

Terbit : 27 September 2024

Perbandingan Metode Klasifikasi dan SMOTE Terhadap Analisa Sentimen Mobil Listrik Indonesia

¹Jefina Tri Kumalasari, ²Indah Puspitorini

^{1,2} Universitas Bina Sarana Informatika, Indonesia

¹jefina.jtk@bsi.ac.id, ²Indah.itn@bsi.ac.id

ABSTRAK

Pesatnya perkembangan penyimpanan energi terutama teknologi baterai membuat inovasi pada kendaraan. Perkembangan transportasi beralih dari kendaraan berbahan bakar fosil ke energi listrik dimana memiliki keuntungan ramah lingkungan dan tidak bising. Namun pada daya terima masyarakat Indonesia berbeda-beda ada yang positif dan negatif. Dinamika di masyarakat dapat di gali melalui komentar yang terdapat di Youtube. Dimana Youtube sebagai situs layanan menonton dan berbagi video yang mengalami penetrasi signifikan. Saat ini Indonesia berada di posisi ke empat sebagai pengguna Youtube sehingga situs ini sesuai digunakan sebagai sumber data. Analisa mengenai berdasarkan sentimen mengenai mobil listrik yang berasal dari salah satu kanal Youtube dimana terdapat mendapatkan hasil 65% positif dan sisanya 35% berpendapat negatif. Data diolah dengan menggunakan metode Knowledge Discovery in Data Base dengan metode CRISP-DM. Komentar awal berjumlah 1265 yang diolah dengan menggunakan metode klasifikasi pada data mining. Dengan cara membandingkan algoritma KNN, algoritma Naïve Bayes dan Algoritma SVM. Kemudian guna menyeimbangkan data dan meningkatkan akurasi peneliti mengkombinasikan dengan algoritma SMOTE. Hasil yang diperoleh integrasi algoritma SVM dan SMOTE menghasilkan keakurasian 94.65%, gabungan Naïve Bayes dan SMOTE menghasilkan akurasi 84.82% dan KNN dengan SMOTE menghasilkan 62.80%. Berdasarkan hasil perbandingan algoritma klasifikasi yang telah dilakukan didapat kombinasi algoritma SVM dan SMOTE menghasilkan prediksi terbaik.

Kata kunci : Klasifikasi, SMOTE, Mobil Listrik

PENDAHULUAN

Sejalan derngan mingkatnya jumlah kendaraan bermotor khususnya di kota besar akan berkontribusi pada polusi udara. Meskipun bukan satu-satunya penyebab polusi udara kendaraan bermotor di Indonesia memberikan kontribusi signifikan terhadap peningkatan polusi. Semakin banyak kendaraan yang ada akan menimbulkan kemacetan sehingga resiko pencemaran udara meningkat.

Dalam penyelesaiannya, masalah polusi udara tidak hanya melibatkan pemerintah saja, masyarakat luas juga harus ikut berpartisipasi. Dalam membangun ekonomi sosial suatu negara masyarakatnya sangat bergantung pada transportasi berbasis bahan bakar fosil seperti minyak bumi, batu bara

Menurut BPS (Badan Pusat Statistik) 2023 data sementara jumlah kendaraan bermotor di Indonesia 157.080.504 unit, 60 persen berada di Pulau Jawa. Angka tersebut mencakup kendaraan pribadi yaitu 127.976.339-unit sepeda motor (87 persen) dan 19.177.264 mobil pribadi dan sisanya merupakan angkutan barang dan orang. Menurut Gaikindo (Gabungan Industri Kendaraan Bermotor Indonesia) Jumlah ini bakal terus bertambah seiring dengan pabrikan otomotif yang konsisten meluncurkan produk baru khususnya pascapandemi Covid-19. Data tersebut menunjukkan bahwa Indonesia merupakan target pasar menarik bagi para produsen kendaraan bermotor khususnya mobil listrik. Cepat atau lambat banyaknya kendaraan bermotor akan mulai memberi dampak negatif. Pesatnya pertumbuhan jumlah kendaraan bermotor khususnya milik pribadi tak mungkin tanpa sebab. Masyarakat membutuhkan transportasi yang efisien, praktis, dan tentunya murah. Oleh karenanya, kendaraan pribadi menjadi pilihan karena dapat menunjang mobilitas yang tinggi. Seiring pertambahnya penjualan kendaraan maka emisi karbon CO₂ juga meningkat.

Masyarakat dapat memulainya dari hal-hal kecil seperti tertib lalu lintas dan berpartisipasi untuk mengurangi kemacetan di jalan dengan lebih memilih untuk menggunakan transportasi umum dibandingkan kendaraan pribadi.

Gabungan Industri Kendaraan Bermotor Indonesia (GAIKINDO) mengeluarkan informasi terbaru mengenai perkembangan penjualan mobil nasional. Pada Agustus 2024, penjualan *whole sales* (pabrik ke dealer) mobil nasional mencapai 76.304 unit. Ini berarti tumbuh 2,8 *persen month to month* (MtM) dibanding bulan Juli 2024 yakni 74.229 unit. Kinerja penjualan mobil listrik (*battery electric vehicle*, BEV) di Indonesia. Berdasarkan data Gabungan Industri Kendaraan Bermotor Indonesia (GAIKINDO), penjualan *whole sales* (pabrik ke dealer) mobil listrik nasional mencapai 23.045-unit pada Januari-Agustus 2024.

Bertambahnya variasi model dan merek akan menjadi faktor pendorong penjualan mobil listrik nasional selama tahun 2024 berjalan. Saat ini, terdapat 18 merek yang memasarkan mobil listrik di Indonesia dimana sebagian besar berasal dari pabrikan China, baik yang diproduksi lokal maupun impor CBU *completely built up*. Mobil listrik lokal memiliki kualitas yang tidak kalah tetapi dalam penjualan masih kalah saing dengan produk luar.

TINJAUAN PUSTAKA

Analisis Sentimen

Analisis sentiment merupakan bagian dari *text analytics* proses untuk mengambil pendapat, emosi dan evaluasi seseorang yang tertulis mengenai suatu topik dengan memanfaatkan teknik pemrosesan bahasa (Cahyaningtyas et al., 2021). Tujuan dari text analisis merekrontuksi dokumen dalam bentuk text menjadi database yang terstruktur. Dimana saat sentiment analisis digunakan sebagai media informasi berupa opini objektif dan konsisten dari banyak data berupa teks seperti email, layanan dukungan pelanggan, percakapan media social. Sosial media biasa digunakan untuk mengumpulkan teman, bertukar informasi dan data dengan memanfaatkan jaringan internet (Pasek et al., 2022)

Menurut databoks.katadata.co.id pada tahun 2024 Indonesia terdapat 242 juta penetrasi internet sedangkan pengguna media sosial Youtube terdapat 139 juta pengguna atau 53,8% dari populasi. Youtube merupakan salah satu *media social* yang banyak digunakan sebagai sarana hiburan, berita, informasi dan lain sebagainya. Youtube memungkinkan pengguna untuk mengunggah, menonton atau berbagi video. Pengguna Youtube juga bervariasi dari anak sampai dewasa. Menurut GoodStats Youtube memiliki 2,49 Milyar pengguna, dan di Indonesia mencapai 139 Juta Orang Indonesia menonton konten video. Youtube rata-rata empat jam per hari dengan menggunakan perangkat *smart TV* maupun *gadget*. Data internal Google, total durasi konten yang diunggah ke Youtube oleh akun-akun Indonesia tumbuh lebih dari 85 persen dalam periode Juni 2023-Juni 2024. Sehingga dapat digunakan sebagai sumber informasi.

Teknik yang digunakan pengambilan data dengan menggunakan metode *crawling*. Web crawling sendiri Teknik yang digunakan adalah cara untuk menghasilkan atau menyalin suatu *big data* dari dunia nyata lalu dilakukan penyimpanan ke dalam file local computer atau kedalam file lokal komputer ke dalam format tabel. Data data berasal dari channel Youtube bertema mobil listrik VS mobil bensin mulai tanggal 24-11-2017 sampai tanggal 10-10-2024 yang diambil menggunakan *crawling method* dari API google kemudian diolah dengan Bahasa Phyton. Sentimen atau comment yang diolah berjumlah 1262 data.

Klasifikasi

Teknik yang digunakan untuk mengolah data menggunakan metode Klasifikasi dimana akan menggunakan operator Union dan Naïve Bayes untuk mengisi sentiment yang masih kosong. Naive Bayes digunakan sebagai salah satu algoritma banyak dipakai dalam metode klasifikasi memiliki kinerja yang baik dan mudah digunakan. Guna penentuan metode yang terbaik dalam klasifikasi penulis membandingkan algoritma Naïve Bayes, KNN dan SVM kemudian kombinasikan dengan algoritma SMOTE guna mendapatkan akurasi terbaik.

Algoritma KNN (K-Nearest Neighbor) algoritma yang mengklasifikasi objek berdasarkan mayoritas dari kategori. Algoritma KNN memiliki beberapa kelebihan diantaranya ketangguhan terhadap data training yang memiliki banyak noise dan data dalam jumlah yang besar (Cholil et al., 2021).

$$\text{Euclidean Distance } (x, y) = \sqrt{\sum_{i=1}^t (x_i - y_i)^2}$$

Keterangan :

t = Jumlah kata

x_i = Dokumen uji

y_i = Dokumen latih

Naïve Bayes salah satu metode dari *machine learning* yang digunakan untuk mencari suatu untuk probabilitas. Dalam kata lain, Naïve, Bayes, merupakan metode untuk klasifikasi text dengan kecepatan pemrosesan yang tinggi (Puspita & Widodo, 2021). Naïve Bayes mempunyai ciri independensi dari masing-masing kondisi atau kejadian, memiliki performa yang cepat dalam melatih data. Pendekatan Nave Bayes memiliki keuntungan karena hanya menggunakan sedikit data training untuk mendapatkan estimasi parameter yang diperlukan dalam proses klasifikasi. (Pebdika et al., 2023)

SVM (*Support Vector Mechine*) merupakan salah satu teknik yang efektif dalam melakukan prediksi baik dalam pengklasifikasian maupun regresi. Memiliki prinsip Minimalkan Risiko Struktural (SRM) dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua kelas dalam ruang input (Hermanto et al., 2020) SVM sendiri Penggunaan SVM dalam pembelajaran telah diperkenalkan sejak tahun 1992 oleh Boser, Guyon, dan Vapnik.

Synthetic Minority Oversampling Technique (SMOTE) merupakan metode Oversampling dimana bertujuan mengatasi permasalahan ketidakseimbangan pada data. Prinsip yang digunakan adalah membangkitkan data buatan untuk menambah jumlah data kelas minoritas agar setara dengan kelas mayoritas (Septian Pradana & Nooraeni, 2023)

Penelitian Terkait

Beberapa peneliti sebelumnya Analisa Sentimen pada program merdeka belajar – kampus merdeka menggunakan twitter sebagai sumber data dan SVM (Support Vector Machine) sebagai algoritmanya, dari nilai F-measure didapat kelas positif sebesar 94.8% dan kelas negatif sebesar 95%. (Ni'ma Kholila, 2021).

Analisis sentiment pada pengguna layanan BPJS dilakukan dengan membandingkan algoritma KNN, Decision Tree dan Naïve Bayes. (Puspita & Widodo, 2021) hasil pengolahan dengan metode KNN mencapai tingkat akurasi 95.58% dengan class precision untuk prediksi negative 45.00%, prediksi positive adalah 0.00%, dan prediksi neutral adalah 96.83%. Pada algoritma Decision Tree tingkat akurasinya mencapai 96.13% dengan class precision untuk prediksi negative adalah 55.00%, prediksi positive adalah 0.00%, dan prediksi neutral adalah 97.28%. Dan yang terakhir adalah metode Naïve Bayes yang mencapai akurasi 89.14% dengan class precision untuk prediksi negative adalah 16.67%, prediksi positive 1.64%, dan prediksi neutral adalah 98.40%. Dalam penelitian ini dapat ditarik kesimpulan bahwa Decision Tree memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan metode yang lain. (Fauzan, 2023). Menggunakan multinominal Naïve Bayes mendapatkan hasil akurasi senilai 86.11% pada 359 dataset yang terbagi menjadi dua tipe data yaitu 182 komentar positif dan 177 komentar negatif dengan menganalisa sentimen komentar Youtube mengenai program Kampus Merdeka Berbasis Web.

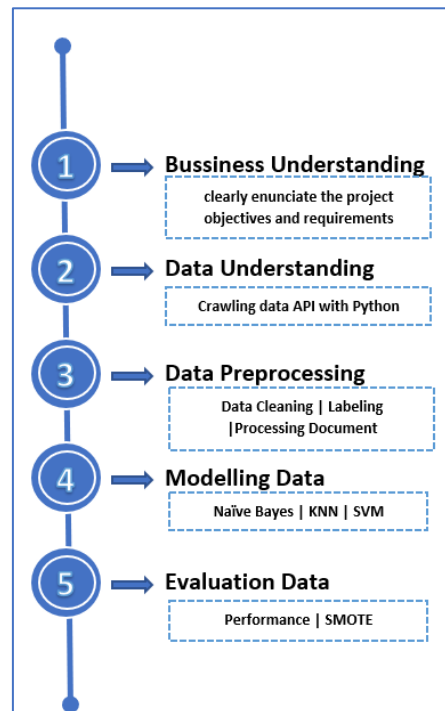
(Rochadiani, 2023) Melakukan Analisa komentar pada Youtube mengenai Chat GPT menggunakan pendekatan Text Blob dan Vader. Katagori yang digunakan positif, netral dan negative. Analisa tersebut menghasilkan 46.6%, 22.9%, 30.4% untuk Text Blob dan Vader 49.5%, 26.3%, 24.2%. Hasil akhir terbesar adalah sentiment positif pada penggunaan Chat GPT dengan menggunakan VADER

METODE PENELITIAN

Metodologi Penelitian

Kerangka penelitian dibuat agar penelitian lebih terstruktur dimana setiap tahapnya akan berhubungan dengan proses sebelumnya. Kerangka ini memiliki acuan pada CRISP (*Cross Industri Standart Process for Data Mining*) merupakan proses standard untuk pemecahan masalah bisnis atau

penelitian dimana memiliki tahapan yang konsisten. (Larose & Larose, 2014). Penelitian ini memiliki tahapan:



Gambar 1. Metodologi Penelitian

Uraian kerangka penelitian sebagai berikut:

1. *Bussiness Understanding*

Pada proses ini ditentukan tujuan yaitu guna mengetahui bagaimana sentimen di masyarakat mengenai keberadaan mobil listrik. Mengetahui mengenai potensi dan hambatan memasarkan mobil listrik berdasarkan komentar dari Youtube. Komentar ini dibagi menjadi komentar positif dan komentar negatif

2. *Data Understanding*

Guna mempersingkat waktu pada proses pengumpulan data maka digunakan teknik crawling. Dengan menggunakan Google API Console lalu mengekstrak data yang diperlukan pada saluran Youtube dengan tema perbandingan mobil listrik dan mobil bensin kemudian di proses dengan menggunakan bahasa Phyton. Data awal yang diperoleh adalah 1265 data. Data yang di dapat kemudian diekstraksi dapat disimpan dalam bentuk spreadsheet Microsoft Excel dengan tipe file CSV. Data *review* yang telah diambil akan dijadikan sebagai dataset. (Cahyaningtyas et al., 2021). Data diambil dengan terlebih dahulu melakukan *crawling data* dengan memanfaatkan API Google kemudian dengan menggunakan Google Colab dengan bahasa Phytion berhasil mendapatkan 1265 data mentah.

3	2024-05-03T16:17:31Z	@Andry559	Lantaran uji kecepatan juga tenaga yg unggul i...	1
4	2024-07-13T05:03:22Z	@blackred5155	Kalo di pasang tramisi atau gearbox bisa saja ...	0
...	...	...	...	...
1260	2018-04-29T14:20:59Z	@muhammadrezaakbarstrpar4561	Ilmu Rekayasa gila animasi lu best banget.. G...	5
1261	2018-05-09T13:46:05Z	@supri6891	Sing murah ono ra,wes nyoba rung	1
1262	2018-08-15T13:22:05Z	@asupanhariananda1648	good job bro, thanks your created new chanel w...	2
1263	2018-08-25T05:30:15Z	@ArifRahman-fz3ux	Ilmu Rekayasa tolong jelaskan cara kerja bitco...	1
1264	2017-11-24T08:00:36Z	@iskandardenny9052	Akhir,a yg d tunggu dataang . Hehee	5

1265 rows × 4 columns

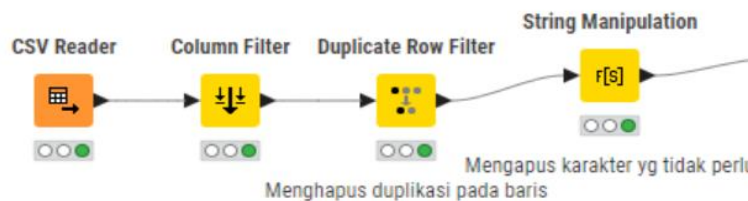
Gambar 2 Data Mentah dari Colabs

HASIL DAN PEMBAHASAN

Tahapan *Preprocessing* merupakan tahapan yang panjang guna mempersiapkan data agar siap untuk langkah selanjutnya sehingga dapat menghasilkan pengetahuan pada beberapa teknik data mining (Mayasari et al., 2024) pada tahap ini mengubah data mentah menjadi format terstruktur yang mudah dipahami (Barus, 2022).

Cleaning Data

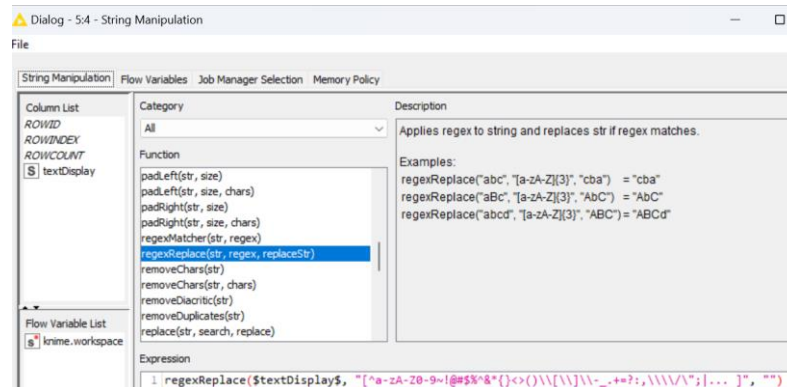
Pada tahap ini dimulai dari pembersihan data yaitu dengan menggunakan aplikasi KNIME



Gambar 3 Proses Cleaning

Tahapan proses cleaning meliputi:

1. **Coloumn Filter:** Menghilangkan kolom yang tidak diperlukan sehingga hanya menyisakan kolom sentimen saja
2. **Duplicate Row Filter:** berguna untuk menghilangkan baris yang sama
3. **String Manipulation** digunakan untuk menghilangkan karakter yang tidak diperlukan dengan bantuan *regex Replace* akan menjadi lebih cepat.



Gambar 4 Proses Penghapusan Karakter

4. Pada tahap pembersihan ini juga lakukan penghapusan sentimen yang tidak berhubungan dengan tema.
5. Untuk proses selanjutnya peneliti menggunakan aplikasi Rapidminer.
6. Setelah data bersih maka disimpan pada file bertipe csv. Tahapan selanjutnya adalah melakukan *labeling* hal ini dilakukan secara manual dari 1265 data awal berkurang menjadi 1189 examples dan data yang telah diberi label sebanyak 658 terdiri dari 432 sentimen positif 226 sentimen negatif.

Label		Least	Most	Values
 sentimen	Polynomial 0	negatif (226)	positif (432)	positif (432), negatif (226)

Gambar 5 Hasil Pembersihan Data

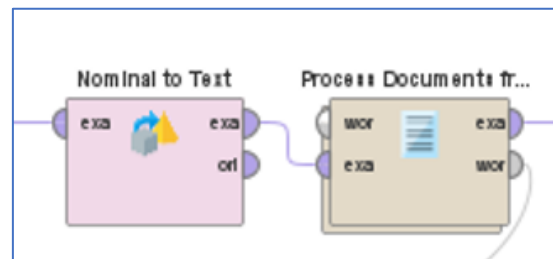
Untuk data yang belum memiliki label akan terisi tanda tanya pada sentimennya. Seperti dibawah ini:

126	positif	kalo pengeboran minyak bumi atau penggalian fosil juga menghasilkan limbah
127	?	Kalo semua mobil listrik jadi ga seru kalo nonton rally flemans dili
128	?	Mobil batremobil listrik tidak ada min
129	negatif	MOBIL LISTRIK RAWAN MELEDAKY
130	positif	Tapi nggak pernah denger bahwa mobil listrik meledak
131	?	Mobil listrik tdak ngam utk budak motorsport Tiada bunyi ekzos

Gambar 6 Tampilan Data Sudah Diberi Label Dan Belum Diberi Label

Processing Document

Langkah selanjutnya adalah *Processing Document*



Gambar 7 Menu Process Document

Pada Process Documents memerlukan operator Nominal to text yang berguna untuk mengubah tipe data nominal menjadi text. Di dalam Process Document From data terdapat beberapa tahapan, yaitu:

1. Tokenize, Proses membagi teks menjadi bagian-bagian yang lebih kecil. Memisahkan beberapa kalimat menjadi satu kata yang memiliki atribut sendiri. Pada proses ini terdapat proses TF - IDF (*Term Frequency - Inverse Document Frequency*) yaitu metode yang menentukan bobot suatu kata dalam dokumen atau artikel.
2. Transform Cases penggunaan node ini untuk menyamakan tipe teks diaman sebelumnya terdapat huruf besar dan kecil diubah menjadi huruf kecil semua atau huruf besar semua
3. Filter Stopwords Dictionary, berfungsi menghapus kata-kata yang tidak penting, data berasal dari dictionary Bahasa Indonesia dimana list kata diambil dari Kaggle.com berbahasa Indonesia.
4. Filter Tokens by Length untuk memfilter atau menghapus kata-kata yang jumlah karakternya terlalu sedikit atau terlalu banyak dimana secara default minimal karakter 3 dan maksimal karakter 25.

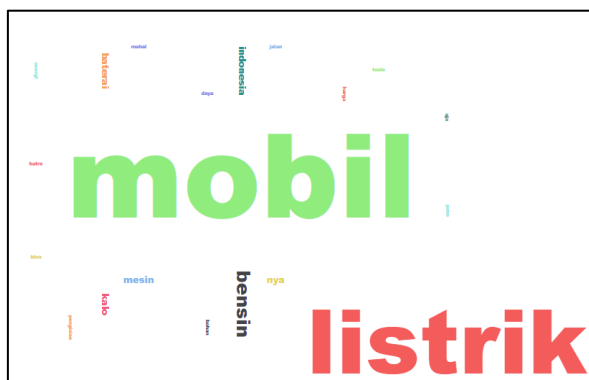
Tabel 1 Data Sebelum Dan Sesudah Proses

Data Awal	Data Setelah Pemrosesan
P5 sma 24	jarak cocok mahal pakai baterai cepat rusak
Ora cocok untuk jarak jauh, ora cocok nggo wong desa, harga masih mahal kalau jarang di pakai baterai malah cepat rusak belum lagi masalah infrastuktur yang ada di desa masih jauh kayaknya untuk untuk meninggalkan mesin bahan bakar ðŸ˜…	lantaran kecepatan tenaga unggul mobil mobil listrik mobil
Lantaran uji kecepatan juga tenaga yg unggul itu cuma mobil BBM. Mobil listrik mobil untuk santai juga jalan santai	mobil listrik cocok keterbatasan energi baterai perapian
Kalo di pasang tramisi atau gearbox bisa saja saingin mobil bbm, mengingat torsinya besar dan konstan, memang untuk saat ini mobil listrik tidak cocok, karena keterbatasan kepadatan energi baterai, itu lah kebanyakan krl dan kereta cepat seperti kcjb pakai motor listrik, dari pada mesin diesel! karena sumber listrik berasal dari pln	mobil unggul irit mobil listrik bensin

Data Awal	Data Setelah Pemrosesan
Mobil BBM yg lebih unggul tapi yg paling irit mobil listrik juga paling bising	dijadikan mobil offroad mobil listrik alias kena lumpur konslet
Apakah bisa untuk dijadikan mobil OFF-ROAD dari mobil listrik ini alias kalau kena air lumpur tak kan konslet YAA? SOALNYA KELISTRIKAN RAWAN AKAN AIR APALAGI BERCAMPUR DENGAN LUMPUR ðŸ˜~... #JawabanDitunggu	mobil listrik ramah lingkungan
mobil gasol fuel menang top speed ev car menang di akselerasi tapi kalo soal tenaga ev car tak bisa disepelein juga karena ada ev car dgn tenaga buas utk offroad yg bisa saingi fuel gasol car	pakai listrik bawa kampung habis baterai

Setelah proses ini maka akan muncul kalimat baku, lebih sederhana, kata yang mengandung karakter kurang dari 4 karakter akan terhapus.

Guna mengetahui data atau kata yang sering muncul sering dibahas digunakan *Wordlist to data*. Dengan ini dapat diketahui apa yang paling banyak dibicarakan pada masalah mobil listrik. Datanya sebagai berikut:



Gambar 8 Wordcloud pada Analisa Mobil Listrik

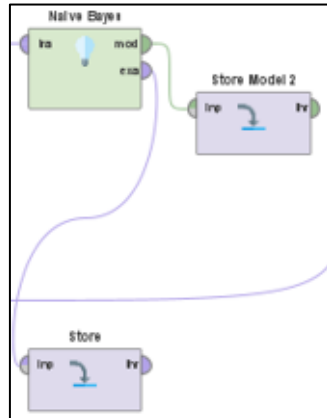
Lebih detailnya adalah sebagai berikut:

Row No.	word	in documents	total
1	mobil	764	972
2	listrik	759	863
3	baterai	222	278
4	bensin	175	197
5	pakai	102	118
6	mesin	84	106
7	kalo	95	103
8	indonesia	76	86
9	banjir	79	81
10	tesla	62	66
11	energi	54	64
12	pengisian	54	62
13	harga	58	60

Gambar 9 List Wordcloud Data

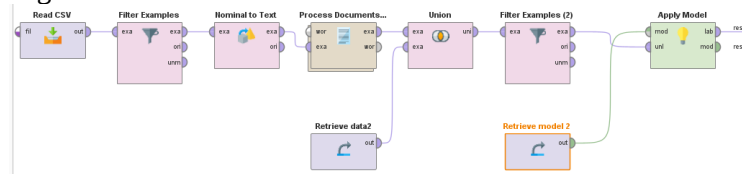
Labeling

Berdasarkan data diawal dapat diketahui bahwa terdapat *missing value* dimana data tersebut belum ditentukan sentimennya oleh karena itu perlu proses analisa berdasarkan data sebelumnya dengan menggunakan operator Union dan algoritma Naïve Bayes yang digunakan sebagai *Mechine learning*. Langkah awal Naïve Bayes akan mempelajari pola kata yang sudah ada lalu membagi data menjadi data latih dan data model.



Gambar 10 Hasil Klasifikasi Data Latih dan Data Model

Dengan menggunakan data sebelumnya dan operator Union akan dilakukan penggabungan pada data prosesnya adalah sebagai berikut:



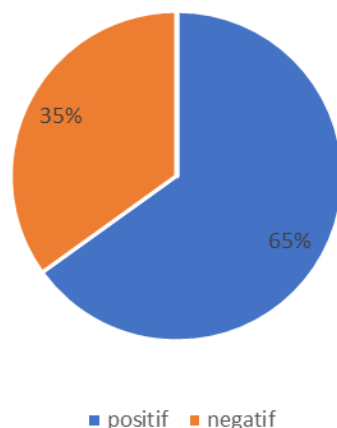
Gambar 11 Proses Union

Setelah proses diatas maka selesai sudah proses penentuan sentimen analisis. Pada analisis mobil listrik mendapatkan hasil sebagai berikut:

Label	Binominal	0	Negative	Positive	Values
sentimen			positif	negatif	positif (879), negatif (444)

Gambar 12 Hasil Prediksi Sentimen Analysis

Data sentimen Mobil Listrik



Gambar 13 Grafik Hasil Prediksi Sentimen Mobil Listrik

Berdasarkan data kita dapat mendapatkan sentimen positif terhadap mobil listrik sebesar 65% dan 35% adalah sentimen negatif. Hal ini sejalan dengan informasi dari GAIKINDO bahwa Mobil listrik kini berkontribusi 4,11 persen terhadap total penjualan *whole sales* mobil nasional yang berjumlah 560.619-unit hingga Agustus 2024. Menurut Suwandi Wiratno Ketua Umum Asosiasi Perusahaan Pembiayaan Indonesia (APPI) “Pasar mobil BBM masih lebih bagus daripada mobil listrik setidaknya hingga tahun 2030 mendatang”. Salah satu tantangan mobil listrik itu charging station belum banyak. Jadi pemerintah harus benar-benar menyediakan infrastrukturnya”.

Hasil dan pembahasan mengenai sentimen analisis terhadap perkembangan mobil listrik di Indonesia dengan menggunakan metode SVM, Naïve Bayes dan KNN berdasarkan nilai acurasi confusion matrix adalah sebagai berikut:

1. Algoritma SVM

Tabel 2 Hasil Akurasi algoritma SVM

accuracy: 81.03% +/- 3.27% (micro average: 81.03%)

	true positif	true negatif	class precision
pred. positif	794	166	82.71%
pred. negatif	85	278	76.58%
class recall	90.33%	62.61%	

Dari tabel diatas dapat diketahui model berhasil mengklasifikasi dengan benar sebesar 81,03% dari total data yang diuji. +/- 3.27% menunjukkan interval ketidakpastian hal ini dihitung pada saat melakukan pengujian berulang. Prediksi true positif berjumlah 796 dan true negatif berjumlah 278.

2. Algoritma KNN

Tabel 3 Hasil Akurasi Algoritma KNN

accuracy: 80.95% +/- 2.85% (micro average: 80.95%)

	true positif	true negatif	class precision
pred. positif	797	170	82.42%
pred. negatif	82	274	76.97%
class recall	90.67%	61.71%	

Algoritma KNN berhasil mengklasifikasi dengan akurasi 80,95%, interval ketidakpastian sebesar 2,85% dengan prediksi true positif sebesar 797 dan true negatif 274. Class recall 90,67% artinya menunjukkan model sangat baik dalam mendeteksi data positif tetapi ada sebagian data positif yang tidak terdeteksi. Sedangkan class recall 61,71% menunjukkan model mengidentifikasi data negatif meskipun belum sempurna karena hanya 61,71% data negatif yang berhasil diteksi dengan benar.

3. Algoritma Naïve Bayes

Tabel 4 Hasil Akurasi Algoritma Naïve Bayes

accuracy: 83.98% +/- 3.45% (micro average: 80.98%)

	true positif	true negatif	class precision
pred. positif	601	9	98.52%
pred. negatif	278	435	61.01%
class recall	68.37%	97.97%	

Algoritma Naïve Bayes berhasil mengklasifikasi dengan akurasi 83,98% dengan interval ketidakpastian sebesar 3,45% dengan prediksi true positif sebesar 601 dan true negatif 435. Class Recall menunjukkan

true positif sebesar 68,67 hal ini menunjukkan model mengidentifikasi data positif hanya sebesar 68,37% dan sebaliknya model sangat baik dalam mendeteksi data negatif yaitu sebesar 97.97%.

Guna peningkatan kinerja algoritma diperlukan algoritma SMOTE sebagai penyeimbang pada kumpulan data pada teknik klasifikasi data mining guna menemukan hasil terbaik.(Shuja et al., 2020)

Tabel 5 Hasil Akurasi Algoritma SVM dan SMOTE

accuracy: 94.65% +/- 2.08% (micro average: 94.65%)

	true positif	true negatif	class precision
pred. positif	801	16	98.04%
pred. negatif	78	863	91.71%
class recall	91.13%	98.18%	

Dari tabel diatas dapat diketahui bahwa terjadi peningkatan akurasi menjadi 94,65% dengan class recall positif 91,13% dan class recall negatif 98,18%.

Tabel 6 Hasil Akurasi Algoritma Naïve Bayes dan SMOTE

accuracy: 84.82% +/- 2.97% (micro average: 84.81%)

	true positif	true negatif	class precision
pred. positif	615	3	99.51%
pred. negatif	264	876	76.84%
class recall	69.97%	99.66%	

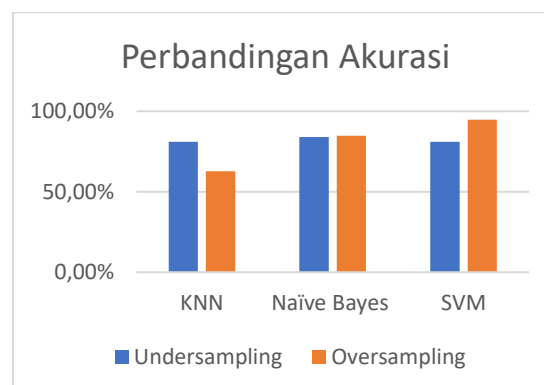
Berdasarkan hasil prediksi diatas didapat sedikit peningkatan akurasi setelah menggunakan algoritma SMOTE yaitu sebesar 84,82% yang semula 83.98%.

Tabel 7 Hasil Akurasi Algoritma KNN dan SMOTE

accuracy: 62.80% +/- 2.45% (micro average: 62.80%)

	true positif	true negatif	class precision
pred. positif	234	9	96.30%
pred. negatif	645	870	57.43%
class recall	26.62%	98.98%	

Berdasarkan hasil prediksi diatas KNN mengalami penurunan akurasi yang semula 80.95% menjadi 62,80%



Gambar 12 Perbandingan Akurasi Algoritma

Dari gambar dapat grafik di dapat bahwa algoritma naïve bayes memiliki performa yang lebih stabil walaupun telah dilakukan oversampling sedangkan algoritma SVM mengalami peningkatan setelah dilakukan oversampling sedangkan algoritma KNN mengalami penurunan pada tingkat akurasi.

KESIMPULAN

Mobil Listrik sebagai mobil yang digerakkan oleh baterai telah lama hadir di Indonesia mulai tahun 2012 oleh menteri BUMN saat itu Dahlan Iskan. Segala upaya telah dilakukan pemerintah seperti memberikan intensif pajak, pengurangan bea masuk dan dorongan pembangunan Stasiun Pengisian Kendaraan Listrik Umum (SPKLU) namun hingga saat ini penjualan mencapai 4,11 persen penjualan *whole sales* mobil nasional. Penerapan teknologi data mining dapat membantu memprediksi dan mengidentifikasi bagaimana sentimen, permasalahan yang terdapat pada masyarakat terhadap mobil listrik. Berdasarkan hasil analisa sentimen dihasilkan 64% positif dan 35% memberikan sentimen negatif, jadi secara umum masyarakat senang dan menyetujui mobil listrik. Algoritma yang digunakan dalam penelitian ini adalah Naïve Bayes, SVM dan KNN. Tingkat keakurasian data pada algoritma SVM 81.03% dengan interval ketidakpastian +/- 3.27%. Tingkat keakurasian algoritma Naïve Bayes 83.98%, dan interval ketidakpastian sebesar +/- 3.45%, sedangkan KNN memiliki keakurasian 80.95% +/- 2.85%. Guna meningkatkan keakurasian penelitian menggunakan algoritma SMOTE sebagai penyeimbang. Integrasi algoritma SVM dan SMOTE menghasilkan keakurasian 94.65%, gabungan Naïve Bayes dan SMOTE menghasilkan akurasi 84.82% dan t KNN dengan SMOTE menghasilkan 62.80%. Berdasarkan hasil perbandingan algoritma klasifikasi yang telah dilakukan didapat kombinasi algoritma SVM dan SMOTE menghasilkan prediksi terbaik. Penelitian ini diharapkan dapat membantu pemerintah dan perusahaan yang bergerak pada mobil listrik dapat digunakan sebagai bahan evaluasi guna meningkatkan penjualan dan infrastruktur.

REFERENSI

- Barus, S. G. (2022). *KLASIFIKASI SENTIMEN DATA TIDAK SEIMBANG MENGGUNAKAN ALGORITMA SMOTE DAN K-NEAREST NEIGHBOR PADA ULASAN PENGGUNA APLIKASI PEDULILINDUNGI*.
- Cahyaningtyas, C., Nataliani, Y., & Widiarsari, I. R. (2021). Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE. *AITI: Jurnal Teknologi Informasi*, 18(Agustus), 173–184.
- Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). IJCIT (Indonesian Journal on Computer and Information Technology) Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. In *IJCIT (Indonesian Journal on Computer and Information Technology)* (Vol. 6, Issue 2).
- Fauzan, R. A. (2023). (SENAFTI) 30 Agustus 2023-Jakarta. In *Nasional.Mahasiswa.Fakultas.Teknologi.Informasi* (Vol. 2, Issue 2). <https://senafiti.budiluhur.ac.id/index.php/senafiti/article/view/929/563>
- Hermanto, Kuntoro, A. Y., Asra, T., Pratama, E. B., Effendi, L., & Ocanitra, R. (2020). Gojek and Grab User Sentiment Analysis on Google Play Using Naive Bayes Algorithm and Support Vector Machine Based Smote Technique. *Journal of Physics: Conference Series*, 1641(1). <https://doi.org/10.1088/1742-6596/1641/1/012102>
- Larose, D. T., & Larose, C. D. (2014). *DISCOVERING KNOWLEDGE IN DATA An Introduction to Data Mining Second Edition Wiley Series on Methods and Applications in Data Mining*.
- Mayasari, Y., Nasution, R., & Sumatera Utara, N. (2024). Post-Election Sentiment Analysis 2024 via Twitter (X) Using the Naïve Bayes Classifier Algorithm. *Journal of Dinda Data Science, Information Technology, and Data Analytics*, 4(2), 123–134. <http://journal.ittelkom-pwt.ac.id/index.php/dinda>

-
- Ni'ma Kholila. (2021). Merdeka belajar twitter. *Jurnal Ilmiah Teknik Informatika*, 15(2), 252–261. <https://doi.org/10.35457/antivirus.v15i2.1866>
- Pasek, P., Mahawardana, O., Sasmita, G. A., Agus, P., & Pratama, E. (2022). Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap “Figure Pemimpin” Menggunakan Python. In *JITTER-Jurnal Ilmiah Teknologi dan Komputer* (Vol. 3, Issue 1).
- Pebdika, A., Herdiana, R., & Solihudin, D. (2023). KLASIFIKASI MENGGUNAKAN METODE NAIVE BAYES UNTUK MENENTUKAN CALON PENERIMA PIP. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 1).
- Puspita, R., & Widodo, A. (2021). Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS. *Jurnal Informatika Universitas Pamulang*, 5(4), 646. <https://doi.org/10.32493/informatika.v5i4.7622>
- Rochadiani, T. H. (2023). Sentiment Analysis of YouTube Comments Toward Chat GPT. *Jurnal Transformatika*, 21(1). <https://doi.org/10.26623/transformatika.v21i2.7033>
- Septian Pradana, R., & Nooraeni, R. (2023). PENERAPAN SMOTE PADA DATA TIDAK SEIMBANG DALAM PEMODELAN STATUS NEET PENDUDUK USIA MUDA DI PROVINSI BANTEN TAHUN 2022 THE APPLICATION OF SMOTE TO IMBALANCED DATA ON NEET STATUS MODELLING OF YOUTH IN BANTEN PROVINCE 2022. 18. <https://doi.org/10.47441/jkp.v18i1.325>
- Shuja, M., Mittal, S., & Zaman, M. (2020). *Effective Prediction of Type II Diabetes Mellitus Using Data Mining Classifiers and SMOTE* (pp. 195–211). https://doi.org/10.1007/978-981-15-0222-4_17