PROSIDING Seminar Nasional SISFOTEK ISSN 2597-3584 (Media Online)
Sistem Informasi dan Teknologi
website : www.seminar.iaii.or.id | email : seminar@iaii.or.id

# Analisis Sentimen Pandangan Masyarakat Terhadap *Uji Emisi* di Twitter Menggunakan Metode *Support Vector Machine*

Aziz Musthafa[1], Dihin Muriyatmoko[2], Anggi Fauzan[3*]
[1,2,3]Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Darussalam Gontor Ponorogo
anggifauzan04@mhs.unida.gontor.ac.id

***Abstract***

*Air pollution caused by motor vehicles is a major problem today. To reduce the dangers of air pollution and pollution caused by increasing motor vehicle users, the government immediately enacted legal action against vehicles that did not conduct or pass emission tests. From these government actions there are various kinds of public opinions, especially on social media twitter. This study aims to classify public sentiment towards emission tests expressed on twitter social media by collecting comment data and labeling it as positive, negative, and neutral. The method used in this research with data amounting to 3459 comments uses Support Vector Machine (SVM). Meanwhile, to measure the performance of the SVM method using Confusion Matrix. The results of testing using the SVM method resulted in an accuracy of 76%, and the results of the classification obtained more positive class values. This states that the policies made by the government related to emission testing have received a positive response from the public.*

*Keywords: air pollution, emission test, public sentiment, social media twitter, support vector machine.*

**Abstrak**

Polusi udara yang disebabkan oleh kendaraan bermotor adalah masalah utama saat ini. Untuk mengurangi bahaya polusi udara dan pencemaran yang disebabkan oleh pengguna kendaraan bermotor yang meningkat, pemerintah segera memberlakukan Tindakan hukum terhadap kendaraan yang tidak melakukan atau tidak lulus uji emisi. Dari tindakan pemerintah tersebut terdapat berbagai macam opini Masyarakat khususnya di media sosial twitter. Penelitian ini memiliki tujuan untuk klasifikasi sentimen Masyarakat terhadap uji emisi yang diungkapkan di media sosial twitter dengan mengumpulkan data komentar dan melabelinya menjadi positif, negatif, dan netral. Metode yang digunakan dalam penelitian ini dengan data sebesar 3459 komentar menggunakan *Support Vector Machine* (SVM). Sedangkan untuk mengukur kinerja metode SVM menggunakan *Confusion Matrix*. Hasil dari pengujian menggunakan metode SVM menghasilkan akurasi sebesar 76%, serta hasil dari klasifikasi didapatkan nilai kelas positif lebih banyak. Hal tersebut menyatakan bahwa kebijakan yang dibuat oleh pemerintah terkait uji emisi mendapati respon yang positif dari masyarakat.

Kata kunci: polusi udara, uji emisi, sentimen mayarakat, media sosial twitter, support vector machine.

## 1. Pendahuluan

Lingkungan hidup merupakan tempat manusia, hewan, dan tumbuhan. Habitatnya terdiri dari daratan, perairan, dan udara dan sangat penting bagi keberlangsungan makhluk hidup khususnya manusia. Pemerintahan dan Masyarakat harus terus mempertimbangkan faktor penting perlindungan lingkungan. Namun permasalahan lingkungan hidup masih sering terjadi. pencemaran udara yang disebabkan oleh kendaraan merupakan masalah utama saat ini[1]

Jumlah kendaraan terus meningkat seiring waktu. Pada tahun 2021, jumlah kendaraan naik sekitar 4% dari 136 juta kendaraan pada tahun 2020. Pada tahun 2021, total kendaraan di Indonesia berjumlah 141.992.573, terdiri dari mobil penumpang 16.413.348, bus 237.566, mobil barang 5.299.361, dan sepeda motor 120.042.298, menurut Badan Pusat Statistika (BPS).

Pada industri transportasi, sulit untuk menentukan seberapa besar polusi udara yang dihasilkan oleh kendaraan bermotor. Hal ini disebabkan fakta bahwa alat untuk melacak Tingkat emisi kendaraan tidak tersedia, dan jika ada, biayanya akan mahal, membuatnya tidak ekonomis. Jika pajak emisi, salah satu pajak lingkungan, dapat diberlakukan, hal itu akan mendorong orang untuk membeli mobil baru karena mereka akan menggunakan bahan bakar minyak (BBM) dengan lebih efisien, menggunakan BBM yang lebih bersih, dan mengemudi lebih sedikit[2].

Dalam Upaya mengurangi polusi udara dan ancaman pencemaran akibat meningkatnya penggunaan kendaraan bermotor, pemerintah akan segera mengambil Tindakan terhadap kendaraan yang gagal dalam uji emisi. Sepeda motor, mobil, truk dan kendaraan pribadi adalah jenis yang wajib memenuhi persyaratan uji emisi pasal 47(2) undang- undang lalu lintas jalan no.22 tahun

2009. Menurut pasal tersebut, semua kendaraan harus lolos uji emisi. Namun undang – undang tersebut belum sepenuhnya diterapkan[3].

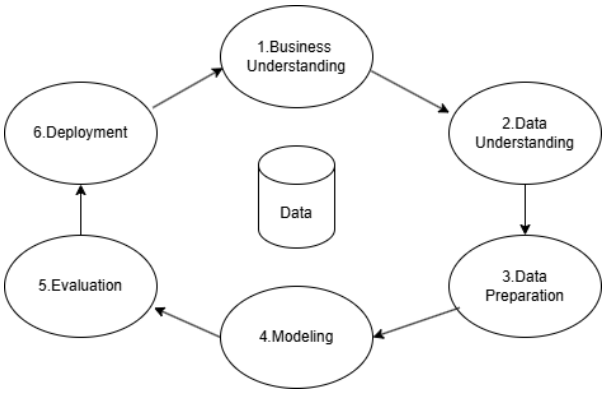
Saat ini pemanfaatan teknologi informasi berkembang sangat pesat. Pemanfaatan teknologi sangat diperlukan untuk mengetahui pandangan sosial Masyarakat Indonesia terhadap uji emisi. Media sosial sebagai salah satu teknologi yang dapat digunakan untuk mendapatkan informasi adalah media sosial twitter. dengan informasi yang pembaruan informasi yang cukup cepat, banyak tweet Masyarakat yang dipublikasikan mengenai uji emisi. Salah satu cara untuk mengetahui presepsi Masyarakat Indonesia terhadap uji emisi adalah dengan menggunakan analisis sentimen.

Penelitian tentang polusi udara dan dampaknya terhadap Kesehatan Masyarakat telah banyak dilakukan. Polusi udara di kota Jakarta yang setiap hari nya meningkat menyebabkan banyak Masyarakat nya bisa terjangkit penyakit gangguan pernapasan, jantung dan menyebabkan kanker[4]. Upaya pemerintah dalam mengatasi masalah ini memberikan kebijakan pajak kendaraan bermotor[5]. dalam permasalahan analisis data, teknik *support vector machine* (SVM) telah terbukti handal dalam klasifikasi sentimen[6].

Tujuan dari penelitian ini adalah untuk mengklasifikasikan pendapat Masyarakat Indonesia tentang uji emisi. Apakah pendapat ini lebih positif, negatif atau netral terhadap kebijakan yang telah dibuat. Data sentimen diambil dari twitter. Hasil penelitian sebelumnya menunjukan bahwa peneliti tertarik menggunakan metode *Support Vector Machine* (SVM) untuk menganalisis opini publik.

hasil dari penelitian ini dapat memberikan kontribusi bagi instansi yang bersangkutan untuk meninjau kebijakan yang akan di buatnya dengan melihat sentimen Masyarakat terlebih dahulu. Dalam konteks keilmuan penelitian ini dapat menambah wawasan terkhusus pada metode *support vector machine*.

## 2. Metode Penelitian




Gambar 1. Tahapan CRISP-DM

Pada penelitian ini menggunakan *Cross Industry Standard Proces for Data Mining* (CRISP-DM). CRISP-DM adalah kerangka kerja umum yang digunakan dalam pemrosesan data untuk membantu penggunaan dalam memahami dan mengelola proyek data mining secara sistematis. Model ini memiliki 6 tahapan seperti yang ditunjukan pada Gambar 1 [7].

### 2.1. *Business Understanding*

Kebijakan uji emisi yang dibuat oleh pemerintahan memberikan Solusi untuk mengurangi polusi udara. Akan tetapi hal tersebut mendapati banyak komentar Masyarakat di media sosial twitter. Komentar – komentar tersebut menciptakan berbagai pandangan dan presepsi yang berbeda – beda. Sebagian ada yang mendukung karena sangat baik manfaatnya, ada juga yang menolak nya karena aturan uji emisi dirasa kurang tepat sasaran dan ada Sebagian yang berada di Tengah – Tengah. Tujuan penelitian ini untuk mengklasifikasikan sentimen Masyarakat terhadap uji emisi berdasarkan komentar twitter. Penelitian ini bisa menjadi acuan bagi pembuat kebijakan berdasarkan analisis data.

### 2.2. *Data Understanding*

Pada tahapan data *understanding*, data dikumpulkan dengan menggunakan *crawling* pada media sosial twitter dengan python. Data yang digunakan dalam penelitian adalah data yang berasal dari media sosial twitter dengan key search “uji emisi”. Data yang berhasil terkumpul berjumlah 5000. Dan data ini memiliki atribut created_at, id_str, full_text, quote_count, reply_count, retweet_count, favorit_count, lang, user_id_str, conversation_id_str, username, tweet_url. Dan memiliki kelas positif, negative, netral.

### 2.3. *Data Preparation*

Pada tahap Data Preparation, Data tweet yang telah dikumpulkan akan melalui beberapa pemrosesan teks (Preprocessing Data) yang terdiri dari *Case folding, Normalisasi*, *Tokenize*, *Stopwords*, dan *Stemming*. Pada tahapan awal dari preprocessing, *case folding* menjadi hal yang pertama dilakukan dalam preprocessing ini karena *Case Folding* merubah semua karakter dalam teks menjadi huruf kecil (*lower case*) untuk memastikan konsistensi dan memfasilitasi perbandingan teks tanpa memperhatikan perbedaan antara huruf besar dan kecil. Setelah melalui tahapan *Case Folding* data kotor akan melanjutkan pada tahapan Normalisasi. Pada tahapan Normalisasi data mentah atau kotor diubah menjadi data yang lebih konsisten dan standar untuk memudahkan analisis dan proses lebih lanjut. Pada Tahapan preprocessing selanjutnya data kotor akan di proses lagi dengan *Tokenizing*. Dalam proses *Tokenizing* teks di pecah menjadi unit – unit yang lebih kecil, yang disebut token agar lebih mudah diproses dalam analisi. Pada tahapan preprocessing selanjutnya data kotor akan diproses dengan *Stopwords*. Pada proses *Stopwords* ini kata – kata umum yang sering muncul dalam teks tetapi tidak memberikan banyak nilai informasi untuk tugas analisis atau pemrosesan teks akan dihilangkan. Menghapus *stopwords* adalah Langkah penting dalam preprocessing teks untuk mengurangi dimensi data dan

meningkatkan efesiensi serta akurasi model analisis. setelah melakukan *Stopwords* lalu dilakukan *Stemming*, Dalam tahapan stemming ini kata – kata dalam teks diubah ke kata dasarnya atau akarnya dengan menghilangkan akhiran atau awalan tertentu. Tujuan dari Stemming adalah mengurangi variasi kata – kata atau entitas yang sama.

Setelah dilakukannya *Preprocessing Data*, data diberi label secara manual menjadi tiga kelas yaitu : *positive*, *negative*, dan *netral* dan melakukan pembobotan dengan menggunakan TF – IDF(*Term Frequency – Invers Document Frequency*). Proses yang pertama adalah *Term Frequency* (TF), Menghitung jumlah kata yang sering muncul dan membagi kata tersebut dengan total kata dalam dokumen ini [8]. nilai TF(*Term Frequency*) dapat dihitung dengan Rumus 1.

$$TF = \frac{(jumlah\ kemunculan\ kata\ dalam\ dokumen)}{(jumlah\ kata\ dalam\ dokumen)} \quad (1)$$

Proses kedua adalah *Inverse Document Frequency* (IDF), IDF memberikan bobot pada setiap kata dalam korpus berdasarkan seberapa unik kata tersebut. Nilai IDF (*inverse document frequency*) dapat dihitung dengan Rumus 2.

$$IDF = log\frac{N}{n} \quad (1)$$

N adalah jumlah dokumen dalam kumpulan dokumen dan n adalah jumlah dokumen yang mengandung kata tersebut.

### 2.4. *Modelling*

Tahapan modeling pada penelitian ini menggunakan *Support Vector Machine*(SVM) karena *support vector machine* berusaha menemukan *Hyperplane* dengan memaksimalkan jarak antar kelas. dengan cara ini, *Support Vector Machine* dapat menjamin kemampuan generalisasi yang tinggi untuk data – data yang akan datang. *Support Vector Machine* hanya menemukan satu *Hyperplane* yang posisinya tepat ditengah - tengah antar dua kelas [9].

*Support vector machine*(SVM) menawarkan beberapa keuntungan dalam analisis sentimen, terutama karena kinerjanya yang baik pada data berdimensi tinggi dan kemampuannya untuk menemukan solusi optimal bahkan ketika data tidak sepenuhnya dapat dipisahkan secara linear. Selain itu, SVM(*support vector machine*) sangat efektif pada ruang fitur yang sangat besar karena hanya sejumlah kecil titik data (*Support Vector*) yang digunakan untuk menentukan *Hyperplane* [10].

Data yang terdapat pada himpunan data latih dinotasikan sebagai $x_i \in R^d$ sedangkan label kelas dinyatakan sebagai $y_i \in \{-1, +1\}$ untuk $i = 1, 2, \ldots, l,$.

$l$ adalah jumlah data. Misalkan kedua kelas -1 dan +1 diasumsikan dapat dipisahkan secara sempurna oleh hyperplane berdimensi d, yang didefinisikan[11].

$$w.x + b = 0 \quad (2)$$

Sebuah data $x_i$ diklasifikasikan sebagai kelas -1 jika data sesuai dengan Rumus 4

$$w.x + b \leq -1 \quad (3)$$

Dan diklasifikasikan sebagai kelas +1 jika seperti Rumus 5

$$w.x + b > 1 \quad (4)$$

w adalah vektor bobot yang menentukan orientasi hyperplane, x adalah vektor fitur dari data input, dan b adalahbias atau intersep

### 2.5. Evaluation

Pengujian pada penelitian ini menggunakan *confusion matrix* untuk mengevaluasi model yang telah dibuat. *Confusion Matrix* adalah Tabel yang berguna untuk mengevaluasi keakuratan prediksi model. Tabel ini memberikan informasi tentang berapa banyak data yang benar – benar diprediksi dengan benar dan berapa banyak data yang diprediksi salah. Dengan menggunakan *Confusion Matrix*, model bisa dioptimalkan dan mengidentifikasikan kesalahan model. Hasil prediksi *confusion matrix* akan disajikan dengan melakukan perbandingan hasil prediksi dan hasil yang diharapkan[12].

Tabel 1. Tabel confusion matrix

| | | Actual | |
|---|---|---|---|
| | | Positive | Negative |
| Prediction | Positive | TP | FP |
| | Negative | FN | TN |

Seperti yang di tunjukan pada Tabel 1 *confusion matrix* memiliki 4 istilah yang penting: TP atau *True Positive* adalah jumlah nilai actual positif yang dilabeli benar oleh classifer; TN atau *True Negative* adalah jumlah nilai actual negative yang dilabeli benar oleh classifer; FP atau *False Postive* adalah jumlah nilai actual negative yang salah dilabeli oleh classifer; FN atau *False Negative* adalah jumlah nilai actual positif yang salah dilabeli oleh classifer.

Melalui proses klasifikasi didapatkan hasil accuracy, precision, recall dan f1 – score. Adapun penjelasan dari masing masing hasil sebagai berikut:

Akurasi adalah hasil jumlah rasio prediksi benar dengan jumlah total prediksi yang ada. Rumus akurasi ditunjukan pada Rumus 6.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

*Recall* adalah perbandingan jumlah item yang relevan diidentifikasi benar. Rumus untuk recall dapat dilihat pada Rumus 7.

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

*Precision* adalah perbandingan jumlah item yang diidentifikasi positif. Rumus untuk precision dapat dilihat pada Rumus 8.

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

*f1 – score* adalah penggabungan dari recall dan precision yang memberikan keseimbangan antara keduanya. Rumus untuk f1 – score dapat dilihat pada Rumus 9.

$$f1 - score = 2 \times \frac{precision \times recall}{precision \times recall} \quad (8)$$

## 3. Hasil dan Pembahasan

Pada tahapan ini akan disampaikan hasil dan pembahasan serta implementasi yang telah di lakukan. Penelitian ini bertujuan untuk mengklasifikasikan opini publik di media sosial Twitter terhadap uji emisi kendaraan dengan menggunakan metode *Support Vector Machine.* Data yang didapatkan dari media social Twitter dikumpulkan menggunakan teknik *crawling* data. Data yang telah terkumpul dibagi menjadi Tiga Atribut *Positive*, *Negative*, dan *Netral*. Acuan evaluasi kinerja model dilakukan dengan melihat nilai akurasi yang dihasilkan dari perhitungan *Confusion Matrix.*

### 3.1. Pengumpulan data

Pada penelitian ini pengumpulan data dilakukan pada google collab dengan memanfaatkan library tweet harvest dengan Bahasa python, data yang digunakan diambil dari media social twitter dengan kata kunci "uji emisi". Data diambil secara acak dari tanggal 1 November 2023 hingga 14 Mei 2024 sebanyak 5000 data dan setelah melalui proses pembersihan secara manual dengan membuang tweet yang duplikat menggunakan Microsoft excel didapati 3459 data pelatihan model dan 50 baru untuk menguji model yang telah di latih. Data yang telah berhasil dikumpulkan bisa dilihat pada Tabel 2.

Tabel 2. hasil pengumpulan data

| Created_at | Tweet | Username |
|---|---|---|
| Thu Jun 20 07:39:15 +0000 2024 | @rapharay Dibawah 1 rich. Tp umumnya mobil di set ke 0 97-98. Klo uji emisi HC sekitar 0-20ppm Klo diatas 1 mesin jd gampang panas. | innovacommunity |
| Wed Jun 19 13:54:15 +0000 2024 | Yang dilakukan untuk mengurangi polusi: Water mist & uji emisi. Padahal ada satu solusi lagi yang bisa sekalian mengurangi kemacetan: Kembalikan ERP Duit buat ERP bisa dialokasikan untuk ongkos transum. Cuma ya gitu kalau eksekusinya ga bener cuma nyusahin warga. https://t.co/Maweaa5Z9a | jhonaldooo |
| Wed Jun 19 12:42:43 +0000 2024 | (5) MOT Test -- kedaraan wajib test uji kelayakan & emisi ke bengkel terdekat tiap tahun yang gagal & gak punya bisa kena denda Rp19 juta! Kebijakan nomor (3) & (4) harusnya bisa jadi quick win buat DKI Jakarta itu juga kalau mau Pak Bud! | dimasku |

### 3.2. Pelabelan data

Tahapan pelabelan data dilakukan setelah proses pengumpulan data. Data yang telah dikumpulkan akan di labelin menjadi 3 label positif, negative, dan netral, Label data ini di tentukan dari tweet atau baris full_text. Untuk tweet yang mendukung atas adanya uji emisi akan di labelin dengan positif, untuk tweet yang menolak atas adanya uji emisi dan mengandung kata – kata yang tidak sopan akan dilabelin dengan negatif, dan untuk tweet yang tidak mengandung hal positif dan hal negatif akan di labelin dengan label netral. Proses pelabelan ini dilakukan secara manual dengan menggunakan aplikasi Microsoft Excel. Dan contoh dari data yang sudah dilabeli bisa di lihat pada Tabel 3.

Tabel 3. Data yang sudah diberi label

| full_text | Sentiment |
|---|---|
| wahai akun yang usernamenya kayak mencet keyboard dari kiri ke kanan... ga sekalian aja gurunya cek uji emisi kah? ini guru-guru udah underpaid mesti ngurus administrasi dll harus juga kah ambil jobdesc verifikator kelayakan armada | negatif |
| trend cumi darat ini keknya harus di uji emisi dah | positif |
| @JantanBinal @innovacommunity emang lu ngerasa pas bayar pajak di suruh uji emisi gas karbon ?? | negatif |

Proses pelabelan data ini memakan waktu selama 3 minggu, selain melakukan pelabelan dilakukan juga pembersihan data secara manual dengan menggunakan Microsoft excel dengan menghapus tweet yang terduplicate. Data yang telah selesai akan dilanjutkan untuk melakukan proses preprocessing.

Tabel 4. jumlah data perkategori kelas

| No | Kelas | Jumlah data |
|---|---|---|
| 1 | Positif | 1002 |
| 2 | Negatif | 1671 |
| 3 | Netral | 786 |
| Total | | 3459 |

Pada Tabel 4 diperlihatkan kelas Positif memiliki 1002 data, kelas Negatif memiliki 1671 data, dan kelas Netral memiliki 786 data.

### 3.3. *Preprocesing*

Data yang telah dilabeli Positif, Negatif, Netral akan melakukan tahapan *preprocessing* terlebih dahulu sebelum diklasifikasikan. Dalam tahapan preprocessing data akan melalui tahapan *Case folding*, *Normalisasi*, *Tokenizing*, *Stopwords*, dan *Stemming.* Proses ini

dilakukan dengan menggunakan google colab dengan memasukan data yang telah diberi label. Berikut adalah detail proses preprocessing yang dilakukan.

*Case folding:* Pada tahapan awal dari *preprocessing*, *case folding* menjadi hal yang pertama dilakukan dalam *preprocessing* ini karena *Case Folding* merubah semua karakter dalam teks menjadi huruf kecil (*lower case*) untuk memastikan konsistensi dan memfasilitasi perbandingan teks tanpa memperhatikan perbedaan antara huruf besar dan kecil.

*Normalisasi:* Setelah melalui tahapan *Case Folding* data kotor akan melanjutkan pada tahapan Normalisasi. Pada tahapan *Normalisasi* data mentah atau kotor diubah menjadi data yang lebih konsisten dan standar untuk memudahkan analisis dan proses lebih lanjut.

*Tokenizing:* Pada Tahapan *preprocessing* selanjutnya data kotor akan di proses lagi dengan *Tokenizing*. Dalam proses *Tokenizing* teks di pecah menjadi unit – unit yang lebih kecil, yang disebut token agar lebih mudah diproses dalam analisi.

*Stopwords:* Pada tahapan *preprocessing* selanjutnya data kotor akan diproses dengan *Stopwords*. Pada proses *Stopwords* ini kata – kata umum yang sering muncul dalam teks tetapi tidak memberikan banyak nilai informasi untuk tugas analisis atau pemrosesan teks akan dihilangkan. Menghapus stopwords adalah Langkah penting dalam preprocessing teks untuk mengurangi dimensi data dan meningkatkan efesiensi serta akurasi model analisis.

*Stemming:* Tahapan Selanjutnya atau tahapan terakhir dalam *preprocessing* adalah *Stemming*. Dalam tahapan *stemming* ini kata – kata dalam teks diubah ke kata dasarnya atau akarnya dengan menghilangkan akhiran atau awalan tertentu. Tujuan dari *Stemming* adalah mengurangi variasi kata – kata atau entitas yang sama. Tabel 5 menunjukkan hasil dari *preprocessing* data yang telah diberikan label.

Tabel 5. Data yang telah di-preprocesing

| No | Full_text | sentiment | casefolding | tokenizing | normalize | stopwords | stemming | Tweet_clean |
|---|---|---|---|---|---|---|---|---|
| 1 | wahai akun yang usernamenya kayak mencet keyboard dari kiri ke kanan... ga sekalian aja gurunya cek uji emisi kah? ini guru-guru udah underpaid mesti ngurus administrasi dll harus juga kah ambil jobdesc verifikator kelayakan armada | negatif | wahai akun yang usernamenya kayak mencet keyboard dari kiri ke kanan ga sekalian aja gurunya cek uji emisi kah ini guru guru udah underpaid mesti ngurus administrasi dll harus juga kah ambil jobdesc verifikator kelayakan armada | ['wahai', 'akun', 'yang', 'usernamenya', 'kayak', 'mencet', 'keyboard', 'dari', 'kiri', 'ke', 'kanan', 'ga', 'sekalian', 'aja', 'gurunya', 'cek', 'uji', 'emisi', 'kah', 'ini', 'guru', 'guru', 'udah', 'underpaid', 'mesti', 'ngurus', 'administrasi', 'dll', 'harus', 'juga', 'kah', 'ambil', 'jobdesc', 'verifikator', 'kelayakan', 'armada'] | wahai akun yang usernamenya seperti mencet keyboard dari kiri ke kanan tidak sekalian saja gurunya cek uji emisi kah ini guru guru sudah underpaid mesti ngurus administrasi dan lain-lain harus juga kah ambil jobdesc verifikator kelayakan armada | akun usernamenya mencet keyboard kiri kanan gurunya cek uji emisi kah guru guru underpaid mesti ngurus administrasi lain-lain kah ambil jobdesc verifikator kelayakan armada | akun usernamenya mencet keyboard kiri kanan guru cek uji emisi kah guru guru underpaid mesti ngurus administrasi lain kah ambil jobdesc verifikator layak armada | akun usernamenya mencet keyboard kiri kanan guru cek uji emisi kah guru guru underpaid mesti ngurus administrasi lain kah ambil jobdesc verifikator layak armada |
| 2 | trend cumi darat ini keknya harus di uji emisi dah | positif | trend cumi darat ini keknya harus di uji emisi dah | ['trend', 'cumi', 'darat', 'ini', 'keknya', 'harus', 'di', 'uji', 'emisi', 'dah'] | trend cumi darat ini kayaknya harus di uji emisi sudah | trend cumi darat kayaknya uji emisi | trend cumi darat kayak uji emisi | trend cumi darat kayak uji emisi |
| 3 | @JantanBinal @innovacommunity emang lu ngerasa pas bayar pajak di suruh uji emisi gas karbon ?? | negatif | jantanbinal innovacommunity emang lu ngerasa pas bayar pajak di suruh uji emisi gas karbon | ['jantanbinal', 'innovacommunity', 'emang', 'lu', 'ngerasa', 'pas', 'bayar', 'pajak', 'di', 'suruh', 'uji', 'emisi', 'gas', 'karbon'] | jantanbinal innovacommunity emang kamu merasa pas bayar pajak di suruh uji emisi gas karbon | jantanbinal innovacommunity emang pas bayar pajak suruh uji emisi gas karbon | jantanbinal innovacommunity emang pas bayar pajak suruh uji emisi gas karbon | jantanbinal innovacommunity emang pas bayar pajak suruh uji emisi gas karbon |

## 3.4 *Wordcloud*

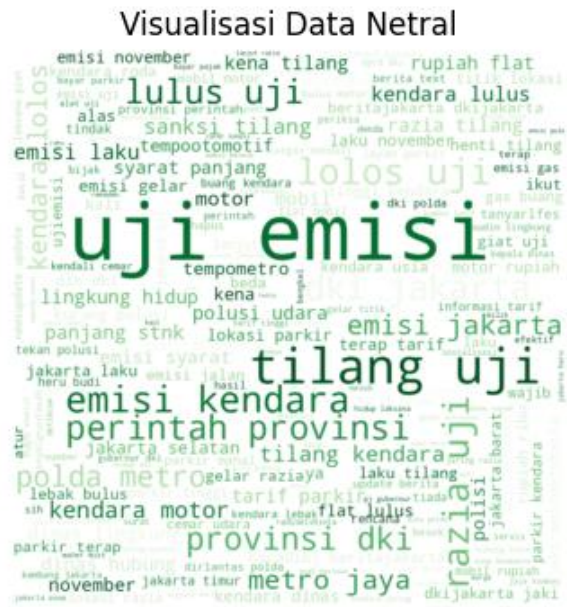
Data yang telah melalui proses *preprocessing* akan direpresentasikan dengan menyoroti seberapa pentingnya kata – kata itu dalam data. Kata – kata yang sering muncul akan tampak lebih pekat warnanya dari lain nya. Representasi ini akan melibatkan 3 kelas yang digunakan dalam penelitian ini yaitu positif, negative, dan juga netral. Hasil dari representasi *Wordcloud* positif ditunjukan pada Gambar 2.

Pada Gambar 2 kendaraan motor memiliki warna yang cukup pekat menandakan bahwa kalimat tersebut sering muncul pada kelas positif. Hasil dari representasi *Wordcloud* negatif ditunjukkan pada Gambar 3


Gambar 2 Visualisasi Data positif


Gambar 3. Visualisasi data negatif

Pada Gambar 3 uji emisi menjadi warna yang paling pekat sekaligus kalimat yang cukup besar pada kelas negatif menandakan kata ini sering muncul pada kelas negatif. Hasil dari representasi Wordcloud netral ditunjukkan pada Gambar 4.

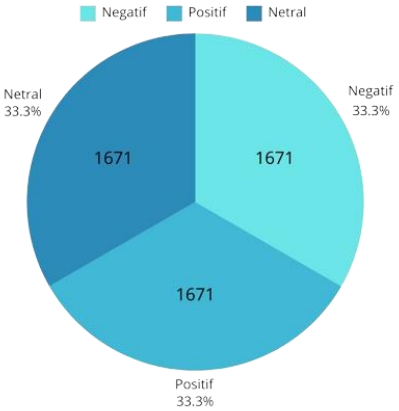



Gambar 4 Visualisasi data netral

Pada Gambar 4 uji emisi dan tilang uji emisi memiliki warna yang cukup pekat menandakan kalimat ini sering muncul pada kelas netral.

3.5 *Oversampling*

Setelah selesainya proses *preprocessing* data akan di klasifikasikan akan tetapi karena data yang ada tidak seimbang maka data akan di proses dengan Teknik *OverSampling*. Teknik *OverSampling* dilakukan untuk mengatasi ketidak seimbangan kelas dalam data. Ketidak seimbangan dapat menyebabkan model pembelajaran lebih cenderung memprediksi kelas mayoritas dan mengabaikan kelas minoritas. Teknik ini meningkatkan jumlah contoh dalam kelas minoritas sehingga distribusi kelas menjadi lebih seimbang. Untuk hasil dari *OverSampling* ini disajikan pada Gambar 5.




Gambar 5. Hasil oversampling

3.6 Pembobotan

Data yang telah melalui proses *preprocessing* dan *OverSampling* akan di beri bobot dengan menggunakan teknik pembobotan TF-IDF. Teknik ini digunakan dalam pemrosesan teks untuk menilai seberapa penting sebuah kata dalam suatu dataset. Teknik ini Merupakan teknik yang umum untuk pemrosesan kata dan penambangan data. Teknik ini dapat mengidentifikasi kata penting dalam data dengan menurunkan bobot kata – kata umum yang sering muncul dalam data sehingga dapat memfokuskan analisis pada kata – kata yang lebih bermanfaat. Adapun hasil dari pemrosesan TF-IDF dapat dilihat pada Gambar 6.

Matrix menunjukan hasil dari Pemrosesan TF-IDF dari berbagai kata – kata yang ada di dalam dokumen. Untuk rincian dari penjelasannya seperti berikut: Angka pertama dalam setiap tuple menunjukan indeks dokumen; Angka kedua dalam setiap tuple menunjukan indeks fitur (kata) dalam matriks fitur; Angka ketiga menunjukan hasil dari TF-IDF dari kata tersebut dalam dokumen.

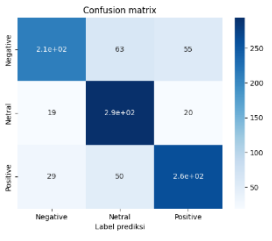
Pada baris pertama matriks menunjukan (0, 6075) 0.20851289461774888. ini menunjukan bahwa Dokumen ke – 0, Kata dalam indeks 6075 dalam daftar semua kata, dan Nilai TF-IDF 0.20851289461774888. Pada penelitian ini terdapat 5013 jumlah dokumen dan 6298 jumlah kata dalam kosa kata yang digunakan dalam pemodelan.

```
TfidfTransformer()
  (0, 6075)     0.20909566376188055
  (0, 6045)     0.20909566376188055
  (0, 6000)     0.20909566376188055
  (0, 5971)     0.02381177140235392
  (0, 3874)     0.19267596709209442
  (0, 3507)     0.15339484067222264
  (0, 3474)     0.20909566376188055
  (0, 3138)     0.1277492631027544
  (0, 3077)     0.1301603541576913
  (0, 2869)     0.20909566376188055
  (0, 2839)     0.20909566376188055
  (0, 2647)     0.18739000054245618
  (0, 2606)     0.28537587821789584
  (0, 2533)     0.20909566376188055
  (0, 1997)     0.5984722708105189
  (0, 1625)     0.023802275064299945
  (0, 1035)     0.11549617270560263
  (0, 340)      0.17346615344201255
  (0, 196)      0.13997771635264283
  (0, 140)      0.17625627042230826
  (0, 54)       0.17346615344201255
  (1, 5971)     0.0693284238511093
  (1, 5847)     0.6087859889869285
  (1, 2707)     0.3971828335652088
  (1, 1625)     0.06930077508283773
  :     :
  (5010, 3109)  0.2107106575295096
  (5010, 2811)  0.28930213387068787
  (5010, 2383)  0.31505171488951
  (5010, 1848)  0.35029595242219624
  (5010, 1783)  0.33832951680444007
  (5010, 1625)  0.041795623724581225
  (5010, 1187)  0.29263159371461606
  (5010, 655)   0.18139759371958444
  (5011, 5971)  0.04894805656099069
  (5011, 5936)  0.2152374100614095
  (5011, 5585)  0.2325208978591487
  (5011, 3533)  0.32019571942842867
  (5011, 2419)  0.27930415099847067
  (5011, 1625)  0.04892853565747058
  (5011, 1541)  0.356581244907204
  (5011, 396)   0.6936192136551225
  (5011, 84)    0.325611839848864
  (5012, 5971)  0.08739299703981668
  (5012, 3931)  0.3428987703110046
  (5012, 3099)  0.6877511929978065
  (5012, 3081)  0.3037825506909465
  (5012, 2785)  0.18041612422347192
  (5012, 2186)  0.39262926164561346
  (5012, 1625)  0.08735814396528406
  (5012, 1388)  0.33934966556066865
(5013, 6298)
```

Gambar 6. Hasil TF - IDF

### 3.7 Hasil Metode Support Vector Machine

Data yang telah dibersihkan, diseimbangkan, dan diberikan bobot akan diklasifikasikan menggunakan metode *support vector machine*(SVM). klasifikasi ini menggunakan data training sebanyak 80% yang berjumlah 4010 data dan data testing sebanyak 20% 1003 data. Data yang sudah siap ini digunakan untuk melatih model. Setelah melakukan pelatihan model, model di evaluasi menggunakan *Confusion Matriks*. Hasil dari *Confusion Matrix* bisa dilihat pada Gambar 7.




Gambar 7. Hasil evaluasi confusion matrix

Pada Gambar 7 menunjukkan bahwa True Negative memiliki jumlah 210, True Netral memiliki jumlah 290, dan true Positive memiliki jumlah 260. Dan juga False Negative pada kelas negative memiliki jumlah 118, pada kelas Netral memiliki jumlah 39, pada kelas Positive memiliki jumlah 79.

Tabel 6. Hasil accuracy confusion matrix

| Kelas | Precision | Recall | F1-score |
|---|---|---|---|
| Negative | 0.82 | 0.64 | 0.72 |
| Positive | 0.78 | 0.77 | 0.77 |
| Netral | 0.72 | 0.88 | 0.79 |
| AVG | 0.77 | 0.76 | 0.76 |

Tabel 6 memperlihatkan rincian hasil *confusion matrix*. Dengan data kelas negative 330, kelas netral 333, dan kelas positif 340 dapat dihasilkan akurasi sebesar 76%.

## 4. Kesimpulan

Pada penelitian ini dapat disimpulkan bahwa metode *Support Vector Machine* memiliki akurasi yang cukup baik dengan akurasi 76% dengan menggunakan data latih sebanyak 4010 atau 80% dari keseluruhan data dan data uji 1003 atau 20% dari keseluruhan data. Dalam proses pengujian model dari 3 data baru yang telah divalidasi oleh validator, di dapati 2 hasil benar dan 1 hasil salah yang menunjukan model ini cukup baik Ketika digunakan untuk klasifikasi. Diharapkan hasil dari penelitian ini dapat memberikan dampak yang baik bagi Masyarakat dan pembuat kebijakan.

## Daftar Rujukan

[1] I. Safira, S. W. Adi, A. Winanti, U. Pembangunan, and N. Veteran, “Efektivitas Peraturan Gubernur Jakarta Tentang Uji Emisi Terhadap Pencemaran Udara Di Dki Jakarta,” *Triwikrama J. Ilmu Sosia*, vol. 01, no. 08, pp. 40–50, 2023.

[2] S. P. Dewi *et al.*, “Pajak Lingkungan Sebagai Upaya Pengendalian Pencemaran Udara Dari Gas Buang Kendaraan Bermotor Di Indonesia,” *J. Ilm. Ekon. dan Pajak*, vol. 2, no. 1, pp. 7–13, 2022, [Online]. Available: https://ojs-ejak.id/index.php/

[3] A. Usman, “SANKSI TIDAK MELAKUKAN ATAU TIDAK LULUS UJI EMISI KENDARAAN BERMOTOR,” Kementerian Hukum dan HAM RI. [Online]. Available: https://bpsdm-dev.kemenkumham.go.id/informasi-publik/publikasi/pojok-penyuluhan-hukum/sanksi-tidak-melakukan-atau-tidak-lulus-uji-emisi-kendaraan-bermotor

[4] J. J. Elizabeth Michelle, Melvin Jusuf, “EFEKTIVITAS PELAKSANAAN KEBIJAKAN BERDASARKAN PERGUB NO 66 TAHUN 2020 TENTANG UJI EMISI KENDARAAN BERMOTOR DI JAKARTA,” *ADIL J. Huk.*, vol. Vol.12 No., no. 66, 2021, [Online]. Available: https://academicjournal.yarsi.ac.id/index.php/Jurnal-ADIL/article/view/1920

[5] M. H. Lazuardi, “Kebijakan Pajak Kendaraan Bermotor, Dikaji Dari Prinsip Pencemar Membayar,” *J. Huk. Lingkung. Indones.*, vol. 7, no. 2, pp. 171–196, 2021, doi: 10.38011/jhli.v7i2.317.

[6] I. Rasyidin Muqsith Rizqi Prasetyo and A. Musthafa, “Comparison between naive bayes method and support vector machine in sentiment analysis of the relocation of the Indonesian capital,” *J. Mantik*, vol. 7, no. 2, pp. 186–193, 2023.

[7] L. Earlene Rendhiva, M. Daivany Nur Auilya Saleh, and R. Koesyan Dipo, “Implementasi Regresi Linier Untuk Waktu Pengiriman Barang,” pp. 600–605, 2023.

[8] R. Wati, S. Ernawati, and H. Rachmi, “Pembobotan TF-IDF Menggunakan Naïve Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH,” *J. Manaj. Inform.*, vol. 13, no. 1, pp. 84–93, 2023, doi: 10.34010/jamika.v13i1.9424.

[9] N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector

Machine," vol. 6, no. 3, 2021, [Online]. Available: http://situs.com

[10] Dr.Suyanto, *DATA MINING UNTUK KLASIFIKASI DAN KLASTERISASI DATA Data Edisi Revisi*, Edisi Revi. Bandung: Informatika Bandung, 2019.

[11] I. W. B. Suryawan, N. W. Utami, and K. Q. Fredlina, "Analisis Sentimen Review Wisatawan pada Objek Wisata Ubud Menggunakan Algoritma Support Vector Machine," *J. Inform. Teknol. dan Sains*, vol. 5, no. 1, pp. 133–140, 2023.

[12] I. Daqiqil Id, *MACHINE LEARNING: Teori, Studi Kasus dan implementasi Menggunakan Pyhton*. Riau,Indonesia: UR PRESS, 2021.