

ANALISIS SENTIMEN PENGGUNA *TWITTER* TERHADAP KEBIJAKAN PEMERINTAH DALAM MENANGANI COVID-19 MENGUNAKAN METODE *SUPPORT VECTOR MACHINE* DAN *NAIVE BAYES* DENGAN *LAPLACE ESTIMATOR*

Sulton Ibrahim¹, Widodo², Murien Nugraheni³

¹ Mahasiswa Prodi Pendidikan Teknik Informatika dan Komputer, FT – UNJ

^{2,3} Dosen Prodi Pendidikan Teknik Informatika dan Komputer, FT – UNJ

¹sultonibrahim_1512618078@mhs.unj.ac.id, ²widodo@unj.ac.id, ³muriennugraheni@unj.ac.id

Abstrak

Pemerintah Indonesia telah melakukan berbagai penanganan untuk mencegah semakin meluasnya penyebaran virus COVID-19 dengan mengeluarkan kebijakan-kebijakan. Tetapi kebijakan pemerintah dalam menangani COVID-19 ini menuai banyak pro dan kontra dari masyarakat di Indonesia. Masyarakat di Indonesia sering memberikan komentar terhadap kebijakan pemerintah dalam menangani masalah virus COVID-19 melalui media sosial, salah satunya adalah Twitter. Komentar masyarakat baik secara positif, netral maupun negatif terhadap kebijakan pemerintah dapat dianalisis dengan menggunakan analisis sentimen. Dalam analisis sentimen ini akan menggunakan dua metode yaitu Support Vector Machine dan Naive Bayes dengan Laplace Estimator. Penelitian ini bertujuan untuk mendapatkan perbandingan tingkat akurasi dari hasil kedua metode tersebut. Dari hasil penelitian didapatkan hasil akurasi untuk Support Vector Machine 86% dan Naive Bayes dengan Laplace Estimator 85% dengan menggunakan sampel sebanyak 2194 data.

Kata kunci : COVID-19, Kebijakan Pemerintah, *Twitter*, Analisis Sentimen, *Support Vector Machine*, Naive Bayes, *Laplace Estimator*

1. Pendahuluan

Pada awal tahun 2020, dunia dikejutkan dengan merebaknya virus baru yaitu coronavirus jenis baru (SARS-CoV-2) dan penyakitnya disebut Coronavirus disease 2019 (COVID-19) (Yuliana, 2020: 187). Virus baru dan penyakit yang disebabkan ini tidak dikenal sebelum mulainya wabah di Wuhan, Tiongkok, bulan Desember 2019. Pada tanggal 16 Maret 2020, keberadaan COVID-19 kemudian ditetapkan menjadi pandemi karena penyebarannya yang telah mencapai sebagian besar negara di dunia (Naraswati, dkk., 2021: 229). Kasus COVID-19 menjadi salah satu permasalahan kompleks yang dihadapi oleh hampir seluruh negara di dunia. Dari hal tersebut diharapkan setiap negara dapat membuat kebijakan untuk virus COVID-19 ini agar meminimalisasi korban dan mengupayakan pencegahan terhadap penyebarannya agar tidak meluas.

Kasus positif COVID-19 di Indonesia pertama kali dideteksi pada tanggal 2 Maret 2020, ketika dua orang terkonfirmasi tertular dari seorang warga negara Jepang. Pada tanggal 9 April, pandemi sudah menyebar ke 34 provinsi dengan DKI Jakarta, Jawa Barat dan Jawa Tengah sebagai provinsi paling terpapar virus corona di Indonesia (Kementerian Kesehatan RI, 2021). Menurut data dari Gugus Tugas Percepatan Penanganan COVID-19 (2022) saat ini di Indonesia pasien yang terinfeksi COVID-19 telah menyebar ke seluruh Provinsi, bahkan sudah tercatat 4.353.370 kasus yang terkonfirmasi positif dan 144.320 kasus meninggal pertanggal 31 Januari 2022. Pemerintah sudah melakukan berbagai penanganan melalui kebijakan-kebijakan yang telah diberlakukan salah satunya adalah memastikan dilaksanakannya protokol kesehatan oleh seluruh masyarakat guna mencegah semakin meluasnya penyebaran virus dan meningkatnya jumlah korban (Naraswati, dkk., 2021: 229). Kebijakan yang digalakkan pemerintah diantaranya Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM), aturan wajib masker, vaksinasi, dan penerapan jam malam. Kebijakan dari pemerintah tersebut diharapkan dapat memastikan masyarakat tetap melaksanakan protokol kesehatan.

Kebijakan pemerintah tersebut dalam penanganan COVID-19 ini menuai banyak pro dan kontra dari masyarakat di Indonesia. Kebijakan yang sudah ditetapkan pemerintah tidak seluruhnya dilaksanakan oleh

masyarakat sehingga masih dapat ditemukan pelanggaran atas kebijakan yang berlaku. Sebagian masyarakat merasa pemerintah sudah melakukan hal yang benar dalam menangani COVID-19, tetapi sebagian masyarakat lainnya merasa pemerintah masih belum serius dalam menangani COVID-19 di Indonesia (Sabrila, dkk., 2021: 70). Keadaan juga diperparah dengan keberadaan beberapa pihak anggota masyarakat yang meragukan efektivitas kebijakan yang ada.

Situasi saat ini yang disebabkan oleh lockdown di beberapa bagian dunia dan adanya penerapan social distancing, membuat penggunaan media sosial secara global telah meningkat (Yulita, dkk., 2021: 1). Penggunaan media sosial yang semakin meningkat menjadikan media sosial sebagai salah satu platform bagi masyarakat untuk menceritakan kehidupan sehari-hari, mengutarakan keluhan kesah (Sabrila, dkk., 2021: 70). Data digital yang berasal dari media sosial sebagian besar didapatkan dalam bentuk yang tidak terstruktur dan membuatnya sulit untuk diproses. Tetapi walaupun begitu media sosial telah menciptakan data dalam jumlah yang besar, salah satunya adalah aspirasi masyarakat mengenai apa yang terjadi di sekitar mereka. Dari hal inilah membuat masyarakat sering memberikan komentar terhadap kebijakan pemerintah dalam menangani masalah virus COVID-19 melalui berbagai media sosial, salah satunya yang populer yaitu *Twitter*. *Twitter* merupakan salah satu media sosial paling populer di Indonesia dan Indonesia merupakan negara dengan pengguna *Twitter* kelima teraktif di dunia (Widowati dan Sadikin, 2020). Pengguna *Twitter* dapat membuat posting yang disebut sebagai *tweet*, dengan maksimal 280 karakter. Melalui *tweet*, pengguna *Twitter* dapat saling berinteraksi dan berbagi pendapat atau opini terhadap suatu kejadian.

Komentar masyarakat baik secara positif, netral maupun negatif terhadap kebijakan pemerintah dalam penanganan COVID-19 merupakan hal yang penting untuk dianalisis sehingga pemerintah dapat mengetahui respon masyarakat terhadap kebijakannya dalam penanganan COVID-19. Salah satu cara yang dapat digunakan yang dapat dilakukan untuk menganalisis data ini adalah dengan menggunakan analisis sentimen. Analisis Sentimen adalah suatu teknik mengekstrak data teks untuk mendapatkan informasi tentang sentimen bernilai positif maupun negatif. Analisis sentimen diberikan oleh pengguna internet pada media sosial untuk memberikan suatu penilaian atau opini pribadi (Sari dan Wibowo, 2019: 681).

Analisis sentimen memiliki banyak metode yang dapat digunakan antara lain *Support Vector Machine*, *K-Nearest Neighbor* dan *Naïve Bayes*. Dari beberapa metode tersebut dipilih *Support Vector Machine* dan *Naïve Bayes* karena memiliki nilai akurasi tinggi dibanding dengan metode lain. Seperti pada penelitian “Analisis Sentimen Opini Publik Mengenai Covid-19 pada *Twitter* Menggunakan Metode *Naïve Bayes* dan *KNN*” oleh Muhammad Syarifuddin, metode *Naïve Bayes* mendapatkan nilai *accuracy* sebesar 63.21% sedangkan *KNN* memiliki nilai *accuracy* 58.10%. Selanjutnya pada penelitian “Perbandingan *Naïve Bayes*, *SVM* dan *KNN* untuk Analisis Sentimen Gadget Berbasis Aspek” oleh Jessica Widyadhana Iskandar dan Yessica Natalian, hasilnya menunjukkan rata-rata *accuracy* dari *Naïve Bayes* adalah 83.54%, *SVM* 96.43% dan *KNN* 59.68%. Lalu pada penelitian selanjutnya yaitu mengenai “Analisis Sentimen *Twitter* Terhadap Tokoh Publik dengan Algoritma *Naive Bayes* dan *Support Vector Machine*” oleh Tanthy Tawaqalia Widowati dan Mujiono Sadikin menunjukkan *accuracy* dari *Naive Bayes* adalah 91.48%, sedangkan *SVM* 85.47%. Dari beberapa penelitian dibuktikan bahwa *Naïve Bayes* dan *SVM* memiliki nilai akurasi yang tinggi dan dari beberapa penelitian tersebut perbedaan akurasinya tidak terlalu jauh dibandingkan metode lainnya. *Naïve Bayes* menggunakan perhitungan probabilitas pada data set dan terkadang menemukan nilai probabilitas sama dengan nol (0). Pada data yang sangat besar, probabilitas = 0 dapat mengakibatkan perhitungan *Naïve Bayes* menjadi tidak seimbang (Jollyta, dkk., 2020: 110). Salah satu cara untuk menghindari hal tersebut yaitu digunakanlah *Laplace Estimator* pada penelitian ini untuk menghindari nilai probabilitas 0 (nol).

Berdasarkan permasalahan tersebut, maka akan dilakukan penelitian mengenai analisis sentimen pengguna *Twitter* dari data *tweet* yang mengandung topik tentang kebijakan pemerintah dalam menangani COVID-19 dengan menggunakan dua jenis metode klasifikasi yaitu, metode *Support Vector Machine* (*SVM*) dan *Naive Bayes* dengan *Laplace Estimator*. Hasil dari penelitian ini akan memberikan gambaran mengenai kebijakan yang dilakukan oleh pemerintah dalam menangani COVID-19 cenderung ke opini yang positif, netral atau negatif dan juga untuk membandingkan tingkat akurasi dari metode *Support Vector Machine* (*SVM*) dan *Naive Bayes* dengan *Laplace Estimator*.

2. Dasar Teori

Pada bagian ini diuraikan landasan teoritis yang berhubungan dengan penelitian atau perancangan yang dilakukan.

2.1. Text Mining

Text Mining dan data *mining* dipersepsikan dalam keilmuan yang sama dikarenakan algoritma sama dapat digunakan dalam kedua konsep *mining* tersebut (Fauziyyah dan Gautama, 2020: 32). Tujuan *Text Mining* hampir mirip dengan tujuan data *mining* yaitu menemukan pola pada data agar dapat dimanfaatkan manusia untuk

membantu pekerjaannya, yang berbeda adalah sumber data yang akan digunakan, pada proses *Text Mining* sumber datanya adalah teks atau dokumen analisis (Afif, dkk., 2021: 200). *Text Mining* merupakan proses penambangan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru, dan tidak diketahui sebelumnya, atau menemukan kembali informasi yang tersirat secara implisit (Sari dan Wibowo, 2019: 682). Data *Text Mining* biasanya didapatkan dari dokumen untuk mencari kata-kata yang dapat mewakili isi dokumen sehingga dapat dilakukan analisis (Nursyiah, dkk., 2021: 118).

2.2. Analisis Sentimen

Analisis sentimen adalah suatu bidang yang berlangsung dalam penelitian berbasis teks dan merupakan kajian tentang cara untuk memecahkan masalah dari opini masyarakat, sikap dan emosi suatu entitas tersebut yang dapat mewakili suatu individu (Nursyiah, dkk., 2021: 118). Analisis sentimen digunakan untuk digunakan untuk mengekstrak data opini, memahami serta mengolah tekstual data secara otomatis untuk melihat sentimen yang terkandung dalam sebuah opini (Sari dan Wibowo, 2019: 682).

Analisis sentimen merupakan bidang penelitian berkelanjutan yang berada diantara berbagai bidang seperti *Data Mining*, *Natural Language Processing (NLP)* dan *Machine Learning* yang berfokus pada ekstraksi sentimen dalam sebuah kalimat berdasarkan isinya (Tuhuteru, 2020: 8). Analisis sentimen berkaitan dengan bidang yang lebih luas seperti pengolahan bahasa alami, komputasi linguistik, dan *Text Mining* untuk tujuan menganalisis, sentimen, pendapat, sikap, evaluasi, penilaian dan emosi seseorang berkenaan pada topik, produk, layanan, organisasi, atau kegiatan tertentu (Attabi, 2018: 12).

Analisis sentimen dapat bermanfaat untuk mengetahui bagaimana tanggapan masyarakat baik itu positif, netral ataupun negatif terhadap kebijakan yang dikeluarkan oleh pemerintah dan dapat dijadikan sebagai bahan evaluasi untuk pemerintah dalam membuat suatu kebijakan. Pada penelitian kali ini akan menggunakan pendekatan dalam machine learning yaitu *Support Vector Machine (SVM)* dan *Naive Bayes* dengan *Laplace Estimator* dan khusus pada teks yang berbahasa Indonesia.

2.3. Twitter dan Twitter API

Twitter adalah layanan bagi teman, keluarga, dan teman sekerja untuk berkomunikasi dan tetap terhubung melalui pertukaran pesan yang cepat dan sering (*Twitter*, 2022). *Twitter* merupakan salah satu media sosial paling populer di Indonesia dan Indonesia merupakan negara dengan pengguna *Twitter* kelima teraktif di dunia (Widowati dan Sadikin, 2020). Pengguna *Twitter* dapat membuat posting yang disebut sebagai *tweet*, dengan maksimal 280 karakter. *Twitter* juga termasuk sebuah microblog yang merupakan alat komunikasi online yang memungkinkan para pengguna dapat memperbarui status tentang mereka yang sedang memikirkan atau melakukan sesuatu, seperti pendapat dan komentar pada suatu objek dan fenomena tertentu (Attabi, 2018: 16).

Ada beberapa cara untuk mendapatkan data *tweet* pada *Twitter* salah satunya adalah dengan menggunakan *Twitter API*. *API (Application Programming Interface)* merupakan salah satu cara komputer untuk “berbicara” satu sama lain sehingga dapat memesan dan mengantar informasi (Harijati, 2019: 16). Untuk dapat memesan informasi maka pengguna harus memperbolehkan aplikasi perangkat lunak mengakses sebuah informasi spesifik, seperti nomor telepon dimana setiap pengguna memiliki nomor yang unik. Agar dapat mengakses *Twitter API* maka pengguna harus mendaftarkan diri terlebih dahulu pada halaman *Twitter*. Setelah mendaftar maka pengguna akan mendapatkan kode yang akan digunakan untuk melakukan pengambilan data di *tweet* di *Twitter*.

2.4. Text Preprocessing

Text Preprocessing atau pemrosesan text merupakan cara untuk mempersiapkan sebuah teks menjadi data yang siap diolah lebih lanjut dan bertujuan untuk menghilangkan *noise*, serta mengambil fitur penting yang ada pada dokumen teks (Wulandari, dkk., 2021: 9009). Ada beberapa tahapan dalam *Preprocessing* yaitu :

- a. *Cleansing*: *Cleansing* merupakan proses yang dilakukan untuk membuang karakter dan tanda baca yang tidak diperlukan seperti titik, koma, tanda seru dan tanda tanya. Pada tahapan ini juga akan menghilangkan URL, menghapus emoji, mention, hashtag dan menghilangkan simbol atau huruf yang tidak relevan.
- b. *Case Folding*: *Case Folding* adalah proses mengubah setiap huruf di *tweet* menjadi *lowcase* atau semuanya dijadikan huruf kecil.
- c. *Tokenization*: *Tokenization* adalah proses yang dilakukan untuk memecah konten tekstual menjadi kata-kata, istilah, simbol, atau beberapa elemen bermakna lainnya. Hasil pemecahan ini kemudian disebut dengan token.
- d. *Filtering*: *Filtering* memiliki tujuan untuk menghilangkan stop words atau kata-kata yang tidak memiliki makna. Penghapusan stop words perlu dilakukan agar dapat fokus pada kata-kata lain yang lebih penting.
- e. *Normalization*: Proses *Normalization* berguna untuk mengubah kata yang tidak baku atau tidak sesuai ejaan menjadi sebuah kata baku dalam Bahasa Indonesia, pada tahap ini dilakukan dengan menggunakan database kamus kata bahasa baku dan tidak baku.
- f. *Stemming*: *Stemming* merupakan proses yang bertujuan mengubah kata yang berimbuhan menjadi kata dasar.

2.5. Term Frequency – Inverse Document Frequency (TF-IDF)

Term Frequency – Inverse Document Frequency (TF-IDF) adalah algoritma pembobotan yang tersusun dari dua nilai berasal dari dua algoritma dengan pembobotan yang berbeda, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Pada dasarnya TF-IDF bekerja dengan menghitung frekuensi relatif dari suatu kata yang muncul pada sebuah dokumen dibandingkan dengan proporsi *inverse* dari kata tersebut yang muncul pada seluruh kumpulan dokumen dan juga secara intuitif, perhitungan ini dapat digunakan untuk mengetahui seberapa relevan kata tersebut pada sebuah dokumen tertentu (Widowati dan Sadikin, 2020).

2.6. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah teknik supervised learning yang mempunyai tingkat akurasi dan kualitas yang baik sehingga membuatnya menjadi sangat diminati di antara algoritma yang lainnya (Laurensz dan Sedyono, 2021: 119). Metode ini berakar dari teori pembelajaran statistik yang hasilnya sangat menjanjikan untuk memberikan hasil yang lebih baik dari metode yang lain. SVM juga bekerja dengan baik pada set data dengan dimensi yang tinggi, bahkan SVM yang menggunakan teknik kernel harus memetakan data asli dari dimensi asalnya menjadi dimensi lain yang relatif lebih tinggi. Pada Metode SVM, hanya sejumlah data terpilih saja yang berkontribusi untuk membentuk model yang akan digunakan dalam klasifikasi yang akan dipelajari. SVM hanya menyimpan sebagian kecil dari data latih yang akan digunakan pada saat prediksi. Hal inilah yang menjadikannya kelebihan dari SVM karena tidak semua data latih akan dipandang untuk dilibatkan dalam setiap iterasi pelatihannya. Data-data yang berkontribusi tersebut disebut *support vector* sehingga metodenya juga disebut *Support Vector Machine*. *Support vector machine* digunakan secara luas untuk mengklasifikasikan informasi linear dan nonlinier dan menginisiasi data berdasarkan teori statistik (Mohan dkk., 2020).

2.7. Naive Bayes Classifier

Naïve Bayes Classifier merupakan metode pengklasifikasi sederhana berdasarkan penerapan teorema Bayes (Statistik Bayesian), biasanya digunakan dalam permasalahan klasifikasi teks (Attabi, 2018: 16). Naïve Bayes Classifier mengadopsi ilmu statistika yaitu dengan menggunakan teori kemungkinan (Probabilitas) untuk menyelesaikan sebuah kasus Supervised Learning, artinya dalam himpunan data terdapat Label, Class atau Target sebagai acuan (Nofriansyah dan Nurcahyo, 2015: 34).

2.8. Laplace Estimator

Perhitungan probabilitas pada data set, terkadang menemukan nilai probabilitas sama dengan nol (0). Pada data yang sangat besar, probabilitas = 0 dapat mengakibatkan perhitungan Naïve Bayes menjadi tidak seimbang dan salah satu untuk menghindari hal tersebut yaitu dengan menggunakan *Laplace Estimator*. *Laplace Estimator* adalah suatu cara untuk menangani nilai probabilitas 0 (nol). Dari sekian banyak data di training set, pada setiap perhitungan datanya ditambah 1 (satu) dan tidak akan membuat perbedaan yang berarti pada estimasi probabilitas sehingga bisa menghindari kasus nilai probabilitas 0 (nol) (Jollyta, dkk., 2020: 111).

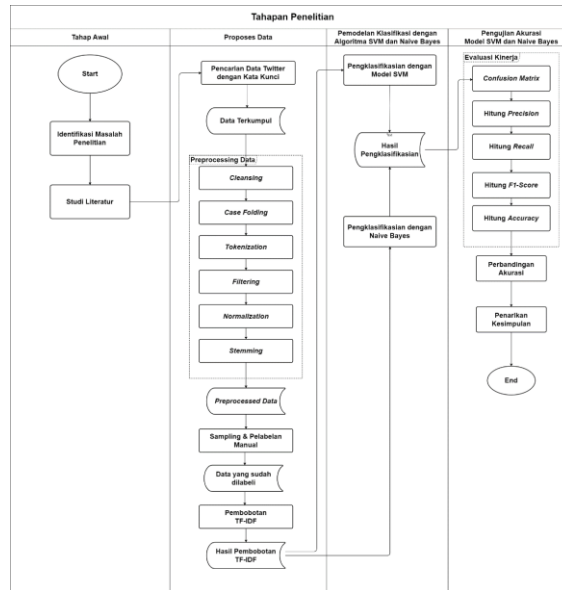
2.9. Evaluasi Kinerja

Evaluasi Kinerja atau performansi adalah suatu tahapan untuk mengukur performansi dari classifier dalam melakukan klasifikasi yang digunakan dalam penelitian (Wulandari, dkk., 2021: 111). Dalam membuat evaluasi performansi matrik perlu adanya *confusion matrix*. Menurut Karsito dan Susanti (2019: 44) *Confusion matrix* adalah metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. *Confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya.

Pada pengukuran kinerja menggunakan *Confusion matrix*, terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Istilah tersebut adalah *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)* dan *False Negative (FN)*. Nilai *True Negative (TN)* merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Positive (FP)* merupakan data negatif namun terdeteksi sebagai data positif. Dengan adanya *Confusion matrix*, performa sebuah model akan dihitung dari seberapa banyak model tersebut melakukan prediksi dengan benar.

- Precision*: *Precision* adalah tingkat ketepatan suatu informasi sistem dalam memprediksi target positif terhadap jumlah data yang diprediksi positif.
- Recall*: *Recall* adalah tingkat ketepatan suatu informasi sistem dalam memprediksi target positif terhadap jumlah data yang riil positif.
- F1-Score*: *F1-Score* adalah salah satu perolehan dari konstanta dua dikali nilai *recall* dan presisi, dan dibagi jumlah keduanya.
- Accuracy*: *Accuracy* adalah keberhasilan suatu algoritma dengan seberapa akurat sistem dapat mengklasifikasikan secara benar.

3. Metodologi



Gambar 3.1. Diagram Alir Penelitian

3.1. Identifikasi Masalah Penelitian

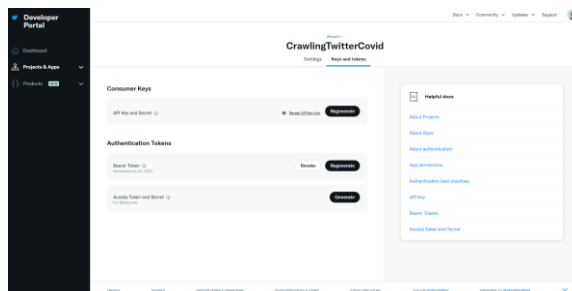
Penelitian ini dimulai dengan melakukan identifikasi masalah mengenai penelitian. Identifikasi ini perlu dilakukan untuk merumuskan masalah yang akan dibahas dalam penelitian. Identifikasi masalah yang didapatkan adalah bagaimana mengelompokkan dan mengklasifikasi opini pengguna *Twitter* terhadap kebijakan pemerintah dalam menangani Covid-19.

3.2. Studi Literatur

Studi literatur digunakan untuk memahami konsep dan dasar teori yang akan diterapkan dalam penelitian “Analisis Sentimen Pengguna *Twitter* terhadap Kebijakan Pemerintah dalam Menangani Covid-19 Menggunakan Metode *Support Vector Machine* dan *Naive Bayes* dengan *Laplace Estimator*”.

3.3. Pengumpulan Data Penelitian

Tahap awal yang akan dilakukan adalah mengumpulkan data *tweet* berbahasa Indonesia berupa kebijakan pemerintah dalam menangani COVID-19 dengan menggunakan *Twitter API*. Kata kunci yang bersangkutan mengenai kebijakan pemerintah adalah “PPKM” dan “Vaksinasi Covid”. Untuk mengambil data dari *Twitter* diharuskan mendaftar terlebih dahulu di *Twitter Developer* untuk mendapatkan akses. Tampilan awal pada website *Twitter Developer* ditunjukkan pada Gambar 4.1 dibawah. Setelah selesai melakukan proses pendaftaran akan diberikan *API Key*, *API Key Secret*, *Access Token* dan *Access Token Secret* yang ditunjukkan pada Gambar 4.2 dibawah dan akan digunakan untuk mengakses data pada *Twitter*.



Gambar 3.2. *Twitter Developer Portal*

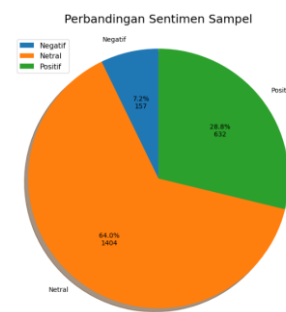
Data diambil pada tanggal 1 Agustus 2022 - 30 Agustus 2022. Data diambil secara bertahap dalam 1 bulan dengan jangka waktu tujuh hari. Data yang berhasil didapatkan sebanyak 35669 *tweet*. Data yang terkumpul disimpan dalam bentuk elektronik dengan format data *Comma Separated Values* (CSV).

3.4. Preprocessing

Data yang sudah terkumpul sebelum diklasifikasi data tersebut akan dilakukan *Preprocessing* yang terdiri dari *Cleansing*, *Case Folding*, *Tokenization*, *Filtering*, *Normalization* dan *Stemming* untuk mendapatkan data yang bersih dari karakter yang tidak dibutuhkan yang berpotensi akan mengganggu akurasi klasifikasi.

3.5. Sampling dan Pelabelan Manual

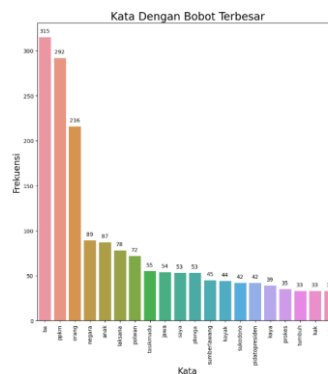
Data bersih yang didapatkan akan diambil hanya 2194 data sebagai data sampel karena jika menggunakan seluruh data yang di *crawl* untuk dilakukan *labeling* secara manual akan membutuhkan waktu dan juga tenaga yang besar sehingga hanya diambil sebagian data saja. Data sampel tersebut akan diberikan label data dalam kategori kelas yang berbeda yaitu kelas positif, netral atau negatif. Pelabelan data *tweet* dilakukan secara manual. Pelabelan dilakukan karena data hasil crawling tidak memiliki nilai sentimen maka harus dilakukan pelabelan secara manual.



Gambar 3.3. Visualisasi Perbandingan Sentimen Sampel

3.6. Pembobotan TF-IDF

Pembobotan dengan algoritma TF-IDF digunakan untuk menghitung frekuensi relatif dari suatu kata yang muncul pada sebuah dokumen dibandingkan dengan proporsi *inverse* dari kata tersebut yang muncul pada seluruh kumpulan dokumen dan juga secara intuitif, perhitungan ini dapat digunakan untuk mengetahui seberapa relevan kata tersebut pada sebuah dokumen tertentu. Pembobotan TF-IDF menggunakan library scikit-learn dengan class *TfidfVectorizer* yang dapat menghitung dari nilai TF-IDF.



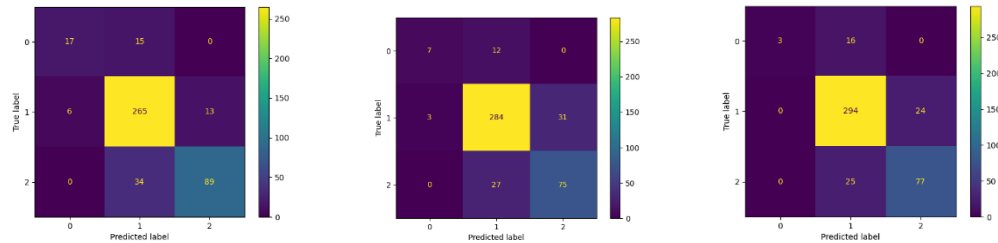
Gambar 3.4. Visualisasi Bobot Kata Terbesar

3.7. Pengklasifikasian dengan Model SVM dan Naive Bayes

Setelah proses *Preprocessing* dan pembobotan, proses selanjutnya adalah melakukan klasifikasi data dengan menggunakan data sampel yang telah dilabeli. Data yang sampel telah diberikan label akan dibagi menjadi data latih sebanyak 80% dan data uji sebanyak 20% dari jumlah data sampel. Data latih adalah data yang digunakan untuk melatih mesin agar dapat mengenali pola, sedangkan data uji adalah data yang digunakan untuk menguji hasil pelatihan yang telah dilakukan terhadap mesin. Model klasifikasi yang digunakan adalah model *Support Vector Machine* (SVM) dan *Naive Bayes* dengan *Laplace Estimator*.

3.8. Evaluasi Kinerja

Pada tahap ini akan berguna untuk mengukur kinerja dari classifier dalam melakukan klasifikasi yang digunakan dalam penelitian. Dalam membuat evaluasi kinerja matrik perlu adanya *confusion matrix*. *Confusion matrix* adalah metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. *Confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya. Kinerja yang akan dihitung yaitu Precision, *Recall*, *F1-Score* dan Accuracy.



Gambar 3.5. *Confusion Matrix SVM, Naive Bayes dan Naive Bayes Laplace Estimator*

Tabel 3.1. Hasil Evaluasi Kinerja

Evaluasi	SVM	Naïve Bayes	Naïve Bayes dengan Laplace Estimator
Akurasi	84%	83%	85%
Presisi	84%	83%	85%
Recall	84%	83%	85%
F1-Score	84%	83%	84%

3.9. Perbandingan Akurasi

Setelah melakukan evaluasi kinerja akan dilanjutkan perbandingan akurasi di antara model algoritma *Support Vector Machine* (SVM) dan *Naive Bayes* dengan *Laplace Estimator*.

3.10. Penarikan Kesimpulan

Tahap terakhir ini akan menyimpulkan hasil yang telah didapat setelah melakukan perbandingan akurasi di antara model algoritma *Support Vector Machine* (SVM) dan *Naive Bayes* dengan *Laplace Estimator*.

4. Hasil dan Analisis

Hasil pengujian ini menunjukkan metode *Naïve Bayes* menggunakan *Laplace Estimator* lebih unggul 2% pada Akurasi, Presisi dan *Recall* dibanding tanpa menggunakan *Laplace Estimator*. Untuk *F-1 Score* memiliki hasil yang lebih unggul 1% daripada *Naïve Bayes* tanpa menggunakan *Laplace Estimator*. Metode *Naïve Bayes* menggunakan *Laplace Estimator* juga memiliki hasil yang lebih unggul dibandingkan dengan *Support Vector Machine* dengan hasil yaitu sebesar 1% pada Akurasi, Presisi dan *Recall* tetapi untuk hasil *F-1 Score* pada kedua metode tersebut memiliki hasil yang sama.

5. Kesimpulan dan Saran

5.1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa penelitian mengenai Analisis Sentimen pengguna *Twitter* dengan topik mengenai kebijakan pemerintah dalam menangani Covid-19 dengan menggunakan metode *Support Vector Machine* dan *Naive Bayes* dengan *Laplace Estimator* telah berhasil dilakukan dengan menggunakan sampel sebanyak 2194 data. Data sampel tersebut setelah dilabeli secara manual mendapatkan hasil sentimen yang berisi 632 data positif, 1404 data netral dan 157 data negatif. Hasil akurasi yang didapatkan setelah melakukan pengujian yaitu untuk *Support Vector Machine* 84% dan *Naïve Bayes* dengan *Laplace Estimator* 85%. Hasil akurasi dari *Naïve Bayes* meningkat sebesar 2% setelah menggunakan *Laplace Estimator*.

5.2. Saran

Pada penulisan penelitian ini penulis menyadari bahwa masih banyak kekurangan yang dialami. Saran yang dapat dilakukan untuk penelitian selanjutnya antara lain adalah :

1. Memperbanyak sampel data yang sudah dilabeli agar data yang didapatkan lebih akurat.
2. Menerapkan Algoritma *Smote* untuk memperbaiki data yang tidak seimbang pada dataset.

Daftar Pustaka:

- Afif, M. S. (2021). *Text Mining* Untuk Mengklasifikasi Judul Berita Online Studi Kasus Radar Banjarmasin Menggunakan Metode Naïve Bayes. *KLIK-KUMPULAN JURNAL ILMU KOMPUTER*, 8(2), 199-208.
- Attabi, A. W. (2018). Penerapan Analisis Sentimen untuk Menilai Suatu Produk pada *Twitter* Berbahasa Indonesia dengan Metode Naïve Bayes Classifier dan Information Gain (Doctoral dissertation, Universitas Brawijaya).
- Fauziyyah, A. K. & Gautama, D. H. (2020). Analisis sentimen pandemi Covid19 pada streaming *Twitter* dengan *Text Mining* Python. *Jurnal Ilmiah SINUS*, 18(2), 32.
- Harijiatno S. D. (2018). Analisis Sentimen *Twitter* Menggunakan Multinomial Naïve Bayes (Skripsi, Universitas Sanata Dharma).
- Iskandar, J. W., & Nataliani, Y. (2021). Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(6), 1120-1126.
- Jollyta, D., Ramdhan, W., & Zarlis, M. (2020). Konsep Data *Mining* Dan Penerapan. Deepublish.
- Karsito, K., & Susanti, S. (2019). Klasifikasi Kelayakan Peserta Pengajuan Kredit Rumah Dengan Algoritma Naïve Bayes Di Perumahan Azzura Residencia. *Jurnal SIGMA*, 9(3), 43-48.
- Kementerian Kesehatan RI. (2021). Data Sebaran. Retrieved from Satuan Tugas Penanganan COVID-19: <https://covid19.go.id/>
- Laurensz, B., & Sedyono, E. (2021). Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 10(2), 118-123.
- Mohan, L., Pant, J., Suyal, P., & Kumar, A. (2020, September). Support vector machine accuracy improvement with classification. In *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)* (pp. 477-481). IEEE.
- Naraswati, N. P. G., Nooraeni, R., Rosmilda, D. C., Desinta, D., Khairi, F., & Damaiyanti, R. (2021). Analisis Sentimen Publik dari *Twitter* Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan *Naive Bayes* Classification. *Sistemasi: Jurnal Sistem Informasi*, 10(1), 222-238.
- Nursyi'ah, S. Y., Erfina, A., & Warman, C. (2021). Analisis Sentimen Pembelajaran Daring Pada Masa Pandemi Covid-19 Di *Twitter* Menggunakan Algoritma Naïve Bayes. In *Seminar Nasional Sistem Informasi dan Manajemen Informatika Universitas Nusa Putra* (Vol. 1, No. 01, pp. 117-123).
- Prasetyo, E. (2019). Data *Mining* Mengolah Data menjadi Informasi Menggunakan Matlab. Andi.
- Sabrila, T. S., Sari, V. R., & Minarno, A. E. (2021). Analisis Sentimen Pada *Tweet* Tentang Penanganan Covid-19 Menggunakan Word Embedding Pada Algoritma *Support Vector Machine* Dan K-Nearest Neighbor. *Fountain of Informatics Journal*, 6(2), 70.
- Santoso, G. T. (2018). Penerapan Analisis Sentimen untuk Menilai Suatu Produk pada *Twitter* Berbahasa Indonesia dengan Metode Naïve Bayes Classifier dan Information Gain (Doctoral dissertation, Universitas Brawijaya).
- Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd. Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 10(2), 681-686.
- Syarifuddin, M. (2020). Analisis Sentimen Opini Publik Mengenai Covid-19 Pada *Twitter* Menggunakan Metode Naïve Bayes Dan Knn. *Inti Nusa Mandiri*, 15(1), 23-28.
- Tuhuteru, H. (2020). Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma *Support Vector Machine*. *Journal Information System Development (ISD)*, 5(2).
- Widowati, T. T., & Sadikin, M. (2020). Analisis Sentimen *Twitter* terhadap Tokoh Publik dengan Algoritma *Naive Bayes* dan *Support Vector Machine*. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 11(2), 626-636.
- Wulandari, D. A., Saedudin, R. R., & Andreswari, R. (2021). Analisis Sentimen Media Sosial *Twitter* Terhadap Reaksi Masyarakat Pada Ruu Cipta Kerja Menggunakan Metode Klasifikasi Algoritma *Naive Bayes*. *eProceedings of Engineering*, 8(5).
- Yuliana. (2020). Corona Virus Diseases (Covid -19); Sebuah Tinjauan Literatur. *Wellness And Healthy Magazine*, 2(1), 187.
- Yulita, W & Nugroho E. D. (2021). Analisis Sentimen Terhadap Opini Masyarakat tentang Vaksin COVID-19 Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal Data Mining dan Sistem Informasi*, 2(2), 1.