasifikasikan dengan algoritma Naive Bayes untuk mengelompokkan sentimen dan emosi. Hasil eksperimen menunjukkan bahwa sentimen positif yang mendominasi”.
2. Penelitian oleh (Mutmainah, 2022) dengan judul “Kemungkinan Depresi Dari Postingan Pada Media Sosial” penelitian ini mengangkat permasalahan “Analisis kemungkinan depresi dari postingan media sosial menggunakan teknik Natural Language Processing (NLP) Pemodelan dilakukan dengan teknik klasifikasi yang sering dipakai dalam sentiment analisis dan pada kasus yang sama yaitu prediksi depresi. hasil terbaik didapatkan oleh metode SVM dengan akurasi 95.56%, menggunakan metode SVM dalam sentiment analisis dan analisis depresi. Pemodelan menggunakan metode Neural network akurasinya 93.79%”.
3. Penelitian oleh (Yan et al., 2019) dengan judul “Analisis Sentimen Komentar Netizen Twitter Terhadap Kesehatan Mental Masyarakat Indonesia” penelitian ini mengangkat permasalahan “Hasil yang diperoleh setelah menganalisis sentimen komentar pengguna media sosial Twitter terhadap kesehatan mental masyarakat Indonesia adalah dari 2.369 data komentar yang telah dianalisis, sebanyak 50,8% negatif, 45,1% positif, dan 4,1% netral. Jadi, dapat disimpulkan bahwa analisis sentimen komentar pengguna media sosial terhadap kesehatan mental masyarakat Indonesia cenderung negatif. Metode Naïve Bayes digunakan saat melakukan analisis sentimen dan hasil akurasinya adalah 0.7961165048543689.
4. Penelitian oleh (Neighbor et al., 2021) dengan judul “Sentiment Analysis of Depression Detection on Twitter Social Media Users Using the K-Nearest Neighbor Method” penelitian ini mengangkat “Pada penelitian kali ini penulis bertujuan untuk menganalisis penerapan dari metode K- Nearest Neighbor dalam mendeteksi tingkat depresi pengguna media sosial Twitter dan melihat nilai akurasinya. Berdasarkan pengujian pada klasifikasi KNN menggunakan tahapan confusion matrix hasil akurasi yang didapatkan sebesar 78.18%”
5. Penelitian oleh (Nugroho et al., 2021) dengan judul “Deteksi Depresi Dan Kecemasan Pengguna Twitter Menggunakan Bidirectional Lstm” penelitian ini mengangkat “Mendeteksi pengguna dengan potensi gangguan depresi dan kecemasan melalui data tekstual tidaklah mudah karena mereka tidak secara eksplisit berbicara tentang kondisi mentalnya. Hal ini dapat dicapai dengan teknik klasifikasi teks. Salah satu pendekatan yang dapat digunakan adalah LSTM. Berdasarkan hasil pengujian, akurasi tertinggi yang diperoleh BiLSTM mencapai 94.12%. Penelitian ini telah berhasil mengembangkan model untuk deteksi depresi dan kecemasan pengguna twitter.”

2.2 Mental

Menurut (Chaplin, 1989) Mental adalah hal yang berkaitan pikiran, akal, ingatan atau proses yang berasosiasi dengan kesadaran manusia. Sedangkan Tekanan mental dapat diartikan sebagai Depresi yang secara medis berdampak negative pada perasaan atau pikiran manusia. Tekanan mental dapat terjadi lebih dari banyak faktor salah satunya adalah perasaan tertekan akan keadaan lingkungan yang sedang dialami serta perasaan sedih yang berlebihan.

2.3 Sentimen Analisis

Sentimen Analisis merupakan proses untuk menjabarkan sentimen yang terjadi sehingga dapat dikelompokkan dari data sumber seperti text ke dalam dokumen atau kalimat sehingga dapat dikategorikan sebagai sentiment positif, netral dan negative (Yan et al., 2019b). Analisis Sentimen atau analisis pendapat merupakan suatu pembelajaran untuk melakukan analisis pendapat setiap individu terhadap sebuah topik yang terjadi. Secara garis besar analisis sentiment melihat pendapat setiap individu menjadi dua jenis yaitu sentiment positif dan negatif (Aditya et al., 2020). Untuk mengukur sentiment analisis dengan berbagai algoritma, antara lain algoritma *Naïve Bayes*, *Support Vector Machine*, *K-Nearest Neighbour*, *Decision Tree*, *K-Medoid* dan *Backpropagation Neural Network* (Giovani et al., 2020). Analisis sentiment ini dapat dipergunakan untuk melihat respon masyarakat terhadap bidang ekonomi, sosial, politik dan agama. Terdapat 3 tingkatan pada Analisis Sentimen yakni (Firmansyah Sulaeman et al., 2019) :

1. Level Dokumen
2. Level Kalimat
3. Level Aspek

2.4 Twitter

Menurut Wikipedia Twitter dirilis pertama kali pada tahun 2006 oleh Jask Dorsey yang berbasis di San Francisco, California. Twitter merupakan layanan jejaring sosial yang memungkinkan pengguna untuk mengirimkan dan membaca pesan berbasis teks hingga 280 karakter yang dikenal dengan sebutan *tweet*. Pada tahun 2022 saham Twitter dibeli oleh Elon Musk. Dalam Twitter terdapat beberapa fitur yaitu dikutip dari (Aditya et al., 2020) yaitu :

1. *Tweet*, adalah sebuah kicauan yang ditulis oleh pengguna Twitter.
2. *Hashtag*, adalah sebuah penanda yang diberikan oleh pengguna Twitter untuk mengklasifikasikan sebuah topik untuk sebuah kejadian di kehidupan sehari-hari.

3. *Trending Topic*, adalah topik yang sering dibicarakan oleh orang yang saling bergantian secara realtime, sehingga *Trending Topic* akan selalu berada di paling atas.
4. *Retweet*, adalah mengirimkan kembali *tweet* dari pengguna lain.
5. *Mention*, adalah cara untuk menandai orang dengan menggunakan tanda @
6. *Reply*, merupakan cara membalas *tweet* dari pemilik akun Twitter lainnya.

2.5 Crawling

Proses di mana mesin pencari mengirimkan tim robot (*crawler* atau *spider*) untuk menemukan konten baru atau yang diperbarui. Konten yang dimaksud dapat berupa halaman web, gambar, video, dokumen, dll (Rofiqoh et al., 2017). Alat yang digunakan dalam *web crawling* juga dikenal sebagai *web spider*. Alat ini menemukan dan mengindeks konten terkini dalam bentuk database URL dengan mengidentifikasi dan merekam semua tautan yang ditemukan pada halaman yang dipindai. Ketika datang ke perayapan, perayap web mulai bekerja berdasarkan daftar tautan halaman yang dipindai sebelumnya dari peta situs. Dari daftar tautan, temukan lebih banyak tautan yang berserakan. Pada titik ini, ia merangkak ke tautan terbaru dan mengulangi prosesnya dengan tautan berikutnya. Data memiliki 3 jenis data yaitu Terstruktur, Semi Terstruktur dan Tidak Terstruktur. Dalam mendapatkan data di internet perlu menggunakan teknik *Crawling data* yaitu proses mengambil data dari sebuah website dengan bantuan Bahasa pemrograman tertentu. Pada penelitian ini *Crawling data* akan mengambil data di Twitter dengan memasukkan kata kunci untuk mencari *tweet* yang diinginkan.

2.6 Preprocessing

Data *preprocessing* adalah salah satu fase dari data mining. sebelum memasuki tahap pengolahan. Data set diproses terlebih dahulu. Data *preprocessing* biasanya dilakukan dengan membuang data yang tidak sesuai. Proses ini juga dapat mengubah data menjadi format yang lebih mudah dipahami oleh sistem (Malini et al., 2022).

Preprocessing data merupakan langkah untuk menghilangkan beberapa masalah yang dapat mengganggu pengolahan data. Ini karena banyak data yang diformat secara tidak konsisten. *Preprocessing* data adalah teknik tertua yang mendahului data mining. Namun, data *preprocessing* melibatkan beberapa proses seperti pembersihan, konsolidasi, transformasi, dan pengurangan data. *Preprocessing* data memungkinkan proses penambangan berjalan lebih

efektif dan efisien. Hal ini dikarenakan data preprocessing merupakan data yang telah mengalami beberapa tahap penyempurnaan. Tahapan dalam *preprocessing* adalah:

1. *Data Cleaning*

Langkah pertama yang harus dilakukan saat melakukan *preprocessing* data adalah membersihkan data. Artinya data mentah yang diperoleh harus diseleksi kembali. Kemudian hapus atau hilangkan data yang tidak lengkap, tidak relevan dan tidak akurat. Prosedur ini membantu menghindari kesalahpahaman saat menganalisis data

2. *Data Integration*

Karena data preprocessing akan menggabungkan beberapa data dalam suatu dataset, maka kita harus mengecek data-data yang datang dari berbagai sumber tersebut supaya memiliki format yang sama.

3. *Data Transformasi*

Proses berikutnya yang harus dilakukan adalah transformasi data. Seperti yang telah dijelaskan di atas, data akan diambil dari berbagai sumber yang kemungkinan memiliki perbedaan format. Kita harus menyamakan seluruh data yang terkumpul supaya dapat mempermudah proses analisis data.

4. *Data Reduction*

Tahap terakhir yang perlu dilakukan adalah mengurangi jumlah data (*data reduction*). Maksudnya adalah kita harus mengurangi sampel data yang diambil, tetapi dengan catatan, tidak akan mengubah hasil analisis data.

2.7 *Machine Learning*

Machine Learning merupakan sebuah cabang ilmu dari *Artificial Intelligence* dengan menggunakan pemrograman komputer dalam menjadikan proses secara otomatis dengan berperilaku seperti manusia dan dapat meningkatkan pemahamannya melalui pembelajaran pengalaman secara otomatis. *Machine Learning* memiliki fokus pada pengembangan sistem yang dapat melakukan belajar secara mandiri untuk menarik kesimpulan dari suatu masalah tanpa harus melakukan perintah berulang kali oleh manusia. *Machine Learning* bekerja dengan menggunakan data sebagai masukan untuk dianalisis terhadap kumpulan data sehingga menemukan pola yang tersembunyi. Didalam *Machine Learning* dikenal data training dan data testing, data training merupakan data untuk melatih algoritma *Machine Learning* dan data testing adalah data yang digunakan untuk

mengetahui performa dari algoritma dalam *Machine Learning* yang telah dilakukan pelatihan. (Retnoningsih & Pramudita, 2020). Bahasa pemrograman yang sering digunakan dalam *Machine Learning* adalah Python dan R, serta tools yang dapat digunakan adalah Rapid Miner dan Weka.

Menurut (Roihan et al., 2019) berdasarkan teknik pembelajarannya tipe *Machine Learning* dapat dibedakan menjadi 3 tipe yaitu :

1. *Supervised Learning*

Merupakan salah satu teknik *Machine Learning* yang mempergunakan dataset (*Data Training*) yang sudah memiliki label data untuk melakukan proses pembelajaran pada *Machine Learning*, sehingga *Machine Learning* mampu melakukan identifikasi data input yang dimasukkan untuk dapat memprediksi maupun klasifikasi data yang dimasukkan. *Supervised learning* memiliki beberapa algoritma populer seperti *Back-propagation Linear regression*, *Random Forest*, *Support Vector Machines*, *Naive Bayesian*, *Metode Rocchio*, *Decision Tree*, *k-Nearest Neighbor*, *Neural Network Logistic Regression*, dan *Neural Network* (Roihan et al., 2019).

2. *Unsupervised Learning*

Merupakan teknik dengan menarik kesimpulan berdasarkan dataset yang merupakan input data *label*. Sistem diberikan beberapa input data tetapi tidak diberikan *output* data yang diinginkan. Sehingga sistem akan melakukan clustering dan asosiasi untuk mencari struktur kesamaan dari data yang ada dengan hasil pola yang tersembunyi. *Unsupervised learning* memiliki beberapa algoritma populer seperti *k-means*, *Apriori*, *Independent Subspace Analysis (ISA)*, dan *DBSCAN* (Roihan et al., 2019).

3. *Reinforcement Learning*

Merupakan teknik yang mempelajari aturan kontrol dengan cara berinteraksi dengan data atau lingkungan yang asing, sehingga sistem akan mendapatkan sebuah hukuman jika keputusan salah dan mendapatkan hadiah bila benar. Pengalaman yang sudah dilakukan tersebut tersimpan sehingga program dapat mengambil kesimpulan dikemudian hari menggunakan pola-pola yang telah dipelajarinya sebelumnya. Contoh algoritmanya adalah algoritma *Temporal Difference* dan algoritma *Q-learning* (Roihan et al., 2019).

2.9 Text Mining

Text Mining merupakan proses menggali pengetahuan dari sebuah dokumen dari waktu tertentu menggunakan beberapa teknik *Text Mining*, seperti *information extraction*, dan *natural language processing*. *Text Mining* bertujuan untuk mencari kata-kata yang dapat menjadi kata kunci didalam suatu dokumen sehingga dapat dilakukan analisis hubungan antar dokumen (Statistika et al., 2020). Proses dalam *Text Mining* yaitu :Text, tahap awal yaitu data akan ditampung dalam sebuah database yang besar.

1. *Text Preprocessing*, untuk melakukan pembersihan teks sehingga menjadi text yang bersih dari beberapa karakter yang tidak diinginkan, tahapan ini seperti *text clean up*, *Case Folding*, *Tokenization*, dan *part of speech*
2. *Text Transformation*, merubah teks dan membentuk teks menjadi sebuah atribut yang mengacu pada proses yang digunakan biasanya dilakukan merubah menjadi matriks.
3. *Feature Selection*, melakukan pengurangan kata yang dianggap tidak relevan. Tahapan yang digunakan adalah Stop Word Removal dan Stemming.

2.10 TF-IDF

TF-IDF atau *Term Frequency Inverse Document Frequency* adalah suatu proses untuk melakukan transformasi data dari data teks menjadi data numerik untuk dilakukan pembobotan pada tiap kata pada tiap dokumen yang ada. TF merupakan frekuensi munculnya kata pada setiap dokumen yang diberikan menunjukkan seberapa bagusnya dan pentingnya kata di dalam tiap dokumen tersebut. Sedangkan DF adalah frekuensi dokumen yang mengandung kata tersebut mempresentasikan seberapa umum kata tersebut. Dan IDF merupakan pemberian bobot pada setiap kata kunci disetiap kategori kata kunci yang menjadi kemiripan dengan kategori yang ada (Aditya et al., 2020).

Dengan demikian rumus dari TF-IDF adalah penggabungan dari rumus TF dan IDF dengan melakukan operasi perkalian nilai, berikut adalah perhitungan bobot TF-IDF.

$$W_{ij} = tf_{ij} \times \log \frac{D}{df_i}$$

Keterangan :

- W_{ij} : Bobot TF-IDF
 idf_i : *Inverse Documen Frequency*
 tf_{ij} : Frekuensi suatu kata

2.11 Support Vector Machine

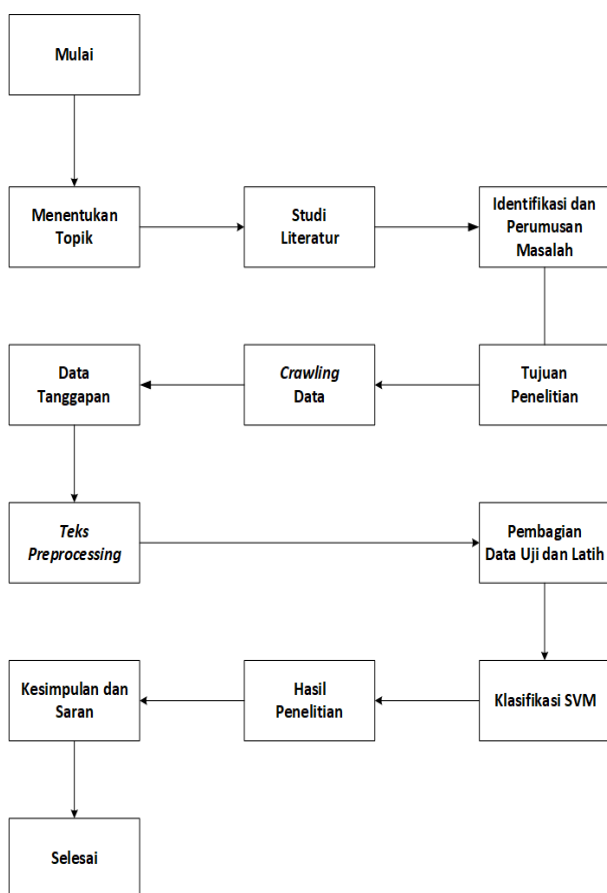
Support Vector Machine adalah metode klasifikasi jenis (*Supervised*) karena ketika proses pelatihan memerlukan goal atau target pembelajaran tertentu. SVM merupakan algoritma yang bekerja menggunakan pemetaan nonlinier untuk mengubah data pelatihan menjadi dimensi yang tinggi. Pada dimensi baru akan mencari hyperplane untuk memisahkan secara linier dan dengan pemetaan nonlinier yang tepat kedimensi yang lebih tinggi lagi. SVM muncul pertama kali pada tahun 1992 oleh Vladimir Vapnik Bersama dengan Bernhard Boser dan Isabell Guyon. Pada dasarnya SVM difungsikan untuk mengklasifikasikan data dengan 2 kelas dan linier separable. Namun masalah yang muncul kebanyakan non-linear. untuk mengatasi persoalan tersebut maka digunakan cara kernel trick pada data training yang bertujuan untuk mengelompokan setiap data pada ruang input menjadi ruang vektor baru dengan dimensi yang lebih tinggi. SVM hanya menggunakan beberapa titik data terpilih yang berkontribusi untuk membentuk model yang akan dipergunakan dalam proses klasifikasi (Handayanto & Herlawati, 2020). Sebuah hyperplane dapat ditulis sebagai persamaan SVM, ilustrasi metode SVM ditunjukkan pada persamaan, Adapun fungsi kernel yang biasanya digunakan dalam SVM yaitu :

Kernel Linear	$X^T X$
Kernel Polynomial	$(X^T X_i + 1)^p$
Kernel RBF	$\exp\left(-\frac{1}{2\sigma^2} x - x_i ^2\right)$
Kernel Sigmoid	$\tanh(\hat{O}(x.y) + x)$

3. Metodologi Penelitian

3.1. Kerangka Penelitian

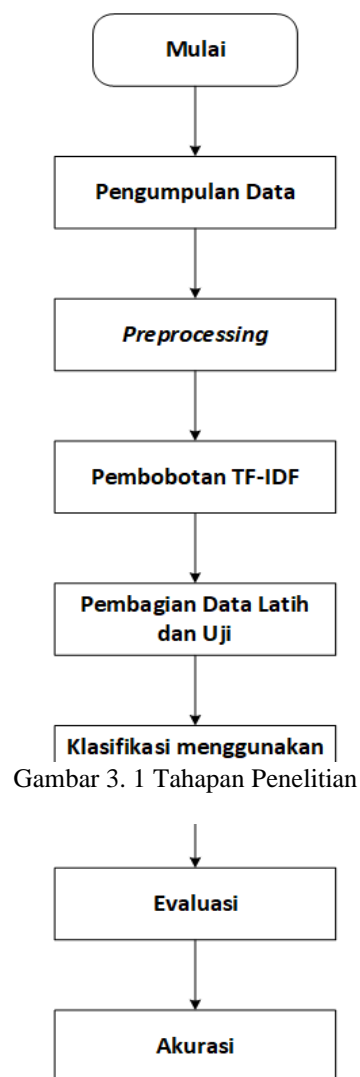
Pada penelitian ini akan melakukan analisis data mendeskripsikan data tentang analisis sentiment opini masyarakat terhadap tekanan mental pada twitter, penelitian ini menggunakan metode SVM. Penelitian ini termasuk dalam penelitian text mining, karena dalam penelitian ini dilakukan *crawling* data pada komentar twitter untuk mendapatkan data penelitian, yang dapat dilakukan proses analisis sentiment. Berikut adalah kerangka penelitian sebagai berikut :



Gambar 3. 2 Kerangka Penelitian

3.2. Tahapan Penelitian

Berikut adalah tahapan penelitian :



Gambar 3. 1 Tahapan Penelitian

3.3. Pengumpulan Data

Pada tahapan ini menjelaskan bagaimana cara data terkumpul dan digunakan dalam penelitian ini. Berikut rincian datanya sebagai berikut :

1. Kategori Data
Untuk penelitian ini data yang digunakan akan dibagi menjadi 2 kategori komentar yaitu komentar data terkena tekanan mental dan tidak terkena tekanan mental.
2. Variable Data
Berdasarkan hasil riset kata kunci dalam penelitian ini ditampilkan dalam tabel 3.1.

Tabel 3. 1 Variable Data

Variable Data
#tekananmental
#tekananhidup
#capek
#kenamental
#galau
#pasrah
#lelashidup

3. Pengumpulan Data dengan Web Crawling
 Pengumpulan data dilakukan dengan crawling data menggunakan Rapid Miner. Selanjutnya komentar yang akan diambil yaitu berbahasa Indonesia saja tanpa emoticon dan komentar akan diambil pada unggahan konten di twitter.

3.4. Preprocessing

Preprocessing untuk mentransformasikan data text tidak terstruktur menjadi data terstruktur.

1. Cleaning

Cleaning adalah proses menghilangkan karakter, symbol, tanda baca, nomor dan URL.

2. Case folding

Case folding adalah proses merubah semua kalimat menjadi bentuk standar dalam huruf kecil atau lowercase.

3. Filtering

Filtering adalah tahapan proses penghilangan kata-kata yang tidak sesuai dengan ejaan, kata hubung, dan ejaan yang disingkat.

4. Stopwords

Stopwords adalah tahapan proses untuk menghilangkan kata yang tidak diperlukan karena tidak memiliki arti dalam proses preprocessing. Proses dalam Stopwords bertujuan untuk pemilihan fitur yang bertujuan untuk dilakuakn pembobotan agar dibuat sebuah dataset.

5. Tokenize

Tokenize adalah proses memecah kalimat pada kalimat pada paragraph dalam sebuah text menjadi perkata yang tidak saling berhubungan sehingga menghasilkan kata yang berdiri sendiri.

3.5. Document Term Matrix Pembobotan TF-IDF

Pembobotan TF-IDF melakukan penentuan frekuensi dari setiap kata yang kemudian dengan bobot kata pada seluruh dokumen. Tahapan pertama untuk mencari pembobotan yang menggambarkan suatu matriks menjadi frekuensi kata. Pembobotan bisa disebut dengan pengujian matematis untuk mengukur seberapa penting sebuah kata pada kumpulan teks. Frekuensi kepentingan kata akan meningkatkan kata tersebut muncul beberapa kali dalam sebuah text.

3.6. Training Data

Pada tahapan training data pada metode SVM dilakukan untuk menentukan dan mencari hasil yang paling optimal pada hyperplane. Pada pembatas margin berfungsi untuk memberikan jarak antara fungsi hyperplane pada masing-masing kelas. Berikut adalah Langkah-langkah untuk melakukan training data sebagai berikut :

1. Melakukan inialisasi parameter.
2. Memasukkan data latih atau training berdasarkan tingkat intensitas kemunculan keyword dalam satu kalimat.
3. Menghitung nilai matrik dengan rumus kernel : $(K(x_i, y_j))$
4. Mencari nilai error dengan rumus : $E_i = \sum_j |1 - \alpha_j|$ Dij
5. Menghitung nilai deltha alpha $\delta \alpha_i = \min\{\max[\gamma(1-E_i) - \alpha_i] C - \alpha_i\}$
6. Menghitung nilai alpha baru dengan rumus $\alpha_i = \alpha_i + \delta \alpha_i$
7. Kernel linier dengan fungsi $K(x, y) = (x \cdot xi)$

3.7. Evaluasi

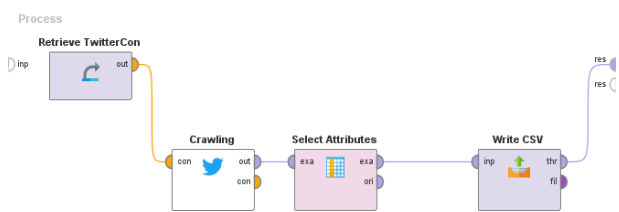
Setelah sistem berhasil melakukan klasifikasi tweet, untuk menentukan seberapa benar atau tepat klasifikasi yang telah dibuat oleh sistem akan dilakukan evaluasi dengan Confussion Matrix. Tabel berikut menunjukkan Confussion Matrix :

4. Hasil dan Pembahasan

Berdasarkan ulasan dan kajian teori dan hasil-hasil penelitian sebelumnya. Pada bab 4 ini akan menjelaskan implementasi dari metode TF-IDF dan SVM untuk menganalisis sentiment opini masyarakat terhadap tekanan mental pada media sosial Twitter akan menggunakan *software* Rapid Miner Studio versi 9.10.

4.1 Pengumpulan Data dengan Web Crawling

Dalam proses penarikan data dari media sosial Twitter maka dilakukan dengan menggunakan teknik *web crawling* dengan mempergunakan Twitter API (*Application Programming Interface*) yang diberikan oleh developer Twitter. API Twitter difungsikan untuk mempermudah program di luar Twitter untuk menarik data dari database Twitter. Unsur dari API Twitter adalah *consumer key*, *consumer secret*, *access key* dan *access secret* yang berfungsi sebagai autentikasi login untuk mendapatkan data Twitter yang dibutuhkan. Dalam penelitian ini data yang diambil adalah ulasan para pengguna twitter tentang kena mental. Total data yang diambil yaitu 5000 *tweet* dan disimpan di dalam file dengan format *.csv*, berikut proses *crawling* data twitter.



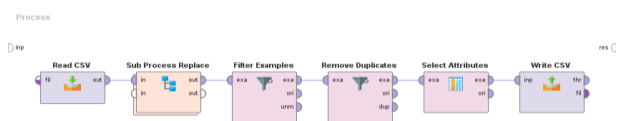
Gambar 4. 1 Proses Crawling Data Twitter

4.2 Pre-processing Data

Untuk mentransformasikan data text tidak terstruktur menjadi data terstruktur. Berikut proses *Preprocessing* yang sudah dilakukan.

1. Cleaning Data

Cleaning adalah proses menghilangkan karakter, simbol, tanda baca, nomor dan URL. Berikut proses *cleaning* data.



Gambar 4. 2 Proses Cleaning Data

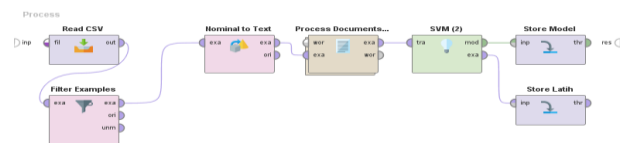
2. Labeling Data

Setelah data set sudah dibersihkan di proses sebelumnya maka akan dilakukan label sentiment dari 5000 *tweet* akan dilakukan label sebesar 300 *tweet*. Terdapat dua jenis label sentiment yaitu kena mental dan tidak kena mental. Berikut data set yang sudah di label sentiment.

4.3 Training Data Latih

Dalam proses ini akan dilakukan pembangunan model klasifikasi menggunakan algoritma SVM dengan tipe kernel Linier, berdasarkan data set latihan yang sudah dimiliki dan sudah diberikan label sentiment. Tujuannya

adalah melakukan pembelajaran dan melatih sebuah model agar dapat mengenali apakah *tweet* bernilai kena

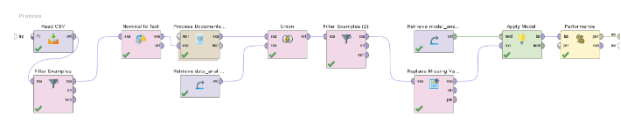


mental atau tidak kena mental. Berikut adalah proses yang dilakukan di tahapan training data latihan.

Gambar 4. 4 Proses training data latihan model

4.4 Pengolahan Data

Dalam proses ini akan dilakukan penerapan model dan data latihan terhadap data uji yang belum memiliki label sentiment sehingga secara otomatis akan diprediksi apakah data *tweet* tersebut bernilai kena mental atau tidak kena mental. Berikut adalah proses pengolahan data.



Gambar 4. 5 Proses Pengolahan Data

Berikut adalah hasil proses data latihan memprediksi :

Row No.	Sentimen	prediction k...	confidence(Tidak Kena Mental)	confidence(Kena Mental)	text	asas
1	?	Tidak Kena M...	0.516	0.484	mental nomor	0
2	?	Kena Mental	0.487	0.513	tidak nomor	0
3	?	Kena Mental	0.423	0.577	aprilia me...	0
4	?	Tidak Kena M...	0.515	0.485	adibawati	0
5	?	Tidak Kena M...	0.534	0.466	concord sig...	0
6	?	Tidak Kena M...	0.542	0.458	wahidula m...	0
7	?	Tidak Kena M...	0.534	0.466	cahaya m...	0
8	?	Tidak Kena M...	0.547	0.453	matier sem...	0
9	?	Tidak Kena M...	0.542	0.458	subswitch p...	0
10	?	Kena Mental	0.436	0.564	kapan sen...	0
11	?	Tidak Kena M...	0.510	0.390	harapacoo	0
12	?	Tidak Kena M...	0.540	0.460	swagman m...	0
13	?	Tidak Kena M...	0.560	0.440	watanda m...	0

Gambar 4. 6 Hasil prediksi sentiment data latihan

4.5 Word Cloud

Kumpulan kata-kata umum dapat dilihat dalam bentuk *word cloud*. Visualisasi *word cloud* mengungkapkan topik kena mental dan tidak kena mental dan kata-kata yang sering digunakan pengguna Twitter untuk memberikan jawaban tentang tekanan mental.



Yang diambil sebanyak 20 *word cloud* teratas, berikut adalah gambaran *word cloud*.

Gambar 4. 6 *Word Cloud*

Dari hasil visualisasi *word cloud* pada gambar 4.14 dapat dilihat bahwa kata yang paling sering muncul dan banyak dibahas pada tweet mengenai tekanan mental masyarakat di Indonesia adalah :

1. “orang” dengan frekuensi kemunculan sebanyak 160.
2. “udah” dengan frekuensi kemunculan sebanyak 157.
3. “kalo” dengan frekuensi kemunculan sebanyak 150.
4. “kuat” dengan frekuensi kemunculan sebanyak 140.
5. “fisik” dengan frekuensi kemunculan sebanyak 100.
6. “banget” dengan frekuensi kemunculan sebanyak 88.
7. “anak” dengan frekuensi kemunculan sebanyak 87,
8. “siapin” dengan frekuensi kemunculan sebanyak 84.
9. “health” dengan frekuensi kemunculan sebanyak 83.
10. “kesehatan” dengan frekuensi kemunculan sebanyak 72.
11. “emang” dengan frekuensi kemunculan sebanyak 70.

4.6 Nilai Akurasi

Klasifikasi data yang telah dilakukan perhitungan membutuhkan evaluasi dalam bentuk nilai akurasi yang digunakan untuk mengukur keberhasilan sebuah sistem. Berikut adalah tabel nilai akurasi.

Tabel 4. 1 Tabel Nilai Akurasi

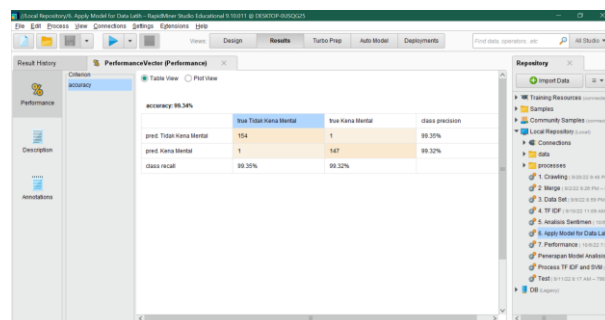
	True Tidak Kena Mental	True Kena Mental	Class Precision
Pred. Tidak Kena Mental	154	1	99,35%
Pred. Kena Mental	1	147	99,32%
Class recall	99,35%	99,32%	

Berikut tampilan nilai akurasi yang dihasilkan oleh Rapid Miner.

Gambar 4. 7 Hasil Perhitungan *Performance Accuracy*

Berdasarkan hasil akurasi diperoleh nilai *accuracy* sebesar 99,34 %. Semakin besar nilai akurasi semakin baik metode tersebut dalam melakukan klasifikasi. Dapat dilihat bahwa nilai akurasi klasifikasi SVM yang didapatkan sebesar

99,34 %. yang artinya bahwa nilai tersebut *good*



classification atau klasifikasi baik.

5. Kesimpulan

5.1 Kesimpulan

Berdasarkan analisis dan pengujian yang dilakukan dengan jumlah data set yang digunakan yaitu 5000 data set dengan pembagian data latih 300 dan data uji 4700 serta menggunakan software Rapid Miner, maka kesimpulan yang dapat diambil adalah hasil dari implementasi algoritma *Support Vector Machine* (SVM). Analisis sentiment dengan algoritma SVM dengan kernel linier berhasil mengklasifikasi sentiment masyarakat di Twitter tentang tekanan mental. Hasil klasifikasi menunjukkan sentiment tidak kena mental lebih tinggi dari pada sentiment kena mental dengan perbandingan 155 sentiment tidak kena mental dan 146 sentiment kena mental. Dengan skor nilai offset yaitu -0,179 dan nilai support vector 301. Hasil dari analisis TF-IDF menunjukkan bahwa kata yang sering muncul dalam topik pembicaraan adalah “orang” dengan nilai 160. Dari hasil analisis dengan menggunakan SVM dan TF-IDF dapat disimpulkan bahwa masyarakat banyak mengalami masalah tekanan mental tetapi masih tetap kuat untuk menghadapi tekanan mental sehingga tidak terlalu masuk ke dalam kondisi tekanan mental secara sikis dan hal yang menjadi penyebab masalah tekanan mental di masyarakat adalah masalah hubungan dengan orang lain. Berdasarkan hasil akurasi diperoleh nilai *accuracy* sebesar 99,34 % artinya bahwa nilai tersebut *good classification* atau klasifikasi baik. Semakin besar nilai akurasi semakin baik metode tersebut dalam melakukan klasifikasi.

5.2 Saran

Berdasarkan hasil analisis dan kesimpulan, peneliti dapat memberikan saran sebagai berikut :

1. Untuk penelitian selanjutnya, untuk mendapatkan dapat memperluas periode waktu agar data didapatkan lebih banyak dan lebih bervariasi. Sehingga diperoleh hasil klasifikasi yang lebih spesifik.
2. Akibat adanya kesamaan antara bahasa melayu dan Indonesia maka pada saat proses crawling

terhadap tweet yang bukan berasal dari user masyarakat twitter Indonesia yang ikut masuk ke dalam data set. Sehingga diharapkan pada penelitian selanjutnya diharapkan dapat bekerja sama dengan ahli bahasa untuk melakukan stemming dan labelling data set sentiment sehingga dapat lebih baik lagi.

3. Selain itu beberapa pengguna twitter akan menggunakan kata tidak baku yang terkadang memiliki makna yang bias dalam membuat hasil sentiment.

Daftar Pustaka.

- [1]. Aditya, H., Ardiansyah, Sidik, & Gata, W. (2020). Analisis Sentimen Penggunaan Twitter Terhadap Penggunaan Cairan Desinfektan Menggunakan Metode Term Frequency – Inverse Document Frequency Dan Support Vector Machine. *Informan's*, 14(2), 167–174.
- [2]. Arriani, A., Pradityas, H., Sofiani, L., & Iwisara, N. G. N. (2020). *Suara Komunitas COVID-19_Aug_#1_Bahasa_FIN.pdf* (pp. 8–9).
- [3]. Aulia, K., & Amelia, L. (2020). Analisis Sentimen Twitter Pada Isu Mental Health Dengan Algoritma Klasifikasi Naive Bayes. *Siliwangi Journal (Seri Sains and Teknologi)*, 6(2), 60–65.
- [4]. BPS RI. (2020). Perilaku Masyarakat Di Masa Pandemi Covid-19. In *Perilaku Masyarakat di Masa Pandemi Covid-19 BPS RI* (Vol. 19, Issue September).
- [5]. Chaplin, C. P. (1989). *Psychology complete dictionary*. Dell Publishing.
- [6]. Firmansyah Sulaeman, A., Afif Supianto, A., & Abdurrachman Bachtar, F. (2019). Analisis Sentimen Opini Mahasiswa Terhadap Saran Evaluasi Kinerja Dosen Menggunakan TF-IDF dan Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(6), 5647–5655.
- [7]. Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/jti.v14i2.679>
- [8]. Handayanto, R. T., & Herlawati. (2020). *Data Mining dan Machine Learning Menggunakan Matlab dan Python*. Informatika.
- [9]. Kominfo. (2013). *Pengguna Internet di Indonesia 63 Juta Orang*. Kominfo. https://www.kominfo.go.id/content/detail/3415/kominfo-pengguna-internet-di-indonesia-63-juta-orang/0/berita_satker
- [10]. Malini, Z., Riskiyanti, C., Gede, I. K., Putra, D., Kt, A. A., & Cahyawan, A. (2022). *Analisis Sentimen Pola Pikir Masyarakat Indonesia Terkait Virus Covid-19 Dalam Media Sosial Twitter Menggunakan Metode Rule Based Leksikon*. 3(1).
- [11]. Mutmainah, S. (2022). Kemungkinan Depresi Dari Postingan Pada Sosial Media. *Jurnal Sains, Nalar, Dan Aplikasi Teknologi Informasi*, 1(2), 17–23.
- [12]. Neighbor, M. K., Tirtopangarsa, A. P., Maharani, W., Informasi, T., & Telkom, U. (2021). Sentiment Analysis of Depression Detection on Twitter Social Media Users Using the K-Nearest Neighbor Method. *Jurnal SEMNASIF*, 247–258.
- [13]. Nugroho, K. S., Akbar, I., Suksmawati, A. N., & Istiadi, I. (2021). Deteksi Depresi dan Kecemasan Pengguna Twitter. *The 4th Conference on Innovation and Application of Science and Technology (CIASTECH 2021)*, Ciastech, 287–296.
- [14]. Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. *Bina Insani Ict Journal*, 7(2), 156. <https://doi.org/10.51211/biict.v7i2.1422>
- [15]. Riyanto, G. P. (2021). *Laporan Keuangan Twitter Kuartal IV (Q4) tahun fiskal 2021*. Kompas. <https://tekno.kompas.com/read/2022/02/14/11010067/berapa-jumlah-pengguna-aktif-twitter-saat-ini?page=all#:~:text=Rata-rata pengguna harian Twitter,211 juta pengguna secara global.>
- [16]. Rofiqoh, U., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan

- Lexion Based Feature. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya*, 1(12), 1725–1732. <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/628>
- [17]. Roihan, A., Abas Sunarya, P., & Rafika, A. S. (2019). IJCIT (Indonesian Journal on Computer and Information Technology) Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82.
- [18]. Statistika, J., Matematika, F., Ilmu, D. A. N., Alam, P., & Indonesia, U. I. (2020). *Tentang Kesehatan Mental Selama Pandemi Covid-19 Di Media Sosial Twitter Menggunakan Naive Bayes Classifier Dan Support Vector*.
- [19]. Styawati, S., & Mustofa, K. (2019). A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(3), 219. <https://doi.org/10.22146/ijccs.41302>
- [20]. Unicef. (2021). Menuju respons dan pemulihan COVID-19 yang berfokus pada anak. *Unicef.Org*, 1–16.
- [21]. Yan, K., Arisandi, D., Studi, P., Informasi, S., & Tarumanagara, U. (2019a). Analisis Sentimen Komentar Netizen Twitter Terdapat Kesehatan Mental Masyarakat Indonesia. *Jurnal Ilmu Komputer Dan Sistem Informasi*.
- [22]. Yan, K., Arisandi, D., Studi, P., Informasi, S., & Tarumanagara, U. (2019b). Analisis Sentimen Komentar Netizen Twitter Terdapat Kesehatan Mental Masyarakat Indonesia. *Jurnal Ilmu Komputer Dan Sistem Informasi*.