

## **SENTIMENT ANALYSIS OF PUBLIC OPINION ON THE RIGHT OF INQUIRY IN INDONESIA IN 2024 USING THE SUPPORT VECTOR MACHINE (SVM) METHOD**

Dicky Fernanda Sebastian<sup>1</sup>, Heni Sulistiani<sup>2</sup>, Auliya Rahman Isnain<sup>\*3</sup>

<sup>1</sup>Information System, Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia, Indonesia

<sup>2</sup>Accounting Information System, Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia, Indonesia

<sup>3</sup>Informatics, Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia, Indonesia

Email: <sup>1</sup>[dicky\\_fernanda\\_sebastian@teknokrat.ac.id](mailto:dicky_fernanda_sebastian@teknokrat.ac.id), <sup>2</sup>[henisulistiani@teknokrat.ac.id](mailto:henisulistiani@teknokrat.ac.id),

<sup>3</sup>[aulyarahman@teknokrat.ac.id](mailto:aulyarahman@teknokrat.ac.id)

(Article received: April 03, 2024; Revision: May 09, 2024; published: July 29, 2024)

### **Abstract**

Research on the right of inquiry refers to public responses on twitter social media related to the 2024 elections. The right of inquiry is a right used in investigations. There are a lot of public opinions about the right of inquiry that are discussed on twitter social media that convey their various opinions or criticisms of government policies towards the 2024 elections. Based on Law No. 17/2014, the right of inquiry of the House of Representatives is regulated in Article 20A of the 1945 Constitution, which regulates the right of inquiry of the House of Representatives. Sentiment analysis is used in this research to determine the accuracy value of public opinion which is categorized into two, namely positive and negative sentiment. In this study, the SVM method is used to identify and find the results of public opinions or responses regarding the issue of the right of inquiry in Indonesia in 2024 which is being widely under the twitter social media platform, so it is necessary to analyze the sentiment. By using the support vector machine (SVM) algorithm and word weighting using TF-IDF (term frequency-inverse document frequency). Data collection using Google Collaboratory tools with the python programming language. The data used were 2,179 tweets with the keywords "inquiry right", "DPR inquiry right", "election inquiry right". The results obtained from the SVM process with an accuracy value of 77%, negative precision value 77%, positive precision value 77%, negative recall value 57%, positive recall value 89%, positive f1-score value 66%, negative f1-score value 82%. The data that has been tested and processed has an adequate accuracy value for SVM algorithm classification using confusion matrix calculation. The results of the research conducted have been effective with the SVM method.

**Keywords:** Hak Angket, Support Vector Machine, Sentiment Analysis, TF-IDF, Twitter.

## **ANALISIS SENTIMEN OPINI MASYARAKAT MENGENAI HAK ANGKET DI INDONESIA TAHUN 2024 MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)**

### **Abstrak**

Penelitian mengenai hak angket merujuk pada tanggapan masyarakat pada media sosial *twitter* terkait pemilu tahun 2024. Hak angket merupakan hak yang digunakan dalam penyelidikan. Banyak sekali opini masyarakat mengenai hak angket yang menjadi pembahasan pada media sosial *twitter* yang menyampaikan berbagai opini atau kritik mereka terhadap kebijakan pemerintah terhadap pemilu 2024. Berlandaskan pada undang-undang nomor 17 tahun 2014, hak angket DPR pada ketatanegaraan diatur dalam UUD 1945 pasal 20A, dimana tugasnya untuk mengatur mengenai hak angket DPR. Analisis sentimen digunakan pada penelitian ini untuk menentukan nilai *accuracy* dari opini masyarakat yang dikategorikan menjadi dua, yaitu sentimen *positif* dan *negatif*. Dalam penelitian ini digunakan metode SVM untuk mengidentifikasi dan mencari hasil dari opini atau tanggapan masyarakat mengenai isu hak angket di indonesia pada tahun 2024 yang sedang banyak di bawah dalam *platform* media sosial *twitter*, sehingga perlu adanya analisis sentimen tersebut dilakukan. Dengan menggunakan algoritma *support vector machine* (SVM) dan pembobotan kata menggunakan TF-IDF (*term frequency-inverse document frequency*). Pengumpulan data dengan menggunakan *tools google collaboratory* dengan bahasa pemrograman *python*. Data yang digunakan sebanyak 2.179 *tweet* dengan kata kunci "hak angket", "hak angket DPR", "hak angket pemilu". Hasil yang didapatkan dari proses SVM dengan nilai *accuracy* 77%, nilai *precision negatif* 77%, nilai *precision positif* 77%, nilai *recall negatif* 57%, nilai *recall positif* 89%, nilai *f1-score positif* 66%, nilai *f1-score negatif* 82%.

Data yang telah diuji dan diproses tersebut memiliki nilai akurasi yang memadai untuk klasifikasi algoritma SVM dengan menggunakan perhitungan *confusion matrix*. Hasil dari penelitian yang dilakukan telah efektif dengan metode SVM.

**Kata kunci:** Analisis Sentimen, Hak Angket, Support Vector Machine, TF-IDF, Twitter.

## 1. PENDAHULUAN

Menurut teori *trias politica* menyatakan bahwa terdapat tiga fungsi utama yang dijalankan dalam kekuasaan negara, yaitu lembaga *eksekutif* yang merupakan fungsi dalam menjalankan kedaulatan rakyat, lembaga *legislatif* yaitu fungsi yang merujuk pada pembuatan aturan-aturan umum yang mengaitkan seluruh warga negara, dan lembaga *yudikatif* adalah lembaga untuk memantau potensi penyimpangan [1].

*Dewan perwakilan rakyat* (DPR) merupakan lembaga negara yang masuk dalam sistem ketatanegaraan Indonesia yang merupakan lembaga perwakilan rakyat. DPR berfungsi sebagai acuan masyarakat yang memiliki tugas sebagai badan pengawas jalannya sistem pemerintahan yang dijalankan oleh lembaga *eksekutif*. Proses pengawasan yang dilakukan berhubungan dengan kebijakan yang di ambil oleh lembaga *eksekutif* [2].

Lembaga perwakilan rakyat masuk dalam lembaga yang mengalami perubahan dan penataan yang menonjol. Kemudian pada lembaga ini secara umum mempunyai dua jenis diantaranya lembaga perwakilan rakyat satu kamar (*unicameral*) dan lembaga perwakilan rakyat dua kamar yaitu (*bicameral*). Kemudian pada lembaga pembentukan *Dewan perwakilan daerah* (DPD) adalah wujud asli diperlukannya kesejajaran dan kontrol di antara lembaga-lembaga negara agar dapat berjalan dengan seimbang, dimana harus berkaitan dengan kebijakan pusat dan di daerah. Pada pembentukan DPD digunakan sebagai bentuk penegasan negara hukum dan melindungi hak asasi (*hak konstitusional*) yang dijamin oleh konstitusi. DPD dibangun agar dipergunakan untuk masalah yang ada dalam ketatanegaraan sebelumnya [3].

Hak angket merupakan hak yang digunakan dalam penyelidikan, kemudian dapat digunakan sebagai proses penyelidikan pada kebijakan pemerintah provinsi, kabupaten, dan kota yang penting dan strategis, dan berdampak pada lingkungan kehidupan masyarakat, daerah, dan negara yang diperkirakan bertentangan pada regulasi peraturan perundang-undangan dengan setelah dilakukan proses penyidikan [4].

Berlandaskan pada undang-undang nomor 17 tahun 2014, dengan digunakannya penelitian *juridis normatif*, diterangkan bahwa hukum konstitusional wewenang hak angket DPR pada ketatanegaraan diatur dalam UUD 1945 pasal 20A, yang mana tugasnya untuk mengatur mengenai hak angket DPR. Hak angket telah diatur dalam materiil Pasal 79 ayat (3) undang-undang nomor 17 tahun 2014,

menyebutkan lembaga-lembaga yang termasuk dalam kategori yang dapat diusulkan angket oleh DPR. Hak angket dijelaskan juga untuk menyelidiki pelaksanaan suatu undang-undang atau kebijakan pemerintah yang memiliki dampak signifikan, strategis, dan melibatkan aspek luas. Penyelidikan dilakukan jika terdapat dugaan tindakan yang bertentangan pada peraturan perundang-undangan [5].

Penelitian sebelumnya pada tahun 2020 yang disusun oleh Harsono et.al mengenai analisis sentimen *stakeholder* atas layanan hai DJPbn pada media sosial *twitter* dengan menggunakan metode yaitu *naïve bayes* dan *support vector machine*, dimana penelitian ini menguji dua metode yaitu membandingkan nilai *accuracy* dan *precision* pada algoritma NB dengan SVM dengan hasil nilai *accuracy* dan *presicion* pada algoritma NB lebih tinggi dibandingkan dengan SVM. Untuk nilai *recall* pada algoritma SVM dan NB memiliki kesamaan, tetapi untuk *margin* pada algoritma SVM lebih besar dibandingkan dengan algoritma NB yaitu +/- 9.72%, dibandingkan dengan +/- 7.07%. Kemudian untuk nilai *recall* maksimal algoritma SVM 103.61%, akan tetapi pada algoritma NB sebesar 100.96% [6].

SVM merupakan metode pembelajaran yang terawasi dengan tingkat akurasi dan kualitas yang tinggi, menjadikannya populer di antara algoritma lainnya. Namun, untuk menerapkannya, diperlukan proses pelatihan secara berurutan dan pengujian yang teliti. Langkah awal dalam proses analisis adalah mengubah data teks menjadi *representasi vektor*, yang kemudian akan diproses dengan metode *term frequency inverse document frequency* (TF-IDF) untuk pembobotan. Salah satu kelebihan SVM adalah kemampuannya dalam mengidentifikasi *hyperplane* yang memisahkan, memungkinkan untuk memaksimalkan *margin* antara kelas yang berbeda. Namun, penting untuk diingat bahwa SVM juga memiliki kelemahan, terutama ketika menghadapi masalah dengan fitur yang serupa, yang dapat secara signifikan memengaruhi tingkat akurasi.

Penelitian mengenai hak angket merujuk pada tanggapan masyarakat pada media sosial *twitter* sehingga kini menjadi pembahasan yang menjadi trending topic pada media sosial *twitter*. Tanggapan tersebut berupa data *textual* yang akan dianalisa dan mendapatkan informasi untuk kemudian dilakukan analisis sentimen masyarakat pada isu hak angket tersebut. Hak angket dapat dilakukan proses analisis sentiman yang dikategorikan menjadi sentimen *negatif* dan sentimen *positif*, kemudian hasil analisis sentimen tersebut dapat digunakan sebagai informasi

dan evaluasi bagi lembaga yang mengangkat hak angket. Analisis sentimen merupakan jenis pengolahan kata untuk mengetahui opini masyarakat mengenai pembahasan tertentu [7].

Penelitian ini dilakukan terkait permasalahan mengenai hak angket di Indonesia tahun 2024 yang selalu menjadi pembahasan trending topik pada media sosial *twitter*, sehingga peneliti melakukan analisis pada opini atau komentar masyarakat tentang hak angket tersebut di *twitter* dengan tujuan untuk mengetahui seberapa besar pengaruh pemahaman tentang hak angket bagi Indonesia dan masyarakat dengan mencari nilai akurasi pada data yang telah dikumpulkan melalui proses *crawling* atau biasa disebut pengumpulan data.

Beberapa langkah yang dilakukan dalam analisis sentimen pada penelitian ini yaitu proses pengumpulan (*crawling*), pemrosesan awal (*preprocessing*), pelabelan (*labelling*), selanjutnya analisis sentimen menggunakan metode *support vector machine* (SVM). Metode SVM dipakai karena memiliki hasil klasifikasi dalam mendeteksi sentimen dari media *twitter* dengan baik. Pada proses pemilu tahun 2024 terdapat konflik isu mengenai pengangkatan hak angket yang menjadi topik pembicaraan di kalangan masyarakat dan menjadikan hal tersebut dapat dibahas pada penelitian ini dengan menentukan analisis sentimen tanggapan masyarakat tersebut [8].

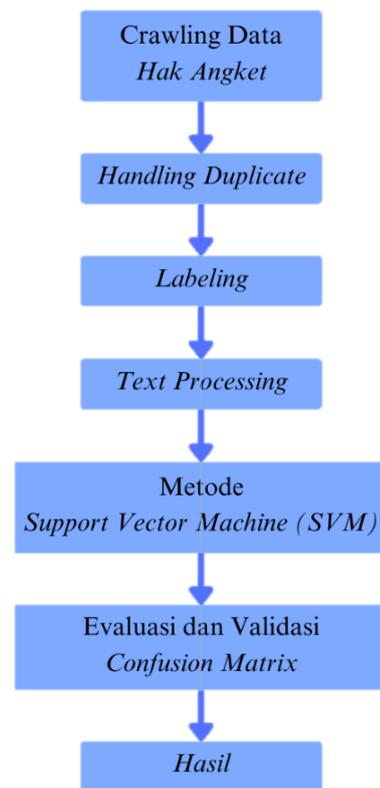
Pada proses penelitian ini data yang telah dikumpulkan sebanyak 2.179 data yang kemudian akan dilakukan proses analisis dengan menggunakan metode SVM. Data tersebut kemudian akan dibagi menjadi data latih dan data uji yang digunakan untuk mengetahui seberapa besar nilai akurasi yang akan didapatkan dalam analisis sentimen ini. Data yang akan diuji untuk mendapatkan nilai akurasi sebesar 80% dari total keseluruhan data yang ada dan 20% menjadi data latih. Proses penelitian ini menggunakan metode SVM mengenai hak angket dengan mengumpulkan data melalui platform media sosial *twitter*. Pengerjaan penelitian ini menggunakan bahasa pemrograman *python* yang dijalankan melalui *google colab*, kemudian akan didapatkan hasil analisis sentimen algoritma SVM yang dikelompokkan menjadi sentimen *positif* dan sentimen *negatif* [9].

## 2. METODE PENELITIAN

Pada penelitian yang dilakukan ini sangat erat hubungannya dengan *text mining*, dimana *text mining* adalah proses pengumpulan pola dari dokumen teks yang tidak terstruktur dikenal sebagai penggalian teks, atau penggalian data teks. Karena *text mining* membutuhkan pemrosesan data *tekstual*, tidak terstruktur, dan ambigu, ini adalah tugas yang menantang dari pada data mining. Proses pengumpulan informasi, analisis teks ekstraksi informasi, pengelompokan, klasifikasi, visualisasi, teknologi basis data, pembelajaran mesin, dan penambangan data semuanya termasuk dalam topik

interdisipliner teknologi penambangan teks, penambangan data, dan pembelajaran mesin. Proses tersebut adalah masalah yang dapat dipcahkan oleh *text mining* [10].

Proses analisis sentimen yang dilakukan pada penelitian ini menggunakan klasifikasi data *tekstual* menjadi dua kelas, yaitu *positif* dan *negatif*. Analisis sentimen dijelaskan sebagai proses pemisahan, pengolahan dan memahami sebuah data text yang tidak terstruktur yang terdapat pada data opini tersebut. Penelitian ini dilakukan untuk menganalisis opini masyarakat terhadap kebijakan hak angket di Indonesia dengan menggunakan metode SVM, dan penelitian dilakukan untuk menentukan model mana yang paling baik. *K-fold cross validation* digunakan untuk validasi model. *Confusion matrix* digunakan untuk mengukur tingkat akurasi model. Metode penelitian yang dilakukan ditunjukkan pada Gambar 1 di bawah ini.



Gambar 1. Metode Penelitian

### 2.1. Pengumpulan Data

Proses setelah melakukan identifikasi topik, langkah selanjutnya adalah proses pengumpulan data dengan menggunakan bahasa pemrograman *python* dengan *emulator google colab*, dengan analisis data mengenai opini masyarakat tentang hak angket [11]. Langkah pengumpulan data dapat dilihat pada Gambar 2.



Gambar 2. Langkah Pengumpulan Data

## 2.2. Labeling Sentimen

Penelitian ini untuk pemrosesan hasil pelabelan digunakan fungsi dari library *vader sentiment* dan *python*. Sebelum digunakan library *vader sentiment* sebuah data *tweet* mengenai hak angkat harus ditranslate dalam bahasa inggris, karena library *vader sentiment* hanya dapat digunakan dengan sebuah data yang menggunakan bahasa inggris pada saat penentuan analisis sentimen yang terbagi menjadi sentimen *positif*, *negatif* dan *netral*.

## 2.3. Text Preprocessing

Proses penelitian ini melakukan beberapa tahapan yang terangkai berawal dari *cleaning*, *case folding*, *tokenizing*, *stopword removal*, *stemming* [12]. Dapat dipahami penjelasan setiap langkah sebagai berikut.

1. **Cleaning:** *Cleaning* merupakan proses untuk membersihkan setiap karakter atau kalimat yang tidak dibutuhkan seperti URL, @, #, https:, *Retweet*, simbol, angka, dan *emoticon*.
2. **Case Folding:** Merupakan tahapan untuk melakukan perubahan pada huruf yang sebelumnya kapital menjadi huruf standar atau disebut *lowercase*.
3. **Tokenizing:** Adalah tahapan yang berguna untuk menghilangkan tanda baca, simbol, dan karakter yang tidak berharga dari isi dokumen.
4. **Stopword Removal:** Jika sebuah kalimat memiliki kata-kata umum dan tidak penting seperti waktu, penghubung, dan sebagainya, maka akan dihilangkan oleh proses *stopword*.
5. **Stemming:** Proses *stemming* merupakan langkah menghapus imbuhan awal dan akhiran agar menghasilkan kata dasar. Proses ini dilakukan dengan library *sastrawi python*.

## 2.4. Analisis Sentimen

Analisis sentimen merupakan langkah dalam penentuan sentimen dan klasifikasi dari suatu dokumen untuk menentukan kategori yang di bagi menjadi tiga yaitu, sentimen *positif*, *negatif*, dan *netral*. Akan tetapi pada penelitian ini analisis sentimen yang digunakan hanya mengambil sentimen *positif* dan *negatif* karena sentimen netral dianggap data yang kosong dan tidak dapat digunakan untuk klasifikasi hasil nilai akurasi. Media sosial *twitter* digunakan pada analisis sentimen berguna untuk menentukan opini publik. Analisis sentimen melibatkan pengumpulan data yang berguna sebagai analisis, ekstraksi data tekstual dalam suatu kelompok data [13].

## 2.5. Algoritma SVM (Support Vector Machine)

Algoritma SVM merupakan metode untuk pemecahan masalah pada klasifikasi data yang handal. Pemecahan masalah pada algoritma SVM menggunakan persamaan *lagrangian* yaitu bentuk

*dual* dari SVM melalui *quadratic* pemrograman. Dimana metode SVM standar yang memiliki hasil lebih baik dan menjadi pilihan alternatif dibandingkan dengan menggunakan SVM tradisional dalam pemrosesan data berkapasitas besar [14].

## 2.6. TF-IDF (Term Frequency-Inverse Document Frequency)

Data yang sudah melalui tahap *preprocessing* harus memiliki data berbentuk *numerik*. Langkah yang dilakukan untuk mengubah data menjadi *numerik* maka dilakukan dengan metode pembobotan TF-IDF (*term frequency-inverse document frequency*) yang merupakan metode untuk menganalisa seberapa jauh ketersambungan kata (*term*) pada data dengan memberikan bobot setiap baris kata [15].

TF merupakan frekuensi kata yang muncul disetiap dokumen dan menunjukkan seberapa penting kata tersebut. Dokumen yang menunjukkan dan mempunyai kata yang umum disebut dengan DF. Nilai DF merupakan inverse nilai IDF. Proses pembobotan dengan metode TF-IDF adalah dengan melakukan perkalian antara TF dikalikan dengan IDF sehingga hasilnya jika bobot kata semakin besar maka peluang dokumen muncul semakin baik dan nilai semakin kecil jika muncul dalam banyak dokumen [16]. Berikut rumus TF-IDF:

$$tf = 0,5 + 0,5 \times \frac{tf}{\max(tf)} \quad (1)$$

$$idf_t = \log\left(\frac{D}{df_t}\right) \quad (2)$$

$$W_{d,t} = TF_{d,t} \times idf_{d,t} \quad (3)$$

Penjelasan mengenai rumus TF-IDF adalah,  $t$  = *term* ke- $t$  dari dokumen,  $W$  = bobot dokumen ke- $d$  terhadap *term* ke- $t$ ,  $tf$  = banyaknya *term*  $i$  pada sebuah dokumen,  $idf$  = *inversed document frequency*,  $df$  = banyak dokumen yang mengandung *term*  $I$  [17].

## 2.7. Evaluasi

Evaluasi data ini memakai tahap *fold-cross* validasi sebesar 10. Penelitian ini menggunakan dua proses validasi yaitu proses *training* dan *testing*. Kemudian pada pemrosesan data *training* menggunakan *tools google colab* yang kemudian akan dihasilkan nilai evaluasi pada akurasi model klasifikasi tersebut yaitu *accuracy*, *precision*, *recall*, dan *f-score*. Berikut merupakan proses rumus pada perhitungan evaluasi model klasifikasi data.

$$Accuracy = \frac{TP+TN}{(TP+FP+FN+TN)} \times 100\% \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \tag{6}$$

$$F - score = \frac{2 \times precision \times recall}{precision + recall} \tag{7}$$

Tahapan klasifikasi yaitu TP (*true positif*) merupakan klasifikasi bernilai benar, TN (*true negatif*) adalah hasil klasifikasi bernilai salah, FP (*false positif*), yaitu klasifikasi yang bernilai benar tetapi kurang tepat, FN (*false negatif*) yang merupakan hasil klasifikasi salah, namun faktanya benar [18].

### 3. HASIL DAN PEMBAHASAN

Pada penelitian ini jumlah data yang dikumpulkan sebanyak 2.179 data bersumber dari *twitter*. Data yang telah dikumpulkan dengan menggunakan cara *crawling* di *python*. Semua data yang telah diperoleh kemudian akan di proses terlebih dahulu agar menjadi data yang tersusun, kemudian dilanjutkan dengan langkah pelabelan dan pembobotan menggunakan perhitungan TF-IDF, dan kemudian dataset tersebut akan dipergunakan sebagai pengujian algoritma SVM untuk menentukan nilai akurasi dan hasil yang optimal.

#### 3.1. Crawling Dataset

Tahap pengumpulan dataset mengenai hak angket yang digunakan dalam penelitian ini bersumber dari media sosial *twitter*. Dataset yang dikumpulkan berupa opini masyarakat mengenai isu hak angket dengan menggunakan bahasa pemrograman *python*. Hasil dari pengumpulan data dapat dilihat pada Tabel 1.

Tabel 1. Hasil *Crawling Dataset*

No	Username	Tweet
1.	heni_eunike	Dia ngerti gak sebenarnya hak angket itu apa? Itu kan hak konstitusional, masa ditolak. Lah lu ngelanggar undang2 namanya neng Di Kuningan Dedi Mulyadi Timpali Soal Hak Angket DPR: Rakyat Maunya Hak Angkut #DediMulyadi #HakAngketHakRakyat
2.	PR_Kuningan	#DediMulyadiHakAngkut #SafariCintaKDM #jagara #Kuningan #Ramadhan #RamadhanKareem #Ramadhan2024 #ramadanmubarak
---	---	---
2020.	JameKakah	Hmmmm dengar kata kata pak Surya bisa diPastikan klaw Nasdem akan Koalisi & batal tuk Hak Angket. Nasdem pastinya Tau Tdk berpikir negatip tuk bangsa ni

#### 3.2. Labeling

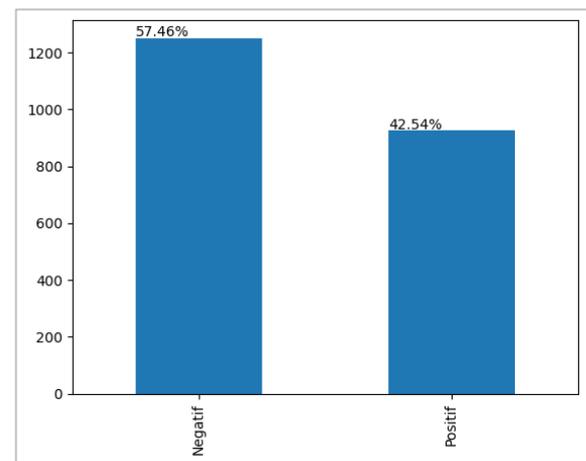
Pada proses labeling dilakukan agar mengetahui hasil komentar tersebut memberikan hasil sentimen *positif* atau *negatif*. Data mentah sebelumnya diperoleh dari hasil *crawling* dan dibersihkan dari data yang memiliki kesamaan atau duplikat dan data

yang tidak berkaitan dengan hak angket, hasil dari proses pelabelan dapat dilihat pada Tabel 2.

Tabel 2. Hasil Proses *Labeling*

No	Tweet	Sentimen
1.	Dia ngerti gak sebenarnya hak angket itu apa? Itu kan hak konstitusional, masa ditolak. Lah lu ngelanggar undang2 namanya neng Di Kuningan Dedi Mulyadi Timpali Soal hak angket DPR: Rakyat Maunya Hak Angkut #DediMulyadi #HakAngketHakRakyat	Positif
2.	#DediMulyadiHakAngkut #SafariCintaKDM #jagara #Kuningan #Ramadhan #RamadhanKareem #Ramadhan2024 #ramadanmubarak	Positif
---	---	---
2020.	@MurtadhaOne1 Ach lebay banget. Drama korea nya kurang keren. Orang yang menolak hak angket belum ada yang turun ke jalan. Itu masa bayaran boneka asing pembenci presiden jokowi. Logika nya ngapain anti hak angket karena yg mau di angket itu suara ra	Positif

Berikut ini merupakan analisis jumlah dari seluruh data yang terbagi menjadi sentimen *positif* dan *negatif*, pada sentimen *positif* berjumlah 1225 tweet, sentimen *negatif* 772 tweet. Dengan sentimen *positif* 57.46%, sentimen *negatif* 42.54%. Hasil analisis jumlah sentimen tersebut dapat dilihat pada Gambar 3.



Gambar 3. Hasil Visualisasi *Labeling*

#### 3.3. Text Preprocessing

Proses *preprocessing* merupakan langkah penting sebelum dilakukannya data mining. Pada *preprocessing* melibatkan beberapa langkah, diantaranya *cleaning*, *case folding (lower case)*, *tokenizing*, *stopword*, *stemming*, dari semua tahapan tersebut saling terhubung untuk menjadikan data agar dapat di proses pada tahap selanjutnya [19]. Data kemudian harus dibersihkan sebelum dapat digunakan untuk melanjutkan proses analisis sentimen.

### 3.3.1. Cleaning

Tahap *cleaning* dikenal sebagai data *cleansing*, ini merupakan tempat dimana data yang tidak lengkap, mengandung data *error* dan tidak konsisten sehingga dibersihkan dari kumpulan data, sehingga data yang telah bersih digunakan sebagai panggilan ulang untuk panggilan pengetahuan (*discovery knowledge*) [20]. Hasil dari tahap *cleaning* dapat dilihat pada Tabel 3.

Tabel 3. Hasil Proses *Cleaning*

Tweet	Cleaning
Dia ngerti gak sebenarnya hak angket itu apa? Itu kan hak konstitusional, masa ditolak. Lah lu ngelanggar undang2 namanya neng	Dia ngerti gak sebenarnya hak angket itu apa itu kan hak konstitusional masa ditolak Lah lu ngelanggar undang namanya neng

### 3.3.2. Case Folding

Dalam komentar *tweet*, pasti terdapat perbedaan huruf, proses ini adalah tahapan untuk merubah bentuk huruf menjadi huruf kecil (*lower case*) [21]. Contoh hasil *case folding* dapat dilihat pada Tabel 4.

Tabel 4. Hasil Proses *Case Folding*

Tweet	Case Folding
Dia ngerti gak sebenarnya hak angket itu apa? Itu kan hak konstitusional, masa ditolak. Lah lu ngelanggar undang2 namanya neng	dia ngerti gak sebenarnya hak angket itu apa itu kan hak konstitusional masa ditolak lah lu ngelanggar undang namanya neng

### 3.3.3. Tokenizing

Proses *tokenizing* melibatkan banyak langkah, salah satunya merupakan mengubah semua huruf kapital menjadi huruf kecil. Langkah selanjutnya adalah penguraian, dimana proses ini membagi teks menjadi kumpulan kata-kata tanpa mempertimbangkan koneksi dan hubungan antar kata, kemudian fungsi dan penempatan dalam kalimat. Karakter disusun berdasarkan abjad dalam kumpulan kata, kata-kata berbahasa Indonesia yang diulang dua kali kemudian akan dipisah menjadi dua kata [22]. Dapat dilihat hasil proses pada Tabel 5.

Tabel 5. Hasil Proses *Tokenizing*

No	Tweet	Tokenizing
1.	Dia ngerti gak sebenarnya hak angket itu apa? Itu kan hak konstitusional, masa ditolak. Lah lu ngelanggar undang2 namanya neng	['dia', 'ngerti', 'gak', 'sebenarnya', 'hak', 'angket', 'itu', 'apa', 'itu', 'kan', 'hak', 'konstitusional', 'masa', 'ditolak', 'lah', 'lu', 'ngelanggar', 'undang', 'namanya', 'neng']

### 3.3.4. Stopword

Proses berikutnya adalah melakukan pemeriksaan *stop word list*, yaitu proses menghilangkan kata-kata yang tidak penting, seperti kata yang tidak memiliki makna, tahap ini dilakukan setelah melewati proses *cleaning* data [23]. Hasil proses *stopword* dilihat pada Tabel 6.

Tabel 6. Hasil Proses *Stopword*

Tweet	Stopword
Dia ngerti gak sebenarnya hak angket itu apa? Itu kan hak konstitusional, masa ditolak. Lah lu ngelanggar undang2 namanya neng	['ngerti', 'gak', 'sebenarnya', 'hak', 'angket', 'apa', 'kan', 'hak', 'konstitusional', 'masa', 'ditolak', 'lah', 'lu', 'ngelanggar', 'undang', 'namanya', 'neng']

### 3.3.5. Stemming

Proses *stemming* merupakan proses ekstraksi kata dari imbuhan untuk menghasilkan kata dasar yang dikenal sebagai *stemming*. Hasil proses *stemming* disebut *stem*, oleh karena itu, metode yang digunakan pada setiap program memiliki tujuan berbeda yang berguna menghasilkan kata dasar [24]. Hasil dari *stemming* dilihat pada Tabel 7.

Tabel 7. Hasil Proses *Stemming*

Tweet	Stemming
Dia ngerti gak sebenarnya hak angket itu apa? Itu kan hak konstitusional, masa ditolak. Lah lu ngelanggar undang2 namanya neng	ngerti gak sebenarnya hak angket apa kan hak konstitusional masa tolak lah lu ngelanggar undang nama neng

### 3.4. TF-IDF (Term Frequency-Inverse Document Frequency)

Proses TF-IDF (*term frequency-inverse document frequency*) merupakan tahap yang digunakan sebagai pembobotan sebuah data [25]. Dilakukan proses pembobotan kata dengan metode TF-IDF ini kemudian akan menjadi data *numerik* yang akan dilanjutkan dalam pemrosesan algoritma SVM [26]. Hasil proses pembobotan kata dapat dilihat pada Tabel 8.

Tabel 8. Sample Hasil TF-IDF

Jumlah Kata	Frekuensi	Nilai TF-IDF
1.062	2.557	0.22321037510309263
1.062	2.653	0.23244083982217392
1.062	2.680	0.06735082882925997
1.062	2.706	0.15446412870224346
1.062	2.729	0.20251569284713
1.062	2.914	0.22321037510309263

### 3.5. Analisis SVM (Support Vector machine)

Setelah tahapan *preprocessing*, pelabelan dan TF-IDF dataset kemudian diproses melalui penerapan model SVM. Pada tahap klasifikasi SVM dengan mencari nilai akurasi data dengan tanggapan masyarakat mengenai hak angket, diketahui bahwa proses pencarian nilai akurasi didapatkan melalui validasi dengan perhitungan *confusion matrix* [27]. Dapat dilihat hasil validasi pada Gambar 4.

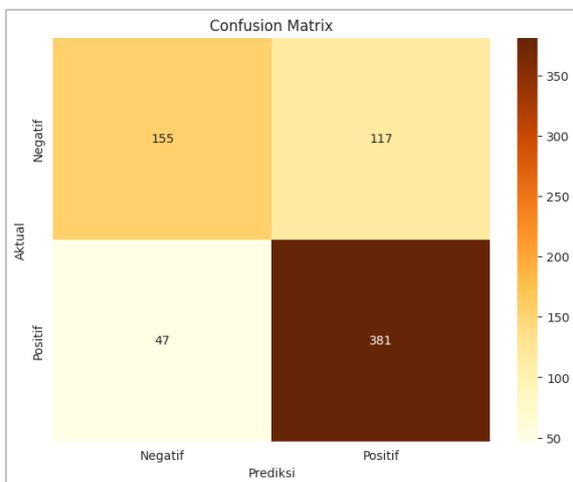
	precision	recall	f1-score	support
Negatif	0.77	0.57	0.65	272
Positif	0.77	0.89	0.82	428
accuracy			0.77	700
macro avg	0.77	0.73	0.74	700
weighted avg	0.77	0.77	0.76	700

Gambar 4. Hasil Perhitungan SVM

Berdasarkan hasil sentimen analisis menggunakan metode SVM (*support vector machine*) dengan memakai *library sklearn* di *python*. Proses pendekatan pada penelitian ini menggunakan kernel *linear*. Diketahui hasil dari klasifikasi metode SVM memiliki tingkat *accuracy* pada kernel *linear* sebesar 77%, dimana tingkat akurasi tersebut sudah baik dan cukup digunakan sebagai penelitian lanjutan, kemudian pada nilai *precision negatif* 77%, nilai *precision positif* 77%, nilai *recall negatif* 57%, nilai *recall positif* 89%, nilai *f1-score positif* 66%, nilai *f1-score negatif* 82%. Data yang telah diuji dan diproses tersebut memiliki nilai akurasi yang memadai untuk klasifikasi algoritma SVM dengan menggunakan perhitungan *confusion matrix*.

### 3.6. Visualisasi Data

Setelah dilakukan pemodelan data dengan menggunakan algoritma SVM kemudian pengujian dengan menggunakan TF-IDF dan validasi perhitungan hasil dilakukan menggunakan *confusion matrix*, ini adalah sebuah alat untuk melakukan evaluasi terhadap keunggulan klasifikasi yang dihasilkan serta membedah karakteristik dari berbagai kelas data yang berbeda. Hasil *confusion matrix* dengan metode SVM dengan hasil kernel *linear* dapat dilihat pada Gambar 5.



Gambar 5. Hasil *Kernel linear SVM*

## 4. DISKUSI

Dalam penelitian ini menggunakan algoritma *support vector machine* (SVM) dalam analisis mengenai hak angket di Indonesia pada tahun 2024. Dengan menerapkan tahapan penelitian seperti pengumpulan data (*crawling*), *preprocessing*, *labelling*, dan analisis sentimen dengan menggunakan metode *support vector machine* dan untuk mencari hasil pembobotan sebuah kata pada data menggunakan term *frequency-inverse document frequency* (TF-IDF). Pada proses *crawling* dataset langkah awal adalah dengan memperoleh kode *token API twitter* dengan cara melihat pada akun setiap pengguna, hasil dari proses *crawling* data terkumpul

sebanyak 2.179 data. Tahap *preprocessing* dilakukan untuk mempersiapkan seluruh data yang akan digunakan dalam analisis sentimen oleh algoritma SVM, sehingga akan mempermudah dalam pencarian nilai akurasi suatu data. Tahapan *preprocessing* terdiri dari *cleaning*, *case folding*, *tokenizing*, *stopword removal*, *stemming*. Proses analisis sentimen dengan menggunakan algoritma SVM diperoleh hasil akurasi sebesar 77%, dimana hasil tersebut sudah cukup baik untuk sebuah nilai akurasi analisis data. Perhitungan akurasi tersebut meliputi proses dari perhitungan *confusion matrix* dan sebelum proses perhitungan tersebut peneliti perlu untuk mencari bobot setiap kata yang ada dalam data dengan menggunakan pembobotan TF-IDF yang berfungsi sebagai tolak ukur pada pencarian nilai akurasi dengan metode yang akan digunakan. Kemudian pada hasil visualisasi kata mendapatkan nilai sentimen positif sebesar 57.46% dan sentimen negatif 42.54%, dapat diketahui bahwa sentimen positif memiliki jumlah yang lebih banyak untuk melakukan proses analisis data dengan SVM.

Penelitian yang dilakukan oleh Indra Yunanto dan Sri Yulianto dengan judul “Analisis Sentimen *Twitter* Aplikasi Pedulilindungi Menggunakan *Naïve Bayes* Dan *Support Vector Machine*”. Analisis mengenai aplikasi pedulilindungi memiliki data yang digunakan sebanyak 4.636 tweet. Data tersebut diperoleh melalui respon komentar pada media sosial *twitter* dengan memakai pencarian kata kunci “PeduliLindungi”. Untuk mengetahui perbandingan akurasi dari metode yang digunakan yaitu *naïve bayes classifier* dan *support vector machine* dengan menggunakan bahasa pemrograman *python*. Hasil akurasi metode *naïve bayes classifier* sebesar 90% dan untuk metode *support vector machine* sebesar 91%. Hasil yang didapat dari dua metode tersebut dapat disimpulkan bahwa *support vector machine* menghasilkan nilai akurasi yang lebih tinggi dari pada *naïve bayes* dengan perbedaan akurasi sebesar 1% [28]. Sehingga dari hasil yang telah didapatkan dapat disimpulkan peneliti perlu melakukan penelitian ini untuk membuktikan seberapa efektif metode SVM untuk menentukan nilai akurasi sebuah data.

## 5. KESIMPULAN

Pada penelitian ini bertujuan untuk mencari hasil nilai akurasi dari sebuah metode yang digunakan yaitu *support vector machine* dengan menggunakan bahasa pemrograman *python* dengan bantuan *tools google colab*. Penelitian ini dilakukan dengan proses tahapan pengumpulan data (*crawling*), *preprocessing*, *labelling*, dan analisis data dengan metode SVM, setelah itu melakukan pembobotan kata dengan TF-IDF. Data yang digunakan untuk penelitian ini memiliki jumlah 2.179 *tweet* dibagi dengan 1.298 data *training* dan 700 data *testing*. Dimana hasil analisis mengungkapkan bahwa nilai dari akurasi metode *support vector machine* sebesar 77%. Dengan tingkat akurasi tersebut sudah baik dan

cukup digunakan sebagai penelitian lanjutan, kemudian pada nilai *precision negatif* 77%, nilai *precision positif* 77%, nilai *recall negatif* 57%, nilai *recall positif* 89%, nilai *f1-score positif* 66%, nilai *f1-score negatif* 82%. Untuk hasil visualisasi kata memiliki hasil sentimen kata positif sebesar 57.46% dan untuk seentimen negatif sebesar 42.54%. pada pembobotan kata dengan TF-IDF memiliki hasil pada jumlah kata 1.062 dengan frekuensi kata sebesar 2.557 dengan nilai TF-IDF sebesar 0.22321037510309263 untuk salah satu contoh hasil pembobotan kata dengan TF-IDF. Analisis menggunakan metode tersebut memiliki hasil cukup baik. Dapat disimpulkan bahwa penelitian menggunakan metode *support vector machine* efektif dalam menentukan analisis data. Sedangkan jika data yang didapat lebih bersih maka hasil nilai akurasi yang didapat akan mencapai maksimum. Sedangkan untuk saran jika ingin mendapatkan nilai akurasi yang lebih akurat maka perlu dilakukan dengan menggabungkan lebih banyak data atau dapat menggunakan metode analisis lain.

#### DAFTAR PUSTAKA

- [1] B. Iskandar, "Analisis Putusan Mahkamah Konstitusi No. 36/PUU-XV/2017 Tentang Hak Angket Dewan Perwakilan Rakyat Terhadap Komisi Pemberantasan Korupsi," *Lex Renaissance*, vol. 4, no. 2, pp. 410–431, Jul. 2019, doi: 10.20885/JLR.VOL4.ISS2.ART12.
- [2] S. Sumartini and J. Arifin, "Fungsi Hak Angket Dewan Perwakilan Rakyat Untuk Melakukan Penyelidikan Terhadap Pelaksanaan Undang-Undang," *Yustitia*, vol. 6, no. 1, pp. 23–44, Apr. 2020, doi: 10.31943/YUSTITIA.V6I1.97.
- [3] H. Sabar, R. Tinambunan, D. Dicky, E. Prasetyo, J. K. Surabaya, and J. Timur, "Rekonstruksi Konstitusi Dalam Regional Representative Dewan Perwakilan Daerah Terhadap Fungsi Legislatif," *Masalah-Masalah Hukum*, vol. 48, no. 3, pp. 266–274, Jul. 2019, doi: 10.14710/MMH.48.3.2019.266-274.
- [4] A. Novilistiana and A. Riwanto, "Pelaksanaan Fungsi Hak Angket Dewan Perwakilan Rakyat Republik Indonesia (Studi Kasus Hak Angket Tentang Komisi Pemberantasan Korupsi)," *Res Publica: Jurnal Hukum Kebijakan Publik*, vol. 4, no. 2, pp. 130–146, Dec. 2020, doi: 10.20961/RESPUBLICA.V4I2.45704.
- [5] I. Aris, I. Amir, and S. Amrianto, "Konstitusionalitas Hak Angket Dewan Perwakilan Rakyat (DPR) Terhadap Komisi Pemberantasan Korupsi (KPK)," *Al-Adalah: Jurnal Hukum dan Politik Islam*, vol. 4, no. 2, pp. 135–158, Oct. 2019, doi: 10.35673/AJMPL.V4I2.436.
- [6] M. L. (Muhammad) Harsono, Y. (Yuris) Alkhalifi, N. (Nurajijah) Nurajijah, and W. (Windu) Gata, "Analisis Sentimen Stakeholder Atas Layanan HAIDJPB pada Media Sosial Twitter dengan Menggunakan Metode Support Vector Machine dan Naïve Bayes," *Infoman*, vol. 14, no. 1, p. 414142, 2020, doi: 10.33481/INFOMANS.V14I1.126.
- [7] S. Y. Pangestu, Y. Astuti, and L. D. Farida, "Algoritma Support Vector Machine Untuk Klasifikasi Sikap Politik Terhadap Partai Politik Indonesia," *Jurnal Mantik*, vol. 3, no. 1, pp. 236–241, Jun. 2019, Accessed: Mar. 25, 2024. [Online]. Available: <https://iocscience.org/ejournal/index.php/mantik/article/view/173>
- [8] I. Yunanto and S. Yulianto, "Twitter Sentiment Analysis Pedulilindungi Application Using Naïve Bayes And Support Vector Machine," *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 4, pp. 807–814, Aug. 2022, doi: 10.20884/1.JUTIF.2022.3.4.292.
- [9] A. Nursalim and R. Novita, "Sentiment Analysis Of Comments On Google Play Store, Twitter And Youtube To The MyPertamina Application With Support Vector Machine," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 6, pp. 1305–1312, Jun. 2023, doi: 10.52436/1.JUTIF.2023.4.6.1059.
- [10] A. M. Pravina, I. Cholisoddin, and P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 3, pp. 2789–2797, Jan. 2019, Accessed: Mar. 16, 2024. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4793>
- [11] S. K. Dirjen *et al.*, "Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 802–808, Aug. 2021, doi: 10.29207/RESTI.V5I4.3308.
- [12] N. Shafiya, N. Salam, A. A. Supianto, and A. R. Perdanakusuma, "Analisis Sentimen Opini Mahasiswa Terhadap Saran Kuesioner Penilaian Kinerja Dosen dengan Menggunakan TF-IDF dan K-Nearest Neighbor," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 6, pp. 6148–6156, Jul. 2019, Accessed: Mar. 20, 2024. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/5649>

- [13] R. Puspitasari, Y. Findawati, and M. A. Rosid, "Sentiment Analysis Of Post-Covid-19 Inflation Based On Twitter Using The K-Nearest Neighbor And Support Vector Machine Classification Methods," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 4, pp. 669–679, Aug. 2023, doi: 10.52436/1.JUTIF.2023.4.4.801.
- [14] E. Fitri, Y. Yuliani, S. Rosyida, and W. Gata, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *Jurnal Transformatika*, vol. 18, no. 1, pp. 71–80, Jul. 2020, doi: 10.26623/TRANSFORMATIKA.V18I1.2317.
- [15] A. Deolika, K. Kusriani, and E. T. Luthfi, "Analisis Pembobotan Kata Pada Klasifikasi Text Mining," (*JurTI*) *Jurnal Teknologi Informasi*, vol. 3, no. 2, pp. 179–184, Dec. 2019, doi: 10.36294/JURTI.V3I2.1077.
- [16] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *INSYST: Journal of Intelligent System and Computation*, vol. 1, no. 1, pp. 43–49, Aug. 2019, doi: 10.52985/INSYST.V1I1.36.
- [17] S. K. Dirjen, P. Riset, D. Pengembangan, R. Dikti, R. A. Supono, and M. A. Suprayogi, "Perbandingan Metode TF-ABS dan TF-IDF Pada Klasifikasi Teks Helpdesk Menggunakan K-Nearest Neighbor," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, pp. 911–918, Oct. 2021, doi: 10.29207/RESTI.V5I5.3403.
- [18] W. Sofiya and E. B. Setiawan, "Fine-Grained Sentiment Analysis In Social Media Using Gated Recurrent Unit With Support Vector Machine," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 3, pp. 511–519, Jun. 2023, doi: 10.52436/1.JUTIF.2023.4.3.855.
- [19] A. Muhadi and A. Octaviano, "Penerapan Data Mining Untuk Prediksi Hasil Keuntungan Lelang Mesin X-Ray Tahun 2020 Dengan Metode K-Nearest Neighbor (Studi Kasus: PT.Ramadika Mandiri)," *Jurnal Informatika Multi*, vol. 1, no. 2, pp. 126–136, Mar. 2023, Accessed: Mar. 19, 2024. [Online]. Available: <https://jurnal.publikasitecno.id/index.php/multi/article/view/19>
- [20] I. Ahmad, S. Samsugi, and Y. Irawan, "Penerapan Augmented Reality Pada Anatomi Tubuh Manusia Untuk Mendukung Pembelajaran Titik Titik Bekam Pengobatan Alternatif," *Jurnal Teknoinfo*, vol. 16, no. 1, p. 46, Jan. 2022, doi: 10.33365/JTI.V16I1.1521.
- [21] S. Hikmawan *et al.*, "Sentimen Analisis Publik Terhadap Joko Widodo terhadap wabah Covid-19 menggunakan Metode Machine Learning," *Jurnal Kajian Ilmiah*, vol. 20, no. 2, pp. 167–176, May 2020, doi: 10.31599/JKI.V20I2.117.
- [22] S. Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti, M. Dwifebri Purbolaksono, M. Irvan Tantowi, and A. Imam Hidayat, "Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 2, pp. 393–399, Apr. 2021, doi: 10.29207/RESTI.V5I2.3008.
- [23] S. K. Dirjen, P. Riset, D. Pengembangan, R. Dikti, S. Khomsah, and A. S. Aribowo, "Text-Preprocessing Model Youtube Comments in Indonesian," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 4, pp. 648–654, Aug. 2020, doi: 10.29207/RESTI.V4I4.2035.
- [24] A. E. Budiman and A. Widjaja, "Analisis Pengaruh Teks Preprocessing Terhadap Deteksi Plagiarisme Pada Dokumen Tugas Akhir," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, no. 3, Dec. 2020, doi: 10.28932/JUTISI.V6I3.2892.
- [25] M. ' Ruf *et al.*, "Data Mining Untuk Klasifikasi Produk Menggunakan Algoritma K-Nearest Neighbor

Pada Toko Online,” *Prosiding SISFOTEK*, vol. 5, no. 1, pp. 141–145, Sep. 2021, Accessed: Mar. 19, 2024. [Online]. Available: <https://seminar.iaii.or.id/index.php/SISFOTEK/article/view/273>

- [26] S. Sarina and A. M. Tanniewa, “Implementasi Algoritma Support Vector Learning Terhadap Analisis Sentimen Penggunaan Aplikasi Tiktok Shop Seller Center,” *Prosiding SISFOTEK*, vol. 7, no. 1, pp. 165–170, Oct. 2023, Accessed: Mar. 19, 2024. [Online]. Available: <https://seminar.iaii.or.id/index.php/SISFOTEK/article/view/404>
- [27] I. G. T. Suryawan and N. M. A. R. Devi, “Implementasi Metode Moora Dalam Menentukan Kelayakan Penerima Bantuan Langsung Tunai Covid-19,” *Jurnal Teknologi Informasi dan Komputer*, vol. 9, no. 2, Jan. 2023, Accessed: Mar. 19, 2024. [Online]. Available: <https://jurnal.undhirabali.ac.id/index.php/jutik/article/view/2349>
- [28] I. Yunanto and S. Yulianto, “Twitter Sentiment Analysis Pedulilindungi Application Using Naïve Bayes And Support Vector Machine,” *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 4, pp. 807–814, Aug. 2022, doi: 10.20884/1.JUTIF.2022.3.4.292..