



# Analisis Sentimen Twitter Terhadap Program MBKM Menggunakan Decision Tree dan Support Vector Machine

Lita Astri Pramesti, Nunik Pratiwi\*

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta  
Jl. Tanah Merdeka No.20, Rambutan, Kec. Ciracas, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia

Email: <sup>1</sup>lita2302@gmail.com, <sup>2,\*</sup>npratiwi@uhamka.ac.id

Email Penulis Korespondensi: npratiwi@uhamka.ac.id

Submitted: 07/07/2023; Accepted: 22/07/2023; Published: 26/07/2023

**Abstrak**—Tujuan dari penelitian ini yaitu untuk menganalisis opini pengguna twitter terhadap program MBKM dengan menggunakan Decision Tree dan Support Vector Machine. Penelitian ini menggunakan 849 data dimana perbandingan datasetnya yaitu 80% data train dan 20% data test. Dalam dataset tersebut terdapat sentimen positif yang berjumlah 524, sedangkan untuk sentiment negatif berjumlah 320. Hal ini menggambarkan bahwa opini para pengguna twitter terhadap program MBKM cenderung positif. Hasil evaluasi penelitian menunjukkan akurasi dari Support Vector Machine sebesar 84.76% lebih tinggi dibandingkan dengan Decision Tree yang hanya menghasilkan akurasi sebesar 72.86%. Berdasarkan hasil tersebut dapat disimpulkan bahwa Algoritma Support Vector Machine lebih unggul dalam melakukan analisis sentimen terhadap program MBKM. Hasil dari penelitian ini diharapkan dapat digunakan sebagai masukan untuk pengembangan program tersebut.

**Kata Kunci:** Analisis Sentimen; MBKM; Kampus Merdeka; Decision Tree; SVM; Rapidminer

**Abstract**—The purpose of this research is to analyze Twitter users' opinions on the MBKM program using Decision Tree and Support Vector Machine. The study utilized 849 data with a dataset ratio of 80% for training and 20% for testing. In the dataset, there were 524 instances of positive sentiment and 320 instances of negative sentiment. This indicates that Twitter users' opinions towards the MBKM program tend to be positive. The research evaluation results showed that the Support Vector Machine achieved an accuracy of 84.76%, which is higher than the accuracy of 72.86% obtained by the Decision Tree. Based on these results, it can be concluded that the Support Vector Machine algorithm outperforms the Decision Tree in sentiment analysis of the MBKM program. The findings of this research are expected to provide input for the development of the program.

**Keywords:** Sentiment Analysis; MBKM; Kampus Merdeka; Decision Tree; SVM; Rapidminer

## 1. PENDAHULUAN

Kemajuan teknologi yang semakin pesat mampu memberikan banyak solusi dan kemudahan dalam memecahkan banyak masalah di banyak bidang, salah satunya adalah bidang Pendidikan. Mahasiswa sebagai agen perubahan harus mampu memanfaatkan teknologi agar dapat menjadi lebih adaptif dan fleksibel untuk terus mengembangkan diri. Untuk itu Kementerian Pendidikan dan Budaya (Kemendikbud) melakukan inisiatif dengan membuat program Merdeka Belajar Kampus Merdeka (MBKM) [1]

Program MBKM dikembangkan oleh Kemendikbud untuk mengatasi permasalahan yang muncul pada pembelajaran daring tidak efektif akibat Virus Corona pada tahun 2019. Kemendikbud mengambil kebijakan untuk mengurangi penyebaran virus dengan cara New Normal, Work From Home (WFH), dan juga PSBB [2]. Oleh karena itu, Kemendikbud membuat program MBKM sebagai solusi untuk memberikan pilihan belajar yang lebih fleksibel bagi siswa selama pandemi.

Melalui MBKM, mahasiswa dapat mengakses materi pembelajaran dan mengerjakan tugas secara online serta melakukan evaluasi yang juga dilakukan secara online. Program ini cocok bagi mahasiswa yang ingin belajar dengan fleksibilitas waktu dan lokasi yang lebih baik sesuai dengan kebutuhan mereka. Namun karena program ini baru saja dibuat maka tentunya terdapat pro dan kontra tentang program ini, salah satunya yaitu isu yang terkait dengan kurikulum dan mekanisme konversi SKS [1]. Dengan demikian, analisis sentimen merupakan salah satu sarana untuk melakukan survey terkait dengan opini pengguna twitter terhadap MBKM agar MBKM dapat berkembang lebih baik lagi. Penulis memanfaatkan teknik text mining dalam melakukan survei untuk mengumpulkan data.

Text mining adalah salah satu bidang dalam kecerdasan buatan (artificial intelligence) yang bertujuan untuk mengambil informasi yang berharga dari data teks yang tidak terstruktur dengan cara mengidentifikasi dan menganalisis pola-pola yang ada [3]. Text mining mempunyai tujuan utama, yaitu untuk mengambil informasi yang bernilai dari sejumlah dokumen. Text mining melibatkan pengolahan dataset yang besar dan kompleks dimana sebagian besar informasinya tidak relevan dan mengganggu [4]. Sebagian besar data yang tersedia di internet atau dalam lingkungan perusahaan memiliki format yang tidak teratur. Oleh karena itu, text mining diperlukan untuk mengolah data tersebut dengan melakukan pengkategorian dan pengelompokan sehingga dapat diambil informasi yang berguna.

Analisis sentimen merupakan bidang ilmu yang melakukan analisis berdasarkan sentimen, opini, pendapat, sikap, emosi, evaluasi, dan penilaian seseorang tentang suatu produk, individu, atau organisasi, suatu isu, peristiwa, atau topik tertentu [5]. Hasil dari analisis sentimen yang sudah dilakukan dapat diklasifikasikan menjadi dua yakni positif dan negatif. Ada beberapa algoritma yang dapat digunakan untuk mengklasifikasi data contohnya adalah



Support Vector Machine, Naïve Bayes, Decision Tree, dan K-Nearest Neighbor. Peneliti memilih metode Decision Tree karena metode pembelajaran mesin ini telah terbukti menghasilkan akurasi yang baik dalam tugas-tugas klasifikasi dan prediksi [6]. Selain itu, peneliti juga memilih metode Support Vector Machine (SVM) sebagai pembanding karena SVM adalah algoritma yang efisien dalam analisis sentimen karena dapat mengklasifikasikan sentimen baru dengan menggunakan sedikit titik penting (Support Vectors). Dengan menghilangkan pola-pola yang tidak memberikan informasi penting, SVM membantu meningkatkan akurasi dan efisiensi dalam analisis sentimen [7]. Dengan membandingkan hasil dari metode Decision Tree dan SVM, peneliti dapat memperoleh pemahaman yang lebih menyeluruh tentang performa dan keefektifan keduanya dalam konteks analisis sentimen.

Adapun Penelitian sebelumnya yang serupa mengenai Support Vector Machine [8] yang dilakukan oleh Styawati, Nirwana Hendrastuty, Ari Yanti Rahmadhani, dan Auliya Rahman Isnain Menggunakan Metode Support Vector Machine untuk Analisis Sentimen Masyarakat tentang Program Kartu Prakerja Pada Twitter untuk mengklasifikasikan 2000 data twitter. Berdasarkan hasil dari evaluasi yang dilakukan akurasi terbaiknya yaitu dengan kernel linear menghasilkan accuracy sebesar 98.34% dan sentimen terhadap program kartu prakerja dimasa pandemi oleh masyarakat pengguna twitter lebih mengarah ke netral.

Penelitian sebelumnya yang serupa tentang analisis sentimen menggunakan metode Naïve Bayes, K-Nearest Neighbors, dan Decision Tree [9] juga dilakukan oleh Abdul Rozaq, Eka Resty Novieta Sari, Yessi Yunitasari, Restyono Ilham Syahputra, dan Kelik Sussolaikah. Penelitian yang dilakukan terhadap implementasi dari MBKM. Terdapat 475 jumlah data, perbandingan antara data latih dan data ujinya adalah 80:20. Hasil akurasi sebesar 37,21% untuk Decision Tree, K-Nearest Neighbors sebesar 96,90% dan Naïve Bayes sebesar 99,22%.

Penelitian sebelumnya yang serupa [10] dilakukan menggunakan metode Support Vector Machine dan Naïve Bayes tentang analisis sentimen terhadap program kampus merdeka oleh Irma Putri Rahayu, Jamaludin Indra, dan Ahmad Fauzi. Dataset yang digunakan berjumlah 1118 data dimana sentimen positifnya berjumlah 618 data dan sentimen negatifnya berjumlah 500 data. Hasil akurasi dari penelitian ini sebesar 97.95% dan 86.30% menggunakan data latih sebanyak 783 data dan data uji sebanyak 335 data menggunakan Naïve Bayes. Kemudian menggunakan SVM kernel linear menghasilkan akurasi sebesar 99.74% dan 86.01% dengan data latih dan data uji yang sama.

Penelitian serupa [11] tentang analisis sentimen juga dilakukan oleh Muhammad Syarifuddin. Dengan menggunakan metode Decision Tree, Naïve Bayes, dan KNN penelitian tentang opini public terhadap efek psbb pada twitter menghasilkan Decision Tree sebagai algoritma tertinggi dari dua algoritma yang lainnya dengan accuracy sebesar 83.3%, precision 79%, dan recall 87.17%.

Penelitian yang serupa [12] yang dilakukan oleh Erina Undamayanti, Ismi Kaniawulan, dan Teguh Iman Hermanto dengan metode Naïve Bayes Berbasis Particle Swarm Optimization tentang analisis sentimen terhadap pelaksanaan program MBKM. Sebanyak 428 data digunakan dalam penelitian ini. Hasil penelitian adalah sentimen positif sebesar 61.92%. Meskipun terdapat sekitar 38.08% sentimen negatif yang muncul, dapat disimpulkan bahwa masyarakat pengguna Twitter terutama mahasiswa dapat menerima program MBKM dengan baik.

Penelitian sebelumnya yang dilakukan dalam bidang analisis sentimen terkait dengan Program Merdeka Belajar Kampus Merdeka (MBKM) hanya fokus pada pengukuran tingkat kepuasan pengguna Twitter terhadap program tersebut. Namun, penelitian ini akan mengevaluasi opini pengguna Twitter terhadap program MBKM dengan menggunakan metode klasifikasi Decision Tree dan Support Vector Machine. Dengan demikian, hasil penelitian ini akan memberikan gambaran yang lebih menyeluruh tentang opini pengguna Twitter terhadap program MBKM dan dapat digunakan sebagai masukan untuk pengembangan program tersebut.

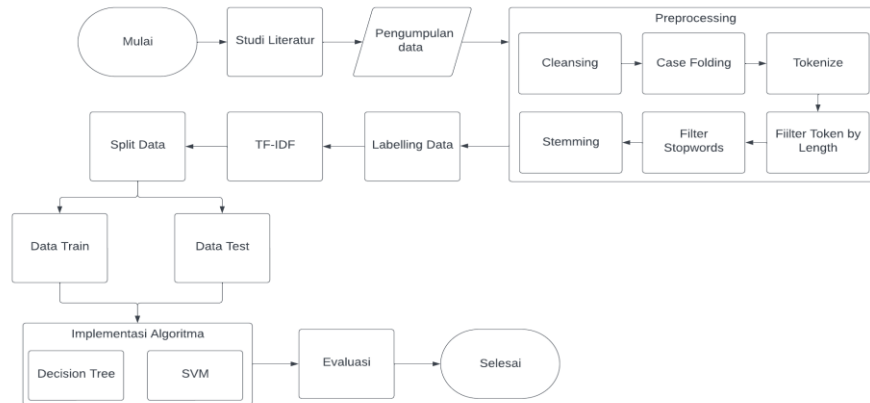
## **2. METODOLOGI PENELITIAN**

### **2.1 Tahapan Penelitian**

Terdapat beberapa tahapan untuk melakukan sebuah penelitian Analisis Sentimen menggunakan Decision Tree dan Support Vector Machine dengan RapidMiner. Tahapannya sebagai berikut:

- a. Studi literatur: Tahap pertama adalah melakukan studi literatur tentang topik yang akan diteliti. Ini melibatkan membaca dan mempelajari penelitian terkait, metode analisis sentimen, algoritma Decision Tree dan Support Vector Machine yang digunakan dalam konteks analisis sentimen.
- b. Selanjutnya, data tweet yang relevan dengan topik penelitian dikumpulkan. Jumlah data yang diperoleh dalam penelitian ini sebanyak 849 tweet.
- c. Preprocessing data: Data tweet yang telah diberi label kemudian diproses dengan melakukan preprocessing. Langkah-langkah preprocessing ini mencakup cleansing, tokenizing, filtering (stopwords removal), filter token by length, dan stemming.
- d. Labelling data: Data tweet kemudian diberi label menjadi dua kelas, yaitu positif dan negatif. Proses ini melibatkan mengidentifikasi sentimen dalam setiap tweet dan menentukan label yang sesuai berdasarkan apakah sentimen tersebut positif atau negatif. Labelling data sangat penting karena mempengaruhi hasil dari akurasi [13].

- e. Model klasifikasi dengan Algoritma Support Vector Machine dan Decision Tree diimplementasikan menggunakan RapidMiner. Model tersebut kemudian diuji menggunakan data pengujian untuk memprediksi sentimen dari tweet yang belum pernah dilihat sebelumnya.
- f. Akhirnya, hasil analisis dievaluasi menggunakan confusion matrix untuk mengukur akurasi model klasifikasi. Confusion matrix memberikan informasi tentang seberapa baik model dapat memprediksi dengan benar kelas sentimen positif dan negatif.



**Gambar 1.** Alur penelitian

## 2.2 Preprocessing

Preprocessing merupakan proses yang bertujuan untuk mengurangi noise pada data [14], meningkatkan kualitas data, dan memperoleh representasi teks yang lebih terstruktur. Terdapat beberapa tahap dalam preprocessing data [15], yaitu:

- a. Cleansing Data : menghapus elemen-elemen yang tidak diperlukan pada data.
- b. Case Folding : merubah seluruh huruf menjadi huruf kecil.
- c. Tokenizing : memecah kalimat menjadi kata yang dilakukan berdasarkan spasi yang ada diantara kata-kata dalam kalimat.
- d. Filtering (stopwords removal) : menghapus kata-kata yang tidak bermakna seperti kata ganti orang ketiga dan kata sambung.
- e. Filter token by length : menghapus kata sesuai dengan jumlah huruf yang sudah ditentukan [16].
- f. Stemming : merubah kata yang mengandung imbuhan menjadi kata dasar.

## 2.3 Decision Tree

Dalam data mining untuk mengelola jumlah data yang besar, algoritma klasifikasi dapat digunakan.. Decision Tree merupakan salah satu metode yang digunakan untuk mengklasifikasikan data dan banyak digunakan dalam berbagai hal misalnya machine learning, pattern identification, dan image processing [17]. Algoritma Decision Tree mencoba meningkatkan keakuratan dengan menghilangkan cabang-cabang pohon yang merefleksikan noises dalam data. Decision Tree menjelaskan berbagai opsi potensial untuk mengatasi masalah, mempertimbangkan faktor-faktor yang dapat mempengaruhi opsi-opsi ini dan memberikan perkiraan hasil untuk setiap kemungkinan [18].

$$\text{Entropy (A)} = \sum_{i=1}^n -p_i \log_2 p_i \quad (1)$$

## 2.4 Support Vector Machine

Support Vector Machine (SVM) diperkenalkan oleh Vapnik pada tahun 1992 [19]. SVM adalah metode pembelajaran linier yang mencari hyperplane optimal untuk memisahkan dua kelas yang berbeda, yaitu kelas positif dan negatif [20]. Pada awalnya, prinsip kerja SVM adalah sebagai metode klasifikasi linier (linear classifier). Namun, metode ini kemudian dikembangkan untuk dapat menangani permasalahan klasifikasi non-linier dengan memanfaatkan fungsi kernel.

Fungsi kernel memungkinkan SVM untuk bekerja dalam ruang kerja dengan dimensi yang tinggi, di mana data dapat diproyeksikan ke ruang fitur yang lebih tinggi untuk memisahkan kelas-kelas yang sulit dipisahkan secara linier. Dengan menggunakan fungsi kernel, SVM dapat menangani data dengan pola yang kompleks dan memungkinkan pemisahan yang optimal antara kelas-kelas yang berbeda [15].

## 2.5 Confusion Matrix

Confusion matrix merupakan sebuah tabel yang digunakan untuk menganalisis sejauh mana metode klasifikasi mampu memprediksi dengan akurat kelas-kelas data [21]. Confusion Matrix dan rumus perhitungannya [22] dapat dilihat sebagai berikut :

**Tabel 1.** Confusion Matrix

| Pred. Kelas | Kelas   |         |
|-------------|---------|---------|
|             | Negatif | Positif |
| Negatif     | TN      | FN      |
| Positif     | FP      | TP      |

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FN+FP)} \times 100\% \tag{2}$$

$$\text{Precision} = \frac{TP}{(TP+FP)} \times 100\% \tag{3}$$

$$\text{Recall} = \frac{TP}{(TP+FN)} \times 100\% \tag{4}$$

Keterangan [23]:

TP : True Positive, adalah jumlah data positif yang diprediksi benar oleh model

FN : False Negative, adalah jumlah data positif yang diprediksi salah oleh model

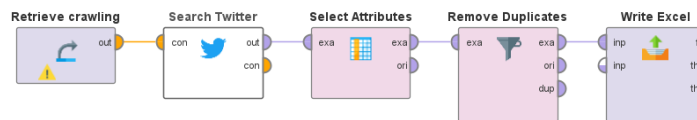
TN : True Negative, adalah jumlah data negatif yang diprediksi salah oleh model

FP : False Positive, adalah jumlah data negatif yang diprediksi benar oleh model

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pengumpulan Data

Pada penelitian ini, pengumpulan data dilakukan menggunakan teknik crawling dengan menggunakan tool RapidMiner. Tahap pertama adalah menghubungkan koneksi Twitter API ke dalam RapidMiner. Selanjutnya, penulis menuliskan query "mbkm" dan "kampus merdeka". Dari proses crawling yang dilakukan, berhasil didapatkan sebanyak 849 data yang akan digunakan dalam penelitian ini. Data tersebut akan menjadi dasar analisis sentimen terkait dengan Program Merdeka Belajar Kampus Merdeka (MBKM). Setelah itu, dilakukan proses select attributes untuk menghilangkan atribut yang tidak digunakan, sehingga hanya atribut Text yang dipertahankan. Langkah terakhir adalah export file dalam format excel menggunakan operator Write Excel. Gambar 2 menampilkan tampilan dan hasil dari proses crawling data pada software RapidMiner.



**Gambar 2.** Proses Crawling Data Pada Rapidminer

Setelah semua operator dan parameter berhasil di eksekusi maka akan keluar hasil dari crawling data berupa text yang berisi cuitan para pengguna twitter yang dapat dilihat pada gambar 3.

| Row No. | Id              | Text   |
|---------|-----------------|--|
| 1       | 167668428856... | RT @lyaaserahdeh: @schfess Kalo aku dari semester 1 sampai semester 2 ini, aku libas semua. Mulai dari lomba, Volunteer, organisasi or himp... |
| 2       | 167663830136... | @schfess jangan lupa * MBKM*   |
| 3       | 167662699581... | Um! Yang mbkm ikut war KRS juga atau tidak??   |
| 4       | 167662078058... | Alhamdulillah project mbkm bisa di gaskan 🙏  |
| 5       | 167661634670... | @lemkoala halo kak mjb aku mau tanya kamu ikut mbkm kewirausahaan itu ngapain aja ya kak? 🙏  |
| 6       | 167661563534... | @leucids @unudmenfess kak mjb mau tanya, kamu pernah ikut mbkm kewirausahaan kah?  |
| 7       | 167661224584... | Halo #TemanMamin,  |
| 8       | 167661091408... | mbkm tai ga adil bgt   |
| 9       | 167660921615... | @collegemenfess Ngga diprodiku, bahkan ngga semua mata kuliah bisa direkognisi, di prodiku program mbkm cuman bisa diambil disemester terter   |
| 10      | 167659304361... | [cm] guys sender lagi bingung, menurut kalian bagus magang dimana? Sender bingung antara bap*eda atau bp5, atau kalian ada rekomendasi t       |

**Gambar 3.** Hasil dari Crawling Data

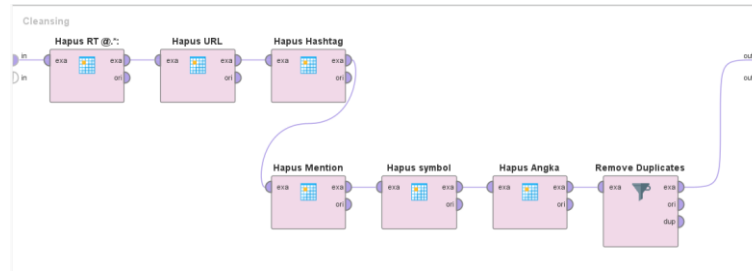
#### 3.2 Preprocessing Data

Setelah pengumpulan data sejumlah 849 data, langkah selanjutnya dalam penelitian ini adalah preprocessing data. Preprocessing data bertujuan untuk meningkatkan kualitas data dan menghapus kebisingan data. Preprocessing data dilakukan dengan beberapa tahap, yaitu:

a. Cleansing Data

Untuk melakukan cleansing data, pertama penulis menggunakan operator replace dan kemudian parameternya diubah sesuai dengan data yang ingin penulis cleansing seperti data yang double (duplikat), data yang tidak

memiliki nilai (missing value), angka [0-9], mention, hashtag, kata "RT" (Retweet), tautan URL, dan simbol-simbol seperti tanda baca dan emoji.



**Gambar 4.** Cleansing Data pada Rapidminer

Hasil dari cleansing data dapat dilihat pada tabel 2 dimana pada tabel tersebut terdapat data sebelum di cleansing dan data sesudah di cleansing.

**Tabel 2.** Hasil dari Cleansing Data

| Sebelum  | Sesudah  |
|--|--|
| emang bener yh ikut mbkm sgt mendongkrak ipk<br>🤔😅🙏<br>@hrdbacot Merelakan lulus sarjana 4 tahun, padahal sebenarnya bisa 3.5 tahun,<br>di kuliah ada tuh seleksi studi independennya MBKM, disini gua nemu mitra yang bikin gua kepingin masuk ke sana pada pandangan pertama. dan yaa bener aja gua batiba dilolosin di sana padahal katanya cukup susah buat diterima di sana | emang bener yh ikut mbkm sgt mendongkrak ipk<br><br>Merelakan lulus sarjana tahun padahal sebenarnya bisa tahun<br>di kuliah ada tuh seleksi studi independennya MBKM disini gua nemu mitra yang bikin gua kepingin masuk ke sana pada pandangan pertama. dan yaa bener aja gua batiba dilolosin di sana padahal katanya cukup susah buat diterima di sana |

b. Case Folding

Pada tahap Case Folding, penulis merubah seluruh huruf alfabet yang sudah di cleansing menjadi huruf kecil (lower case).

**Tabel 3.** Hasil dari Case Folding

| Sebelum  | Sesudah  |
|--|--|
| emang bener yh ikut mbkm sgt mendongkrak ipk<br>Merelakan lulus sarjana tahun padahal sebenarnya bisa tahun<br>di kuliah ada tuh seleksi studi independennya MBKM disini gua nemu mitra yang bikin gua kepingin masuk ke sana pada pandangan pertama. dan yaa bener aja gua batiba dilolosin di sana padahal katanya cukup susah buat diterima di sana | emang bener yh ikut mbkm sgt mendongkrak ipk<br>merelakan lulus sarjana tahun padahal sebenarnya bisa tahun<br>di kuliah ada tuh seleksi studi independennya mbkm disini gua nemu mitra yang bikin gua kepingin masuk ke sana pada pandangan pertama. dan yaa bener aja gua batiba dilolosin di sana padahal katanya cukup susah buat diterima di sana |

c. Tokenizing

Pada tahap ini, proses pemecahan dilakukan berdasarkan spasi di antara kata-kata dalam kalimat. Setiap kata dalam kalimat dianggap sebagai sebuah token atau unit terpisah.

**Tabel 4.** Hasil dari Tokenizing

| Sebelum   | Sesudah  |
|---|--|
| emang bener yh ikut mbkm sgt mendongkrak ipk  | ['emang', 'bener', 'yh', 'ikut', 'mbkm', 'sgt', 'mendongkrak', 'ipk']  |
| merelakan lulus sarjana tahun padahal sebenarnya bisa tahun   | ['merelakan', 'lulus', 'sarjana', 'tahun', 'padahal', 'sebenarnya', 'bisa', 'tahun']   |
| di kuliah ada tuh seleksi studi independennya mbkm disini gua nemu mitra yang bikin gua kepingin masuk ke sana pada pandangan pertama. dan yaa bener aja gua batiba dilolosin di sana padahal katanya cukup susah buat diterima di sana | ['di', 'kuliah', 'ada', 'tuh', 'seleksi', 'studi', 'independennya', 'mbkm', 'disini', 'gua', 'nemu', 'mitra', 'yang', 'bikin', 'gua', 'kepingin', 'masuk', 'ke', 'sana', 'pada', 'pandangan', 'pertama', 'dan', 'yaa', 'bener', 'aja', 'gua', 'batiba', 'dilolosin', 'di', 'sana', 'padahal', 'katanya', 'cukup', 'susah', 'buat', 'diterima', 'di', 'sana'] |

d. Filtering (stopwords removal)

Pada tahap ini penulis menghapus kata-kata yang tidak mempunyai makna dan tidak penting untuk mempermudah proses selanjutnya.

**Tabel 5.** Hasil dari Stopwords Removal

| Sebelum  | Sesudah   |
|--|---|
| ['emang', 'bener', 'yh', 'ikut', 'mbkm', 'sgt', 'mendongkrak', 'ipk']  | ['emang', 'bener', 'yh', 'mbkm', 'sgt', 'mendongkrak', 'ipk']   |
| ['merelakan', 'lulus', 'sarjana', 'tahun', 'padahal', 'sebenarnya', 'bisa', 'tahun']   | ['merelakan', 'lulus', 'sarjana', 'tahun', 'sebenarnya', 'tahun']   |
| ['di', 'kuliah', 'ada', 'tuh', 'seleksi', 'studi', 'independennya', 'mbkm', 'disini', 'gua', 'nemu', 'mitra', 'yang', 'bikin', 'gua', 'kepingin', 'masuk', 'ke', 'sana', 'pada', 'pandangan', 'pertama', 'dan', 'yaa', 'bener', 'aja', 'gua', 'batiba', 'dilolosin', 'di', 'sana', 'padahal', 'katanya', 'cukup', 'susah', 'buat', 'diterima', 'di', 'sana'] | ['kuliah', 'tuh', 'seleksi', 'studi', 'independennya', 'mbkm', 'gua', 'nemu', 'mitra', 'bikin', 'gua', 'kepingin', 'masuk', 'pandangan', 'pertama', 'yaa', 'bener', 'aja', 'gua', 'batiba', 'dilolosin', 'susah', 'diterima'] |

e. Filter Token by Length

Tahap ini berfungsi untuk mengeliminasi kata dengan jumlah huruf yang sudah di tentukan. Parameter yang digunakan dalam operator ini yaitu minimal sebanyak empat huruf dan maksimal sebanyak dua puluh lima huruf.

**Tabel 6.** Hasil dari Filter token by length

| Sebelum   | Sesudah   |
|---|---|
| ['emang', 'bener', 'yh', 'mbkm', 'sgt', 'mendongkrak', 'ipk']   | ['emang', 'bener', 'mbkm', 'mendongkrak']   |
| ['merelakan', 'lulus', 'sarjana', 'tahun', 'sebenarnya', 'tahun']   | ['merelakan', 'lulus', 'sarjana', 'tahun', 'sebenarnya', 'tahun']   |
| ['kuliah', 'tuh', 'seleksi', 'studi', 'independennya', 'mbkm', 'gua', 'nemu', 'mitra', 'bikin', 'gua', 'kepingin', 'masuk', 'pandangan', 'pertama', 'yaa', 'bener', 'aja', 'gua', 'batiba', 'dilolosin', 'susah', 'diterima'] | ['kuliah', 'seleksi', 'studi', 'independennya', 'mbkm', 'nemu', 'mitra', 'bikin', 'kepingin', 'masuk', 'pandangan', 'pertama', 'bener', 'batiba', 'dilolosin', 'susah', 'diterima'] |

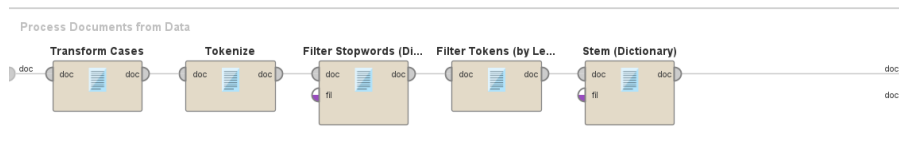
f. Stemming

Pada tahap ini penulis merubah kata yang mengandung imbuhan menjadi kata dasar.

**Tabel 7.** Hasil dari Stemming

| Sebelum   | Sesudah   |
|---|---|
| ['emang', 'bener', 'mbkm', 'mendongkrak']   | ['emang', 'bener', 'mbkm', 'mendongkrak']   |
| ['merelakan', 'lulus', 'sarjana', 'tahun', 'sebenarnya', 'tahun']   | ['merelakan', 'lulus', 'sarjana', 'tahun', 'sebenarnya', 'tahun']   |
| ['kuliah', 'seleksi', 'studi', 'independennya', 'mbkm', 'nemu', 'mitra', 'bikin', 'kepingin', 'masuk', 'pandangan', 'pertama', 'bener', 'batiba', 'dilolosin', 'susah', 'diterima'] | ['kuliah', 'seleksi', 'studi', 'independennya', 'mbkm', 'nemu', 'mitra', 'bikin', 'kepingin', 'masuk', 'pandangan', 'pertama', 'bener', 'batiba', 'dilolosin', 'susah', 'diterima'] |

Pada gambar 5 dapat dilihat keseluruhan proses preprocessing data mulai dari cleansing, case folding, filtering (stopwords removal), filter tokens by length, dan stemming dengan menggunakan rapidminer.



**Gambar 5.** Proses Preprocessing data di Rapidminer

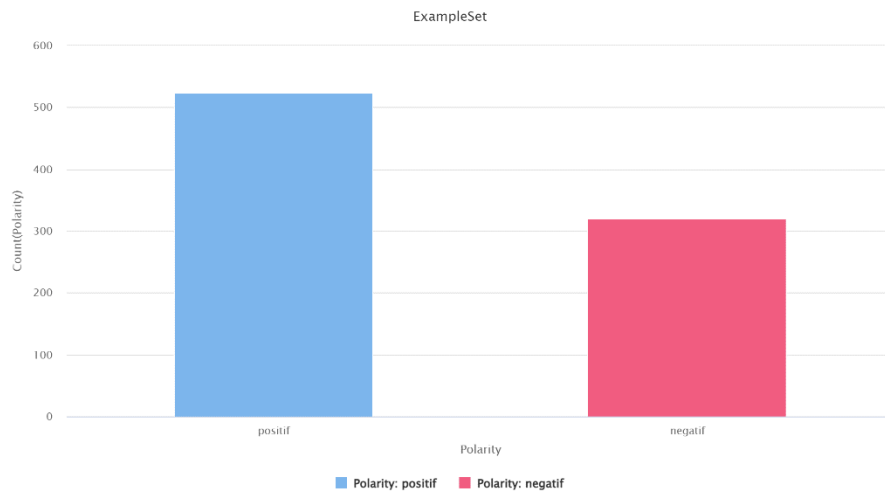
### 3.3 Labelling Data

Pada tahap berikutnya dilakukan pelabelan data. Pelabelan data dilakukan untuk membagi data menjadi dua kelas, yaitu positif dan negatif. Tujuan dari pelabelan ini adalah membedakan sentimen atau opini yang bersifat positif dan negatif terkait dengan Program MBKM. Dalam pelabelan data, setiap data akan diberikan label berdasarkan sentimen atau opini yang terkandung dalam teks. Data yang mengandung sentimen positif akan diberi label

"positif", sedangkan data yang mengandung sentimen negatif akan diberi label "negatif". Penulis melakukan pelabelan data secara manual dengan membaca dan memahami konteks dan makna dari setiap data, kemudian memberikan label yang sesuai.

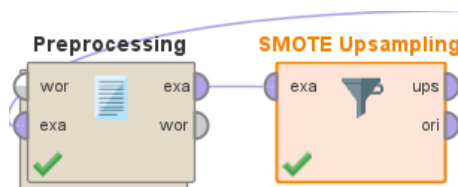
**Tabel 8.** Pelabelan Data

| Text  | Polarity |
|---|----------|
| emang bener mbkm mendongkrak  | positif  |
| merelakan lulus sarjana tahun sebenarnya tahun mbkm skripsian                             | negatif  |
| kuliah seleksi studi independennya mbkm nemu mitra bikin kepingin masuk pandangan pertama | positif  |
| bener batiba dilolosin susah diterima   |          |



**Gambar 6.** Perbandingan antara banyaknya Label Positif dan Negatif

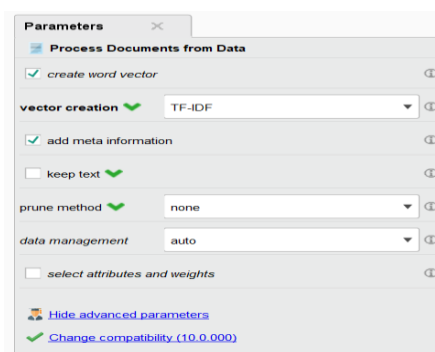
Setelah proses pelabelan, ditemukan bahwa terdapat 524 data berlabel positif dan 320 data berlabel negatif. Hal ini menunjukkan adanya ketidakseimbangan dalam distribusi kelas sentimen dalam dataset yang digunakan. Untuk mencegah agar tidak terjadinya overfitting dan underfitting maka penulis menggunakan operator SMOTE upsampling. Operator ini berfungsi untuk menyeimbangkan dataset.



**Gambar 7.** Operator dari SMOTE Upsampling

### 3.4 TF-IDF

Langkah selanjutnya yaitu Pembobotan Kata dengan TF-IDF. Semakin banyak dokumen yang akan diproses, maka semakin banyak pula fitur yang akan dihasilkan dalam pembobotan kata. Dengan menggunakan TF-IDF, penulis dapat mengidentifikasi kata-kata yang memiliki bobot tinggi dalam suatu dokumen, yang dapat memberikan indikasi pentingnya kata tersebut dalam konteks dokumen tersebut [14].



**Gambar 8.** Pembobotan kata dengan TF-IDF pada Rapidminer

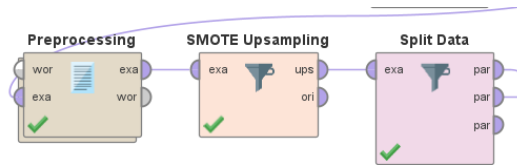
Selanjutnya untuk memberikan visualisasi terhadap kata-kata yang sering muncul pada penelitian ini penulis menggunakan wordcloud.



**Gambar 9.** Visualisasi dari TF-IDF berdasarkan Total Occurance

### 3.5 Pembagian Dataset

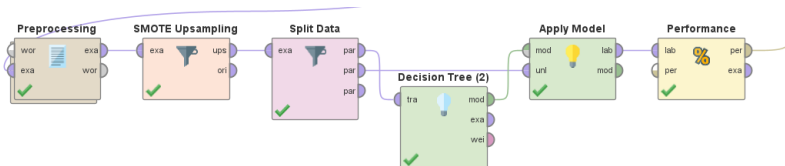
Sebelum melakukan klasifikasi, penulis membagi dataset menjadi data train dan data test. Data train sebagai pelatih model, sedangkan data test digunakan sebagai pengujian kinerja model yang telah dilatih. Hal ini dilakukan agar model yang dihasilkan dapat digeneralisasikan ke data yang belum pernah model pelajari. Untuk melakukan pembagian data pada rapidminer, penulis menggunakan operator Split Data dengan rasio partisi 0.8:0.2 atau data train sebanyak 80% dan data test sebanyak 20%.



**Gambar 10.** Pembagian dataset dengan operator Split Data

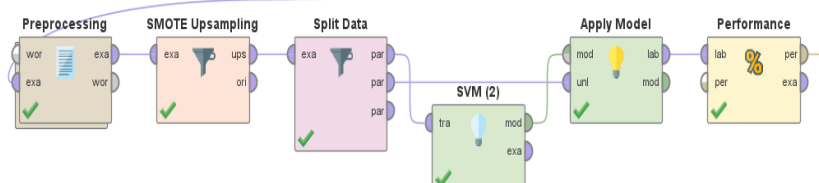
### 3.6 Implementasi Algoritma

Tahap selanjutnya yaitu implementasi klasifikasi menggunakan Algoritma Decision Tree dan Support Vector Machine (SVM). Untuk melakukan klasifikasi menggunakan Decision Tree penulis menghubungkan konektor data train ke operator decision tree untuk melatih model, kemudian penulis menghubungkan data test ke operator apply model. Setelah itu model yang telah dilatih di data train akan diuji ke data test. Parameter yang digunakan di operator decision tree menggunakan criteria accuracy agar akurasi yang dihasilkan oleh algoritma ini dapat maksimal.



**Gambar 11.** Implementasi Algoritma Decision Tree

Hal yang sama dilakukan pada implementasi algoritma Support Vector Machine, yaitu dengan menghubungkan data train ke operator SVM (LibSVM) dan menghubungkan data test ke apply model. Setelah itu model yang sudah dilatih akan di evaluasi. Kernel yang digunakan dalam penelitian ini adalah kernel poly.



**Gambar 12.** Implementasi Algoritma Support Vector Machine

### 3.7 Evaluasi

Tahap yang terakhir yaitu evaluasi, dalam tahap ini penulis dapat mengukur bagaimana performa dari model untuk mengklasifikasikan hasil sentimen. Evaluasi juga membantu dalam menilai kinerja dan keakuratan model. Di gambar 12 terlihat hasil confusion matrix dari decision tree dengan accuracy 72.86%, precision positif 68.75%, precision negatif 79.27%, recall positif 83.81%, dan recall negatif 61.90%,

accuracy: 72.86%

|               | true positif | true negatif | class precision |
|---------------|--------------|--------------|-----------------|
| pred. positif | 88           | 40           | 68.75%          |
| pred. negatif | 17           | 65           | 79.27%          |
| class recall  | 83.81%       | 61.90%       |                 |

**Gambar 13.** Hasil Confusion Matrix dari Decision Tree

Hasil confusion matrix dari Support Vector Machine ditampilkan pada gambar 13 dimana accuracy sebesar 84.76%, precision positif 78.74%, precision negatif 93.98%, recall positif 95.24% dan recall negatif 74.29%.

accuracy: 84.76%

|               | true positif | true negatif | class precision |
|---------------|--------------|--------------|-----------------|
| pred. positif | 100          | 27           | 78.74%          |
| pred. negatif | 5            | 78           | 93.98%          |
| class recall  | 95.24%       | 74.29%       |                 |

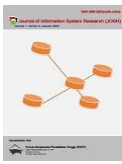
**Gambar 14.** Hasil Confusion Matrix dari Support Vector Machine

## 4. KESIMPULAN

Hasil dari pemrosesan dataset berjumlah 849 untuk mengetahui bagaimana opini pengguna twitter tentang MBKM, tetapi setelah dilakukan proses preprocessing jumlah dataset berkurang menjadi 844 dikarenakan terdapat replika data yang sama. Dalam dataset tersebut terdapat sentimen positif yang berjumlah 524, sedangkan untuk sentiment negatif berjumlah 320. Hal ini menunjukkan bahwa opini para pengguna twitter terhadap program MBKM cenderung positif. Akurasi tertinggi dihasilkan oleh Support Vector Machine sebesar 84.76%, sedangkan akurasi dari algoritma Decision Tree 72.86%. Selain itu Precision dan juga Recall dari Support Vector Machine menunjukkan hasil yang lebih baik dibanding Decision Tree, dimana Precision Positif Support Vector Machine sebesar 82.24% sedangkan Precision Positif Decision Tree hanya mencapai 68.75%. Sementara itu untuk Precision negatif Support Vector Machine sebesar 93.98%, sedangkan Decision Tree mencapai 79.27%. Recall positif dari kedua algoritma sama-sama sebesar 83.81%, namun recall negatif Support Vector Machine lebih tinggi yaitu sebesar 74.29% dibandingkan dengan Decision Tree yang hanya sebesar 61.90%. Berdasarkan hasil tersebut dapat disimpulkan bahwa Algoritma Support Vector Machine lebih unggul dalam melakukan analisis sentimen terhadap program MBKM. Support Vector Machine memberikan kinerja dan juga Akurasi yang lebih baik dalam memprediksi sentimen positif dan negatif dibandingkan dengan Decision Tree. Namun, tetap perlu mempertimbangkan faktor-faktor lain seperti tingkat kerumitan model, jumlah dataset, dan juga proses pelabelan data.

## REFERENCES

- [1] S. Jurnalis Pipin and H. Kurniawan, "Analisis Sentimen Kebijakan MBKM Berdasarkan Opini Masyarakat di Twitter Menggunakan LSTM," J. SIFO Mikroskil, vol. 23, no. 2, pp. 197–208, 2022, doi: 10.55601/jsm.v23i2.900.
- [2] T. Wiratama Putra, A. Triayudi, and A. Andrianingsih, "Analisis Sentimen Pembelajaran Daring Menggunakan Metode Naïve Bayes, KNN, dan Decision Tree," J. JTIK (Jurnal Teknol. Inf. dan Komunikasi), vol. 6, no. 1, pp. 20–26, 2022, doi: 10.35870/jtik.v6i1.368.
- [3] F. Hashfi, D. Sugiarto, and I. Mardianto, "Sentiment Analysis of An Internet Provider Company Based on Twitter Using Support Vector Machine and Naïve Bayes Method," Ultim. J. Tek. Inform., vol. 4, no. April, pp. 1–6, 2022, doi: 10.31937/ti.v14i1.2384.
- [4] M. Hofmann and R. Klinkenberg, Data Mining and Knowledge Discovery Series Edited by. 2014.
- [5] H. Sibyan and N. Hasanah, "Analisis Sentimen Ulasan Pada Wisata Dieng Dengan Algoritma K-Nearest Neighbor (K-Nn)," J. Penelit. dan Pengabd. Kpd. Masy. UNSIQ, vol. 9, no. 1, pp. 38–47, 2022, doi: 10.32699/ppkm.v9i1.2218.
- [6] A. Harun and D. P. Ananda, "Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan Naïve Bayes dan Decision Tree," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 1, no. 1, pp. 58–63, 2021.
- [7] M. Kantardics, Data mining: Concept, models, methods, and algorithms. 2020.
- [8] N. Hendrastuty et al., "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," J. Inform. J. Pengemb. IT, vol. 6, no. 3, pp. 150–155, 2021, [Online]. Available: <http://situs.com>
- [9] A. Rozaq, Y. Yunitasari, K. Sussolaikah, E. R. N. Sari, and R. I. Syahputra, "Analisis Sentimen Terhadap Implementasi



- Program Merdeka Belajar Kampus Merdeka Menggunakan Naïve Bayes, K-Nearest Neighbors Dan Decision Tree,” *J. Media Inform. Budidarma*, vol. 6, no. 2, p. 746, 2022, doi: 10.30865/mib.v6i2.3554.
- [10] I. P. Rahayu, A. Fauzi, and J. Indra, “Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine,” vol. 4, pp. 296–301, 2022, doi: 10.30865/json.v4i2.5381.
- [11] M. Syarifuddin, “Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada Twitter Dengan Algoritma Decision Tree, Knn, Dan Naïve Bayes,” *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 87–94, 2020, doi: 10.33480/inti.v15i1.1433.
- [12] E. Undamayanti et al., “Analisis Sentimen Menggunakan Metode Naive Bayes Berbasis Particle Swarm Optimization Terhadap Pelaksanaan Program Merdeka Belajar Kampus Merdeka,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 2, pp. 916–930, 2022.
- [13] A. Wibowo, Firman Noor Hasan, Rika Nurhayati, and Arief Wibowo, “Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naïve Bayes Classifier,” *J. Asimetrik J. Ilm. Rekayasa Inov.*, vol. 4, pp. 239–248, 2022, doi: 10.35814/asimetrik.v4i1.3577.
- [14] J. A. Septian, T. M. Fachrudin, and A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepkabolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor,” *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.
- [15] M. N. Ardhiyansyah, R. Umar, and Sunardi, “Analisis Sentimen pada Twitter Menggunakan Metode Support Vector Machine,” *Semin. Nas. Teknol. Fak. Tek. Univ. Krisnadipayana*, vol. 1, no. 1, pp. 739–742, 2019, [Online]. Available: <https://jurnal.teknikunkris.ac.id/index.php/semnastek2019/article/view/343/342>
- [16] I. R. Afandi, F. N. Hasan, A. A. Rizki, N. Pratiwi, and Z. Halim, “Analisis Sentimen Opini Masyarakat Terkait Pelayanan Jasa Ekspedisi Anteraja Dengan Metode Naive Bayes,” *J. Linguist. Komputasional*, vol. 5, no. 2, pp. 63–70, 2022, [Online]. Available: <https://t.co/2HADwgl drL>
- [17] I. P. Wibina, K. Gumi, and A. Syafrianto, “Perbandingan Algoritma Naïve Bayes dan Decision Tree Pada Sentimen Analisis,” vol. 1, pp. 1–15, 2022.
- [18] K. A. Rokhman, B. Berlilana, and P. Arsi, “Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online,” *J. Inf. Syst. Manag.*, vol. 3, no. 1, pp. 1–7, 2021, doi: 10.24076/joism.2021v3i1.341.
- [19] A. Suhairy Batubara, “Analisa Prediksi Penjualan Jersey World Cup 2018 Terlaris Dengan Menggunakan Metode Support Vector Machine (Studi Kasus: Toko One Sport Football Distro),” *J. Inf. Sist. Res.*, vol. 1, no. 4, pp. 246–257, 2020.
- [20] F. F. Irfani, “Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine,” *JBMI (Jurnal Bisnis, Manajemen, dan Inform.)*, vol. 16, no. 3, pp. 258–266, 2020, doi: 10.26487/jbmi.v16i3.8607.
- [21] O. I. Gifari, M. Adha, F. Freddy, and F. F. S. Durrand, “Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine,” *J. Inf. Technol.*, vol. 2, no. 1, pp. 36–40, 2022, doi: 10.46229/jifotech.v2i1.330.
- [22] M. I. Fikri, T. S. Sabrila, and Y. Azhar, “Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter,” *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020, doi: 10.32664/smatika.v10i02.455.
- [23] F. Sidik, I. Suhada, A. H. Anwar, and F. N. Hasan, “Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier,” *J. Linguist. Komputasional*, vol. 5, no. 1, p. 34, 2022, doi: 10.26418/jlk.v5i1.79.