

ANALISIS SENTIMEN PADA TWITTER TERHADAP MOBIL LISTRIK MENGUNAKAN ALGORITMA NAIVE BAYES

Putri Gea Aryanti¹, Imam Santoso²

¹ Program Studi Sarjana Sistem Informasi, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika

² Dosen Universitas Teknologi Muhammadiyah Jakarta

aryanti@stikomcki.ac.id¹, imam.santoso@utmj²

ABSTRACT

Analisis sentimen twitter merupakan teknik untuk mengidentifikasi sentimen atau pendapat dalam tweet dan kemudian mengategorikannya ke dalam tweet positif atau tweet negatif salah satu topik yang dibahas pada social media twitter adalah mobil listrik, mobil listrik memiliki beberapa kelebihan dibandingkan dengan mobil bahan bakar fosil. Mobil listrik ini menuai banyak komentar dari masyarakat sehingga menimbulkan pro dan kontra di sosial media twitter. Penelitian ini dilakukan tujuannya untuk mengetahui pendapat masyarakat terhadap mobil listrik. Perlu dilakukan penelitian untuk melihat seberapa besar dampak dari kehadiran produk tersebut. Peneliti memilih analisis sentimen karena merupakan tehnik yang tepat untuk pengolahan dataset. Apakah pendapat tersebut lebih mengarah ke positif atau negatif dan untuk mengetahui nilai accuracy, AUC dari penggunaan algoritma *Naive Bayes* pada Software RapidMiner Studio. Sehingga penelitian yang telah dilakukan mendapatkan hasil akurasi sebesar 87.43% dan AUC 0,518.

Kata Kunci: Analisis Sentimen, Text Mining, Naive Bayes, Mobil Listrik.

ABSTRACT

Twitter sentiment analysis is a technique to identify sentiments or opinions in tweets and then categorize them into positive tweets or negative tweets One of the topics discussed on social media twitter is electric cars, electric cars have several advantages compared to fossil fuel cars. This electric car reaped many comments from the public, causing pros and cons on social media twitter. This research was conducted to find out the opinion of the public towards electric cars. Research needs to be done to see how much impact the presence of the product is. Researchers chose sentiment analysis because it is the right technique for dataset processing. Is the opinion more positive or negative and to find out the accuracy value, AUC from the use of the Naive Bayes algorithm in RapidMiner Studio Software. So that the research that has been done gets accuracy results of 87.43% and AUC 0.518.

Keyword: Sentiment Analysis, Text Mining, Naive Bayes, Electric Cars.

I. PENDAHULUAN

Sumber energi dunia telah mengalami beberapa kali perubahan, dari yang awalnya sebagian besar menggunakan biomassa untuk memenuhi kebutuhan energi, menjadi bahan bakar fosil [1]. Energi fosil adalah energi yang tak terbarukan dan akan habis pada beberapa tahun yang akan datang. Mobil listrik merupakan salah satu sarana transportasi yang bisa mengurangi pemakaian energi fosil agar memenuhi kebutuhan mobilitas masyarakat namun tetap ramah lingkungan karena tidak memiliki polusi atau emisi gas buang [2]. Mobil listrik memiliki beberapa kelebihan yang potensial yaitu dapat mengurangi emisi gas rumah kaca [3]. Kehadiran mobil listrik menimbulkan pro dan kontra di kalangan masyarakat seperti yang rame di bincangkan di media sosial twitter. Sebagian kalangan masyarakat berpendapat mobil listrik diklaim efektif menjadi solusi defisit migas tetapi sebagiannya lagi berpendapat bahwa untuk menuju kendaraan listrik perlu persiapan yang matang, khususnya sari segi

infrastruktur. Banyaknya respon dari masyarakat indonesia dalam menilai mobil listrik menjadi salah satu tolak ukur peneliti untuk melakukan analisis sentimen masyarakat mengenai mobil listrik pada sosial media twitter.

Menurut penelitian yang dilakukan oleh Ahmad Santoso, Agung Nugroho, dan Aswan Supriyadi Sunge dengan judul “Analisis Sentimen Tentang Mobil Listrik Dengan Metode Support Vector Machine Dan Feature Selection Particle Swarm Optimization”. Menurut peneliti dalam pengecekan akurasinya, menggunakan feature selection Particle Swarm Optimization pada metode support vector machine dapat meningkatkan nilai akurasi yang didapat. Dimana nilai akurasi yang didapat pada awalnya sebesar 82,51% meningkat menjadi 86,07% [4]. Menurut penelitian yang dilakukan oleh Aditta Agustian, Tukino, dan Fitria Nurapriani dengan judul “Penerapan Analisis Sentimen dan Naive Bayes Terhadap Opini Penggunaan Kendaraan Listrik di

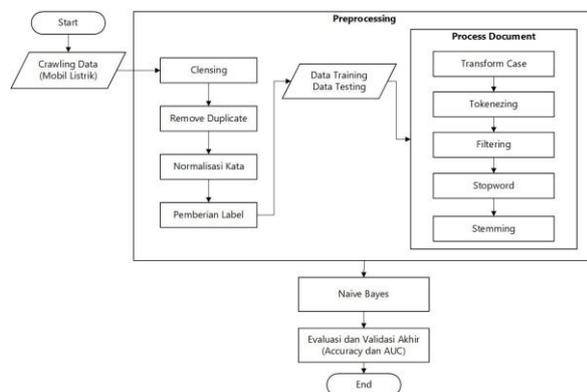
Twitter”. Pada penelitian ini peneliti mendapatkan akurasi 80% pada penelitiannya [5].

Pada penelitian yang dilakukan oleh Bagja Nugraha yang berjudul “Metode Klasifikasi Analisis Sentimen pada Media Sosial”. Peneliti melakukan penelitian terkait metode klasifikasi dengan menggabungkan Support Vector Machine (SVM), Principal Component Analysis dan Genetic Algorithm, dimana pada penelitian ini tingkat akurasi yang didapat menggunakan metode SVM sebesar 86,00% sedangkan untuk metode SVM dengan mengoptimalkan penggunaan Principal Component Analysis dan Genetic Algorithm memiliki tingkat akurasi lebih tinggi yaitu sebesar 97% [6].

Penelitian ini dilakukan tujuannya untuk mengetahui pendapat masyarakat Indonesia terhadap mobil listrik. Apakah pendapat tersebut lebih mengarah ke positif atau pun negatif. Data sentimen yang digunakan berasal dari sosial media twitter. Berdasarkan hasil penelitian terdahulu, peneliti tertarik untuk menggunakan metode *Naive Bayes* sebagai metode yang akan peneliti gunakan dalam menganalisis sentimen masyarakat, dikarenakan *Naive Bayes* memiliki kelebihan diantaranya adalah dalam menentukan jarak menggunakan *Naive Bayes* proses komputasi menjadi cepat.

II. METODE

Dalam penelitian ini menggunakan algoritma Naive Bayes, dimana dilakukan beberapa langkah seperti yang dijelaskan pada gambar di bawah ini :



Gambar 1. Metode Penelitian

Tahap pada penelitian ini dimulai mengumpulkan dataset, tahap selanjutnya adalah dengan proses preprocessing data, kemudian dilakukan pengujian dengan menggunakan algoritma *Naive Bayes* dan tahap akhir yang dilakukan adalah dengan evaluasi dan validasi akhir yang didapat dari penelitian ini.

A. Pengumpulan dan Pengolahan Data

Pengambilan data twitter menggunakan operator search twitter di perangkat lunak RapidMiner. Data tweet diambil dari pendapat-pendapat masyarakat

perihal mobil listrik yang terhubung dengan akun twitter. Parameter yang terkandung pada operator search twitter diantaranya adalah parameter connection, query, result type, limit, dan language. Proses pengambilan data pada penelitian ini terkoneksi dengan akun twitter dengan query “mobil listrik”. Jenis pencarian yang digunakan adalah recent or popular, data yang di crawling sebesar 700 data tweet berbahasa indonesia.

Setelah melakukan pengumpulan data, dilakukan proses pengolahan data atau *preprocessing*, tahapan ini mencakup kegiatan membangun data dan dilanjutkan kegiatan pembersihan data agar dapat dilakukan pengelolaan ketahap selanjutnya. Berikut tahapan *preprocessing* data, antara lain :

1. Transform Cases.

Mengubah huruf kapital yang masih ada di dataset menjadi huruf - huruf kecil. Hal ini bertujuan agar terjadi keseragaman text pada model klasifikasi dan tidak terjadi kesalahan pada proses tokenize.

2. Tokenizing.

Memecah sekumpulan karakter atau kalimat menjadi sebuah potongan karakter atau kata – kata sesuai dengan kebutuhan, biasa juga disebut tokenisasi.

3. Filtering.

Menghilangkan kata – kata dengan panjang karakter tertentu, biasanya kata yang memiliki hanya 2 karakter tidak memiliki arti.

4. Stopword.

Membuang kata – kata yang diabaikan pada sentimen analisis, biasanya yang berupa kata sambung dan kata keterangan.

5. Stemming.

Tahapan ini berfungsi mengubah kata perkata menjadi sebuah kata dasar, dengan cara menghilangkan imbuhan, baik awalan ataupun akhiran.

B. Algoritma

Peneliti menggunakan Naive Bayes karena merupakan algoritma terbaik dalam data mining, dan sangat baik dalam melakukan proses pengklasifikasian data.

C. Evaluasi dan Validasi

Pada tahap pengujian metode, peneliti menggunakan *RapidMiner* dalam melakukan proses eksperimen, data *training* yang digunakan adalah tweet twitter mengenai pro dan kontra pada social media twitter adalah mobil listrik. Kemudian *dataset* dikelompokkan menjadi 2 bagian yaitu opini positif dan negatif.

Peneliti menggunakan *10 Fold Cross Validation* dalam melakukan validasi, dan untuk mengukur

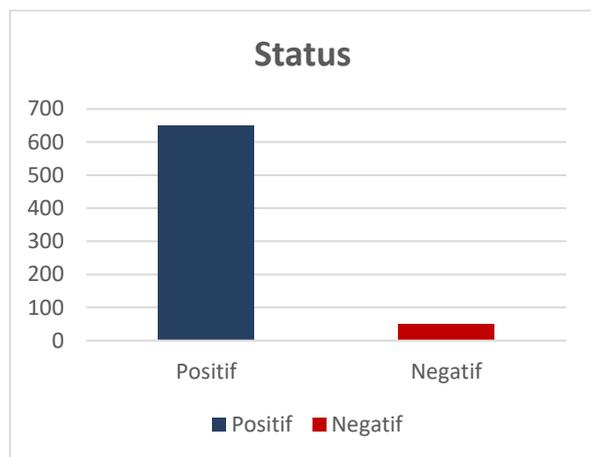
akurasi dilakukan dengan *Confusion Matrix*, yaitu akurasi *Naive Bayes*. Kemudian Kurva *ROC* digunakan untuk mengukur nilai *AUC*.

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data opini atau tweet yang digunakan pada penelitian ini dikumpulkan menggunakan metode crawling pada RapidMiner Studio. Data tweet yang akan digunakan berkaitan dengan kata kunci “Mobil Listrik”. Jumlah data yang telah dikumpulkan dan telah dilakukan penghapusan duplikat sebanyak 700 data tweets dalam bentuk file dengan format XLSX. Pelabelan data pada penelitian ini dilakukan secara manual untuk memastikan ketepatan sentimen pada data yang ada, setelah itu data dikelompokkan menjadi 2 bagian yaitu data positif dan data negatif.

Pengelompokan data ini dilakukan oleh beberapa koresponden, sehingga kemudian dihasilkan 650 opini Positif dan 50 opini negatif.



Gambar 2. Statistik hasil pengumpulan data.

B. Pengolahan Data

Setelah pengumpulan data dan membagi data menjadi komentar positif dan komentar negatif, selanjutnya dilakukan proses pengolahan data atau *preprocessing*. Berikut ini adalah tahapan *preprocessing* :

1. Transform Case.

Mengubah huruf kapital yang masih ada di dataset menjadi huruf - huruf kecil. Hal ini bertujuan agar terjadi keseragaman text pada model klasifikasi dan tidak terjadi kesalahan pada proses *tokenize*.

Tabel 1. Hasil *Transform Case*.

Sebelum	Sesudah
Mobil listrik itu baru dilembang teknologi bersih kalau sumber listriknya pake renewable energy. Tapi kalau sumber pake bbm dan	mobil listrik itu baru bilang teknologi bersih kalau sumber listrik pake renewable energy tapi kalau sumber pake bbm dan

batubara ya tambah kotor. Efisiensi makin menyusut	batubara ya tambah kotor efisiensi makin susut
--	--

2. Tokenizing.

Menghilangkan kata – kata dengan panjang karakter tertentu, biasanya kata yang memiliki hanya 2 karakter tidak memiliki arti.

3. Filtering.

Tahapan ini mengambil kata – kata penting dari hasil token.

Tabel 2. Hasil *Filtering*.

Sebelum	Sesudah
padahal yang design mobil listrik sudah jauh lebih paham mengenai safety dari pada yang asal bunyi	padahal design mobil jauh lebih sudah jauh lebih paham mengenai safety asal bunyi

4. Stopwords.

Membuang kata – kata yang diabaikan pada sentimen analis, biasanya yang berupa kata sambung dan kata keterangan.

Tabel 3. Hasil *Stopword*.

Sebelum	Sesudah
harganya bikin jangkau dong pak kan sudah produk dalam negeri bukan cbu lagi	harga bikin jangkau produk negeri cbu

5. Stemming.

Tahapan ini berfungsi mengubah kata perkata menjadi sebuah kata dasar, dengan cara menghilangkan imbuhan, baik awalan ataupun akhiran.

Tabel 4. Hasil *Stemming*.

Sebelum	Sesudah
saya masih tidak ngerti mobil listrik disebut ramah lingkungan padahal listriknya didapat dari hasil membakar jutaan ton batu bara yang mulai dari penambangan sampai pembakarannya merusak lingkungan	saya masih tidak ngerti mobil listrik sebut ramah lingkung padahal listrik dapat dari hasil bakar juta ton batu bara yang mulai dari tambang sampai bakar rusak lingkung

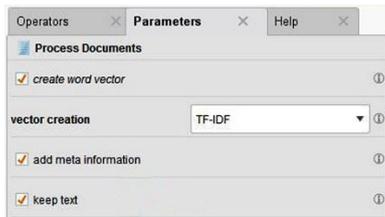
Berikut gambar proses pengolahan data pada *tools rapidminer*.



Gambar 3. Proses Pengolahan Data *Rapidminer*.

Data hasil *preprocessing* yang berupa kata akan diubah ke dalam bentuk angka dengan dilakukan proses pembobotan kata yang bertujuan untuk menghitung bobot pada masing-masing kata yang

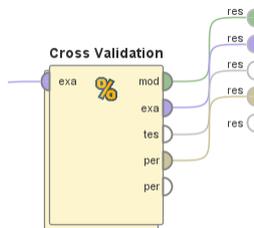
akan digunakan sebagai fitur, semakin banyak dokumen yang akan diproses maka semakin banyak fitur. Pada tahapan ini terdapat dua bagian proses yaitu *TF* (*Term Frequency*) dan *IDF* (*Inverse Document Frequency*), *TF* adalah jumlah kemunculan tiap kata pada sebuah dokumen semakin banyak kata muncul pada tiap dokumen maka semakin besar nilai *TF*. *Df* adalah jumlah nilai dokumen pada tiap kata yang berbanding terbalik yaitu apabila suatu kata jarang muncul pada sebuah dokumen maka nilai *IDF* lebih besar daripada kata yang sering muncul.



Gambar 4. Proses *Tf Idf*.

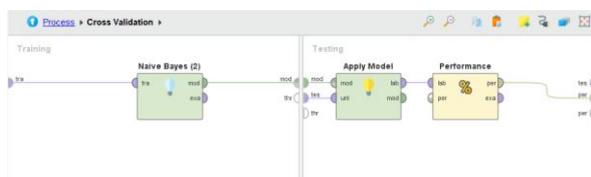
C. Proses Validasi Algoritma.

Selanjutnya adalah penggunaan algoritma *Naive Bayes* untuk melakukan klasifikasi data yang dihubungkan dengan pengujian *10-fold cross validation* dimana proses ini untuk mengevaluasi proses kerja algoritma tersebut dengan membagi data secara acak ke dalam *10 fold* untuk mendapatkan 10 data yang sama, kemudian data tersebut digunakan 9 *fold* untuk data *training* dan 1 *fold* untuk data *testing*.



Gambar 5. Proses *Cross Validation* dengan *Rapidminer*.

Tahapan selanjutnya adalah proses Evaluasi, ialah mengevaluasi kinerja terhadap permodelan Untuk mengestimasi performa dari model Algoritma yang telah dipilih digunakan *Cross Validation* sehingga dapat menghasilkan nilai akurasi (Proses *Cross Validation* dapat dilihat pada Gambar 5). Dibawah ini adalah model pengujian *Naive Bayes* dengan menggunakan tools *Rapidminer*.



Gambar 6. Model Pengujian *Naive Bayes* dengan *RapidMiner*

E. Evaluasi

Penelitian ini menggunakan algoritma *Naive Bayes* untuk melakukan analisis sentimen. Kemudian dalam menentukan evaluasinya peneliti menggunakan *Accuracy* dan *AUC* (*Area Under Curve*).

Dari tahapan - tahapan pengujian yang sudah dilakukan dengan menggunakan *dataset* sebanyak 700 respon tweet pengguna *twitter* mengenai mobil listrik, maka hasil Akurasi Algoritma *Naive Bayes* yaitu sebesar 87,43 %, sedangkan untuk nilai *AUC* sebesar 0.518.

Berikut ini adalah Tabel *Confusion Matrix* Algoritma *Naive Bayes*.

Tabel 5. *Confusion Matrix* Algoritma *NB*.

	True Positive	True Negative	
Pred Positive	582	2	99.66%
Pred Negative	68	48	41.38%
	89.54%	96.00%	

Pada tabel 5 *Confusion Matrix* dapat dilihat, sebanyak 582 data diprediksi *class* positif ternyata sesuai, yaitu masuk kedalam *class* positif, sebanyak 68 data yang diprediksi *class* negatif ternyata termasuk kedalam prediksi *class* positif.

Dan sebanyak 2 data yang diprediksi *class* positif ternyata masuk dalam *class* negatif, kemudian 48 data di prediksi *class* negatif sesuai yaitu termasuk kedalam prediksi *class* negatif.



Gambar 7. Grafik *Area Under Curve* (*AUC*) *NB*.

IV. KESIMPULAN

Berdasarkan hasil penelitian dapat disimpulkan bahwa analisis sentimen masyarakat terhadap mobil listrik di media sosial *twitter* dari total 700 data tweets yang di uji, pengujian hanya menggunakan Algoritma *Naive Bayes* hasilnya terdapat 650 data tweets mengandung sentimen positif dan 50 data tweets mengandung sentimen negatif. Dalam penelitian ini mendapat nilai akurasi sebesar 87,43% dan *AUC* 0,518%.

DAFTAR PUSTAKA

- [1] Setyono, A. E., & Kiono, B. F. T. (2021). Dari energi fosil menuju energi terbarukan: potret kondisi minyak dan gas bumi Indonesia tahun 2020–2050. *Jurnal Energi Baru dan Terbarukan*, 2(3), 154-162.
- [2] Aziz, M., Marcellino, Y., Rizki, I. A., Ikhwanuddin, S. A., & Simatupang, J. W. (2020). Studi analisis perkembangan teknologi dan dukungan pemerintah Indonesia terkait mobil listrik. *TESLA: Jurnal Teknik Elektro*, 22(1), 45-55.
- [3] Adhan Efendi, M. F. (2020). Electric Systems in Sula Electric Cars in Sula Electric Cars' Subang State Polytechnic. *Journal of Mechanical Engineering Education*, 5(1), 47-58.
- [4] A. Santoso, A. Nugroho, and A. S. Sunge, "Analisis Sentimen Tentang Mobil Listrik Dengan Metode Support Vector Machine Dan Feature Selection Particle Swarm Optimization", *jpcs*, vol. 2, no. 1, pp. 24-31, Jul. 2022, doi : <https://doi.org/10.37366/jpcs.v2i1.1084>
- [5] A. Agustian, T. Tukiro, dan F. Nurapriani, "Analisis Sentimen, Text Mining Penerapan Analisis Sentimen Dan Naive Bayes Terhadap Opini Penggunaan Kendaraan Listrik Di Twitter", *TIKA*, vol. 7, no. 3, hlm. 243-249, 2022, doi: <https://doi.org/10.51179/tika.v7i3.1550>
- [6] B. Nugraha, "Metode Klasifikasi Analisis Sentimen pada Media Sosial", *Syntax J. Inf.*, vol. 9, no. 2, pp. 119–123, Dec. 2020, doi: <https://doi.org/10.35706/syji.v9i2.3593>
- [7] Setiawan, K., Rahmatullah, B., Burhanuddin, B., Paryanti, A. B., & Fauzi, F. (2020). Komparasi Metode Naive Bayes Dan Support Vector Machine Menggunakan Particle Swarm Optimization Untuk Analisis Sentimen Mobil Esemka. *JISAMAR (Journal of Information System, Applied, Management, Accounting and Research)*, 4(3), 102-111.
- [8] Pratama, Y., Murdiansyah, D. T., & Lhaksana, K. M. (2023). Analisis Sentimen Kendaraan Listrik Pada Media Sosial Twitter Menggunakan Algoritma Logistic Regression dan Principal Component Analysis. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 7(1), 529-535. doi : <http://dx.doi.org/10.30865/mib.v7i1.5575>
- [9] Sodik, F., & Kharisudin, I. (2021). Analisis Sentimen dengan SVM , NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter. *Prisma*, 4, 628– 634.
- [10] Taufiq, A., Studi, P., Informatika, T., Bangsa, U. P., & Sentimen, A. (2020). Analisis Sentimen Terhadap Pemindahan Ibu Kota Pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization, 10, 173–182.