

Analisis Sentimen Terhadap Rangka E-SAF Honda Pada Media Sosial X Dengan Algoritma Naïve Bayes

Akbar Cleary Syafi'i, Ade Davy Wiranata*

Fakultas Teknologi Industri dan Informatika, Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

Email: ¹syaficleary26@gmail.com, ^{2,*}adedavy@uhamka.ac.id,

Email Penulis Korespondensi: adedavy@uhamka.ac.id

Abstrak—Sepeda motor adalah kendaraan terbaik untuk bepergian ketika lalu lintas padat karena sepeda motor memungkinkan orang menghemat waktu saat melakukan perjalanan sehari-hari karena ukurannya yang kecil dan kemampuan untuk bergerak di jalanan yang sempit. Komponen penting dalam sepeda motor adalah rangka sepeda motor, rangka sepeda motor adalah suatu bagian yang berguna untuk menopang bobot komponen-komponen tersebut dalam sistem kendaraan sepeda motor. Namun di isukan bahwa suatu rangka sepeda motor dengan jenis E-SAF memiliki kualitas kurang baik maka diperlukanlah analisis sentimen. Riset ini bertujuan untuk menghimpun jumlah komentar baik bermuatan positif maupun bermuatan negatif dari pengguna media sosial X tentang rangka E-SAF, dan juga mengetahui keakuratan penerapan metode Naive Bayes. Dataset yang di himpun dari media sosial X berjumlah 756 dataset. Kemudian setelah melewati tahapan pembersihan data seperti cleansing, tokenize, dan filter stopword, data yang dapat digunakan untuk penelitian ini berjumlah 696 dataset. Tahapan berikutnya adalah pelabelan data yaitu dengan membagi dataset dengan perbandingan 60:40 yaitu data latih sebesar 60% berjumlah 417 dataset yang sudah dilabeli secara manual dengan hasil data bermuatan negatif sebanyak 224 data, data bermuatan positif sebanyak 193 data sedangkan data uji sebesar 40% dengan berjumlah 279 dataset yang nantinya dilabeli secara otomatis dengan implementasi metode Naive Bayes. Tahapan selanjutnya adalah data uji melewati tahapan processing data supaya data uji tersebut siap di implementasikan ke dalam metode Naive Bayes. Setelah mengimplementasikan metode Naive Bayes, hasil akurasi yang diperoleh sebesar 70,27% dengan jumlah precision sebesar 76% dan juga jumlah recall sebesar 79,17%. Di dapatkan juga data true Positif sebesar 57 dan data true Negatif sebesar 21. Dalam Visualisasi Data juga menampilkan kata yang sering muncul dalam dataset. Disini menunjukkan bahwa metode Naive Bayes cukup efektif untuk klasifikasi analisis sentimen.

Kata Kunci: Analisis Sentimen; Rangka; E-SAF; Naive Bayes; Rapidminer

Abstract—Motorcycles are the best vehicles for traveling when traffic is heavy because motorcycles allow people to save time while going about their daily commute due to their small size and ability to move on narrow streets. An important component in a motorcycle is the motorcycle frame, the motorcycle frame is a useful part to support the weight of these components in the motorcycle vehicle system. However, it is rumored that a motorcycle frame with the E-SAF type has poor quality, so a sentiment analysis is needed. This research aims to collect the number of comments, both positive and negative, from social media users X about the E-SAF framework, and also to determine the accuracy of the application of the Naive Bayes method. The datasets collected from social media X amounted to 756 datasets. Then after going through the stages of data cleaning such as cleansing, tokenize, and stopword filters, the data that can be used for this research amounted to 696 datasets. The next stage is data labeling, namely by dividing the dataset with a ratio of 60:40, namely 60% of the training data totaling 417 datasets that have been manually labeled with the results of 224 negatively charged data, 193 positively charged data while the test data is 40% with a total of 279 datasets which will later be automatically labeled with the implementation of the Naive Bayes method. The next stage is that the test data goes through the data processing stage so that the test data is ready to be implemented into the Naive Bayes method. After implementing the Naive Bayes method, the accuracy obtained was 70.27% with a precision of 76% and also a recall of 79.17%. There was also a true Positive data of 57 and a true Negative data of 21. Data Visualization also displays words that appear frequently in the dataset. Here it shows that the Naive Bayes method is quite effective for the classification of sentiment analysis.

Keywords: Sentiment Analysis; Frame; E-SAF; Naïve Bayes; Rapidminer

1. PENDAHULUAN

Kebutuhan transportasi sepeda motor di Indonesia tumbuh dari tahun ke tahun. Ini adalah kesempatan yang bagus bagi perusahaan yang bekerja dalam sektor kendaraan roda dua atau sepeda motor untuk memasarkan barang mereka dan meningkatkan statistik penjualan mereka dengan berbagai cara. Banyak bisnis menggunakan taktik untuk menjual produk kendaraan beroda dua ini. Beberapa metodenya termasuk penetapan harga, strategi kualitas, dan inovasi produk. Memahami kebutuhan pelanggan dalam barang-barang yang mereka butuhkan telah menjadi prioritas dalam berbagai bisnis, termasuk industri otomotif. Saat ini, kebutuhan kendaraan, khususnya sepeda motor maju dengan pesat [1]. Rangka merupakan komponen terpenting dari sebuah kendaraan bermotor karena memegang andil bagi banyak komponen serta menopang bobot komponen-komponen tersebut dalam sistem kendaraan. Menghitung kekuatan suatu rangka sepeda motor itu penting dilakukan untuk memastikan bahwa rangka dapat menahan beban. Perhitungan kekuatan rangka sepeda motor dapat dilakukan secara manual (tangan) atau menggunakan sistem analisis elemen hingga (FEA) [2].

Rangka e-SAF Enhanced Smart Architecture Frame (eSAF) menggunakan teknik manufaktur mutakhir seperti pengelasan press dan laser. Konstruksinya dirancang untuk meningkatkan stabilitas penanganan, membuat sepeda motor lebih mudah dikendarai, lebih ringan, dan lebih nyaman saat menikung. Kapasitas area bagasi 14 L dan kapasitas tangki bahan bakar 4,2 L menunjukkan bahwa rangka eSAF ini dapat memberikan penggunaan ruang yang lebih efektif. Struktur ini tidak lagi terbuat dari pipa, tetapi dari pelat baja yang ditebuk dan didorong sebelum dilas laser untuk membuat lipatan atau tulang di sisinya. Penggabungan ini menghapus 34 komponen, membuat rangka tampak lebih sederhana. Rangka hanya terdiri dari rangka utama (underbones) dan dua tulang tambahan di belakangnya (sub-kerangka). Hasilnya, rangkanya 8% lebih ringan dibandingkan rangka skutik Honda sebelumnya. “Rangka yang lebih ringan membuat

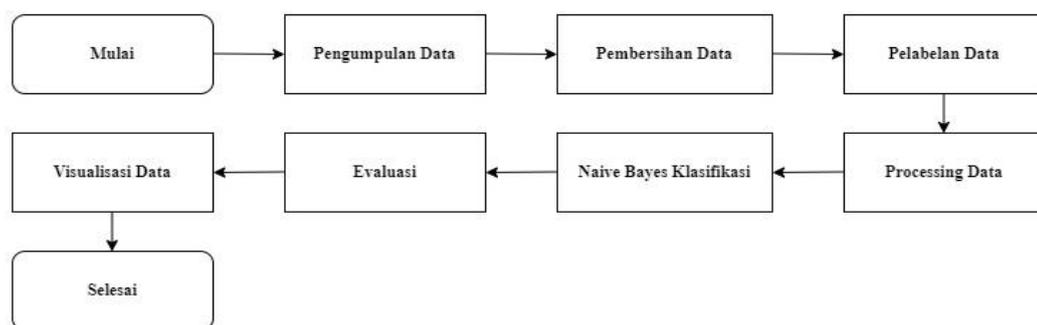
berkendara semakin nyaman dan mudah, terutama saat bermanuver.” “Bobot motor juga berkurang atau tidak seberat itu,” kata Endro Sutarno, Technical Service Department PT Astra Honda Motor[3]. Media sosial telah berkembang menjadi saluran vital bagi industri otomotif, terutama yang memproduksi sepeda motor, untuk berkomunikasi dengan pelanggan dan mempelajari bagaimana perasaan mereka tentang kendaraan yang dipakai.

Analisis sentimen adalah proses pengumpulan, pengolahan, dan penilaian pendapat publik tentang suatu hal atau topik tertentu[4]. Metode pemecahan Naive Bayes dipilih karena metode pembagian terstruktur ini memiliki formula yang sederhana dan mudah dipraktikkan dan juga cukup tepat dibandingkan dengan metode lainnya[5]. Di dalam hal ketepatan dan komputasi, Naive Bayes dianggap sebagai metode pemrosesan klasifikasi data tersusun yang mungkin lebih baik daripada metode klasifikasi tersusun lainnya[6]. Studi terdahulu yang dilakukan oleh Gusti Agung Indrawan, Ayu Indah Cahya Dewi, dan Ida Ayu Putu Ananda Wisdantini menemukan bahwa akurasi pengujian pemecahan metode Naive Bayes Classifier adalah 88,01 persen[7]. Dan juga Penelitian yang dilakukan oleh Dueti Putri dengan judul yaitu ”Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier” mendapatkan hasil akurasi sebesar 80%[8]. Selanjutnya Penelitian yang dilakukan oleh Irfan Ricky dengan judul yaitu ” Analisis Sentimen Opini Masyarakat Terkait Pelayanan Jasa Ekspedisi Anteraja Dengan Metode Naive Bayes” mendapatkan hasil akurasi sebesar 85,06%[9]. Hal ini membuktikan bahwa metode Naive bayes dapat memberikan akurasi tinggi dalam Evaluasi sentimen.

Analisis sentimen media sosial memberikan industri kemampuan untuk menilai kebahagiaan konsumen, memahami opini publik, dan dengan cepat dan efektif mengatasi masalah yang muncul[10]. Dengan cara yaitu mengidentifikasi pola sentimen yang dominan terkait dengan Rangka e-SAF Honda di Media Sosial X, termasuk aspek positif, dan negatif yang paling menonjol. Produsen sepeda motor terkemuka brand Honda menyadari nilai riset sentimen media sosial dalam melacak reputasi merek dan meningkatkan kepuasan konsumen. Dibandingkan dengan melakukan survei secara langsung, pengumpulan data melalui media sosial dinilai lebih efisien dalam semua aspek. Efisiensi ini mencakup mengurangi biaya yang diperlukan untuk mendapatkan data, memiliki kemampuan untuk mendapatkan data secara *real time*, dan menghasilkan data yang lebih rinci yang menggambarkan pendapat masyarakat[11]. Maka dari itu, analisis sentimen di rasa sangat efektif dalam kepuasan konsumen sepeda motor honda dikarenakan adanya isu terkait rangka yang digunakan dalam sepeda motor honda memiliki kualitas yang kurang baik meskipun pihak honda telah klarifikasi bahwa pihak Honda memberikan garansi selama 5 tahun tanpa batasan jarak kilometer motor Honda tersebut digunakan. Sebuah penelitian dilakukan tentang analisis sentimen berdasarkan latar belakang yang telah dijelaskan sebelumnya. Dengan menggunakan teknik Naive Bayes, peneliti akan melakukan penelitian tentang analisis sentimen rangka E-SAF Honda di media sosial X.

2. METODOLOGI PENELITIAN

Dalam penelitian ini. Metode yang digunakan adalah Naive Bayes untuk analisis setntimen ini. Berdasarkan penerapan teorema Bayes, atau aturan Bayes, dan dengan asumsi naif, atau independensi, yang kuat, naive Bayes adalah metode prediksi sederhana yang berbasis probabilitas.[12]. Berikut tahapan metode penelitian yang ditampilkan pada Gambar 1 berikut.



Gambar 1. Tahapan Metode Penelitian

Gambar 1 menunjukkan proses penelitian. Proses penelitian dimulai dengan pengumpulan data menggunakan website Google Colab yang terhubung ke media sosial X dengan Kata kunci "esaf" digunakan untuk mengumpulkan dataset. Tahapan berikutnya adalah Pembersihan Data, data yang telah dikumpulkan kemudian dibersihkan melalui fase pembersihan data yang terdiri dari beberapa tahapan yaitu menyambungkan operator replace seperti Replace URL link,, Replace @nama pengguna, Replace simbol dan terakhir adalah Trim (Pengurangan jarak berlebih antar teks). Selanjutnya, proses pelabelan manual pada data latih digunakan sebagai pembelajaran mesin untuk klasifikasi. Data Uji yang telah dibersihkan kemudian disiapkan melalui fase processing data yang terdiri dari beberapa tahap tambahan, seperti tokenize, transform cases, filter stopword dan filter token. Setelah dataset tersebut disiapkan, dataset tersebut kemudian dikelompokkan menjadi dua bagian yaitu data latih dan data uji, sementara pelabelan pada data uji tersebut dilakukan dengan mesin, yaitu dengan mengklasifikasikan data sentimen menjadi positif atau negatif dengan menggunakan metode Naive Bayes. Evaluasi dan visualisasi data adalah langkah akhir dari penelitian ini, nantinya hasilnya berupa accuracy,

precision, recall dan juga kata kata yang sering keluar dari dataset tersebut. Dan berikut penjelasan tahapan metodologi penelitian sebagai berikut.

2.1 Pengumpulan Data

Pada tahap ini, peneliti menggunakan metode crawling data pada platform media sosial X. Pengumpulan data adalah proses mengumpulkan atau mengambil data dari website atau platform media sosial melalui API (Application Program Interface) yang disediakan oleh platform tersebut[13]. langkah pertama adalah memperoleh kunci API web developer X dan memasukkannya ke dalam script file crawling yang berisi kode menggunakan bahasa pemrograman python dengan menggunakan kata kunci "e-saf". API X berfungsi untuk menghubungkan sistem X. Dan dataset tersebut di simpan ke dalam format file .csv ekstensi.

2.2 Pembersihan Data

Pembersihan data atau *cleansing* data merupakan langkah penting dalam memastikan kualitas data sebelum analisis lebih lanjut. RapidMiner mencakup sejumlah alat dan operator untuk membantu pengguna menemukan, memperbaiki, dan memvalidasi data dengan cepat[14]. Membersihkan data memungkinkan pengguna untuk meningkatkan kualitas dan keandalan hasil analitik mereka, membuat keputusan berdasarkan data lebih akurat dan mendasar.

2.3 Pelabelan Data

Pelabelan data dilakukan setelah melewati preprocessing data. Tujuan pelabelan dataset adalah untuk membuat representasi dari objek data yang dimiliki oleh ulasan dan membantu dalam pemahaman tentang cara mengidentifikasi sentimen data baik manual maupun otomatis dengan aplikasi rapidminer. Pelabelan data harus sesuai dengan rasio yang ditentukan dengan data latih dan data uji.

2.4 Processing Data

Processing data adalah langkah penting dalam siklus analisis data yang memastikan data mentah diubah menjadi format yang sesuai untuk analisis lebih lanjut[15]. RapidMiner menyediakan berbagai alat dan operator untuk mendukung setiap langkah dalam processing data, dari pembersihan dan transformasi data hingga penggabungan dan pengurangan dimensi. Pengguna dapat memastikan bahwa data yang dianalisis berkualitas tinggi dan siap memberikan wawasan yang akurat dan bermakna dengan melakukan proses pengolahan data yang efektif.

2.5 Naive Bayes Klasifikasi

Naive Bayes adalah algoritma klasifikasi berbasis probabilistik yang sangat efektif dan efisien, terutama untuk tugas-tugas dengan data berukuran besar[16]. Dalam aplikasi RapidMiner, proses membangun dan mengaplikasikan model Naive Bayes sangat mudah dan intuitif, berkat berbagai operator yang mendukung setiap tahap proses klasifikasi. Dengan menggunakan Naive Bayes di RapidMiner, pengguna dapat dengan cepat membangun model yang akurat dan melakukan prediksi pada data baru, membantu pengambilan keputusan berbasis data yang lebih baik.

2.6 Evaluasi

Evaluasi data adalah langkah krusial dalam siklus analisis data yang melibatkan penilaian kinerja model berdasarkan metrik tertentu[17]. Untuk memastikan bahwa model yang dibangun memiliki akurasi yang tinggi dan mampu membuat prediksi yang andal ketika diterapkan pada data baru, evaluasi data dilakukan. Evaluasi ini membantu dalam memahami kelebihan dan kekurangan model serta dalam membuat keputusan untuk perbaikan model lebih lanjut. Dalam penelitian ini, ada 3 rumus yang digunakan yaitu *Precision*, *Recall*, dan *Accuracy* berikut rumusnya

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

2.7 Visualisasi Data

Visualisasi data adalah tahap terakhir dari penelitian ini. Visualisasi data dimaksudkan untuk menganalisis sentimen kata-kata yang sering muncul dalam dataset[18]. Ini pasti akan menjadi lebih mudah untuk menganalisis pendapat pengguna sepeda motor Honda tanpa harus membaca semua ulasan satu-satu.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pada tahap ini, saya sebagai peneliti mengumpulkan dataset menggunakan aplikasi google colab dan kami menggunakan teknik crawling data. Melalui proses *crawling*, *Google Colab* membantu mendapatkan dataset menggunakan bahasa

Python. [19]. Data total yang dikumpulkan mencapai 756 dataset. Berikut gambar 2. yang menjelaskan alur dari pengumpulan data penelitian ini.



Gambar 2. Alur Pengumpulan Dataset

Pada gambar 2 tersebut, peneliti harus mengakses situs media sosial “X” untuk mendapatkan token dari akun tersebut, setelahnya token tersebut di masukan kedalam google collab supaya website google collab bisa menarik data tersebut menggunakan akun sosial media X. Dan tahapan selanjut nya, Kata kunci digunakan oleh peneliti yaitu ”esaf” untuk menyaring dataset yang diperoleh dari platform media sosial X. Riset ini menggunakan data yang dihimpun yang mencakup periode dari 1-7-2023 hingga 1-11-2023 dan dataset tersebut menggunakan bahasa Indonesia. sesuai yang terdapat pada gambar 3 berikut ini.

```

▶ data = 'data_esaf.csv'
  search_keyword = 'esaf lang:id until:2023-11-01 since:2023-07-01'
  limit = 800

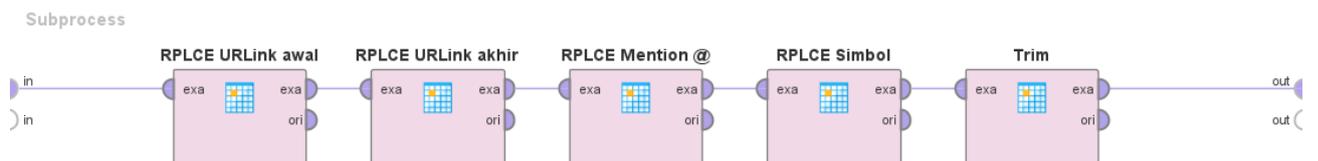
!npx --yes tweet-harvest@2.2.8 -o "{data}" -s "{search_keyword}" -l {limit} --token ""
    
```

Gambar 3. Pengambilan dataset menggunakan Google Colab

Gambar 3 tersebut dapat dijelaskan bahwa variabel kode “data” berguna untuk menentukan nama file untuk disimpan setelah hasil crawling. Variabel kode “search_keyword” berguna untuk menemukan data terkait rangka e-SAF sebagai kata kunci pencarian, keterangan “esaf lang:id” berfungsi menetapkan kata kunci pencarian sebagai esaf dengan dataset yang menggunakan bahasa Indonesia. Variabel “limit” berguna untuk mengontrol jumlah dataset yang akan dikumpulkan dari internet. Selain itu, baris terakhirnya berisi kode yang menjelaskan bahwa program yang digunakan adalah “tweet-harvest” dengan versi 2.2.8.

3.2 Pembersihan Data

Setelah tahap pengumpulan data. Peneliti melanjutkan proses selanjutnya yaitu Pembersihan Data yang dilakukan untuk mengolah data dengan beberapa tahapan menjadi data yang lebih siap digunakan untuk sesuai kebutuhan pada tahap lanjut. Pada tahapan ini, peneliti menggunakan aplikasi Rapidminer. Selain itu, beberapa operator digunakan ditunjukkan pada gambar 4 berikut. :



Gambar 4. Operator untuk Pembersihan data

Pada gambar 4 tersebut, tahap pembersihan data dilakukan dengan empat operator replace dan satu operator trim. Pertama, operator replace digunakan untuk menghapus URL yang ada di awal kalimat; kedua, operator replace 2 menghapus URL di akhir kalimat; ketiga, operator replace 3 menghapus semua keberadaan @nama pengguna dalam teks; dan keempat, Operator replace 4 sangat berguna dengan menghapus simbol atau karakter yang tidak relevan dari teks. Terakhir, Untuk menghilangkan jarak baris yang berlebih dalam teks, peneliti dapat menggunakan operator trim, teks dapat diubah sesuai kebutuhan. Setelah pembersihan data selesai, Table 1 menunjukkan proses pembersihan tweet yang dilakukan RapidMiner.

Tabel 1. Hasil Pembersihan Data

Sebelum	Sesudah
Halo @AstraHondaCare , rangka motor saya mendadak Krispy dan karatan. Apakah saya dapat ganti rangka gratis yang lebih bagus kualitasnya? Saya tidak mau diganti rangka Esaf Krispy.. @honda @HondaJP @KNKT_RI https://t.co/mztBFKDs1	Halo rangka motor saya mendadak Krispy dan karatan Apakah saya dapat ganti rangka gratis yang lebih bagus kualitasnya? Saya tidak mau diganti rangka Esaf Krispy.
@afiiitz10 @welovehonda yaelahhh, update bangg, jangan cuma nongki nungguin honda posting trus hujat, KNKT dah jelasin klo esaf dan rangkanya dah aman	yaelahhh update bangg jangan cuma nongki nungguin honda posting trus hujat KNKT dah jelasin klo esaf dan rangkanya dah aman,

@dimzz78 @AndyHuskyyy Rangka esaf ku kupakai buat kesana kemari. Kadang sama suamiku dipakai buat bawa peralatan kerja. Dan hingga sekarang aman2 aja tuh.	Rangka esaf ku kupakai buat kesana kemari Kadang sama suamiku dipakai buat bawa peralatan kerja Dan hingga sekarang aman2 aja tuh
--	---

Tabel 1 kolom bagian sebelah kiri merupakan dataset yang belum melewati proses pembersihan data, terlihat masih adanya @nama pengguna dan juga URL. Sedangkan pada Tabel 1 di kolom bagian sebelah kanan merupakan dataset yang telah melewati proses pembersihan data, terlihat bahwa hilangnya @nama pengguna dan link URL di komentar dataset tersebut.

3.3 Pelabelan Data

Tahap pembersihan data sudah dilakukan. Selanjutnya masuk ke tahap pelabelan data yaitu sentiment untuk rangka "e-saf" Honda. Data uji dan data latih adalah dua komponen data ini. Data latih dilabelkan sebagai ilustrasi untuk pengolahan data uji berikutnya[20]. Dari 756 data yang telah dihimpun yang bersumber dari dunia maya, hanya sebanyak 696 yang akan digunakan untuk tujuan penelitian ini. Data yang tidak sesuai dengan konteks tidak akan digunakan untuk meningkatkan akurasi penelitian. Dari 696 data, peneliti membagi 417 sebagai data latih dan 279 sebagai data uji. Jika data latih mengandung pendapat positif, mereka diberi label positif, dan jika pendapat negatif, mereka diberi label negatif, seperti yang ditunjukkan dalam tabel 2 berikut:

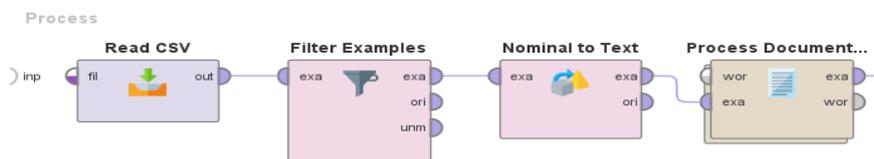
Tabel 2. Pelabelan Data

Text	Sentimen
Gak perlu takut kali sama rangka esaf tidak seperti yang diberitakan Buktinya motor aku baik2 aja tuh	Positif
Kalau merasa punya malu ya recall dulu rangka esaf kemarin jangan merasa market leader semua gue	Negatif
kata siapa rangka esaf kopong dan rapuh bg coba lu googling bang terkait rangka esaf KNKT udh bilang aman kali dan udah sesuai standar global	Positif

Pada tabel 2 diatas menunjukkan bahwa dataset komentar yang bermuatan positif diberi sentimen positif, sedangkan komentar yang bermuatan negatif diberi sentimen negatif. Setelah dilakukan pelabelan data latih secara manual didapatkan dari 417 dataset, dataset yang bermuatan sentimen positif berjumlah 193 data sedangkan dataset yang bermuatan negatif berjumlah 224 data.

3.4 Processing Data

Selanjutnya, sisa datanya sebanyak 279 diubah menjadi data uji untuk diproses oleh RapidMiner, seperti yang ditunjukkan pada gambar 5:

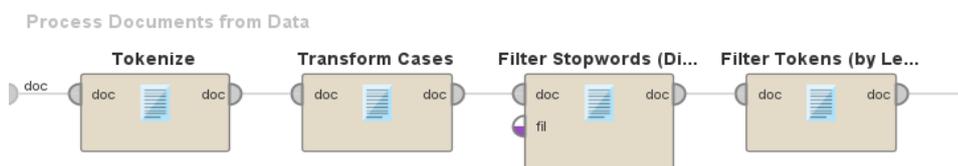


Gambar 5. Operator Processing Data

Pada gambar 5 terdapat beberapa operator yang digunakan. Pertama, untuk membuka dan memindai file CSV yang telah disiapkan, gunakan operator "Read CSV". Data latih disimpan dalam file CSV yang digunakan. Selanjutnya dihubungkan dengan operator "Filter Example". Operator ini bertujuan untuk memilah data sesuai kebutuhan. Untuk mencapai hal ini, kondisi digunakan memilah sentimen yang tidak hilang dengan menggunakan opsi "is not missing". Oleh operator "Filter Examples" selanjutnya akan dikaitkan oleh operator "Process Document from Data", tetapi diperlukan operator "Nominal to Text" Sebelum mengubah nilai nominal menjadi teks, untuk operator "Process Document from Data" mampu membaca data dan beroperasi. Di dalam operator "Process Document from Data" ada beberapa operator yang digunakan, misalnya dapat dilihat pada gambar 6 berikut. Penjelasan gambar 6 dalam poin – poin berikut:

3.4.1 Tokenize

Tokenize adalah suatu operator yang berfungsi untuk memilah kalimat menjadi kata - kata yang bersifat individual yang berfungsi untuk memudahkan dalam proses *Filter Stopwords*[21]. Tabel 3 berikut menampilkan sebelum proses tokenize dan sesudah proses tokenize



Gambar 6. Operator di dalam Process Document from Data

Tabel 3. Hasil Dari Tokenize

Sebelum Proses Tokenization	Sesudah Proses Tokenization
Soal rangka eSAF Honda yang viral tidak perlu khawatir Hasil penelitian Ditjen Hubdat dan KNKT ketebalan teknik coating secara dipping sudah sesuai standar manufaktur global	'Soal', 'rangka', 'eSAF', 'Honda', 'yang', 'viral', 'tidak', 'perlu', 'khawatir', 'Hasil', 'penelitian', 'Ditjen', 'Hubdat', 'dan', 'KNKT', 'ketebalan', 'teknik', 'coating', 'secara', 'dipping', 'sudah', 'sesuai', 'standar', 'manufaktur', 'global'

Pada tabel 3 diatas proses tokenize dilakukan dengan cara yaitu memilahkan bagian per bagian kata dalam satu kalimat. Setuap dataset diproses dengan cara memilah beberapa potongan kata dengan operator tokenize di aplikasi Rapidminer.

3.4.2 Transform Cases

Tahapan selanjutnya adalah Transform Cases. Dengan menggunakan Transform Cases, Teks mengubah semua huruf kapital menjadi huruf kecil[22]. Operator ini diatur kedalam “lower case” supaya berjalan dengan baik. Tabel 4 berikut menampilkan sebelum proses Transform Cases dan sesudah proses Transform Cases.

Tabel 4. Hasil dari Transform Cases

Sebelum Proses Transform Cases	Sesudah Proses Transform Cases
Yang salah rangka esaf coba ngelipet duluan pas masih di rumah pasti nggak jadi insiden Emangnya kenapa kalo rangka eSAF motor yang pake rangka eSAF kuat kuat kok	yang salah rangka esaf coba ngelipet duluan pas masih di rumah pasti nggak jadi insiden emangnya kenapa kalo rangka esaf motor yang pake rangka esaf kuat kuat kok

Dari Tabel 4 tersebut menunjukkan bahwa kolom bagian sebelah kiri merupakan dataset yang belum melewati proses Transform Cases sehingga masih ada huruf kapital di awal kalimat. Sedangkan kolom bagian sebelah kanan merupakan dataset yang sudah melewati proses Transform Cases sehingga huruf di kolom tersebut merupakan huruf kecil semua.

3.4.3 Filter Stopwords

Tahapan berikutnya adalah Filter Stopword. Filter Stopword adalah merupakan tahap dimana menghimpun semua kata sambung dihapus dari dataset. Peneliti berpedoman pada kamus stopwords yang tersedia di situs web kaggle. Tabel 5 berikut menampilkan hasil dari filter stopwords.

Tabel 5. Hasil dari Filter Stopword

Sebelum Proses Filter Stopwords	Sesudah Proses Filter Stopwords
Emang jadi faktor penting juga ya soal perawatan rangka eSAF Karena apapun mereknya kalo gagal merawat bisa dipastikan akan patah juga	Emang jadi faktor penting juga ya soal perawatan rangka eSAF Karena apapun mereknya kalo gagal merawat dipastikan akan patah

Pada tabel 5 menunjukkan bahwa ada perbedaan antara kolom bagian kiri dengan kolom bagian kanan, perbedaan yang paling terlihat adalah hilangnya kata “juga” di akhir kalimat, karena kata “juga” merupakan kata sambung bagian dari kamus filter stopwords.

3.4.4 Filter Tokens

Berikutnya adalah tahapan dengan operator Filter tokens. Menghapus atau menghilangkan sejumlah kata tertentu adalah tugas dari operator Filter Tokens ini[23]. Dalam penelitian ini, peneliti menggunakan maksimal 25 karakter, dengan minimum 4 karakter. Tabel 6 merupakan hasil dari Filter Tokens.

Tabel 6. Hasil Dari Filter Tokens

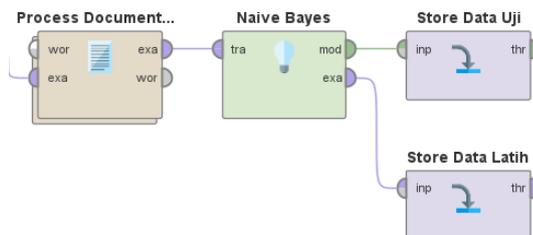
Sebelum Proses Filter Tokens	Sesudah Proses Filter Tokens
kata siapa rangka esaf kopong dan rapuh bg coba lu googling bang terkait rangka esaf KNKT udh bilang aman kali dan udah sesuai standar global	kata siapa rangka esaf kopong dan rapuh coba googling bang terkait rangka esaf KNKT bilang aman kali udah sesuai standar global

Pada tabel 6 di atas menunjukkan ada perbedaan antara kolom bagian kiri dengan kolom bagian kanan, yaitu hilangnya kata tertentu yang jumlahnya kurang dari 4 karakter.

3.5 Naive Bayes Klasifikasi

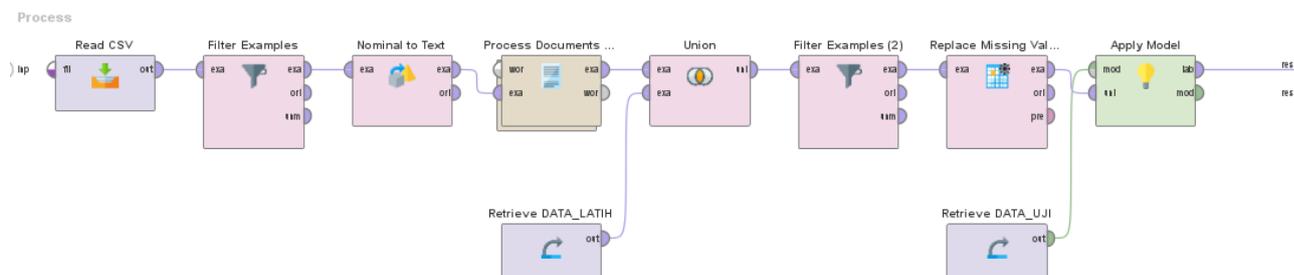
Setelah melakukan tahap processing data. Selanjutnya adalah tahap *Naive Bayes* klasifikasi. Untuk kategorisasi, metode Naive Bayes Classifier digunakan, dimaksudkan untuk menentukan apakah tanggapan dari seseorang tentang sebuah objek atau masalah yang cenderung berkategori negatif atau positif.[24]. Operator Naive Bayes kemudian berhubungan dengan dua operator "Store". Operator "Store" pertama mampu menyimpan hasil proses output operator "Naive Bayes"

yang sebelumnya disebut "Store Data Uji". Operator "Store" kedua disebut "Store Data Latih" untuk menyimpan hasil dari data latih, seperti yang ditunjukkan pada gambar 7 di bawah ini.



Gambar 7. Operator simpan Data Uji & Data Latih

Setelah menyimpan data uji dan data latih seperti di gambar 7 diatas lalu mempersiapkan model Naive Bayes, data uji, dan data latih. Langkah berikutnya adalah menerapkan algoritma Naive Bayes untuk melakukan klasifikasi. Pertama, data uji yang memuat sentimen dari aplikasi X disaring untuk memilih data yang tidak memiliki label sentimen. Setelah itu, proses preprocessing dilakukan dengan memanfaatkan operator "Process Document". Hasil dari proses ini digabungkan dengan data latih dengan operator union. Mengingat bahwa setiap kata dapat berbeda di antara kedua dataset, penggabungan ini dilakukan dengan tujuan untuk menyelaraskan atribut yang ada di antara keduanya. Setelah digabungkan, data disaring lagi untuk memastikan bahwa sentimen tidak ada padanya. Setelah data uji selesai, langkah berikutnya adalah menerapkan prosedur klasifikasi. Dengan memanfaatkan operator apply pada data uji, model Naive Bayes memungkinkan klasifikasi sentimen dilakukan secara otomatis. Algoritma Naive Bayes menghasilkan data yang dilabeli secara otomatis setelah langkah-langkah klasifikasi selesai. Proses pemberian label secara otomatis ditunjukkan pada gambar 8.



Gambar 8. Operator untuk Pelabelan Otomatis (Dengan Algoritma Naive Bayes)

Hasil dari algoritma Naive Bayes untuk otomatisasi label sentimen adalah prediksi sentimen menggunakan algoritma. Data latih yang telah disiapkan sebelumnya oleh peneliti digunakan untuk melatih algoritma Naive Bayes pada tahap pelabelan otomatis. Data latihan ini pada awalnya telah diberi label secara manual. Algoritma Naive Bayes memiliki kemampuan untuk memahami pola data latih setelah menyelesaikan tahap pelatihan. Dengan mengetahui hal ini, algoritma Naive Bayes dapat memperkirakan bahwa sentimen data baru dari dua kelas akan bersifat positif atau negatif. Gambar 9 di bawah ini menunjukkan hasil pemberian label otomatis yang dihasilkan oleh metode algoritma Naive Bayes.

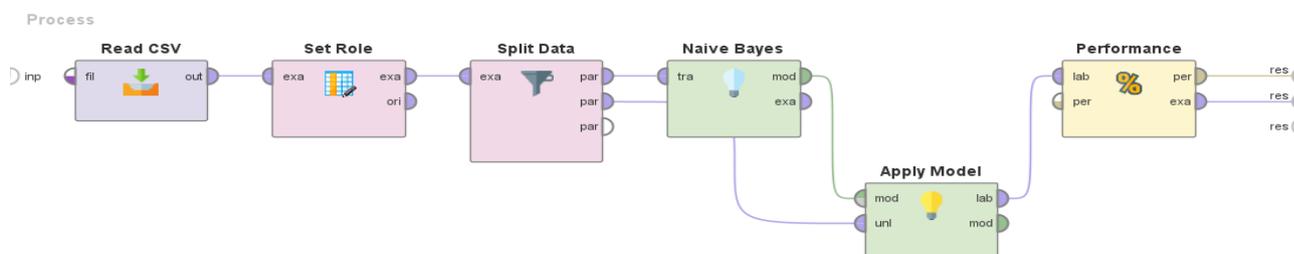
Sentimen	prediction(Sentimen)	text
?	Negatif	garagara esaf kali marc
?	Positif	elegan honda beat insyaf esaf
?	Negatif	twit astra honda care tragedi rangka esaf semoga admin twitternya hidup
?	Positif	rumornya berpisah motor pake rangka esaf
?	Positif	rangka esaf dipaksa balapan crash mulu
?	Negatif	hahaha modaro produksi turun drastis pangan esaf
?	Negatif	marquez pindah rumornya musim pakai rangka esaf emang bener pindah ducati
?	Negatif	korban esaf patah
?	Negatif	kasian honda wkwkw gara esafwkwkw

Gambar 9. Hasil Pelabelan Otomatis Data Uji

Gambar 9 di atas menunjukkan bahwa algoritma naive bayes sangat efektif dalam melakukan pelabelan otomatis seperti dataset pertama merupakan dataset bermuatan sentiment negatif dan juga prediksi sentimen dari naive bayes pun juga memprediksi bahwa sentiment itu bermuatan sentiment negatif.

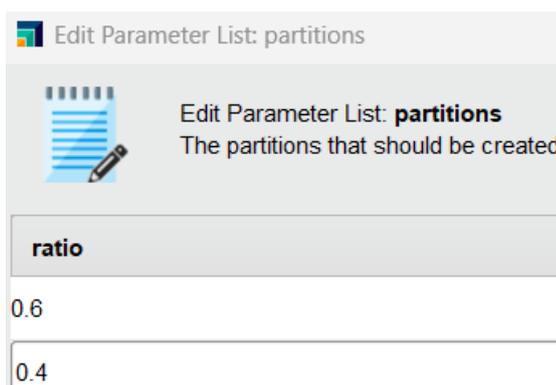
3.5 Evaluasi

Setelah dataset penelitian dievaluasi dengan metode Naive Bayes, pengujian Confusion Matrix dilakukan. Tujuan dari pengujian ini adalah untuk menghitung tingkat akurasi dengan memanfaatkan evaluasi dan validasi. Confusion Matrix dilakukan. Tujuan dari uji ini adalah untuk menghitung tingkat akurasi yang dihasilkan dari evaluasi dan validasi[25]. Confusion Matrix menghitung tingkat akurasi berdasarkan tiga parameter yaitu accuracy, precision, dan recall. Aplikasi RapidMiner lainnya juga digunakan untuk melakukan perhitungan ini. Gambar 10 menunjukkan operator-operator yang digunakan dalam proses evaluasi dan validasi dibawah ini.



Gambar 10. Operator Untuk Confusion Matrix

Dari gambar 10 yang digunakan dalam proses evaluasi diatas. Perhitungan Confusion Matrix dimulai dengan operator "Read CSV", yang memuat data latihan. Kemudian, operator "Set Role" mengubah kolom yang mengubah tipe data hasil sentimen dengan atribut label. Selanjutnya, operator "Split Data" memilah data dengan klasifikasi pembagian 60:40. Pemilahan data Confusion Matrix ditunjukkan pada gambar 11 berikut.



Gambar 11. Memilah Data di Operator Split Data

Setelah memilah data di operator split data seperti pada gambar 11 diatas. Selanjutnya memilih opsi "automatic" untuk parameter "Sampling type", yang menunjukkan bahwa data digunakan secara acak. Setelah itu, disambungkan kembali ke operator "Naive Bayes" dan "Apply Model", dan proses diakhiri dengan menyambungkan operator "Performance" untuk menampilkan hasil pengukuran akurasi. Hasil confusion matrix dengan nilai akurasi sebesar 70.27%, dengan 57 nilai positif asli dan 21 nilai negatif asli, seperti yang ditunjukkan pada gambar 12 di bawah ini.

accuracy: 70.27%

	true Negatif	true Positif	class precision
pred. Negatif	21	15	58.33%
pred. Positif	18	57	76.00%
class recall	53.85%	79.17%	

Gambar 12. Hasil dari Confusion Matrix

Penjabaran penghitungan confusion matrix pada gambar 12 dapat dilihat dibawah ini sebagai berikut:

$$Precision = \frac{TP}{TP+FP} = \frac{57}{57+18} = \frac{57}{75} = 0,76 \text{ atau } 76\%$$

$$Recall = \frac{TP}{TP+FN} = \frac{57}{57+15} = \frac{57}{72} = 0.7916666666666667 \text{ atau } 79,17\%$$

$$Acurracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{57+21}{57+21+18+15} = \frac{78}{111} = 0,7027 \text{ atau } 70,27\%$$

3.6 Visualisasi Data

Visualisasi data disandingkan untuk mendapatkan opini apa saja yang sering diutarakan oleh konsumen dalam menggunakan sepeda motor dengan rangka e-SAF. Hanya 21 kata yang paling sering digunakan dalam penelitian dipilih oleh peneliti. Semakin sering kata tersebut muncul, semakin jelas visualisasinya. Gambar 13 di bawah ini menunjukkan hasil visualisasi data (wordcloud).



Gambar 13. Hasil dari Visualisasi Data Wordcloud

Gambar 13 diatas menyajikan visualisasi data (wordcloud) yang menunjukkan kata dengan runtut dari 1 sampai 21 setelah data dijabarkan secara descending (dari besar ke kecil). Pada wordcloud terlihat kata ‘motor’ adalah kata yang paling banyak digunakan dengan jumlah kemunculan sebanyak 99 kali. Kemudian di posisi kedua yang memiliki jumlah kemunculan 71 kali adalah kata ‘pake’. Dan di posisi ketiga yang muncul sebanyak 51 kali adalah kata ‘aman’. Posisi keempat kata ini muncul sebanyak 43 kali yaitu ‘patah’.

4. KESIMPULAN

Data yang diolah untuk penelitian analisis sentimen ini berasal dari komentar pengguna media sosial “X” dengan menghimpun dataset tersebut melalui teknik crawling data. Hasil crawling data media sosial “X” terkait rangka E-SAF ini sebanyak 756 dataset dalam rentang waktu 1-7-2023 hingga 1-11-2023, setelah melalui tahap preprocessing yang terdiri dari cleansing, tokenizing dan filtering stopwords, data yang siap diteliti untuk penelitian selanjutnya sebanyak 696 dataset. Dari dataset tersebut, peneliti melakukan pelabelan dataset secara manual sebanyak 417 dataset atau 60% dari data keseluruhan. Didapati sentimen bermuatan positif dengan jumlah 193 dan sentimen negatif 224. Tahapan pelabelan dataset ini dilakukan agar Naive Bayes dapat melabeli dataset secara otomatis dengan mempelajari pola dataset yang telah dilabeli secara manual. Hasilnya adalah algoritma Naive bayes ini memberikan hasil akurasi sebesar 70,27% dengan precision sebesar 76% dan juga recall sebesar 79,17%. Visualisasi data tersebut menunjukkan bahwa peringkat kata yang sering muncul di dataset ini adalah “motor” sebanyak 99 kali selanjutnya adalah kata “pake” dengan kemunculan kata sebanyak 71 kali, berikutnya adalah kata “aman” dengan kemunculan data sebanyak 51 kali, dan terakhir adalah kata “patah” dengan jumlah kemunculan 43 kali. Meskipun rangka E-SAF di-isukan rangka sepeda motor dengan kualitas yang kurang begitu baik yang membuat sentimen negatif bermunculan di sosial media, tetapi dengan diberikannya garansi rangka sepeda motor tersebut selama 5 tahun dan tanpa batasan jarak kilometer membuat netizen sosial media meyakini bahwa sepeda motor dengan menggunakan rangka E-SAF aman untuk berkendara yang membuat sentimen positif bermunculan di sosial media.

REFERENCES

- [1] S. Anwar and Mujito, “Faktor-Faktor Yang Mempengaruhi Keputusan Pembelian Motor Merek Yamaha Di Kota Bogor,” JIMKES Jurnal Ilmiah Manajemen Kesatuan, vol. 9, no. 1, pp. 189–202, 2021.
- [2] R. Setiawan, D. Sugiyanto, and ari Daryus, “ANALISIS SIMULASI KEKUATAN DAN PEMBUATAN RANGKA KENDARAAN SEPEDA MOTOR LISTRIK Analysis of Strength Simulation and Frame Fabrication of Electric Motorcycle Vehicle,” Jurnal Konversi Energi dan Manufaktur, vol. 8, no. 1, pp. 58–66, 2023.
- [3] Endro Sutarno, “Mengenal Rangka eSAF Pada Sepeda Motor Honda,” HERONUSA HONDA. Accessed: Nov. 01, 2023. [Online]. Available: <https://heronusahonda.com/?p=1467>

- [4] Y. Akbar and T. Sugiharto, "Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan Naïve Bayes (Yuma Akbar 1*, Tri Sugiharto 2) Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan Naïve Bayes," *Jurnal Sains dan Teknologi*, vol. 5, no. 1, pp. 115–122, 2023, doi: 10.55338/saintek.v4i3.1368.
- [5] A. Safira, A. S. Masyarakat..., and F. N. Hasan, "ANALISIS SENTIMEN MASYARAKAT TERHADAP PAYLATER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER," *Jurnal Sistem Informasi*, vol. 5, no. 1, 2023.
- [6] D. P. Ray, F. N. Hasan, and A. R. Dzikrillah, "KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Sentimen Terhadap KPU 2024 Berdasarkan Tweet Media Sosial Twitter Menggunakan Algoritma Naïve Bayes," *Media Online*, vol. 4, no. 4, pp. 2235–2243, 2024, doi: 10.30865/klik.v4i4.1587.
- [7] I. Gusti, A. Indrawan, D. Ayu, I. Cahya Dewi, I. A. Putu, and A. Wisdantini, "Analisis Sentimen Terhadap Presidensi G20 2022 pada Media Sosial Twitter Menggunakan Metode Naïve Bayes," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 1, pp. 553–561, 2023, doi: 10.30865/klik.v4i1.1104.
- [8] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.
- [9] I. R. Afandi, F. Noor, H. #2, A. A. Rizki, N. Pratiwi, and Z. Halim, "Analisis Sentimen Opini Masyarakat Terkait Pelayanan Jasa Ekspedisi Anteraja Dengan Metode Naive Bayes," *Jurnal Linguistik Komputasional (JLK)*, vol. 5, no. 2, pp. 63–70, 2022, [Online]. Available: <https://t.co/2HADwg1drL>
- [10] R. Hidayat, A. B. Hakim, and R. Nugraha, "Perbandingan Metode Naïve Bayes Dan Decision Tree C4.5 untuk Analisis Sentimen Produk Es Teh Indonesia di Media Sosial Twitter," *Jurnal Sistem Komputer dan Kecerdasan Buatan*, vol. 7, no. 2, pp. 88–98, 2024.
- [11] F. F. Rachman and S. Pramana, "Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter," 2020.
- [12] H. Derajad Wijaya and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," *JURNAL INFORMATIKA*, vol. 7, no. 1, 2020, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [13] Y. Severianus and S. Kolo, "ANALISIS SENTIMEN TERHADAP OPINI MASYARAKAT TERKAIT PERUBAHAN CUACA DI INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE," 2024.
- [14] D. Ismiyana Putri and M. Yudhi Putra, "KOMPARASI ALGORITMA DALAM MEMREDIKSI PERUBAHAN HARGA SAHAM GOTO MENGGUNAKAN RAPIDMINER," *Jurnal Khatulistiwa Informatika*, vol. 11, no. 1, 2023, [Online]. Available: <https://www.kaggle.com/>,
- [15] Z. Setiawan et al., *BUKU AJAR DATA MINING*. Jambi: PT. Sonpedia Publishing Indonesia., 2023.
- [16] R. Ardiansyah Yudhanegara, N. Aliya Hana, S. Yonanda Mahfiridho, and A. Rosadi Kardian, "Perbandingan Resident Set Size dan Virtual Memory Size Algoritma Machine Learning dalam Analisis Sentimen," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 2024, doi: 10.30865/mib.v8i1.7201.
- [17] P. W. Rahayu et al., *Buku Ajar Data Mining, Cetakan Pertama*. Jambi: PT. Sonpedia Publishing Indonesia., 2024.
- [18] S. Syafrizal, M. Afdal, and R. Novita, "Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 10–19, Dec. 2023, doi: 10.57152/malcom.v4i1.983.
- [19] J. S. Gea and H. Budiati, "Analisis Sentimen Masyarakat Terhadap Direktorat Jenderal Pajak," *JURNAL SAINS DAN KOMPUTER*, vol. 8, no. 01, pp. 30–36, Jan. 2024, doi: 10.61179/jurnalinfact.v8i01.466.
- [20] A. T. Mukti and F. N. Hasan, "Analisis Sentimen Warganet Terhadap Keberadaan Juru Parkir Liar Menggunakan Metode Naive Bayes Classifier," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, pp. 644–653, 2024, doi: 10.30865/mib.v8i1.6982.
- [21] E. Dwi Nugraha and G. Gata, "Penerapan Algoritma KNN Pada Twitter Untuk Analisis Sentimen Masyarakat Terhadap Event MotoGP Di Sirkuit Mandalika," *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) Jakarta-Indonesia*, 2022.
- [22] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiasari, "Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE," *AITI: Jurnal Teknologi Informasi*, vol. 18, no. Agustus, pp. 173–184, 2021.
- [23] N. Luh, P. P. Dewi, I. Nyoman Purnama, and N. W. Utami, "Penerapan Data Mining Untuk Clustering Penilaian Kinerja Dosen Menggunakan Algoritma K-Means (Studi Kasus: STMIK Primakara)," *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 16, no. 2, 2022.
- [24] Hartati, H. Deni, M. Akhsanal, Z. Wahyudi, A. Ariyanto, and D. D. Saputra, "Optimasi Analisis Sentimen Pada Twitter Olshop Tokopedia Menggunakan Textmining Dengan Algoritma Naïve Bayes & Adaboost," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 6, no. 2, pp. 821–828, 2022.
- [25] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *Jurnal KomtekInfo*, vol. 10, no. 1, pp. 1–7, Jan. 2023, doi: 10.35134/komtekinfo.v10i1.330.