

Analisis Sentimen Masyarakat Terhadap Kebijakan Pemerintah Selama Pandemi Covid-19 Menggunakan Algoritma Naïve Bayes

Muhammad Aqil Emeraldi ¹, Inna Ekawati ¹, Malikus Sumadyo ^{1,*}

* Korespondensi: e-mail: malikus.sumadyo@gmail.com

¹ Teknik Komputer; Universitas Islam 45; Jalan Cut Meutia No.83 Bekasi Timur, Kota Bekasi; e-mail: syahruli917@gmail.com, inna.ekawati@gmail.com, malikus.sumadyo@gmail.com

Submitted : 31 Maret 2022
Revised : 22 April 2022
Accepted : 5 Mei 2022
Published : 30 Mei 2022

DOI:
<https://doi.org/10.31599/jsr.cs.v3i1.1513>

Abstract

The increase in data is very large, one of the sources comes from social media, especially Twitter which talks a lot about Covid-19. The news through Twitter media regarding the impact of the Covid-19 virus is widely discussed because it causes unrest for the public which has led to the issuance of various government policies with the aim of preventing the spread of Covid-19. Related to this, it is necessary to conduct a sentiment analysis of the text contained in the Twitter media. In this study, a sentiment analysis process was carried out related to public sentiment towards government policies during the Covid-19 pandemic in Indonesia on Twitter social media using the Naive Bayes Classifier method where the data used was classified into 2 sentiment values, namely positive and negative sentiment. The data used are 300 positive tweets data and 300 negative tweets data, where 80% of the total data is used as training data and 20% data is used as test data. Based on the test results, the data with a total of 120 tweets obtained the results of measuring the recall value of 93.33%, precision 93.33%, F-Score 93.33% and an average accuracy of 93.33%.

Keywords: Covid-19, Naive Bayes Classification, Sentiment Analysis.

Abstrak

Pertambahan data yang sangat banyak, salah satu sumbernya berasal dari media sosial khususnya *Twitter* yang banyak membahas *Covid-19*. Pemberitaan melalui media *Twitter* mengenai dampak dari virus *Covid-19* marak dibicarakan karena menimbulkan keresahan bagi masyarakat yang menyebabkan dikeluarkannya berbagai kebijakan pemerintah dengan tujuan untuk mencegah penyebaran *Covid-19*. Terkait hal tersebut perlu dilakukan sentimen analisis terhadap teks yang terdapat pada media *Twitter*. Pada penelitian ini, dilakukan proses sentimen analisis terhadap sentimen masyarakat terhadap kebijakan pemerintah selama pandemi *Covid-19* di Indonesia pada sosial media *Twitter* dengan menggunakan metode *Naive Bayes Classifier* dimana data yang digunakan digolongkan menjadi 2 nilai sentimen yaitu sentimen positif dan negatif. Data yang digunakan sebanyak 300 data *tweets* positif dan 300 data *tweets* negatif, dimana 80% dari keseluruhan data digunakan sebagai data latih dan 20% data digunakan sebagai data uji. Berdasarkan hasil pengujian, data dengan jumlah sebanyak 120 *tweet* diperoleh hasil pengukuran nilai recall 93.33%, precision 93.33%, F-Score 93.33% serta rata-rata akurasi 93.33%.

Kata kunci: Analisis Sentimen, *Covid-19*, Klasifikasi *Naive Bayes*.

1. Pendahuluan

Data teks merupakan salah satu yang dapat dijadikan sumber potensi untuk menggali suatu informasi yang diinginkan, salah satu contohnya adalah data *text* yang diambil dari media sosial *Twitter*. *Twitter* adalah salah satu *platform* sosial media *microblogging* yang dapat digunakan untuk menyebarkan atau membicarakan informasi hingga ratusan karakter yang dikenal sebagai *Tweets* (Olhang et al., 2020). *Tweets* yang sedang ramai saat ini diperbincangkan adalah terkait informasi virus *Covid-19*. *Covid-19* adalah penyakit menular yang baru ditemukan disebabkan oleh jenis *coronavirus*. Saat ini *Covid-19* menjadi wabah di banyak negara sebagai pandemi global telah menginfeksi satu juta manusia lebih di seluruh dunia (Rifai et al., 2021). Respon masyarakat terkait dampak dari virus *Covid-19* yang menimbulkan keresahan dan berbagai kebijakan pemerintah yang bertujuan untuk mencegah penyebaran *Covid-19* juga ramai dibicarakan pada media *Twitter* sebagai media sosial sarana berinteraksi masyarakat. Kebijakan pemerintah saat ini juga menjadi sorotan publik dalam media sosial *Twitter*. Berbagai kebijakan-kebijakan pemerintah untuk menangani *Covid-19* menimbulkan berbagai macam opini publik, mulai dari pujian, kritik, saran, dan sindiran (Krisdiyanto & Nurharyanto, 2021; Yulianita et al., 2020) seperti kebijakan pemberlakuan Pembatasan Sosial Berskala Besar (PSBB), Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) dan juga tindakan Vaksinasi untuk mengatasi *Covid-19*. Terkait kebijakan pemerintah tersebut dapat dikumpulkan data teks berupa opini pada *platform Twitter* dan mengolahnya untuk dilakukan sentimen analisis terkait dengan sentimen masyarakat terhadap kebijakan pemerintah selama pandemi *Covid-19*. Terdapat beberapa penelitian yang sudah pernah dilakukan berkaitan dengan analisis sentimen.

Penelitian pada 2020 oleh Olhang dkk terkait menganalisis sentimen masyarakat terhadap aspirasi yang disampaikan melalui *Twitter* mengenai dampak dari adanya *Covid-19* dengan menggunakan metode *Naïve bayes classifier* untuk mengklasifikasikan sentimen. Penelitiannya berupa *tweet* yang diperoleh dari *Twitter* menggunakan *keyword #coronavirusindonesia* atau *#Covid-19* dengan jumlah data 600 data *tweet*, pengelompokkan sentimen positif dan negatif dari setiap *tweet* yang sudah melewati tahap *preproceessing*, hasil pengujian diperoleh hasil pengukuran akurasi *recall* 32%, *precision* 80%, *F-Measure* 45% serta rata-rata akurasi 36% (Olhang et al., 2020). Penelitian pada 2020 oleh Tuhuteru terkait pelaksanaan Pembatasan Sosial Berskala Besar (PSBB) menjadi polemik bagi sebagian masyarakat Ambon terutama yang tidak memiliki penghasilan tetap. Banyak komentar terhadap rencana penerapan PSBB di Kota Ambon melalui *tweet* dan komentar di *platform* media sosial. Data yang didapatkan pada penelitian sejumlah 1075 *tweet* dan komentar dan dipisahkan antara data *training* sebanyak 350 dan data *testing* sebanyak 725 komentar. Hasil dari penelitian menunjukkan sentimen positif sebesar 28%, sentimen negatif sebesar 27%, dan sentimen netral sebesar 45%. Disimpulkan sentimen masyarakat terhadap rencana penerapan PSBB di Kota Ambon di komentar media sosial cukup berimbang antara sentimen positif, negatif dan dominasi sentimen netral (Tuhuteru, 2020). Penelitian pada 2021 oleh Himawan

dan Eliyani terkait kebijakan pada masa pandemi dengan informasi dalam media sosial seperti *Twitter* sebagai umpan balik untuk mengetahui opini masyarakat terhadap kebijakan publik melalui analisis sentimen *tweet* dari masyarakat dijadikan sebagai salah satu parameter penunjang bagi pemerintah dalam mengevaluasi kebijakan dan pengambilan keputusan mendatang. Penelitiannya untuk mengetahui sentimen data *tweet* terhadap akun *Twitter* resmi Pemerintah Provinsi DKI Jakarta di masa pandemi *Covid-19* dengan *query* pada *tweet* yang mengandung kata atau menyebut *username* @dkijakarta, berdasarkan kelas sentimen negatif, netral, dan positif dengan menggunakan *TF-IDF Vectorizer* diperoleh hasil akurasi sebesar 75,81%, algoritma *naive bayes* dengan hasil akurasi 75,22%, dan algoritma *support vector machine* 77,58%. Dilakukan proses analisis sentimen pada *tweet* dengan presentase hasil negatif, netral, dan positif masing-masing yaitu 8,8%, 83,6%, 7,6% (Himawan & Eliyani, 2021). Penelitian pada 2021 oleh Laurensz & Sedyono terkait vaksinasi untuk mengatasi pandemi *Covid-19*, tanggapan dari masyarakat beragam di media sosial *Twitter*, ada yang mendukung dan ada yang tidak setuju. Data yang digunakan sebanyak 845 *tweet*, dengan menggunakan dua kata kunci, yaitu “vaksinmerahputih” dan “vaksinsinovac”, dengan metode SVM dan *Naïve Bayes* diperoleh hasil pada metode *Naïve Bayes* dengan kata kunci “vaksinsinovac” dan “vaksinmerahputih” memperoleh sentimen positif dari masyarakat sebanyak 66% dan 98%. Metode SVM dengan kata kunci “vaksinsinovac” dan “vaksinmerahputih” mendapatkan sentimen positif 96% dan 98% (Laurensz & Sedyono, 2021).

Berdasarkan beberapa penelitian terdahulu yang telah dilakukan melalui *platform* media sosial *Twitter* terhadap kebijakan pemerintah untuk mengatasi pandemi *Covid-19*, diperoleh sentimen dari masyarakat berupa positif, negatif dan netral. Hasil penelitian terdahulu juga dapat di jadikan pendukung dalam pengambilan keputusan pemerintah sesuai kebijakan yang akan diberlakukan. Demikian juga dengan penelitian ini dilakukan melalui media sosial *Twitter*, untuk menganalisis sentimen masyarakat terhadap kebijakan pemerintah dalam skala lebih luas yang diberlakukan di seluruh daerah Indonesia selama pandemi *Covid-19*, yaitu menggunakan berbagai *keyword* berbahasa Indonesia terkait *Covid-19* seperti #jagajarak, #Covid, #Covid19, #kitajagakita, #pakaimasker, #ingatpesanibu, #cucitangan, #pandemi, #ppkm, #psbb, #dirumahaja, #karantina, #lockdown. Tujuan penelitian untuk mengetahui hasil nilai akurasi dari menggunakan algoritma *Naïve bayes classifier* dalam pengklasifikasian data dari sentimen masyarakat terhadap kebijakan pemerintah selama pandemi *Covid-19*.

Sentimen Analisis

Analisis sentimen merupakan cara mengumpulkan pendapat khalayak umum menggunakan jejaring sosial, dengan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang didalamnya mengandung pelayanan umum, serta isu terkini (Sudiantoro & Zuliarso, 2018; Syarifuddin, 2020). Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan tanggapan terhadap sebuah masalah salah satunya bagaimana menentukan suatu *tweet* bersentimen positif, negatif atau netral (Ardiani et al., 2020). Dalam analisis sentimen terdapat tiga level yaitu level dokumen diekspresikan dalam

positif, negatif atau kelas pendapat, level kalimat diekspresikan dengan subjek atau objek dalam kalimat menjadi positif atau negatif, dan level aspek memiliki entitas di atas level kalimat seperti paragraf (Arsi et al., 2021). Setiap *tweet* merupakan pendapat masyarakat tidak mengandung makna yang selaras, sehingga perlu dilakukan analisis di *Twitter* mengenai *Covid-19* dengan klasifikasi positif dan negatif. Dalam melakukan sentimen analisis kita membutuhkan adanya metode klasifikasi agar dapat mengklasifikasikan sentimen dari opini dokumen *Tweets* yang tersedia, salah satu metode klasifikasi yang memiliki kemampuan yang cukup baik untuk melakukan klasifikasi data teks serta mudah diimplementasikan dengan strukturnya yang cukup sederhana dan memiliki tingkat efektifitas yang tinggi adalah *Naïve bayes classifier* (Marga et al., 2020; Olhang et al., 2020).

Preprocessing Text

Preprocessing digunakan sebagai proses untuk mempersiapkan data mentah sebelum dilakukannya proses lain, *preprocessing* data dilakukan untuk menghilangkan gangguan pada data agar lebih mudah diproses oleh sistem (Syakuro, 2017). Terdapat beberapa macam proses pada *preprocessing*, diantaranya: a) *Cleansing*, proses yang dilakukan untuk menghilangkan karakter-karakter tertentu pada teks yang dapat mengganggu pemrosesan pada data; b) *Case Folding*, proses yang digunakan untuk mengubah data teks keseluruhan ke huruf kecil; c) *Stopword*, proses untuk menghilangkan teks yang tidak memberikan arti sehingga dimensi teks akan berkurang tanpa mengurangi isi sentimen dari data tersebut (Harjanta, 2015); d) *Stemming*, proses untuk merubah data yang berisi teks yang memiliki kata yang berimbuhan dengan menghilangkan awalan atau akhiran untuk memperoleh kelompok kata yang serupa (Syadid, 2019).

2. Metode Penelitian

Metode dalam penelitian akan dijabarkan *Naïve bayes classifier*, tahapan yang dilakukan dalam penelitian, dan penyusunan program.

Naïve bayes classifier

Naive bayes merupakan sebuah metode yang memanfaatkan probabilitas dan statistik, pada pengklasifikasiannya dalam menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari *dataset* yang sudah diberikan sebelumnya. Algoritma yang digunakan merupakan *teorema bayes* dan mengasumsikan semua atribut tidak saling bergantung atau independen terhadap nilai yang diberikan pada variabel kelas. Keuntungan *Naive bayes* hanya membutuhkan jumlah *training data* yang kecil untuk menentukan parameter penilaian dalam proses pengklasifikasian, bekerja lebih baik dalam kebanyakan situasi di dunia nyata yang kompleks. Tahapan dari proses algoritma *naive bayes* menghitung jumlah kelas/label, menghitung jumlah kasus per kelas, kalikan semua variabel kelas, dan bandingkan hasil per kelas (Bajabir, 2018).

Pada metode *naive bayes* terlebih dahulu mencari nilai probabilitas dan *likelihood* maksimum dari masing-masing atribut untuk setiap kelas. Evaluasi performa dilakukan untuk

menguji hasil kinerja dari klasifikasi dengan mengukur nilai performa dari sistem yang sudah dibuat. Tabel 1 menampilkan sebuah *confusion matrix* untuk pengklasifikasian kedalam dua buah kelas (Syakuro, 2017).

Tabel 1. *Confusion Matrix*

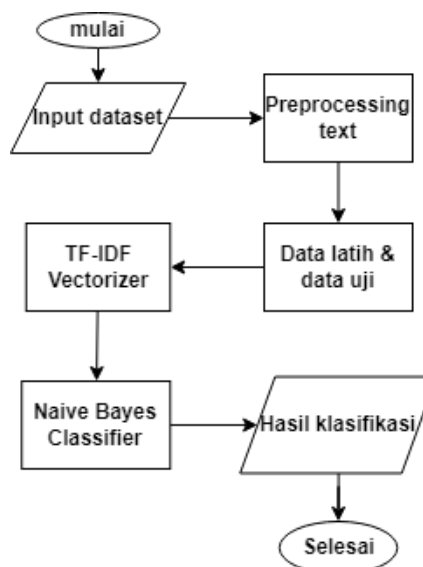
Predicted Class	True Class	
	Positif	Negatif
Positif	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
Negatif	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Sumber: Syakuro (2017)

True Positive (TP) ketika kelas yang diprediksi positif dan faktanya positif, *True Negative (TN)* ketika kelas yang diprediksi negatif, dan faktanya negatif, *False Positive (FP)* ketika kelas yang diprediksi positif dan faktanya negatif dan *False Negative (FN)* ketika kelas yang diprediksi negatif dan faktanya positif. Berdasarkan Tabel 1 nilai dari matriks tersebut dapat dijadikan acuan dalam perhitungan evaluasi (Olhang et al., 2020) yaitu *Precision* perbandingan antara *True Positive (TP)* dengan banyaknya data yang diprediksi positif, *Recall* perbandingan antara *True Positive (TP)* dengan banyaknya data yang sebenarnya positif, *F1-Score* atau dikenal juga dengan nama *F-Measure* merupakan perbandingan rata-rata presisi dan *recall* yang dibobotkan, *Accuracy* merupakan rasio klasifikasi yang benar (positif dan negatif) dengan keseluruhan data.

Tahapan Penelitian

Tahapan penelitian ditujukan untuk mempermudah dalam pengembangan penelitian. Tahap dimulai dengan menyusun rancangan proses yang diperlukan dalam pengembangan penelitian sebagaimana pada Gambar 1.



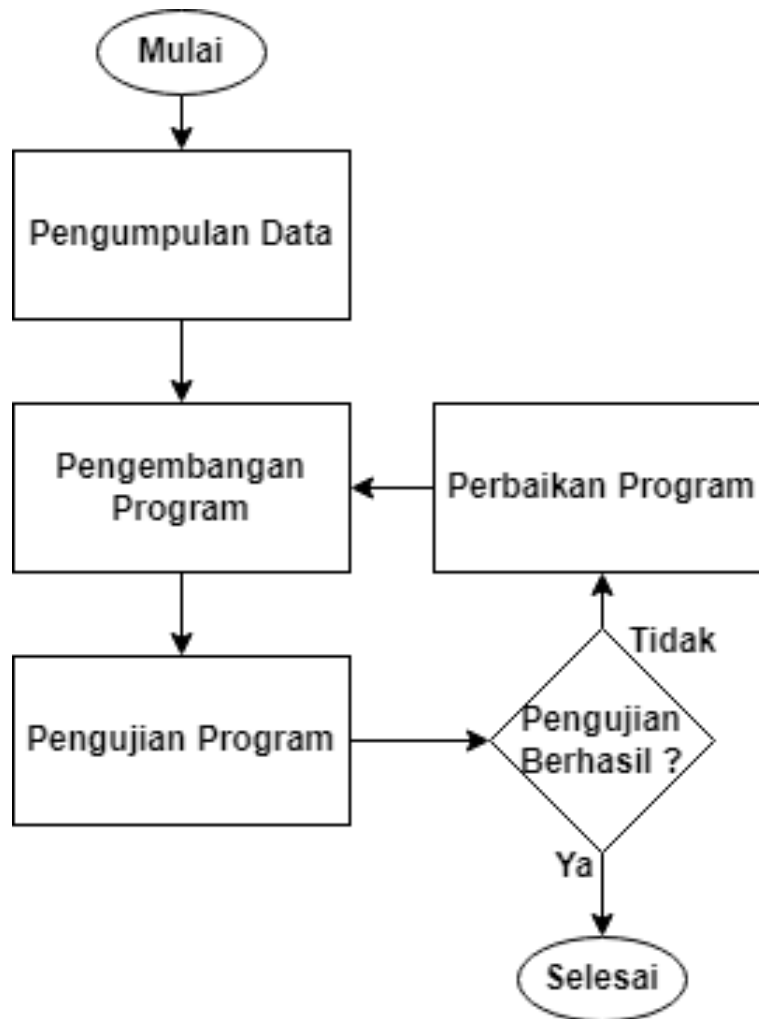
Sumber: Hasil Penelitian (2022)

Gambar 1. *Flowchart* Penelitian

Berdasarkan Gambar 1, diagram alur penelitian yaitu a) Menyiapkan *dataset* yang akan digunakan; b) Memasukkan *dataset* ke proses *preprocessing text* untuk menghilangkan atau membersihkan data dari gangguan seperti merubah semua data ke huruf kecil, menghilangkan tanda baca, menghilangkan angka pada data, menghapus *hyperlink*, *stopword* dan proses *stemming*; c) Memisahkan data kedalam data uji dan data latih; d) Memasukkan data kedalam proses *TF-IDF* untuk pembobotan kata; e) Memasukkan data kedalam proses klasifikasi *naïve bayes* untuk dipelajari; f) Menampilkan hasil klasifikasi.

Pembuatan Program

Dalam penyusunan program analisis sentimen terdapat lima tahapan, ditunjukkan pada Gambar 2, ada pun penjabaran Gambar 2 dijelaskan pada bagian hasil dan pembahasn.



Sumber: Hasil Penelitian (2022)

Gambar 2. *Flowchart* Pembuatan Program

3. Hasil dan Pembahasan

Sebagaimana pada Gambar 2, tahapan pembuatan program berikut uraian dalam penelitian yang dilakukan.

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian terkait sentimen masyarakat terhadap kebijakan pemerintah selama pandemi *Covid-19* di Indonesia yang diambil dari media sosial *Twitter* (<http://Twitter.com>) dan situs Kaggle (<http://www.kaggle.com>). Pada pengambilan data yang pertama diambil dari situs *Twitter* dikumpulkan dengan menggunakan program *scrapping* dengan bantuan *API* milik *Twitter*, cara yang kedua data yang diambil pada situs *kaggle* diambil dengan cara manual (*download*), dimana data yang diambil dari kedua langkah tersebut memiliki kata kunci terkait *Covid-19*. Setelah itu data yang sudah terkumpul akan melewati beberapa langkah, diantaranya: a) Data di-*filter* kembali dengan hanya mengambil data yang memiliki indikasi terkait dengan reaksi masyarakat terhadap kebijakan pemerintah; b) Lalu data yang berupa *tweets* yang sudah di-*filter* sebelumnya belum memiliki nilai sentimen maka data akan dilabeli kedalam dua kelas yaitu sentimen positif dan sentimen negatif.

3.2. Pengembangan Program

Tahap pengembangan program dilakukan dengan cara membuat sekaligus mengembangkan sebuah program analisis sentimen yang mengadaptasi metode ataupun diagram alur yang sudah dirancang dalam tahapan perancangan sebelumnya. Dalam pembuatannya, program dibuat dengan menggunakan bahasa pemrograman *Python* versi 3 serta dalam pembuatannya peneliti menggunakan beberapa *library* untuk mempercepat dan membantu dalam pembuatan program ini, seperti *library pandas* untuk membaca *dataset*, *library Sastrawi* untuk proses *stopword* dan *stemming*, *library sklearn* untuk memisahkan data uji dengan data latih, menggunakan proses pembobotan dengan *tf-idf* dan penggunaan metode klasifikasi dengan *naïve bayes*. Dalam penulisan baris programnya peneliti menggunakan situs penyedia layanan dari *google* yaitu *google colab* sehingga mempermudah dalam mengembangkan program.

3.3. Pengujian Program

Pada tahap pengujian dilakukan uji coba terhadap data yang telah dikumpulkan menggunakan program analisis sentimen dibuat untuk mengetahui seberapa baik hasil klasifikasi dari data yang diuji coba. Proses pengujian ini dilakukan melalui beberapa tahap, yakni *input* data, proses data, dan *output* data. Input data yaitu sebuah tahap meng-*input* data-data yang diperlukan dalam program ini, yakni data *tweets* beserta kelas sentimennya. Setelah data diinput dalam program, lalu data akan melewati tahap proses data. Dimana pada tahap ini data akan dimasukkan dalam proses *preprocessing* untuk membersihkan data dari duplikasi, simbol atau penggunaan kata yang bisa memberikan gangguan pada data sebelum diklasifikasikan, setelah itu data akan dipisah menjadi data uji dan data latih, lalu data yang telah dipisah akan diberi bobot nilai pada proses *TF-IDF* sebelum dimasukkan pada model untuk dianalisa dengan menggunakan klasifikasi *naïve bayes* sehingga akan menghasilkan sebuah hasil akurasi dari model yang telah dibuat. Setelah tahap proses data, selanjutnya data akan ditampilkan melalui tahap *output* data, tahapan ini sebagaimana pada Gambar 3.



Sumber: Hasil Penelitian (2022)

Gambar 3. Blok Diagram Program

Berdasarkan pada Gambar 3, berikut tahapan mulai dari memasukkan data set, *preprocessing*, memisahkan data latih dan data uji, *TF-IDF*, *Naive Bayer* sampai dengan menampilkan hasil.

Input

Pengujian terhadap data yang telah dikumpulkan menggunakan program analisis sentimen yang telah dibangun untuk menghasilkan model probabilitas dari data *training*, serta menguji akurasi model probabilitas dengan data *testing* setelah semua rancangan selesai disusun dan program juga sudah dibangun. Pada pengujian ini digunakan 300 data *tweets* positif dan 300 data *tweets* negatif untuk dilatih dengan *classifier* sehingga total data untuk *classifier* berjumlah 600 data dan telah dilabeli sentimennya, Tabel 2 adalah data yang akan digunakan pada proses pengujian ini.

Tabel 2. Contoh Data *Tweets*

No.	<i>Tweets</i>	Sentimen
1.	Siapa yang jadi korban atas lucu-lucuan pemerintah dalam menangani <i>Covid-19</i> ? Yak rakyat lah! @FraksiRakyat_ID @BEMUNJ_OFFICIAL @aliansibem_si #TolakOmnibusLaw#AtasiVirusCabutOmnibus#LockdownDPR#FokusTanganiCov id19 #MayDay2020 pic.twitter.com/zudNaRoM0M	Negatif
2.	patuhi anjuran pemerintah, JANGAN MUDIK DAHULU, bersama kita cegah penyebaran covid 19. #lebihbaikdirumahaja pic.twitter.com/04UaKyfs5N	Positif
3.	#NetizenBekasi selalu patuhi himbauan pemerintah demi terputusnya mata rantai <i>COVID-19</i> #PSBBBekasi#BersamaLawanCorona#dirumahaja https://twitter.com/Rahmat_Effendi_/status/1256814376167718912	Positif

Sumber: Hasil Penelitian (2022)

Proses

Sesudah data disiapkan maka data akan masuk kedalam beberapa proses, diantaranya:

- a. *Text-preprocessing*, yang meliputi *case folding*, menghilangkan *hashtag*, menghilangkan *mentions*, menghilangkan tanda baca, menghilangkan digit/angka (*cleansing*) yang dijelaskan pada Tabel 3.

Tabel 3. *Sample Text Preprocessing (Cleansing)*

No.	Sebelum <i>cleansing</i>	Setelah <i>cleansing</i>
1.	Siapa yang jadi korban atas lucu-lucuan pemerintah dalam menangan <i>Covid-19</i> ? Yak rakyat lah! @FraksiRakyat_ID@BEMUNJ_OFFICIAL @aliansibem_si #TolakOmnibusLaw#AtasiVirusCabutOmnibus #LockdownDPR #FokusTanganiCovid19 #MayDay2020 pic.twitter.com/zudNaRoM0M	siapa yang jadi korban atas lucu lucuan pemerintah dalam menangan covid yak rakyat lah.
2.	patuhi anjuran pemerintah, JANGAN MUDIK DAHULU, bersama kita cegah penyebaran covid 19. #lebihbaikdirumahaja pic.twitter.com/04UaKyfs5N	patuhi anjuran pemerintah jangan mudik dahulu bersama kita cegah penyebaran covid
3.	#NetizenBekasi selalu patuhi himbauan pemerintah demi terputusnya mata rantai <i>COVID-19</i> #PSBBBekasi #BersamaLawanCorona #dirumahaja https://twitter.com/Rahmat_Effendi_/status/1256814376167718 912ÃfÃfÃ,Ã,ÃfÃ,Ã,Ã ÃfÃfÃ,ÃçÃfÃ,Ã,Ã€ÃfÃ,Ã,Ã!	selalu patuhi himbauan pemerintah demi terputusnya mata rantai covid

Sumber: Hasil Penelitian (2022)

Selanjutnya proses *stopword* untuk menghilangkan kata yang kurang penting seperti *dan*, *yang* dan *dalam* ditunjukkan pada Tabel 4.

Tabel 4. *Sample Stopword*

No.	Sebelum <i>stopword</i>	Setelah <i>stopword</i>
1.	siapa yang jadi korban atas lucu lucuan pemerintah dalam menangan covid yak rakyat lah.	siapa jadi korban atas lucu lucuan pemerintah menangan covid yak rakyat lah.
2.	patuhi anjuran pemerintah jangan mudik dahulu bersama kita cegah penyebaran covid.	patuhi anjuran pemerintah jangan mudik bersama cegah penyebaran covid.
3.	selalu patuhi himbauan pemerintah demi terputusnya mata rantai covid.	selalu patuhi himbauan pemerintah terputusnya mata rantai covid.

Sumber: Hasil Penelitian (2022)

Selanjutnya dilakukan proses *stemming* untuk menghilangkan awalan atau akhiran dari sebuah kata yang memiliki imbuhan, seperti yang ditunjukkan pada Tabel 5.

Tabel 5. *Sample Stemming*

No.	Sebelum stemming	Setelah stemming
1.	siapa jadi korban atas lucu lucuan pemerintah menangani covid yak rakyat lah.	siapa jadi korban atas lucu lucu perintah tangan covid yak rakyat lah.
2.	patuhi anjuran pemerintah jangan mudik bersama cegah penyebaran covid.	patuh anjur perintah jangan mudik sama cegah sebar covid.
3.	selalu patuhi himbauan pemerintah terputusnya mata rantai covid.	selalu patuh himbauan perintah putus mata rantai covid.

Sumber: Hasil Penelitian (2022)

- b. Setelah melalui seluruh rangkaian pada tahap *preprocessing*, data akan dibagi kembali dengan perbandingan 80% data latih dan 20% data uji, sehingga didapatkan 480 data sebagai data latih dan 120 data sebagai data uji.
- c. Proses pembobotan kata dilakukan dengan metode *tf-idf*. Pada proses ini dilakukan perhitungan terhadap banyaknya *keyword* pada *tweets* yang muncul (*TF*), banyaknya *tweets* yang mengandung *keyword* (*DF*), *inverse document frequency* (*IDF*), kemudian mengalikan *TF* dengan *IDF* lalu *smoothing* hasilnya sebagai bobot dari *keyword* pada setiap *tweets*. Hasil pembobotan dapat dilihat pada Tabel 6.

Tabel 6. Hasil Pembobotan Dengan *TF-IDF*

No.	Keyword	Data 1	Data 2	Data 3
1.	Anjur	0.000000	0.370725	0.000000
2.	Atas	0.280632	0.000000	0.000000
3.	Cegah	0.000000	0.370725	0.000000
4.	Covid	0.165746	0.218956	0.235755
5.	himbau	0.000000	0.000000	0.399168
6.	Jadi	0.280632	0.000000	0.000000
7.	Jangan	0.000000	0.370725	0.000000
8.	Korban	0.280632	0.000000	0.000000
9.	Lah	0.280632	0.000000	0.000000
10.	Lucu	0.561265	0.000000	0.000000
11.	Mata	0.000000	0.000000	0.399168
12.	Mudik	0.000000	0.370725	0.000000
13.	Patuh	0.000000	0.000000	0.303578
14.	Perintah	0.165746	0.218956	0.235755
15.	Putus	0.000000	0.000000	0.399168
16.	Rakyat	0.280632	0.000000	0.000000
17.	Rantai	0.000000	0.000000	0.399168
18.	Sama	0.000000	0.370725	0.000000
19.	Sebar	0.000000	0.370725	0.000000
20.	Selalu	0.000000	0.000000	0.399168
21.	Siapa	0.280632	0.000000	0.000000

No.	Keyword	Data 1	Data 2	Data 3
22.	Tangan	0.280632	0.000000	0.000000
23.	Yak	0.280632	0.000000	0.000000

Sumber: Hasil Penelitian (2022)

Berdasarkan hasil pembobotan pada Tabel 6 dapat disimpulkan bahwa kata yang sering muncul dan tersebar di banyak dokumen akan di beri bobot nilai yang lebih kecil dibandingkan kata yang sering muncul pada sedikit dokumen.

- d. Pengujian analisis sentimen dengan jumlah data latih sebanyak 480 data dan data uji sebanyak 120 data dengan menggunakan metode *naïve bayes classifier*.

Setelah proses analisis selesai didapatkan hasil evaluasi yang bertujuan untuk menguji kebenaran hasil klasifikasi dengan menghitung beberapa angka yang diperoleh. Acuan yang digunakan untuk tahap evaluasi yakni dengan perhitungan Tabel kontigensi klasifikasi (*confussion matrix*) seperti yang tertera pada Tabel 7. Berdasarkan Tabel 7, didapatkan data *tweets* dengan prediksi positif dan faktanya positif sebanyak 56 *tweets*, data dengan prediksi positif dan faktanya negatif sebanyak 4 *tweets*, data dengan prediksi negatif dan faktanya negatif sebanyak 56 *tweets*, serta data dengan prediksi negatif dan faktanya positif sebanyak 4 *tweets*.

Tabel 7. Hasil *Confusion Matrix*

Prediksi	True Class		Total
	Positif	Negatif	
Positif	56 (TP)	4 (FP)	
Negatif	4 (FN)	56 (TN)	
Total	60	60	120

Sumber: Hasil Penelitian (2022)

Hasil klasifikasi dengan *naïve bayes* dari data *tweets* sebanyak 120 *tweets* ditunjukkan pada Tabel 8, Berdasarkan hasil nilai yang ditunjukkan pada Tabel 8, dengan data sebanyak 120 *tweets* memperoleh hasil pengukuran nilai *recall* sebesar 93.33%, nilai *precision* sebesar 93.33%, nilai *F-Measure* sebesar 93.33% serta nilai akurasi sebesar 93.33%.

Tabel 8. Hasil Kinerja Algoritma *Naïve Bayes* Dalam Memprediksi Sentimen Masyarakat Terhadap Kebijakan Pemerintah Selama Pandemi *Covid-19*

Measure	Hasil
<i>Precision</i>	0.9333
<i>Recall</i>	0.9333
<i>F1-Score</i>	0.9333
<i>Accuracy</i>	0.9333

Sumber: Hasil Penelitian (2022)

Output

Berdasarkan perhitungan evaluasi hasil klasifikasi pada Tabel 8 dapat disimpulkan bahwa *Naïve Bayes Classifier* sangat baik dalam memprediksi kategori sentimen yang benar

terhadap *tweets* yang diberikan. Nilai rata-rata mendapatkan hasil yang tinggi yaitu 93.33% yang artinya kinerja algoritma *naïve bayes classifier* dapat mengklasifikasi data *tweets* dengan sangat baik.

3.4. Perbaikan Program

Tahap perbaikan, yaitu tahapan untuk memperbaiki dan mengatasi program apabila terdapat kesalahan ataupun kekurangan pada saat dijalankan. Masih terdapat beberapa kekurangan yang dapat mengganggu hasil dari klasifikasi seperti penggunaan fitur *stemming* yang dimana pada penggunaannya digunakan untuk menghilangkan awalan atau akhiran dari kata yang memiliki imbuhan tanpa mengetahui konteks dari kata tersebut berbeda dengan fitur *lemmatization* yang dapat menghilangkan awalan atau akhiran dari kata yang memiliki imbuhan dengan mengetahui konteks dari kata tersebut. Melihat tahun *update* terakhir *library sastrawi* pada 2016 maka diperlukan pemutakhiran pada *library sastrawi*, Pada penelitian ini juga ditemukan masih banyak adanya kemunculan kata yang kurang penting setelah proses *stopword*. Kelemahan yang terakhir yaitu adanya data yang memiliki nilai yang sama (duplikat) yang dapat menyebabkan *overfitting* apabila model klasifikasi yang telah dihasilkan digunakan untuk data yang baru.

4. Kesimpulan

Penelitian ini telah berhasil menerapkan algoritma *naïve bayes* untuk menganalisa sentimen pada data *tweets* mengenai sentimen masyarakat terhadap kebijakan pemerintah selama pandemi *Covid-19*. Hal ini dapat dilihat dari hasil akurasi dengan menggunakan data *tweets* sebanyak 600 data, dimana 480 data adalah data latih dan 120 adalah data uji, hasil menunjukkan bahwa *Naive Bayes Classifier* berhasil mendapatkan nilai akurasi sebesar 93.33% dengan memprediksi kategori sentimen yang benar terhadap *tweets* yang diberikan yang artinya kinerja algoritma *naïve bayes classifier* dapat mengklasifikasi data *tweets* dengan sangat baik.

Daftar Pustaka

- Ardiani, L., Sujaini, H., & Tursina. (2020). Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, 8(2), 183. <https://doi.org/10.26418/justin.v8i2.36776>
- Arsi, P., Kusuma, B. A., & Nurhakim, A. (2021). Analisis Sentimen Pindah Ibu Kota Berbasis Naive Bayes Classifier. *Jurnal Informatika Upgris*, 7(1), 1–6. <https://doi.org/10.26877/jiu.v7i1.7636>
- Bajabir, A. Z. A. M. (2018). Penerapan Metode Naive Bayes untuk Prediksi Menentukan Karyawan Tetap Pada PT. YSP Industries Indonesia [Sekolah Tinggi Teknologi Pelita Bangsa]. <https://ecampus.pelitabangsa.ac.id/pb/AmbilLampiran?ref=22968&jurusan=&jenis=Item&usingId=false&download=false&clazz=ais.database.model.file.LampiranLain>

- Harjanta, A. T. J. (2015). Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining. *Jurnal Informatika UPGRIS*, 1(Juni), 1–9.
- Himawan, R. D., & Eliyani, E. (2021). Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 7(1), 58. <https://doi.org/10.26418/jp.v7i1.41728>
- Krisdiyanto, T., & Nurharyanto, E. M. O. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers. *Jurnal CoreIT.*, 7(1), 32. <https://doi.org/10.24014/coreit.v7i1.12945>
- 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. <https://doi.org/10.22146/jnteti.v10i2.1421>
- Marga, N. S., Isnain, A. R., & Alita, D. (2020). Sentimen Analisis Tentang Kebijakan Pemerintah Terhadap Kasus Corona Menggunakan Metode Naive Bayes. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 2(4), 453–463. <http://jim.teknokrat.ac.id/index.php/informatika/article/view/1602/500>
- Olhang, M. M. M., Achmadi, S., & Ariwibisono, F. . (2020). Analisis Sentimen Pengguna Twitter Terhadap Covid-19 di Indonesia Menggunakan Metode Naïve Bayes Classifier(NBC). (*Jurnal Mahasiswa Teknik Informatika (JATI)*, 4(2), 43–50. <https://doi.org/10.36040/jati.v4i2.2695>
- Rifai, B., Normah, Febryanto, B. D., Yulianto, F., & Reflianah, N. (2021). Analisis Sentimen Opini Publik Terhadap Penerapan Kebijakan Social Distancing Dalam Pencegahan Covid-19. *Paradigma – Jurnal Informatika Dan Komputer*, 23(1), 55–62. <https://doi.org/10.31294/p.v23i1.8756>
- Sudiantoro, A. V., & Zuliarso, E. (2018). Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naïve Bayes Classifier. *Prosiding SINTAK 2018*, 10(2), 398–401.
- Syadid, F. (2019). *Analisis Sentimen Komentar Netizen Terhadap Calon Presiden Indonesia 2019 dari Twitter Menggunakan Algoritma Term Frequency-Invers Document Frequency (TF- IDF) dan Metode Multi Layer Perceptron (MLP) Neural Network* [Universitas Islam Negeri Syarif Hidayatullah Jakarta]. [https://repository.uinjkt.ac.id/dspace/bitstream/123456789/48184/1/FAQI SYADID-FST.pdf](https://repository.uinjkt.ac.id/dspace/bitstream/123456789/48184/1/FAQI%20SYADID-FST.pdf)
- Syakuro, A. (2017). *Pada Media Sosial Menggunakan Metode Naïve Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG)* [Universitas Islam Negeri Maulana Malik Ibrahim]. <http://etheses.uin-malang.ac.id/11706/>
- 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. <https://doi.org/10.33480/inti.v15i1.1347>
- Tuhuteru, H. (2020). Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma Support Vector Machine. *Information System Development (ISD)*, 5(2), 7–13.

Yulianita, T., Utami, T. W., & Haris, M. Al. (2020). Analisis Sentimen Dalam Penanganan Covid-19 di Indonesia Menggunakan Naive Bayes Classifier. *Seminar Nasional Variansi (Venue Artikulasi-Riset, Inovasi, Resonansi-Teori, Dan Aplikasi Statistika)*, 235–243.