

Optimasi Naïve Bayes menggunakan Seleksi Fitur Forward Selection untuk Analisis Sentimen Kendaraan Listrik

Salman Alfarizi ¹, Deni Gunawan ², Hasan Basri ³, Alif Rizqi Mulyawan ⁴, Nurul Ichsan ⁵

^{1,2,3,4,5} Sistem Informasi; Universitas Bina Sarana Informatika

^{1*}e-mail: salman.slz@bsi.ac.id

²e-mail: deni.dee@bsi.ac.id

³e-mail: hasan.hhi@bsi.ac.id,

⁴e-mail: alif.aqm@bsi.ac.id

⁵e-mail: nurul.nrc@bsi.ac.id

Diterima	Direvisi	Disetujui
14-06-2024	22-07-2024	13-08-2024..

Abstrak - Isu pemanasan global menjadi perhatian utama para pemimpin di beberapa negara termasuk Indonesia. Salah satu kontribusi untuk mengurangi emisi karbon yang berakibat pada pemanasan global yaitu dengan menggulirkan kebijakan kendaraan listrik. Pro kontra kebijakan kendaraan Listrik di Indonesia telah merambah ke berbagai ruang diskusi publik termasuk media social X (Twitter). Berdasarkan hal tersebut penelitian ini bermaksud untuk menganalisis sentimen Masyarakat pada media social X tentang kebijakan kendaraan Listrik dilakukan klasifikasi dengan memanfaatkan algoritma naïve bayes berbasis forward selection. Berdasarkan pengujian model klasifikasi naïve bayes ditambahakan forward selection dihasilkan accuracy sebesar 82,11% dengan nilai AUC 0,801. Ada kenaikan nilai accuracy sebesar 2,67% dibandingkan tanpa menggunakan seleksi fitur forward selection yang hanya mencapai accuracy dikisaran 79,43 % dengan nilai AUC 0,639. Peneliti menarik Kesimpulan bahwa penggunaan seleksi fitur menggunakan forward selection dapat berkontribusi meningkatkan algoritma naïve bayes dalam klasifikasi sentimen pada kendaraan yang bertenaga listrik.

Kata Kunci: *Naïve bayes, Analisis sentimen, Forward Selection*

Abstract - The issue of global warming has become a major concern for leaders in several countries, including Indonesia. One of the contributions to reducing carbon emissions that lead to global warming is by rolling out electric vehicle policies. The pros and cons of electric vehicle policies in Indonesia have spread to various public discussion spaces, including social media platform X (Twitter). Based on this, this study aims to analyze public sentiment on social media platform X regarding electric vehicle policies by classifying using the naïve bayes algorithm based on forward selection. Based on the testing of the naïve bayes classification model with added forward selection, an accuracy of 82.11% was achieved with an AUC value of 0.801. There was an increase in accuracy by 2.67% compared to not using forward selection feature selection, which only reached an accuracy of around 79.43% with an AUC value of 0.639. The researchers concluded that the use of feature selection using forward selection can contribute to improving the naïve bayes algorithm in sentiment classification on electric-powered vehicles.

Keywords: *Naïve Bayes, Sentiment analysis, Forward Selection*

PENDAHULUAN

Meningkatnya penggunaan kendaraan berbahan bakar minyak yang berdampak pada pencemaran udara dan iklim global mendorong pemerintah membuat kebijakan untuk beralih menjadi kendaraan berbasis energi listrik (Riyadi et al., 2021). Meski Kendaraan listrik sudah menjadi isu dunia dalam menekan pemanasan global tetapi belum menjadi alternatif pilihan utama. Dampak kebijakan ini

mendapat respon setuju dan tidak setuju dari masyarakat umum yang disebut sebagai sentimen. Diskusi tentang pro kontra tentang kendaraan listrik bisa dijumpai di beberapa platform media sosial diantaranya adalah media sosial X (Twitter) yang dapat diolah menggunakan teknik analisis sentimen (Prawinata et al., 2024). Analisis sentimen itu sendiri merupakan teknik menganalisis opini publik berupa pernyataan tertulis dalam bentuk teks terhadap entitas



berupa opini positif maupun negatif maupun netral menggunakan algoritma klasifikasi (Dyah Anggita & Abdulloh, 2023).

Beberapa hasil penelitian sebelumnya mengenai analisis sentimen telah banyak dilakukan dengan penggunaan seleksi fitur untuk memaksimalkan kinerja algoritma. Diantaranya Penelitian yang dilakukan 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*” menghasilkan 94,25 % respon positif dari pengguna twitter dengan hasil klasifikasi menggunakan Algoritma SVM mencapai *accuracy* sebesar 82,51 % dan setelah ditambahkan fitur seleksi PSO nilai akurasi meningkat menjadi 86,07 % dengan nilai AUC 0,862 % (Santoso et al., 2022).

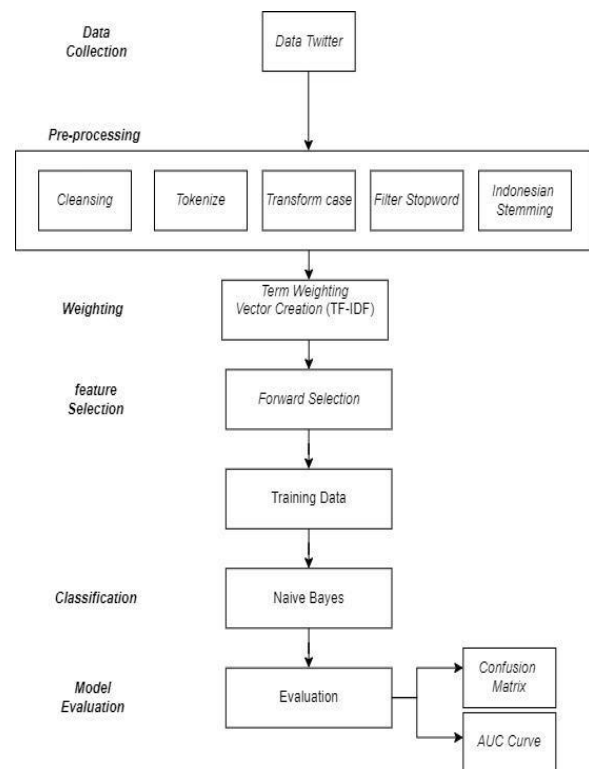
Penelitian yang dilakukan oleh Arif Bijaksana Putra Negara, Hafiz Muhardi dan Indira Melinda Putri dengan judul “Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes Dan Seleksi Fitur Information Gain” menghasilkan nilai rata-rata dari akurasi, presisi dan recall bertambah sebesar 0,81 setelah ditambah Information Gain (Bijaksana et al., 2020).

Penelitian berikutnya dilakukan oleh Juen Ling, I Putu Eka N. Kencana, Tjokorda Bagus Oka dengan judul “Analisis Sentimen Menggunakan Metode Naive Bayes Classifier Dengan Seleksi Fitur Chi Square” bahwa algoritma seleksi fitur Chi Square dapat digunakan untuk seleksi fitur dan membantu meningkatkan akurasi sebesar 83% menggunakan Naive Bayes menggunakan program Java(Ling et al., 2014).

Berdasarkan paparan penelitian terdahulu menggunakan algoritma seleksi fitur terbukti meningkatkan kinerja algoritma. Pada penelitian ini akan menguji salah satu algoritma seleksi fitur yaitu Forward Selection untuk mengeliminasi beberapa atribut yang tidak relevan yang nantinya akan mengetahui apakah berpengaruh terhadap kinerja algoritma Naive Bayes atau tidak sama sekali.

METODE PENELITIAN

Berikut adalah model tahapan *text mining* yang diusulkan pada penelitian ini sesuai standar metode klasifikasi teks. Dengan tambahan tahapan proses seleksi fitur (Faisal et al., 2022):



Sumber: Hasil Penelitian (2024)

Gambar 1. Model Alur Klasifikasi *Text Mining*

Penjelasan tahapan-tahapan Model klasifikasi *Text Mining* pada Gambar 1 adalah sebagai berikut:

Data Collection

Merupakan proses pengambilan data teks berupa *tweet* dari media sosial X (*Twitter*). *Crawling data* dengan memanfaatkan API Twitter menggunakan aplikasi Rapid Miner 10. data disimpan ke dalam bentuk *spreadsheet*.

Labelling

Proses pemberian label pada data *tweet* yang telah dikumpulkan. Label ini berupa kategori atau kelas yang akan diprediksi seperti positif dan negatif. Pemberian label pada seluruh data dilakukan secara manual.(Khomsah & Aribowo, 2017)

Tabel 1. Merupakan contoh komentar yang diberikan status

Text	Status
Luhut Bilang Kendaraan Listrik Kurangi Konsumsi BBM 70 Miliar Liter Per Tahun Selengkapnya Klik Tautan Berikut ini #Jakarta #Luhut	Positive

#Kendaraanlistrik #bbm https://t.co/QyK11vOI5D	
Mendukung produsen mobil listrik masuk ke Indonesia#produsenmobil#mobillistri k#Indonesiaku https://t.co/kuB0vy4pBa	Positive
Kalo di kota besar mungkin ini keren sih. Tapi kalo didesa-desa mungkin gak gw ajurkan sih. Pengisian dayanya masih belum memadai. #motorlistrik #motorbike https://t.co/VX2nA8eGXv	Negative
Pemerintah harusnya mengalokasikan subsidi untuk konversi ketimbang untuk pembelian kendaraan listrik baru. #kumparanBISNIS #bisnisupdate #kendaraanlistrik https://t.co/cRzLVM2IDX	Negative

Sumber: *Dataset* Twitter (2024)

Pre-processing

Merupakan tahapan pembersihan data sebelum data tersebut diproses ke tahapan selanjutnya. Tahapannya meliputi teknik *cleansing*, *tokenize*, *transform case*, *filter stopword*, *Indonesian stemming*. (Khomsah & Aribowo, 2017)

a. Cleansing

Merupakan proses untuk pembersihan teks pada beberapa elemen yang tidak diperlukan seperti menghapus tanda baca diantaranya titik koma dan tanda seru. Kemudian menghapus URL dan karakter khusus seperti notasi @, \$, #.

b. Tokenize

Proses merubah kalimat atau teks menjadi frasa atau disebut token. Contohnya seperti kalimat “Setuju dengan kendaraan listrik” dirubah menjadi [“Setuju”, “dengan”, “kendaraan”, “listrik”]

c. Transform case

Pada tahapan ini teks yang sudah dijadikan token dirubah hurufnya menjadi huruf kecil atau *lowercase*. Hal ini dilakukan agar tidak terjadi perbedaan antara teks seperti “Listrik” dengan “listrik” padahal berasal dari kata yang sama.

d. Filter stopword

Merupakan Proses menghapus kata dengan frekuensi kemunculannya sering tetapi dan tidak memiliki arti seperti “dan”, “atau”.

e. Indonesian stemming

Proses ini adalah mengubah kata yang berbasis indonesia menjadi kata dasarnya seperti kata “mengalokasikan” menjadi kata dasar “alokasi”

Weighting

Memberikan bobot pada fitur-fitur yang dipilih berdasarkan kepentingannya. TF-IDF merupakan gabungan dari dua metrik yaitu *Term Frequency* dan *Inverse Document Frequency* merupakan metode *weighting* dimana frekuensi kata dalam dokumen dan kebalikannya dalam seluruh korpus berguna dalam penentuan bobot. *Term Frequency* (TF) mengukur seberapa sering sebuah kata muncul pada dokumen sedangkan *Inverse Document Frequency* (IDF) digunakan dalam mengukur seberapa penting sebuah kata. Kata yang frekuensi kemunculannya sering akan dianggap tidak penting maka bobotnya 0.(Dharmawan et al., 2020)

Feature Selection

Proses memilih fitur-fitur yang relevan untuk selanjutnya digunakan pada model klasifikasi. Fitur yang dimaksud berupa kata atau frasa. Seleksi fitur yang digunakan yaitu *Forward Selection*. Metode *Forward Selection* merupakan algoritma yang berguna untuk mencari variabel (fitur) yang relevan dari banyaknya variabel yang tersedia sebagai prediktor (Bagus et al., 2023). Langkah algoritma *Forward Selection* diawali dengan model yang kosong yaitu model yang tidak memiliki fitur prediktor kemudian tambahkan satu persatu fitur ke dalam model, pilih fitur terbaik, ulangi proses langkah kedua dan ketiga sampai tidak ada lagi fitur atau variabel yang diprediksi akan menghasilkan akurasi tinggi (Fanani, 2020).

Classification

Pada tahapan ini merupakan proses membangun model klasifikasi menggunakan algoritma *machine learning*. Algoritma yang akan digunakan dalam klasifikasi ini yaitu *Naive Bayes Classifier*. Algoritma *Naive Bayes* merupakan sebuah metode klasifikasi dalam *mechine learning* yang menghitung berdasarkan probabilitas tentang satu kejadian berdasar pada informasi kejadian yang lain. Keuntungan dari penggunaan algoritma *naive bayes* diantaranya bisa mengklasifikasi dengan

dataset yang sedikit (Felicia Watratan et al., 2020).
Persamaan naive bayes adalah sebagai berikut:

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)}$$

Dimana :

A merupakan variabel data yang belum diketahui kelasnya

B yaitu hipotesis data merupakan suatu kelas yang spesifik

$P(A | B)$ merupakan probabilitas A berdasarkan kondisi hipotesis B

$P(B | A)$ merupakan probabilitas B berdasarkan kondisi hipotesis A

$P(A)$ merupakan probabilitas awal dari kondisi A

$P(B)$ merupakan probabilitas awal dari kondisi B

Model Evaluation

Tahap ini proses mengevaluasi model klasifikasi yang digunakan. Evaluasinya menggunakan *Confusion Matrix* dan *AUC Curve*.

HASIL DAN PEMBAHASAN

A. Data Collection (Pengumpulan Data)

Pengambilan data memanfaatkan API Twitter menggunakan kata kunci kendaraan listrik dan juga hastag #mobillistrik #motorlistrik #subsidikendaraanlistrik. Data yang berhasil terkumpul sekitar 680 tweet dan setelah dibersihkan dari data yang ganda jumlah data tweet yang bisa digunakan menjadi 486 data.

B. Labelling (Pemberian Label)

Proses pemberian label dilakukan secara manual yaitu memberikan status “positif” maupun “negatif” pada data *tweet*. Proses ini membutuhkan waktu serta ketelitian dalam memberikan label suatu text berdasarkan opini yang terkandung di dalamnya.

C. Pre-Processing (Pembersihan Data)

1. Cleansing

Proses menghapus tanda baca diantaranya titik dua koma dan tanda seru. Kemudian menghapus URL dan karakter khusus seperti notasi @, \$, #. Kemudian menghilangkan juga URL.

Tabel 2. *Cleansing*

Teks Sebelum <i>Cleansing</i>		Teks Setelah <i>Cleansing</i>	
Presiden menyatakan ekosistem kendaraan bermotor listrik yang terintegrasi dan optimal bisa menjadikan Indonesia sebagai negara maju. Selengkapnya https://t.co/VWycRJikvs	Jokowi	Presiden menyatakan ekosistem kendaraan bermotor listrik yang terintegrasi dan optimal bisa menjadikan indonesia sebagai negara maju selengkapnya Jokowi Presidenjokowi Listrik MotorListrik MobilListrik Electricvehicle Baterai	

Sumber : Hasil Penelitian (2024)

2. Tokenize

Proses ini dilakukan untuk merubah kalimat atau teks menjadi frasa atau disebut token. Pemisahan bisa dilakukan dengan tanda kutip.

Tabel 3. *Tokenize*

Teks Sebelum <i>Tokenize</i>		Teks Setelah <i>Tokenize</i>	
Presiden menyatakan penciptaan ekosistem kendaraan bermotor listrik yang terintegrasi dan optimal bisa menjadikan indonesia sebagai negara maju selengkapnya Jokowi Presidenjokowi Listrik MotorListrik MobilListrik Electricvehicle Baterai	Jokowi	[“Presiden” “menyatakan” “penciptaan” “kendaraan” “listrik” “yang” “terintegrasi” “optimal” “menjadikan” “sebagai” “maju” “Jokowi” “Presidenjokowi” “Listrik” “MotorListrik” “MobilListrik” “Electricvehicle” “Baterai”]	“Jokowi” “ekosistem” “bermotor” “dan” “bisa” “indonesia” “negara” “selengkapnya”

Sumber : Hasil Penelitian (2024)

3. Transform case

Pada tahapan ini teks yang sudah dijadikan token dirubah hurufnya menjadi huruf kecil atau *lowercase*. Dimaksudkan agar kata yang sama tidak dianggap berbeda saat klasifikasi. Seperti pada kata “Kendaraan” dengan “kendaraan”.

Tabel 4. *Transform Case*

Teks Sebelum <i>Transform case</i>	Teks Setelah <i>Transform case</i>
["Presiden"	["presiden"
"Jokowi"	"jokowi"
"menyatakan"	"menyatakan"
"penciptaan"	"penciptaan"
"ekosistem"	"ekosistem"
"kendaraan"	"kendaraan"
"bermotor" "listrik"	"bermotor" "listrik"
"yang"	"yang"
"terintegrasi" "dan"	"terintegrasi" "dan"
"optimal" "bisa"	"optimal" "bisa"
"menjadikan"	"menjadikan"
"indonesia"	"indonesia"
"sebagai" "negara"	"sebagai" "negara"
"maju"	"maju"
"selengkapnya"	"selengkapnya"
"Jokowi"	"jokowi"
"Presidenjokowi"	"presidenjokowi"
"Listrik"	"listrik"
"MotorListrik"	"motorlistrik"
"MobilListrik"	"mobillistik"
"Electricvehicle"	"electricvehicle"
"Baterai"]	"baterai"]

Sumber : Hasil Penelitian (2024)

4. *Filter stopword*

Proses menghapus kata yang sering muncul tetapi tidak memiliki arti seperti “dan”, “atau”.

Tabel 5. *Filter stopword*

Teks Sebelum <i>Filter stopword</i>	Teks Setelah <i>Filter stopword</i>
["presiden"	["presiden"
"jokowi"	"jokowi"
"menyatakan"	"penciptaan"

"penciptaan"	"ekosistem"
"ekosistem"	"kendaraan"
"kendaraan"	"bermotor" "listrik"
"bermotor" "listrik"	"terintegrasi"
"yang"	"optimal" "bisa"
"terintegrasi" "dan"	"menjadikan"
"optimal" "bisa"	"indonesia"
"menjadikan"	"sebagai" "negara"
"indonesia"	"maju"
"sebagai" "negara"	"selengkapnya"
"maju"	"jokowi"
"selengkapnya"	"presidenjokowi"
"jokowi"	"listrik"
"presidenjokowi"	"motorlistrik"
"listrik"	"mobillistik"
"motorlistrik"	"electricvehicle"
"mobillistik"	"baterai"]
"electricvehicle"	
"baterai"]	

Sumber : Hasil Penelitian (2024)

5. *Indonesian stemming*

Proses ini adalah mengubah kata yang berbasis indonesia menjadi kata dasar.

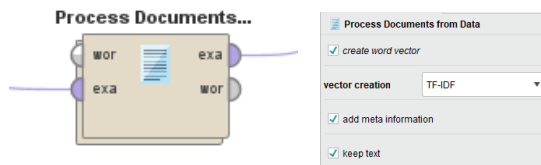
Tabel 6. *Indonesian Stemming*

Teks Sebelum <i>Filter stopword</i>	Teks Setelah <i>Filter stopword</i>
["presiden"	["presiden"
"jokowi"	"jokowi" "cipta"
"penciptaan"	"ekosistem"
"ekosistem"	"kendara" "motor"
"kendaraan"	"listrik" "integrasi"
"bermotor" "listrik"	"optimal" "jadi"
"terintegrasi"	"Indonesia"
"optimal" "bisa"	"negara" "maju"
"menjadikan"	"lengkap"
"indonesia"	"presidenjokowi"
"sebagai" "negara"	"listrik"
"maju"	"motorlistrik"
"selengkapnya"	"mobillistik"
"jokowi"	"electricvehicle"
"presidenjokowi"	"baterai"]
"listrik"	
"motorlistrik"	
"mobillistik"	
"electricvehicle"	
"baterai"]	

Sumber : Hasil Penelitian (2024)

6. Weighting (Pembobotan Atribut)

Setelah *Pre-processing* dilakukan. Langkah selanjutnya yaitu menambahkan modul *Process Document* untuk pembobotan token menggunakan TF-IDF *vector creation*.

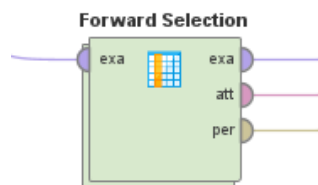


Sumber : Hasil Penelitian (2024)

Gambar 1 Modul *Process Document* TF-IDF

7. Feature Selection (Pemilihan Atribut)

Setelah dilakukannya pembobotan token dengan TF-IDF tahap berikutnya yaitu menambahkan modul *Forward Selection* untuk dilakukan seleksi fitur yang paling relevan diantara beberapa fitur. Setelah dilakukan seleksi fitur maka atribut paling teratas diantaranya **motorlistrik, kendaraanlistrik, cepat, dorong**.

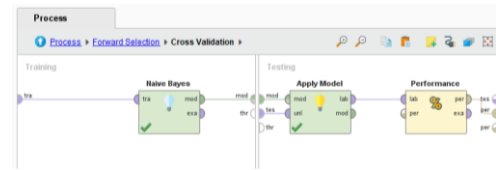


Sumber : Hasil Penelitian (2024)

Gambar 2 Modul *Forward Selection*

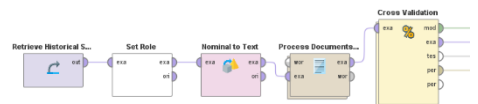
8. Classification (Proses Klasifikasi)

Proses klasifikasi menggunakan algoritma *Naïve Bayes* dengan menggunakan bantuan aplikasi Rapid Miner versi 10. Data latih yang sudah melewati tahap *pre-processing* sebanyak 486 data. Dari data tersebut terdapat data yang berlabel positif berjumlah 206 dan data berlabel negative berjumlah 280 yang selanjutnya data tersebut kita sebut dengan data latih dan akan diujikan pada aplikasi Rapid Miner. Berikut merupakan beberapa model yang dibuat untuk klasifikasi yang dapat dilihat pada gambar 3, gambar 4 dan gambar 5.



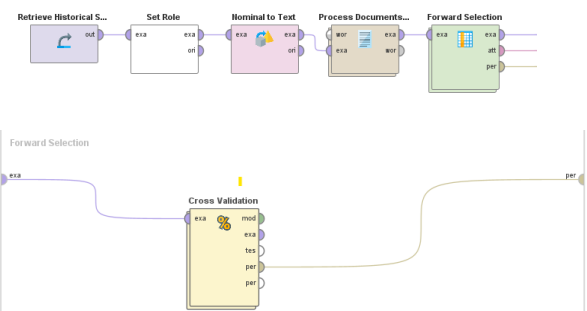
Sumber : Hasil Penelitian (2024)

Gambar 3 Model Klasifikasi Naïve Bayes



Sumber : Hasil Penelitian (2024)

Gambar 4. Model Proses dokumen dan Pembobotan TF-IDF dan K-Fold Cross Validation



Sumber : Hasil Penelitian (2024)

Gambar 5. Model Proses dokumen dan Pembobotan TF-IDF dan *K-Fold Cross Validation* dengan *Forward Selection*

Pengujian algoritma *naïve bayes* dengan K-Fold Cross Validation bernilai K=10 dan pembobotan TF-IDF. Hasil Eksperimen Klasifikasi berdasarkan model pada gambar 3, 4 dan 5.

Tabel 7. Hasil Klasifikasi

Hasil Klasifikasi	Naïve Bayes	Naïve Bayes + Forward Selection
Accuracy	79,43 %	82,10 %

Sumber : Hasil Penelitian (2024)

Hasil klasifikasi menggunakan algoritma *naïve bayes* berdasarkan tabel 7 nilai akurasi jika tanpa seleksi fitur

forward selection yang dihasilkan sebesar 79,43 %. Sedangkan klasifikasi *naïve bayes* dengan *forward selection* sebesar 82,10 % terjadi kenaikan sebesar 2,67 %.

9. Model Evaluation (Evaluasi Model)

A. Accuracy

Tabel 8. Model *Confusion Matrix Naïve Bayes*

Accuracy: 79,43%			
	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 (2024)

Berdasarkan Tabel 8 *Confusion matrix* dari klasifikasi *naïve bayes* dihasilkan *accuracy* sebesar 79,43%. Data *tweet* dengan prediksi negatif dinyatakan *true negative* sebesar 216 dan dinyatakan *true positive* sebesar 36 data. Untuk data yang diprediksi awalnya positif dinyatakan *true positive* sebesar 170 dan dinyatakan *true negative* 64 data.

Tabel 9. Model *Confusion Matrix Naïve Bayes + Forward Selection*

Accuracy: 82,11%			
	True Negative	True Positive	Class precision
pred. Negative	264	71	78,81%
Pred. Positive	16	135	89,40%
Class recall	77,14%	82,52%	

Sumber : Hasil Penelitian (2024)

Berdasarkan Tabel 9 *confusion matrix* klasifikasi *naïve bayes* ditambah seleksi fitur *Forward selection* dihasilkan *accuracy* 82,11%. Terjadi kenaikan akurasi dibanding tanpa seleksi fitur. Data *tweet* yang diprediksi *negative* mendapat hasil *true negative* berjumlah 264 dan 71 data dinyatakan *true positive*. Data yang diprediksi *negative* mengalami kenaikan yang sebelumnya 216. Untuk data yang diprediksi positif 135 dinyatakan *true positive* terjadi penurunan data sebelumnya 170 diprediksi *true positive*.

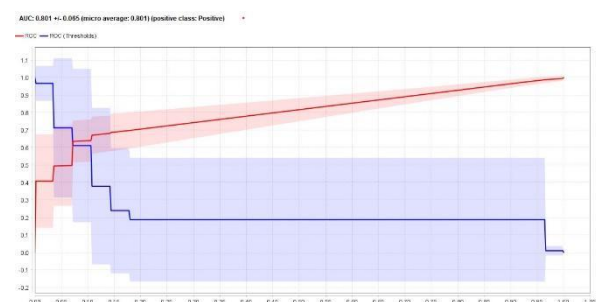
B. AUC (Area Under Curve)



Sumber : Hasil Penelitian (2024)

Gambar 6. AUC (Area Under Curve) yang dihasilkan Klasifikasi *Naïve Bayes*

Berdasarkan gambar 6 yang merupakan kurva ROC (*Receiver Operating Characteristic*) untuk menguji performa model klasifikasi *naïve bayes*. Nilai AUC yang dihasilkan adalah 0,639 dan dikatakan performa algoritma dikategori *poor classification*.



Sumber : Hasil Penelitian (2024)

Gambar 7. AUC yang dihasilkan Klasifikasi *Naïve Bayes + Forward Selection*

Berdasarkan gambar 7 merupakan kurva ROC (*Receiver Operating Characteristic*) untuk menguji

performa model klasifikasi naïve bayes dengan ditambah seleksi fitur *forward selection* terjadi peningkatan nilai AUC (*Area Under Curve*) menjadi 0,801 dan dapat dikatakan performa algoritma dikategori klasifikasi yang *fair*.

KESIMPULAN

Hasil dari klasifikasi menggunakan naïve bayes dengan menggunakan seleksi fitur *forward selection*. Data komentar pada social media X (*twitter*) yang digunakan sebanyak 481 data. Berdasarkan pengujian model klasifikasi naïve bayes dengan ditambahkan seleksi fitur *forward selection* dihasilkan nilai *accuracy* sebesar 82,11% dengan nilai AUC 0,801. Ada kenaikan nilai *accuracy* sebesar 2,67% dibandingkan tanpa menggunakan seleksi fitur *forward selection* hanya dikisaran 79,43 % dengan nilai AUC 0,639. Peneliti menarik kesimpulan bahwa penggunaan seleksi fitur menggunakan *forward selection* dapat meningkatkan kinerja algoritma *naïve bayes* dalam klasifikasi analisis sentimen kendaraan.

REFERENSI

- Bagus, D., Prasetyo, C., Andono, N., & Supriyanto, C. (2023). Metode Naive Bayes Classifier dan Forward Selection Untuk Deteksi Berita Hoaks Bahasa Indonesia. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 7(3), 1541–1550. <https://doi.org/10.30865/mib.v7i3.6459>
- Bijaksana, A., Negara, P., Muhandi, H., & Putri, I. M. (2020). ANALISIS SENTIMEN MASKAPAI PENERBANGAN MENGGUNAKAN METODE NAIVE BAYES DAN SELEKSI FITUR INFORMATION GAIN. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 7(3), 599–606. <https://doi.org/10.25126/jtiik.202071947>
- Dharmawan, L. R., Arwani, I., & Ratnawati, D. E. (2020). Analisis Sentimen pada Sosial Media Twitter Terhadap Layanan Sistem Informasi Akademik Mahasiswa Universitas Brawijaya dengan Metode K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 4(3), 959–965. <http://j-ptiik.ub.ac.id>
- Dyah Anggita, S., & Abdulloh, F. F. (2023). Optimasi Algoritma Support Vector Machine Berbasis PSO Dan Seleksi Fitur Information Gain Pada Analisis Sentimen. *JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST)*, 4(1), 2723–1453. <https://doi.org/10.52158/jacost.524>
- Faisal, M. R., Kartini, D., & Saragih, T. H. (2022). *Belajar Data Science: Text Mining Untuk Pemula I*. Scripta Cendekia. <https://www.researchgate.net/publication/359619425>
- Fanani, M. R. (2020). ALGORITMA NAÏVE BAYES BERBASIS FORWARD SELECTION UNTUK PREDIKSI BIMBINGAN KONSELING SISWA. *Jurnal DISPROTEK*, 11(1), 13–22.
- Felicia Watratan, A., Puspita, A. B., Moeis, D., Informasi, S., & Profesional Makassar, S. (2020). Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia. In *JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST)* (Vol. 1, Issue 1). <http://journal.isas.or.id/index.php/JACOST>
- Khomsah, S., & Aribowo, A. S. (2017). Model Text-Preprocessing Komentar Youtube Dalam Bahasa Indonesia. *JURNAL RESTI*, 1(3), 648–654.
- Ling, J., Putu, I., Kencana, E. N., & Oka, T. B. (2014). ANALISIS SENTIMEN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER DENGAN SELEKSI FITUR CHI SQUARE. *E-Jurnal Matematika*, 3(3), 92–99.
- Prawinata, D. A., Pembangunan, U., Veteran, N., & Timur, J. (2024). Analisis Sentimen Kendaraan Listrik Pada Twitter Menggunakan Metode Long Short Term Memory Ani Dijah Rahajoe I Gede Susrama Mas Diyasa. *SABER: Jurnal Teknik Informatika, Sains Dan Ilmu Komunikasi*, 2(1), 300–313. <https://doi.org/10.59841/saber.v2i1.857>
- Riyadi, A. F., Rahman, F. R., Nofa Pratama, M. A., Khafidli, M. K., & Patria, H. (2021). Pengukuran Sentimen Sosial Terhadap Teknologi Kendaraan Listrik: Bukti Empiris di Indonesia. *EXPERT: Jurnal Manajemen Sistem Informasi Dan Teknologi*, 11(2), 141. <https://doi.org/10.36448/expert.v11i2.2171>
- Santoso, A., Nugroho, A., & Sunge, A. S. (2022). Analisis Sentimen Tentang Mobil Listrik Dengan Metode Support Vector Machine Dan Feature Selection Particle Swarm Optimization. In *Journal of Practical Computer Science* (Vol. 2, Issue 1).