

Analisis Sentimen Kendaraan Listrik Menggunakan Algoritma Naive Bayes dengan Seleksi Fitur *Information Gain* dan *Particle Swarm Optimization*

Salman Alfarizi¹, Eka Fitriani*²

^{1,2}Universitas Bina Sarana Informatika
Email: ¹salman.slz@bsi.ac.id, ²eka.ean@bsi.ac.id

Abstrak

Kendaraan listrik telah menjadi tren global sebagai alternatif kendaraan berbahan bakar fosil. Namun masih terdapat beberapa permasalahan seperti infrastruktur yang belum memadai, harga yang relatif mahal, dan waktu pengisian baterai yang lama. Untuk meningkatkan penggunaan kendaraan listrik, diperlukan pemahaman dan kesadaran masyarakat terhadap kendaraan listrik. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen opini masyarakat terhadap kendaraan listrik dengan menggunakan algoritma Naive Bayes dengan seleksi fitur *Information Gain* dan *Particle Swarm Optimization* dan. Data yang digunakan dalam penelitian ini adalah data komentar dan ulasan pada media sosial (Twitter) yang terkait dengan kendaraan listrik. Data diambil dengan teknik web crawling data menggunakan API Twitter dan dicoba menggunakan software rapidminer. Dalam penelitian ini menggunakan algoritma Naive Bayes dihasilkan nilai akurasi 79,43% dan AUC 0,639. Selanjutnya dilakukan seleksi fitur menggunakan *Information Gain* dan *Particle Swarm Optimization* untuk menganalisis sentimen opini masyarakat dalam penggunaan kendaraan listrik untuk meningkatkan akurasi dan AUC. Hasil akurasi yang didapat setelah menggunakan algoritma Naive Bayes dengan seleksi fitur *Information Gain* dan *Particle Swarm Optimization* adalah 84,54% dan AUC 0,729. Sehingga dapat disimpulkan bahwa penerapan algoritma Naive Bayes menggunakan seleksi fitur *Information Gain* dan *Particle Swarm Optimization* menjadi metode yang baik dalam analisis sentimen opini masyarakat tentang kendaraan listrik. Penelitian ini diharapkan dapat memberikan informasi yang berguna bagi produsen kendaraan listrik dalam mengembangkan produk mereka, serta memberikan masukan bagi pemerintah dalam mengembangkan kebijakan untuk mendukung penggunaan kendaraan listrik di Indonesia.

Kata kunci: Analisis Sentimen, Naive Bayes, PSO, Information Gain, Kendaraan Listrik

Abstract

Electric vehicles have become a global trend as an alternative to fossil fuel vehicles. However, there are still some problems such as inadequate infrastructure, relatively expensive prices, and long battery charging times. To increase the use of electric vehicles, public understanding and awareness of electric vehicles is needed. Therefore, this study aims to analyze public opinion sentiment towards electric vehicles using the Naive Bayes algorithm with *Information Gain* and *Particle Swarm Optimization* feature selection. The data used in this study are comments and reviews on social media (Twitter) related to electric vehicles. The data was retrieved using web crawling data techniques using the Twitter API and tested using rapidminer software. In this study using the Naive Bayes algorithm resulted in an accuracy value of 79.43% and AUC 0.639. Furthermore, feature selection was carried out using *Information Gain* and *Particle Swarm Optimization* to analyze public opinion sentiment in the use of electric vehicles to increase accuracy and AUC. The accuracy results obtained after using the Naive Bayes algorithm with *Information Gain* and *Particle Swarm Optimization* feature selection are 84.54% and AUC 0.729. So it can be concluded that the application of the Naive Bayes algorithm using *Information Gain* feature selection and *Particle Swarm Optimization* is a good method for analyzing public opinion sentiment about electric vehicles. This research is expected to provide useful information for electric vehicle manufacturers in developing their products, as well as provide input for the government in developing policies to support the use of electric vehicles in Indonesia.

Keywords: Sentiment Analysis, Naive Bayes, PSO, Information Gain, Electric Vehicles

1. PENDAHULUAN

Kendaraan listrik telah menjadi topik yang semakin penting dalam industri otomotif selama beberapa tahun terakhir. Kendaraan listrik adalah jenis kendaraan yang menggunakan motor listrik sebagai sumber tenaga utama untuk menggerakkan roda dan bergerak (Agustian et al., 2022). Kendaraan listrik telah menjadi alternatif yang menarik bagi kendaraan konvensional yang berbahan bakar fosil, karena memiliki beberapa keunggulan seperti ramah lingkungan, hemat energi, dan biaya operasi yang lebih rendah. Energi fosil adalah energi yang tak terbarukan dan akan habis pada beberapa tahun yang akan datang (Santoso et al., 2022).

Kendaraan listrik juga memiliki kinerja yang memuaskan, sehingga mereka dapat bersaing dengan kendaraan konvensional dalam hal kecepatan, jarak tempuh, dan daya angkut. Meskipun kendaraan listrik masih belum menjadi pilihan utama bagi sebagian besar konsumen di seluruh dunia, namun, banyak produsen otomotif terkemuka telah berinvestasi dalam teknologi ini dan memperkenalkan model-model yang semakin canggih dan terjangkau. Dengan kemajuan teknologi yang terus berkembang dan permintaan yang semakin meningkat, kendaraan listrik diharapkan akan menjadi alternatif yang lebih populer dan berkelanjutan bagi kendaraan konvensional di masa depan.

Salah satu penelitian yang sudah dilakukan oleh Adittia Agustian yang berjudul "Penerapan Analisis Sentimen Dan Naive Bayes Terhadap Opini Penggunaan Kendaraan Listrik Di Twitter" penelitian ini menjelaskan tentang opini masyarakat dalam menanggapi kendaraan listrik di media sosial Twitter dengan menghasilkan akurasi 80,00% (Agustian et al., 2022). kemudian penelitian Ratih Yulia Hayuningtyas dan Retno Sari dengan judul "Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata TMII Menggunakan Naïve Bayes Dan PSO" penelitian ini menjelaskan tentang opini masyarakat dalam wisata TMII (Hayuningtyas & Sari, 2019).

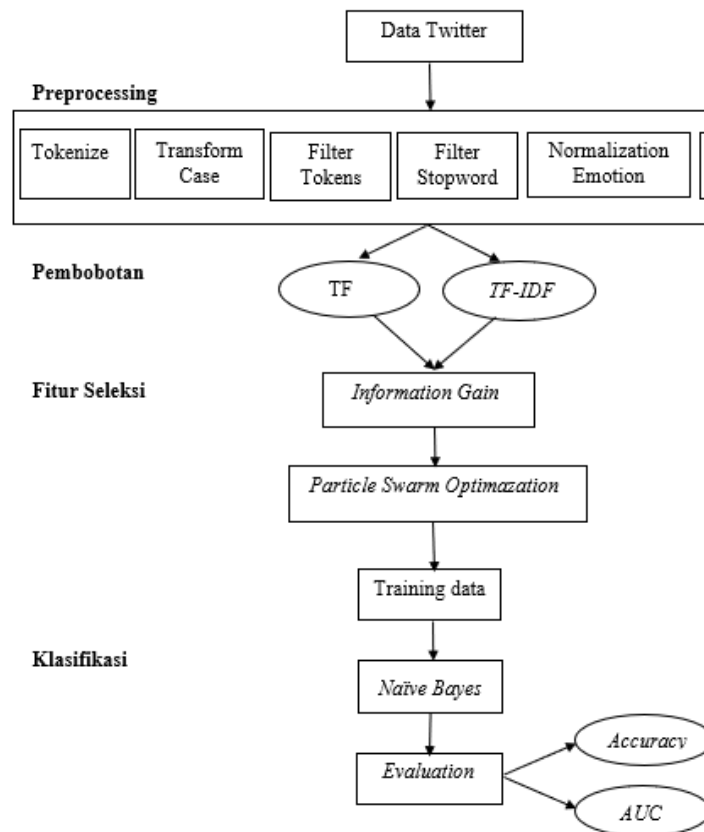
Sebenarnya perkembangan kendaraan listrik bukan merupakan teknologi baru dalam bidang transportasi, tetapi teknologi kendaraan listrik ini sudah ditemukan dan banyak digunakan sejak abad ke 18. Perkembangan selanjutnya, pada abad ke 19, mobil listrik mulai pertama kali diproduksi secara massal dan komersial oleh General Motor (Simbolon et al., 2022). Kehadiran kendaraan listrik ini menimbulkan pro kontra di kalangan masyarakat Indonesia dan jadi bahan perbincangan di media sosial salah satunya yaitu Twitter.

Twitter merupakan salah satu media sosial yang digunakan untuk mengungkapkan ulasan tentang berbagai topik atau isu yang sedang tren melalui kolom tweet (Aryanti et al., 2022). Masyarakat mengungkapkan pada twitter tersebut bahwa dengan diadakannya kendaraan listrik harus disiapkan infrastruktur yang sangat memadai. Dari banyaknya respon masyarakat tersebut maka peneliti melakukan analisis sentimen masyarakat tentang kendaraan listrik dengan mengambil data pada sosial media Twitter.

Analisis sentimen merupakan suatu teknik menghasilkan data berupa teks guna untuk mendapatkan informasi mengenai sentimen bernilai positif, negatif dan netral. Analisis sentimen diberikan oleh user internet pada media sosial untuk memberikan suatu pendapat dan penilaian pribadi (Sari & Wibowo, 2019).

2. METODE PENELITIAN

Text mining merupakan teknik analisis data yang digunakan untuk mengekstraksi informasi dan pengetahuan dari data teks (Olive et al., 2020). Metode penelitian text mining melibatkan beberapa tahapan yang harus dilakukan (Baydhowi et al., 2019). Berikut langkah-langkah yang dilakukan pada penelitian ini :



Sumber : Hasil Penelitian (2023)

Gambar 1. Langkah-Langkah Penelitian

1. Pengumpulan Data

Teks yang akan digunakan dalam penelitian berasal dari data Tweet, kemudian dilakukan dengan cara crawling data ke API Twitter menggunakan Rapid Miner 10. Rapid miner adalah software untuk melakukan analisis terhadap suatu data seperti text mining ataupun data mining (Que et al., 2020).

2. Preprocessing data

Data teks mentah kemudian harus diproses untuk menghilangkan informasi yang tidak relevan dan mengubahnya ke dalam bentuk yang dapat diproses oleh mesin, seperti penghapusan karakter khusus, tokenisasi, penghapusan stop words, stemming dan normalisasi (Deni Gunawan et al, 2022)

3. Metode Klasifikasi

Klasifikasi merupakan proses pencarian dari beberapa model yang menghasilkan dan membedakan kelas data dengan tujuan model tersebut dapat digunakan untuk memprediksi kelas dari suatu objek yang belum diketahui kelasnya (Soepomo, 2014). Setelah data melewati Tahap Preprocessing, tahap selanjutnya adalah pemilihan metode analisis yang sesuai dengan tujuan penelitian yaitu klasifikasi. Metode yang diusulkan adalah Klasifikasi menggunakan algoritma Naïve Bayes dengan seleksi fitur menggunakan particle swarm optimization dan Information Gain.

4. Implementasi dan evaluasi model

Setelah metode analisis dipilih, model yang digunakan dalam penelitian diimplementasikan dan dievaluasi Aplikasi Rapid Miner 10. Evaluasi model dilakukan dengan menggunakan metrik evaluasi seperti akurasi, precision, recall dan AUC.

3. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Penelitian ini menggunakan 600 data tweet mengenai opini publik terhadap kendaraan listrik. Setelah dihilangkan data ganda jumlah data menjadi 486. Terbagi kedalam dua yaitu *tweet* positif dan *tweet* negatif.

B. Preprocessing Data

Preprocessing merupakan proses untuk membersihkan, mengubah dan memformat teks mentah menjadi bentuk yang dapat diproses dan dianalisis lebih (Deni Gunawan et al, 2022). Berikut adalah tahapan preprocessing pada text mining menurut beberapa sumber:

1. Cleansing

Penghapusan karakter khusus dan tanda baca seperti simbol, emotikon, titik, koma, dan tanda tanya. Didalam proses cleansing juga ada proses @Anotation Removal.

Tabel 1. *Cleansing*

Text Sebelum Cleansing	Text Setelah cleansing
@Ripalmaa @sergap_motor @LenIndustri Setuju banget, ini searah seperti keinginan Pak Jokowi, supaya bisa melakukan percepatan pemakaian transportasi Dinas listrik.	Setuju banget ini searah seperti keinginan Pak jokowi supaya bisa melakukan percepatan pemakaian transportasi Dinas listrik

Sumber : Hasil Penelitian (2023)

2. Tokenize

Proses memecah teks menjadi kata-kata atau token. Tokenisasi dapat dilakukan dengan cara memisahkan setiap kata menggunakan spasi, tanda baca, atau simbol tertentu (Julianto, 2022).

Tabel 2. *Tokenize*

Text Sebelum <i>Tokenize</i>	Text Setelah <i>Tokenize</i>
Setuju banget ini searah seperti keinginan Pak jokowi supaya bisa melakukan percepatan pemakaian transportasi Dinas listrik	['Setuju' 'banget' 'ini' 'searah' 'seperti' 'keinginan' 'Pak' 'jokowi' 'supaya' 'bisa' 'melakukan' 'percepatan' 'pemakaian' 'transportasi' 'Dinas' 'listrik']

Sumber : Hasil Penelitian (2023)

3. Transform Case

Mengubah seluruh kata menjadi lowercase (huruf kecil) (Prajamukti et al., 2022).

Tabel 3. *Transform Case*

Text Sebelum <i>Transform Case</i>	Text Setelah <i>Transform Case</i>
Setuju banget ini searah seperti keinginan Pak jokowi supaya bisa melakukan percepatan pemakaian transportasi Dinas listrik	['setuju' 'banget' 'ini' 'searah' 'seperti' 'keinginan' 'pak' 'jokowi' 'supaya' 'bisa' 'melakukan' 'percepatan' 'pemakaian' 'transportasi' 'dinas' 'listrik']

Sumber : Hasil Penelitian (2023)

4. Filter stopword

Penghapusan stop words atau kata-kata umum seperti "dan", "atau", dan "tidak" "ini" "seperti" yang tidak memberikan informasi yang signifikan dalam analisis teks (Ramadhani & Wahyudin, 2022).

Tabel 4. *Filter stopword*

Text Sebelum <i>Transform Case</i>	Text Setelah <i>Transform Case</i>
['setuju' 'banget' 'ini' 'searah' 'seperti' 'keinginan' 'pak' 'jokowi' 'supaya' 'bisa' 'melakukan' 'percepatan' 'pemakaian' 'transportasi' 'dinas' 'listrik']	['setuju' 'banget' 'searah' 'jokowi' 'percepatan' 'pemakaian' 'transportasi' 'dinas' 'listrik']

Sumber : Hasil Penelitian (2023)

5. Indonesian Stemming

Proses untuk mengubah kata-kata ke bentuk dasar. Stemming menghapus awalan atau akhiran kata, sementara lemmatisasi mengubah kata ke bentuk dasarnya (Afdal & Elita, 2022).

Tabel 5. Indonesian Stemming

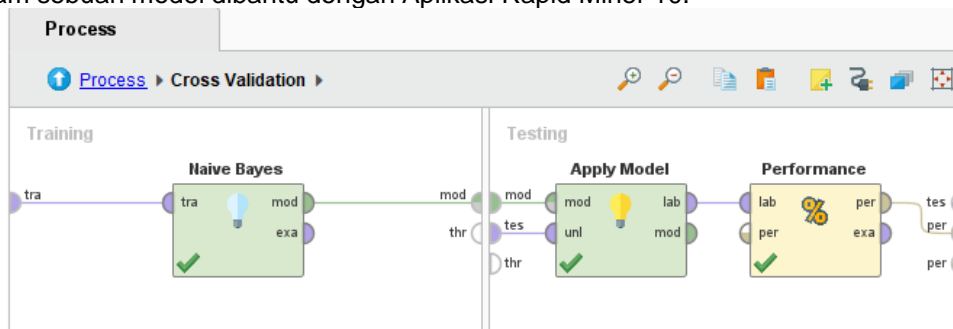
Text Sebelum Indonesian stemming	Text Setelah Indonesian stemming
['setuju' 'banget' 'searah' 'jokowi' 'percepatan' 'pemakaian' 'transportasi' 'dinas' 'listrik']	['setuju' 'banget' 'arah' 'jokowi' 'cepat' 'pakai' 'transportasi' 'dinas' 'listrik']

Sumber : Hasil Penelitian (2023)

C. Proses Klasifikasi

1. Persiapan Klasifikasi

Setelah dilakukan *preprocessing* maka tahap selanjutnya yaitu proses klasifikasi. Penelitian ini menggunakan metode klasifikasi *Naive Bayes* dengan *data training* sebanyak 486 *tweet* tentang opini masyarakat terhadap kendaraan listrik. Berdasarkan kalimat yang ada di dataset *tweet*, data dikelompokkan menjadi dua kelas yaitu label positif dan negatif. Sebanyak 225 *tweet* diberikan label positif dan sisanya 261 *tweet* diberikan label negatif. Selanjutnya data yang sudah diberi label tersebut diujikan dalam sebuah model dibantu dengan Aplikasi Rapid Miner 10.

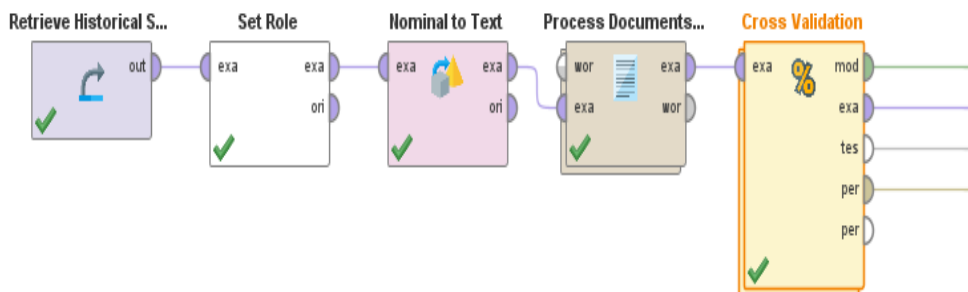


Sumber : Hasil Penelitian (2023)

Gambar 2. Model Klasifikasi Naive Bayes

2. Pembobotan dan seleksi fitur

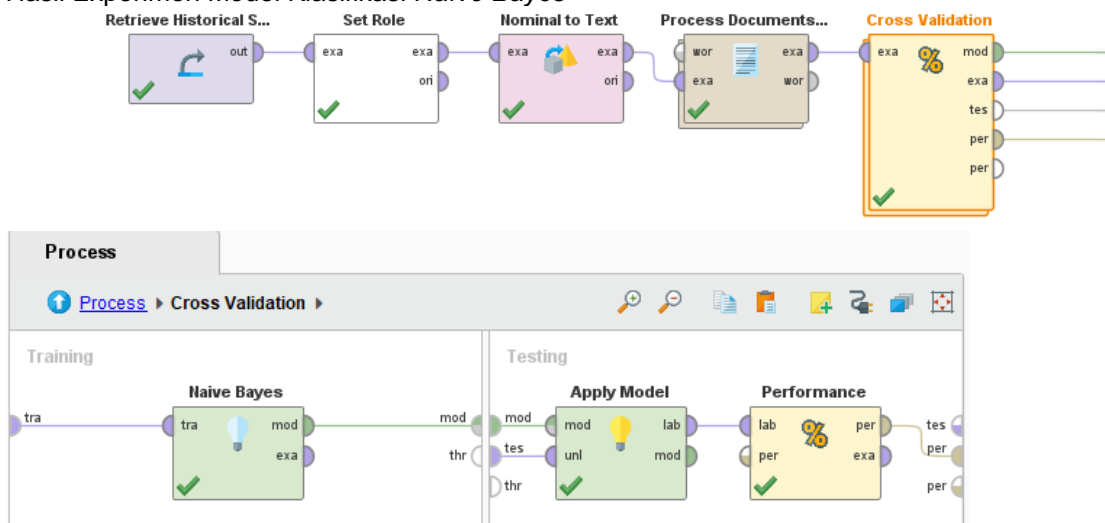
Metode yang digunakan untuk pembobotan yaitu *Term Frequency Invers Document Frequency* (TF-IDF) dan *K-Fold Cross Validation*. Selanjutnya pengujian berikutnya akan menggunakan Pembobotan seleksi fitur *Information Gain* dan *Particle Swarm Optimization* diharapkan tingkat akurasi klasifikasinya meningkat. Model Pembobotan TF-IDF dapat dilihat pada Gambar 3.



Sumber : Hasil Penelitian (2023)

Gambar 3. Model Pembobotan Term Frequency Invers Document Frequency (TF-IDF) dan K-Fold Cross Validation

3. Hasil Experimen Model Klasifikasi *Naïve Bayes*

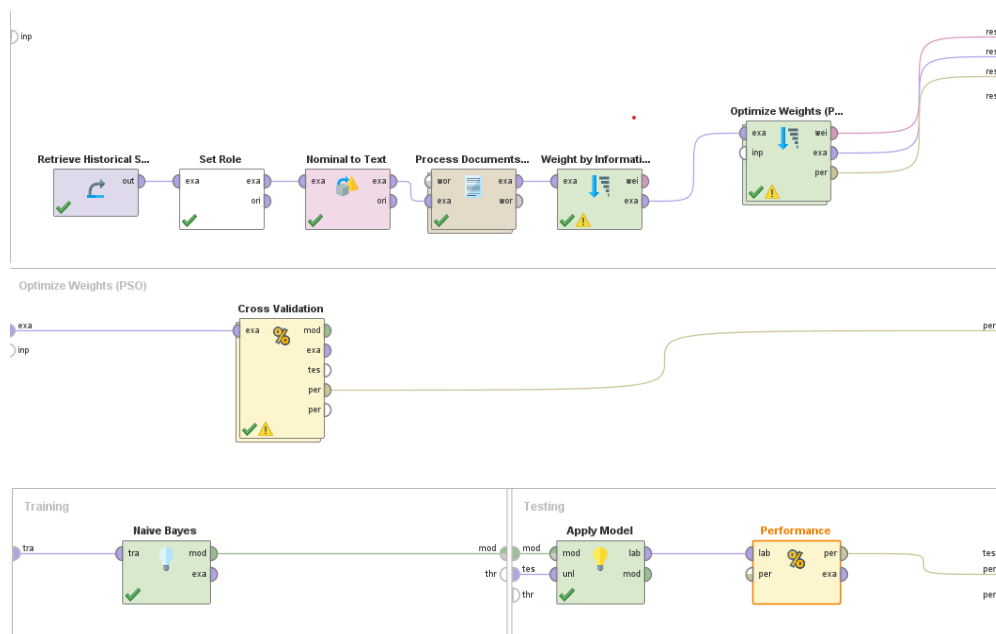


Sumber : Hasil Penelitian (2023)

Gambar 4. Model *Naïve Bayes*

Dari pengujian menggunakan Algoritma *Naïve Bayes* ini dengan nilai $K=10$ pada *K-Fold Cross Validation* menghasilkan *Accuracy* 79,43% dengan *AUC (the area under curve)* 0,639.

4. Hasil Experimen Model Klasifikasi *Naïve Bayes* + *Information Gain* + *PSO*



Sumber : Hasil Penelitian (2023)

Gambar 5. Model *Naïve Bayes* + *Information Gain* + *PSO*

Dari pengujian menggunakan Algoritma *naïve bayes* ini dengan nilai $K=10$ pada *K Fold Cross Validation* menghasilkan *Accuracy* 84,54% dengan *AUC (the area under curve)* 0,729. terjadi peningkatan akurasi dari pengujian sebelumnya yang tanpa Fitur seleksi *Information Gain* dan *PSO*.

D. Evaluasi

1. Accuracy

a. Algoritma Naïve Bayes

Tabel 6. Model Confusion Matrix Naïve Bayes

	true Negative	true Positive	class precision
pred. Negative	216	36	85.71%
pred. Positive	64	170	72.65%
class recall	77.14%	82.52%	

Sumber : Hasil Penelitian (2023)

Tabel 6 dijelaskan bahwa, 216 data diklasifikasikan Negative sesuai dengan prediksi yang dilakukan dengan algoritma Naive Bayes, 36 data diprediksi Negative tetapi hasilnya Positive. 170 data diklasifikasikan Positive dan 64 data diprediksi Positive ternyata Negative. Tingkat akurasi dengan algoritma Naive Bayes didapatkan hasil sebesar 79,42%.

b. Algoritma Naïve Bayes menggunakan seleksi fitur *Information Gain* + PSO

Tabel 7. Model Confusion Matrix Naïve Bayes + Information Gain + PSO

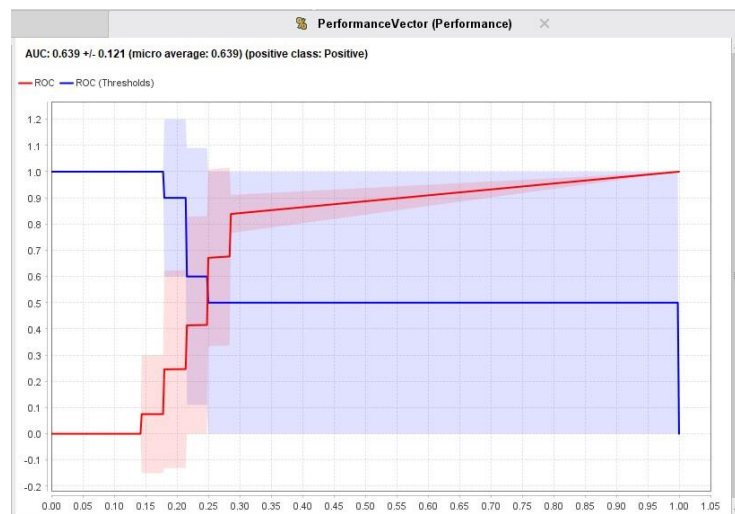
	true Negative	true Positive	class precision
pred. Negative	233	28	89.27%
pred. Positive	47	178	79.11%
class recall	83.21%	86.41%	

Sumber : Hasil Penelitian (2023)

Tabel 7 dijelaskan bahwa, 233 data diklasifikasikan Negative sesuai dengan prediksi yang dilakukan dengan algoritma Naive Bayes dengan seleksi fitur Information Gain dan PSO, 28 data diprediksi Negative tetapi hasilnya Positive. 178 data diklasifikasikan Positive dan 47 data diprediksi Positive ternyata Negative. Tingkat akurasi dengan algoritma Naive Bayes dengan seleksi fitur Information Gain dan PSO didapatkan hasil sebesar 84,57%.

2. AUC

- a. Gambar *confusion Matrix Naïve Bayes* pada Tabel 6 dapat divisualisasikan kedalam kurva ROC berikut ini :

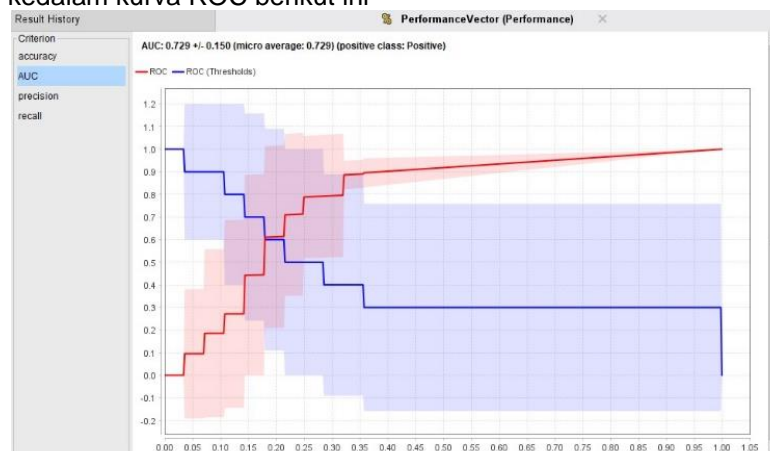


Sumber : Hasil Penelitian (2023)

Gambar 6. Kurva ROC Naïve Bayes

Gambar 6 terdapat grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0,639 dimana hasilnya dapat dinyatakan sebagai *Poor Classification*.

- b. Gambar *confusion Matrix Naïve Bayes* menggunakan seleksi fitur *Information Gain+PSO* pada Tabel 7 dapat divisualisasikan kedalam kurva ROC berikut ini



Sumber : Hasil Penelitian (2023)

Gambar 7. Kurva ROC Naïve Bayes + *Information Gain+PSO*

Gambar 7 terdapat grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0,729 dimana hasilnya dapat dinyatakan sebagai *Fair Classification*.

4. KESIMPULAN

Dalam penelitian ini penulis melakukan klasifikasikan teks terhadap sentimen dari opini masyarakat tentang kendaraan listrik melalui media sosial salah satunya yaitu Twitter. Penelitian ini menerapkan klasifikasi algoritma *Naive Bayes* dengan seleksi fitur *Information Gain* dan *Particle Swarm Optimazation*. Hasil pengujian dengan menggunakan algoritma *Naive Bayes* mendapatkan nilai akurasi 79,43% dan nilai AUC 0,639. Sedangkan setelah menggunakan selelsi fitur *Information Gain* dan *Particle Swarm Optimazation* adanya peningkatan nilai akurasi yang dihasilkan menjadi 84,54% dan nilai AUC 0,729. Sehingga dapat direkomendasikan bahwa algoritma naive bayes dengan seleksi fitur *Information Gain* dan

Particle Swarm Optimazation sebagai algoritma klasifikasi yang baik. Pada penelitian ini juga menggunakan *framework gataframework* dalam membantu pada tahapan *preprocessing text mining* dan data diolah menggunakan aplikasi rapidminer.

REFERENSI

- Afdal, M., & Elita, L. R. (2022). *PENERAPAN TEXT MINING PADA APLIKASI TOKOPEDIA*. 8(1), 78–87.
- Agustian, A., Tukiro, T., & Nurapriani, F. (2022). Penerapan Analisis Sentimen Dan Naive Bayes Terhadap Opini Penggunaan Kendaraan Listrik Di Twitter. *Jurnal TIKTA*, 7(3), 243–249. <https://doi.org/10.51179/tika.v7i3.1550>
- Aryanti, R., Atang Saepudin, Eka Fitriani, Rifky Permana, & Dede Firmansyah Saefudin. (2022). Komparasi Algoritma Naive Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Pengguna Busway. *Jurnal Teknik Komputer AMIK BSI*, 8(2), 174–180. <https://doi.org/10.31294/jtk.v4i2>
- Baydhowi, M., Apriliah, W., & Kurniawan, I. (2019). Klasifikasi Tweet Berdasarkan Keterkaitan Tweet Terhadap Topik Tertentu Pada Twitter Menggunakan Naive Bayes. *Information System for Educators and Professionals*, 4(1), 95–103. <http://ejournal-binainsani.ac.id/index.php/ISBl/article/view/1244>
- Deni Gunawan et al. (2022). Komparasi Algoritma Support Vector Machine Dan Naive Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023. *Jurnal Teknik Komputer AMIK BSI*, 8(2), 174–180. <https://doi.org/10.31294/jtk.v4i2>
- Hayuningtyas, R. Y., & Sari, R. (2019). Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata Tmii Menggunakan Naive Bayes Dan Pso. *Jurnal Techno Nusa Mandiri*, 16(1), 37–42. <https://doi.org/10.33480/techno.v16i1.115>
- Julianto, I. T. (2022). *Analisis Sentimen Terhadap Sistem Informasi Akademik Mahasiswa Institut Teknologi Garut*. 458–465.
- Olive, I., Putra, D., Prilianti, K. R., Lucky, P., & Irawan, T. (2020). *IMPLEMENTASI TEXT MINING UNTUK ANALISIS LAYANAN TRANSPORTASI ONLINE DENGAN ANALISIS FAKTOR*. 8(2).
- Prajamukti, R., Jayanta, & Santoni, M. M. (2022). *KLASIFIKASI DAN ANALISIS SENTIMEN PADA DATA TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES (STUDI KASUS : TIMNAS 1 Pendahuluan 2 Tinjauan Pustaka 3 Metode Penelitian*. 28–29.
- Que, V. K. S., Iriani, A., & Purnomo, H. D. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 9(2), 162–170. <https://doi.org/10.22146/jnteti.v9i2.102>
- Ramadhani, S. H., & Wahyudin, M. I. (2022). *Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naive Bayes dan K-NN*. 6(4).
- Santoso, A., Nugroho, A., & Sunge, A. S. (2022). Analisis Sentimen Tentang Mobil Listrik Dengan Metode Support Vector Machine Dan Feature Selection Particle Swarm Optimization. *Journal of Practical Computer Science*, 2(1), 24–31. <https://doi.org/10.37366/jpcs.v2i1.1084>
- Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naive Bayes Classifier Berbasis Konversi Ikon Emosi. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 10(2), 681–686. <https://jurnal.umk.ac.id/index.php/simet/article/view/3487>
- Simbolon, A. M., Rusli, B., & -, C.-. (2022). Kebijakan Kendaraan Listrik dalam Perspektif Pasar dan Infrastruktur: Studi Reviu Komparasi Bilateral Korea Selatan dan Indonesia. *Jurnal Penelitian Transportasi Darat*, 24(2), 83–91. <https://doi.org/10.25104/jptd.v24i2.1943>
- Soepomo, P. (2014). *PENERAPAN TEXT MINING PADA SISTEM KLASIFIKASI*. 2, 73–83.