ANALISIS SENTIMEN KOMENTAR VIDEO MOBIL LISTRIK DI *PLATFORM YOUTUBE* DENGAN METODE *NAIVE BAYES*

Ayu Karimah[1], Gifthera Dwilestari[2], Mulyawan[3]
[1]Teknik Informatika, STMIK IKMI Cirebon
[2,3]Sistem Informasi, STMIK IKMI Cirebon
Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135
ayukarimah669@gmail.com

ABSTRAK

Dalam era perkembangan teknologi yang pesat, mobil listrik menjadi inovasi utama dalam industri otomotif. Namun, analisis sentimen menunjukkan adanya tantangan, terutama kurangnya pemahaman masyarakat tentang mobil listrik yang mempengaruhi persepsi negatif. Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap mobil listrik di Indonesia melalui konten *YouTube*, menggunakan algoritma *Naive Bayes*. Metode *Knowledge Discovery in Databases* (KDD) digunakan untuk mengumpulkan dan menganalisis opini masyarakat. Evaluasi model menggunakan *RapidMiner* menunjukkan peningkatan signifikan dengan penerapan teknik *SMOTE Upsampling*, meningkatkan akurasi dari 50,70% menjadi 70,69%. Analisis kelas menunjukkan peningkatan *true positive*, *true negative*, dan *true neutral*. Meskipun algoritma *Naive Bayes* memberikan akurasi 70,69%, *presisi* 43.64%, dan *recall* 39.48%, penelitian ini memiliki kekurangan dan disarankan untuk membandingkan algoritma klasifikasi lain, memperluas dataset, dan menguji kombinasi algoritma yang lebih luas. Rekomendasi juga mencakup penerapan hasil penelitian ke mesin klasifikasi dengan pengujian lebih lanjut. Studi ini memberikan wawasan untuk meningkatkan pemahaman dan penerimaan mobil listrik di masyarakat Indonesia.

Kata kunci : *Analisis Sentimen, Naive Baye, Mobil Listrik, Youtube.*

1. PENDAHULUAN

Dalam era modern yang dipenuhi dengan kemajuan teknologi, dampaknya terasa signifikan di berbagai aspek kehidupan, termasuk dalam industri otomotif. Salah satu perkembangan terkini yang menonjol adalah mobil listrik, sebuah kendaraan yang menggunakan motor listrik sebagai penggerak utama dan baterai sebagai sumber energi. Keunggulan mobil listrik meliputi efisiensi energi dan dampak lingkungan yang positif. Meski begitu, di Indonesia, adopsi mobil listrik masih terbatas. Oleh karena itu, diperlukan analisis sentimen untuk memahami pandangan masyarakat terhadap mobil listrik, guna mendukung pengembangan dan peningkatan penggunaannya di Indonesia.

Permasalahan yang muncul terkait analisis sentimen terhadap mobil listrik mencakup kurangnya informasi dan pemahaman masyarakat, yang mengakibatkan pandangan kurang positif terhadap teknologi ini. Selain itu, kendala infrastruktur dan harga yang tinggi turut menjadi hambatan. Analisis sentimen di media sosial, khususnya *YouTube*, dianggap sebagai pendekatan yang relevan untuk menilai pandangan masyarakat terhadap mobil listrik.

Sejumlah penelitian sebelumnya telah menjalankan analisis sentimen dengan menggunakan Hasil penelitian ini diharapkan dapat memberikan kontribusi pada pemahaman pandangan masyarakat terhadap mobil listrik di *YouTube* dan mendukung keputusan terkait pengembangan teknologi ini di masa depan. Implikasi praktisnya mencakup manfaat bagi praktisi dan peneliti di bidang otomotif, serta berbagai metode. Studi oleh [1] menggunakan metode *Support Vector Machine* dan *Feature Selection Particle Swarm Optimization*, yang berhasil meningkatkan akurasi hasil analisis sentimen. Studi lain oleh [2]) menggunakan algoritma *VADER* dan menunjukkan bahwa sebagian besar sentimen publik terhadap mobil listrik adalah positif. Sebaliknya, studi oleh [3] menggunakan algoritma *Logistic Regression* dan *Principal Component Analysis* dengan akurasi yang mencapai 90%.

Tujuan utama penelitian ini adalah menganalisis sentimen masyarakat di *YouTube* terhadap mobil listrik dengan menggunakan algoritma *Naive Bayes*. Data sentimen berasal dari media sosial *YouTube*, dan penelitian ini diharapkan dapat memberikan gambaran tentang persepsi masyarakat terhadap mobil listrik serta membantu dalam pengambilan keputusan terkait pengembangan mobil listrik di masa depan.

Penelitian ini memilih metode analisis sentimen dengan algoritma *Naive Bayes* karena sederhana dan efektif dalam mengklasifikasikan data teks. Algoritma ini memperhitungkan probabilitas kemunculan kata-kata dalam kelas sentimen tertentu, memberikan kontribusi penting dalam memahami pandangan masyarakat secara menyeluruh.
potensi dampak positif pada perkembangan teknologi ramah lingkungan di sektor otomotif. Studi ini juga dapat dijadikan acuan untuk penelitian lebih lanjut dalam analisis sentimen pada *YouTube* terhadap topik tertentu, dengan algoritma *Naive Bayes* sebagai metode yang dapat diandalkan.

2. TINJAUAN PUSTAKA

2.1. *Text mining*

Menurut Charles Joergensen E Munthe *Text Mining* adalah suatu teknik yang dapat digunakan untuk klasifikasi, di mana text mining merupakan variasi dari *data mining* yang berusaha menemukan pola menarik dari sekumpulan data teks yang besar. Selain itu, *text mining* juga dapat diartikan sebagai penerapan konsep dari teknik *data mining* untuk mencari pola dalam teks, dengan tujuan mencari informasi yang bermanfaat untuk tujuan tertentu [4]. *Text mining* merujuk pada proses penambangan data dalam bentuk teks, khususnya dokumen, dengan tujuan mengidentifikasi kata-kata yang terdapat dalam dokumen tersebut. Analisis keterkaitan antara kata-kata ini kemudian dapat dilakukan untuk mendapatkan pemahaman lebih lanjut terhadap isi dokumen tersebut [5].

2.2. *Naive Bayes algorithm*

Naive Bayes, yang diperkenalkan oleh ilmuwan Inggris Thomas Bayes, melakukan prediksi peluang di masa depan berdasarkan pengalaman masa lalu. Metode *Naive Bayes* juga dianggap sangat baik dalam mengklasifikasikan dokumen jika dibandingkan dengan metode klasifikasi lainnya, terutama dalam hal akurasi [6]. Metode Naive Bayes dikenal sebagai teknik klasifikasi yang menggunakan teori probabilitas untuk menentukan probabilitas suatu klasifikasi tertentu berdasarkan frekuensi setiap klasifikasi pada data pelatihan [7]

2.3. *Sentiment analysis*

Analisis sentimen, atau yang juga dikenal sebagai penambangan opini, merupakan bidang studi yang menyelidiki pandangan, sentimen, evaluasi, penilaian, sikap, dan emosi terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atribut yang terkait. Bidang ini melibatkan berbagai aspek dan memiliki berbagai istilah serta fungsi yang sedikit berbeda, termasuk analisis sentimen, penambangan opini, ekstraksi pendapat, sentimen penambangan, analisis subjektivitas, analisis pengaruh, analisis emosi, dan peninjauan penambangan. Meskipun istilah-istilah ini memiliki perbedaan, semuanya dapat dimasukkan ke dalam kategori analisis sentimen atau penambangan opini. Meskipun dalam industri, istilah analisis sentimen lebih umum digunakan, baik di industri maupun di dunia akademis, sering kali istilah analisis sentimen dan penambangan opini digunakan secara bersamaan [8].

2.4. Mobil Listrik

konsep inovasi pertama menggunakan mobil listrik diperkenalkan pada tahun 1828 dan diproduksi untuk pertama kalinya pada tahun 1884. Antara tahun 1897 dan 1900, sekitar 28% dari seluruh kendaraan yang beredar di pasaran merupakan kendaraan listrik. Meskipun mobil listrik lebih populer pada periode tersebut, daya saingnya tergerus oleh kendaraan berbahan bakar minyak karena harga minyak dunia yang masih rendah. Seiring waktu, minat terhadap mobil listrik menurun, dan konsep ini hampir terlupakan hingga munculnya *EV1 General Motors* pada tahun 1996. Meskipun konsep ini awalnya kurang dikenal, namun berhasil menjadi tren yang populer, menjadi salah satu momen penting dalam kebangkitan mobil listrik. Kesuksesan *General Motors* mendorong produsen mobil besar lainnya, seperti Ford, Toyota, dan Honda, untuk menggali potensi mereka dan ikut serta dalam memperkenalkan produk kendaraan listrik masing-masing [9].

2.5. *YouTube*

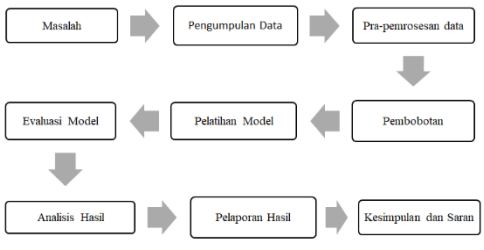
YouTube digunakan untuk berbagai informasi, pendapat, dan perasaan tentang berbagai topik. *YouTube* juga merupakan situs berbagi media yang merupakan salah satu jenis hiburan virtual untuk berbagi media video dan suara. *YouTube* telah menjadi salah satu platform paling populer untuk menonton video saat ini. Oleh karena itu, sangat sulit untuk memahami apakah pengguna internet yang menggunakan situs *YouTube* umumnya positif, negatif, atau netral [10].

2.6. *Rapidminer*

RapidMiner merupakan aplikasi yang berfungsi sebagai sarana pembelajaran ilmu data mining. Basis ini dikembangkan oleh sebuah perusahaan yang berspesialisasi dalam data besar di semua tingkatan di bidang bisnis, penelitian, pembelajaran, dan analisis yang menguntungkan. RapidMiner memiliki ratusan solusi pembelajaran untuk pengelompokan, klasifikasi, asosiasi, dan analisis regresi [11].

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif untuk menilai tingkat akurasi tertinggi serta mengevaluasi sentimen pengguna *Youtube*, apakah bersifat positif atau negatif, yang dapat mempengaruhi persepsi masyarakat terhadap Mobil Listrik.



Gambar 1. Tahapan Metode Penelitian

3.1. Sumber Data

Sumber data ini diperoleh melalui pengambilan sampel data dari Kaggle, *specifically from https://www.kaggle.com/datasets/billycemerson/analisis-sentimen-terkait-intensif-mobil-listrik*. Pengolahan data dilakukan menggunakan algoritma *Naive Bayes*

setelah melakukan pra-pemrosesan pada komentar-komentar yang ada. Meskipun tidak ada sumber data yang secara eksplisit membahas penelitian ini, terdapat beberapa sumber yang membicarakan penggunaan algoritma *Naive Bayes* dalam analisis sentimen. Selain itu, juga terdapat beberapa referensi yang membahas penerapan algoritma *Naive Bayes* dalam analisis sentimen.

3.2. Teknik Analisis Data

Dalam studi yang berjudul " Analisis Sentimen Komentar Video Mobil Listrik Di *Platform Youtube* Dengan Metode *Naive Bayes*", analisis data diterapkan melalui metode pendekatan *Knowledge Discovery in Databases* (KDD). Setelah data terkumpul, analisis dilakukan menggunakan algoritma *Naïve Bayes* untuk meramalkan sentimen masyarakat di media sosial terkait dengan pandangan terhadap mobil listrik. Langkah-langkah dalam menganalisis data dalam penelitian dengan pendekatan KDD mencakup:

3.2.1. Data Selection

Pada fase ini, metode *Knowledge Discovery in Databases* (KDD) diterapkan untuk pemilihan atribut dan pengumpulan data. Data yang dikumpulkan mencakup informasi sentimen yang diberikan oleh pengguna di platform *YouTube*. Dataset ini berasal dari program *open source Kaggle* dan terdiri dari total 1517 entri data dengan tiga atribut: nama akun, komentar, dan label.

3.2.2. Pre-processing

Tahap pembersihan dan perbaikan data merupakan langkah penting dalam mengelola data yang seringkali tidak terstruktur dan mengandung beragam karakter setelah pengumpulan data. Tujuan utama dari *preprocessing* adalah untuk menghilangkan *noise* atau gangguan yang tidak diinginkan dalam data. Dalam tahap ini, beberapa langkah yang diterapkan meliputi:

a. Pembersihan Data
   Proses mengidentifikasi dan menghapus elemen-elemen data yang tidak relevan atau tidak diinginkan, seperti karakter khusus, spasi berlebih, atau entitas yang tidak bermanfaat.
b. Pelipatan Case
   Melibatkan standardisasi huruf dalam data, seperti mengubah seluruh teks menjadi huruf kecil atau huruf besar, untuk memastikan konsistensi dan memudahkan analisis lebih lanjut
c. Tokenisasi
   Memecah teks atau kalimat menjadi unit-unit lebih kecil yang disebut token. Token bisa berupa kata, frasa, atau karakter, memudahkan pemrosesan dan analisis lebih lanjut.
d. Pemfilteran
   Proses menghilangkan atau menyaring elemen-elemen yang tidak relevan atau tidak diinginkan dalam data. Misalnya, menghapus *stop words* (kata-kata umum yang sering tidak memberikan makna tambahan) atau menghapus kata-kata yang tidak sesuai dengan tujuan analisis.
e. *Data Mining*
   Dalam tahap *data mining* ini, digunakan metode algoritma *Naive Bayes* untuk mengklasifikasikan analisis sentimen berdasarkan data yang telah melalui proses *preprocessing* hingga pembobotan kata dengan metode tf-idf. Setelah berhasil melatih data, langkah selanjutnya adalah melakukan pengujian menggunakan data uji untuk mengevaluasi akurasi klasifikasi yang telah dilakukan.
f. Evaluasi
   Hasil dari proses *data mining* menghasilkan pola informasi yang perlu disajikan dalam format yang dapat dimengerti oleh para pemangku kepentingan. Evaluasi dilakukan dengan tujuan memeriksa akurasi hasil klasifikasi, yang melibatkan perhitungan nilai yang telah ditentukan sebelumnya. Tahap evaluasi menggunakan perhitungan tabel matriks klasifikasi (*confusion matrix*) sebagai dasar evaluasi.

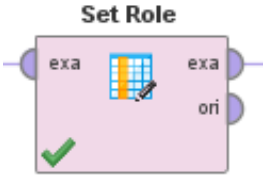
4. HASIL DAN PEMBAHASAN

Penelitian ini memanfaatkan dataset yang diperoleh dari situs resmi Kaggle, yang terdiri dari ulasan pengguna YouTube sebagai respons terhadap kata kunci "Mobil Listrik". Jumlah keseluruhan data dalam dataset ini mencapai 1517 entri. Dataset ini terdiri dari tiga atribut utama, yaitu nama_akun, coment, dan label. Dapat dilihat pada table 3.1 *dataset*

Tabel 1. Dataset

Nama_akun	Coment	Label
Sqn Ldr	saran sih bikin harga ionic sama kayak brio insya alloh laris manis	positif
lushen ace	problem subsidi kualitas diturunin harga dinaikin usaha gitu cari cuan subsidi sebab inflasi paling gede	negatif
Fatih Al-Ayyubi	baik kualitas kembang dulu baik kualitas motor motor pabrikan jepang	positif

4.1. *Data Selection*



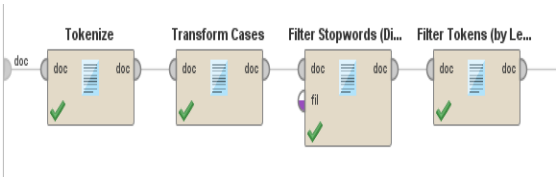
Gambar 2. Set Role

Set Role dalam *RapidMiner* berfungsi untuk menentukan peran dari suatu atribut dalam dataset. Untuk membuat pemodelan prediktif lebih mudah dan efisien, atribut harus diberi peran sehingga

RapidMiner dapat memahami penggunaan atribut ini dalam analisis dan pemodelan data.

4.2. *Process Document*

Dokumen Proses adalah bagian dari alat pengolah kata yang memungkinkan Anda menganalisis dan mengelola data teks. Ini mencakup beberapa atribut lain seperti *Tokenize, transform cases, stopwords*, dan *filter tokens*.



Gambar 3. Process Document

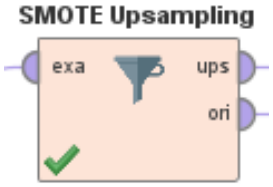
Proses pengolahan data yang mendukung proses pembelajaran algoritma mengambil langkah-langkah penting yang memungkinkan data yang sebelumnya tidak terstruktur diubah menjadi format yang lebih terstruktur.

Pertama, tokenisasi dilakukan untuk memecah teks menjadi unit-unit lebih kecil yang disebut token untuk memudahkan analisis lebih lanjut. Selanjutnya, langkah peka huruf besar-kecil dilakukan untuk mengubah semua karakter dalam teks ke format huruf besar-kecil tertentu, seperti huruf kecil atau huruf besar. Proses ini dilanjutkan dengan pemfilteran kata, yaitu menghilangkan kata-kata tertentu berdasarkan aturan atau kriteria tertentu yang relevan. Selain itu, ada langkah pemfilteran kata berhenti yang menghilangkan kata-kata yang sering muncul dari teks yang kemungkinan besar tidak memberikan informasi penting. Terakhir, Anda dapat menghapus token dan kata-kata yang tidak perlu berdasarkan sejumlah karakter tertentu, sehingga Anda dapat fokus pada informasi yang lebih relevan.

Semua langkah ini merupakan proses penting dalam mengubah data awal menjadi data terstruktur, yang memfasilitasi proses analisis dan pembelajaran algoritma selanjutnya.

4.3. *SMOTE Upsamplimg*

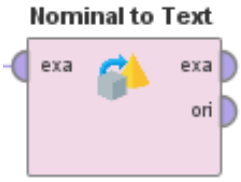
SMOTE, atau *Synthetic Minority Over-sampling Technique*, adalah suatu teknik *oversampling* yang digunakan dalam *machine learning* untuk menangani ketidakseimbangan kelas pada data. Ketidakseimbangan kelas terjadi ketika jumlah sampel dalam satu kelas jauh lebih besar atau lebih kecil dibandingkan dengan kelas lainnya. Dalam konteks *oversampling*, fokusnya adalah pada kelas minoritas. Ini diterapkan dalam tugas klasifikasi dengan tujuan meningkatkan performa model untuk sebagian data yang spesifik.



Gambar 4. SMOTE Upsampling

4.4. *Transformation*

Tahap transformasi merupakan tahap transformasi data ke dalam format yang dapat diolah pada tahap *data mining*, yaitu data diubah menjadi teks nominal dan dari nominal menjadi teks.



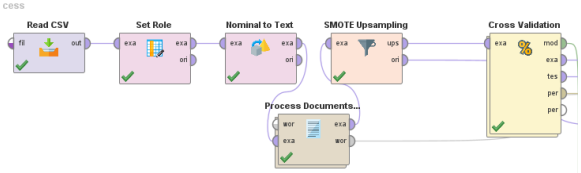
Gambar 5. Nominal to Text

4.5. *Data Mining*

Dalam fase ini, model yang menggunakan metode naive bayes diimplementasikan melalui serangkaian operator dalam platform analisis data. Langkah pertama melibatkan operator Read CSV, yang berfungsi untuk mengakses data pelatihan. Operator ini terhubung dengan Set Role, yang mengubah peran target menjadi label, dan kemudian terhubung dengan Nominal to Text untuk mengubah nilai nominal menjadi teks.

Proses selanjutnya melibatkan operator Document from Data, yang mencakup sub-proses seperti Tokenize, Transform Cases, Stopword, dan Filter Tokens. Langkah-langkah ini dirancang untuk mengolah teks dengan membaginya menjadi token, mengubah format huruf, menghapus kata-kata umum, dan memfilter token berdasarkan kriteria tertentu.

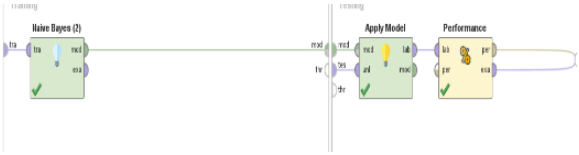
Setelah itu, dilakukan SMOTE upsampling untuk menyeimbangkan data, mengatasi ketidakseimbangan kelas yang mungkin terjadi. Tahap terakhir melibatkan operator Cross Validation untuk menguji model naive bayes secara menyeluruh, menghasilkan evaluasi yang akurat terhadap kinerja model pada data yang belum pernah dilihat sebelumnya. Keseluruhan proses ini dirancang untuk memberikan pemahaman menyeluruh mengenai kualitas dan efektivitas model naive bayes yang telah dikembangkan, dapat dilihat seperti pada gambar 6.



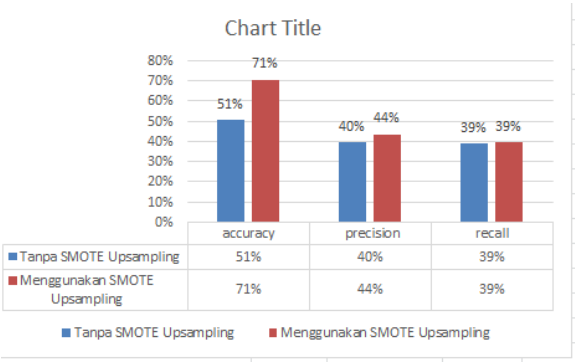
Gambar 6. Proses Model Naïve Bayes

4.6. Evaluasi Model

Evaluasi merupakan tahapan penting dalam berbagai bidang penelitian, pengembangan, dan analisis, termasuk ilmu pengetahuan, teknologi, bisnis, dan pendidikan. Fungsinya untuk mengevaluasi atau mengukur efektivitas, kinerja, atau kualitas suatu program, produk, atau proses berdasarkan kriteria tertentu. Dalam konteks ini dilakukan evaluasi untuk mengukur kinerja model Naive Bayes dengan menggunakan metode validasi silang terkait penerapan dan evaluasi kinerja model untuk menghitung hasil algoritma Naive Bayes.



Gambar 7. Proses validation Model Naïve Bayes



Gambar 8. Hasil perbandingan menggunakan SMOTE Upsampling

Berdasarkan hasil evaluasi model, hasil klasifikasi model *Naive Bayes* mencapai akurasi sebesar 51% tanpa menggunakan *SMOTE Upsampling*, adapun yang menggunakan SMOTE Upsampling mendapat akurasi sebesar 71%.

4.7. Penerapan Algoritma *Naïve Bayes*

Algoritma Naive Bayes berdasarkan teorema Bayes merupakan metode klasifikasi yang efisien dan banyak digunakan dalam analisis data. Algoritma ini mengasumsikan bahwa setiap fitur yang digunakan dalam proses klasifikasi tidak bergantung satu sama lain, namun asumsi ini tidak selalu mencerminkan kondisi data kompleks di dunia nyata. Keuntungan utama dari Naive Bayes adalah permasalahannya dapat diimplementasikan dengan cepat dan mudah. Sebagai bagian dari klasifikasinya, algoritma ini menggunakan teorema Bayes untuk menghitung probabilitas kelas berdasarkan distribusi probabilitas fitur yang diamati.

Penerapan Naive Bayes melibatkan langkah-langkah seperti persiapan data. Contoh pada Tabel 3.1 menunjukkan kumpulan data dengan 1517 data dan tiga atribut. Proses ini juga dapat dilihat pada Gambar 4.5 yang menunjukkan langkah-langkah Rapidminer. Untuk mengimpor data, gunakan beberapa operator seperti "Baca CSV" lalu sambungkan ke "Tetapkan Peran" untuk menentukan peran atribut dalam kumpulan data. Langkah selanjutnya adalah mengubah atribut nominal menjadi teks menggunakan Nominal ke Teks, diikuti dengan upsampling SMOTE untuk mengatasi ketidakseimbangan kelas dalam data.

Proses dokumen digunakan untuk menganalisis dan mengelola data teks dan menyertakan operator seperti tokenisasi, transformasi kasus, stopwords, dan token filter. Tahap pemrosesan dokumen mencakup beberapa operator seperti Tokenize, Transform Cases, Stopword, dan Filter Tokens. Ini termasuk membagi teks menjadi unit-unit yang lebih kecil, mengubah format huruf, menghapus kata-kata umum, dan memfilter token berdasarkan jumlah karakter. Model Naive Bayes dilatih menggunakan data pelatihan yang menghitung kemungkinan munculnya kata dalam kategori sentimen. Selanjutnya, gunakan data pengujian untuk menguji performa model dengan mengklasifikasikan ulasan ke dalam kategori sentimen yang diprediksi. Hasil evaluasi model Naive Bayes menunjukkan presisi sebesar 70,50%, presisi sekitar 43.64%, dan recall sebesar 39,48%.

Langkah selanjutnya adalah menggunakan validasi silang untuk memberikan estimasi performa model Naïve Bayes yang stabil dan konsisten. Naive Bayes banyak digunakan dalam klasifikasi teks dan analisis sentimen, namun algoritme juga diterapkan dalam berbagai situasi lain seperti klasifikasi gambar, pengenalan pola, dan sistem rekomendasi. Naïve Bayes tetap menjadi pilihan populer untuk analisis data dan pembelajaran mesin karena manfaatnya yang ringan dan kemudahan penerapannya.

4.8. Evalusai Kinerja Model Algoritma *Naïve Bayes*

Tahap evaluasi performa model selanjutnya dilakukan dengan menggunakan set pengujian, menggunakan metrik evaluasi seperti akurasi, presisi, dan *recall*. untuk memberikan gambaran yang komprehensif. Selama fase evaluasi model RapidMiner, ditemukan perbedaan signifikan antara pemodelan dengan dan tanpa *SMOTE upsampling*. Tanpa *SMOTE upsampling*, akurasi model adalah 50,70% +/- 3,36% dan skor mikro rata-rata adalah 50,69%. Di tingkat kelas, terdapat 196 true positif, 248 true negative, dan 48 true neutral, dengan tingkat presisi 39,84% dan tingkat *recall* 38,89%.

Namun, penerapan *SMOTE upsampling* secara signifikan meningkatkan kinerja model, meningkatkan akurasi menjadi 70,69% +/- 3,09% dan rata-rata granularitas hasil mikro mencapai 70,70%. Dalam analisis kelas, nilai positif sebenarnya meningkat menjadi 199, nilai negatif nyata meningkat menjadi 253, dan nilai netral nyata meningkat menjadi 856. Tingkat presisi dengan *recall* 43,64 n tiap kelas juga menunjukkan peningkatan signifikan sebesar 39,48%, dapat dilihat pada gambar 4.7.

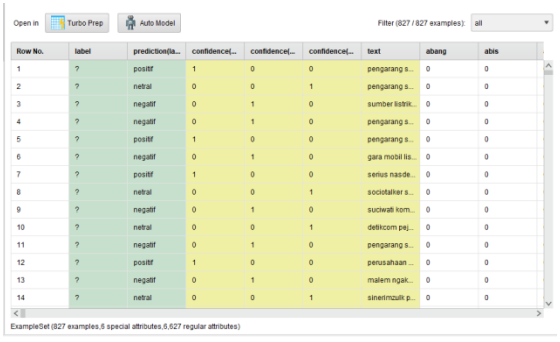
Hasil ini menunjukkan bahwa penerapan *SMOTE upsampling* berdampak positif pada kemampuan model untuk mengatasi ketidakseimbangan kelas, sehingga menghasilkan peningkatan signifikan dalam akurasi prediksi. *SMOTE* memungkinkan model memberikan respons yang lebih baik terhadap kelas minoritas. Hal ini tercermin dari meningkatnya nilai *true positif* dan *recall* pada kelas ini. Sama halnya dengan penelitian yang dilakukan [12], metode algoritma *Naive Bayes* memperoleh nilai akurasi sebesar 50,74% yang menunjukkan tingkat akurasi yang bisa dibilang cukup baik. Setelah menerapkan *SMOTE upsampling*, akurasinya meningkat menjadi 70,50%.

Hal ini menunjukkan bahwa penggunaan optimasi *SMOTE* secara signifikan meningkatkan hasil akurasi metode algoritma *Naive Bayes* dan dapat tergolong hasil yang baik.

4.9. Implementasi

Setelah menyelesaikan tahap evaluasi, penelitian dilanjutkan ke tahap implementasi menggunakan model Naïve Bayes untuk analisis sentimen terhadap mobil listrik. Langkah pertama adalah pemilihan dataset komentar atau ulasan. Dilanjutkan dengan pembersihan data untuk mengoptimalkan fokus pada informasi yang esensial, termasuk penghapusan karakter khusus, tautan, dan elemen tidak relevan.

Setelah pembersihan data, dilakukan tokenisasi untuk memecah teks menjadi kata-kata, diikuti dengan penghapusan stopwords untuk meningkatkan akurasi analisis. Dataset yang telah diproses kemudian dilatih dengan model Naïve Bayes untuk mengklasifikasikan sentimen. Setelah pelatihan, model diuji dengan data baru dengan penyesuaian parameter Union untuk meningkatkan ketepatan analisis sentimen.



Gambar 9. Hasil Implementasi

Proses selanjutnya melibatkan Filter Example untuk memilih atribut tertentu dan Replace Missing Value untuk mengisi nilai yang hilang dalam dataset. Hal ini dilakukan untuk menjaga integritas dan kualitas data. Model kemudian diaplikasikan pada dataset baru menggunakan metode Apply Model untuk membuat prediksi dan menyajikan hasil analisis sentimen. Dengan demikian, kesimpulan dapat ditarik berdasarkan prediksi sentimen dari model terhadap data baru. Hasil implementasi dapat dilihat pada gambar 9

5. KESIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan dapat diambil kesimpulan Penerapan algoritma *Naive Bayes* menghasilkan akurasi 70,69%, presisi 43,64%, dan *recall* 39,48%, menunjukkan kinerja positif dengan akurasi di atas 70%. Meskipun perlu perbaikan, terutama dalam mengurangi kesalahan positif, kinerja *Naive Bayes* menunjukkan potensi besar untuk penerapan praktis. Teknik *SMOTE Upsampling* terbukti efektif menangani ketidakseimbangan kelas, meningkatkan kinerja model. Implementasi *Naive Bayes* dalam analisis sentimen terhadap mobil listrik menunjukkan adaptasi yang baik terhadap ulasan terkini, dengan penyesuaian parameter Union meningkatkan ketepatan analisis sentiment, memberikan hasil yang lebih akurat. Dan menyarankan Lakukan perbandingan hasil pengujian model dengan berbagai algoritme klasifikasi untuk memahami efektivitasnya. Pertimbangkan dan terus implementasikan *SMOTE Upsampling* dalam menangani ketidakseimbangan kelas. Perluas penggunaan model *Naive Bayes* dalam analisis sentimen kendaraan listrik dengan menguji data baru dan memperbarui parameter secara berkala, sambil mempertimbangkan penyesuaian parameter untuk meningkatkan hasil analisis sentimen.

DAFTAR PUSTAKA

[1] A. Santoso, A. Nugroho, and A. S. Sunge, “Sentiment Analysis About Electric Cars With Support Vector Machine Method and Feature Selection Particle Swarm Optimization,” *J. Pract. Comput. Sci.*, vol. 2, no. 1, pp. 24–31, 2022.

[2] P. G. Aryanti and I. Santoso, “Analisis Sentimen Pada Twitter Terhadap Mobil Listrik Menggunakan Algoritma Naive Bayes,” *IKRA-ITH Inform. J. Komput. dan Inform.*, vol. 7, no. 2, pp. 133–137, 2023, [Online]. Available: https://journals.upi-yai.ac.id/index.php/ikraith-informatika/article/view/2821

[3] Y. Pratama, D. T. Murdiansyah, and K. M. Lhaksmana, “Analisis Sentimen Kendaraan Listrik Pada Media Sosial Twitter Menggunakan Algoritma Logistic Regression dan Principal Component Analysis,” *J. Media Inform. Budidarma*, 2023, [Online]. Available: http://www.stmik-budidarma.ac.id/ejurnal/index.php/mib/article/view/5575

[4] C. J. E. Munthe, N. A. Hasibuan, and H. Hutabarat, “Penerapan Algoritma Text Mining Dan TF-RF Dalam Menentukan Promo Produk Pada Marketplace,” *RESOLUSI Rekayasa Tek. Inform. dan Inf.*, vol. 2, no. 3, pp. 110–115, 2022.

[5] A. B. Sasmita, B. Rahayudi, and L. Muflikhah,

"Analisis Sentimen Komentar pada Media Sosial Twitter tentang PPKM Covid-19 di Indonesia dengan Metode Naïve Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 3, pp. 1208–1214, 2022, [Online]. Available: http://j-ptiik.ub.ac.id

[6] D. Wahyuningsih and E. Patima, "Penerapan Naive Bayes Untuk Penerimaan Beasiswa," *J. Telemat.*, vol. 11, no. 1, p. 135, 2018, doi: 10.35671/telematika.v11i1.665.

[7] R. Vincent *et al.*, "Perbandingan Klasifikasi Naive Bayes Dan Support Vector Machine Dalam Analisis Sentimen Dengan Multiclass Di Twitter," vol. 7, no. 4, pp. 2496–2505, 2023.

[8] H. Irsyad, A. Farisi, and M. R. Pribadi, "Klasifikasi Opini Masyarakat Terhadap Jasa ISP MyRepublic dengan Naïve Bayes," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 8, no. 1, p. 30, 2019, doi: 10.22146/jnteti.v8i1.487.

[9] M. Aziz, Y. Marcellino, I. A. Rizki, S. A. Ikhwanuddin, and J. W. Simatupang, "Studi analisis perkembangan teknologi dan dukungan pemerintah indonesia terkait mobil listrik," *TESLA J. Tek. Elektro*, vol. 22, no. 1, pp. 45–55, 2020.

[10] A. A. Ningtyas, A. Solichin, and R. Pradana, "Analisis Sentimen Komentar Youtube Tentang Prediksi Resesi Ekonomi Tahun 2023 Menggunakan Algoritme Sentiment Analysis Of Youtube Comments On Prediction Of Economic Recession In 2023 Using The Naïve Bayes," *Bit (Fakultas Teknol. Inf. Univ. Budi Luhur)*, vol. 20, no. 1, pp. 9–16, 2023.

[11] K. Akmal, A. Faqih, and F. Dikananda, "Perbandingan Metode Algoritma Naïve Bayes Dan K-Nearest Neighbors Untuk Klasifikasi Penyakit Stroke," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 470–477, 2023, doi: 10.36040/jati.v7i1.6367.

[12] R. Ardianto, T. Rivanie, Y. Alkhalifi, F. S. Nugraha, and W. Gata, "Sentiment Analysis on E-Sports for Education Curriculum Using Naive Bayes and Support Vector Machine," *J. Ilmu Komput. dan Inf.*, vol. 13, no. 2, pp. 109–122, 2020, doi: 10.21609/jiki.v13i2.885.