

PENERAPAN ALGORITMA NAIVE BAYES BERBASIS FORWARD SELECTION UNTUK MEMPREDIKSI PENJUALAN MOBIL BEKAS

Haditsah Annur¹, Moh.Efendi Lasulika²

^{1,2}Fakultas Ilmu Komputer, Universitas Ichsan Gorontalo
titintunru@gmail.com¹, fendidsn.ui@gmail.com²

ABSTRAK

Mobil merupakan salah satu kendaraan yang menjadi kebutuhan sehari-hari masyarakat, tidak hanya penggunaan pada mobil baru yang diminati, tetapi sekarang ini mobil bekas juga banyak diminati dikarenakan kualitas dari mobil bekas yang masih bagus dan banyaknya jenis mobil bekas yang dijual di pasaran. Tujuan peneliti ingin meningkatkan minat masyarakat untuk beralih membeli mobil bekas. Penelitian ini menggunakan metode data mining, salah satunya prediksi dengan menggunakan algoritma naive bayes sebagai algoritma yang penggunaan metode probabilistik dan statistik untuk memprediksi masa depan, selain itu juga datanya diolah dengan penggunaan seleksi fitur forward selection yang bertujuan mengurangi tingkat kompleksitas dari sebuah algoritma klasifikasi sekaligus meningkatkan akurasi. Data penelitian yang digunakan sebanyak 2318 record, pada penelitian ini melakukan percobaan dengan hasil akurasi yang didapatkan menggunakan split validation pada algoritma *naive bayes* sebesar 96,98 % dan kemudian dilakukan percobaan kembali dengan mendapatkan hasil akurasi menggunakan *split validation* pada algoritma *naive bayes* berbasis *forward selection* sebesar 97,82 %. Dengan demikian algoritma *naive bayes* berbasis *forward selection* cocok digunakan untuk memprediksi, sekaligus digunakan untuk penanganan dimasa depan bahwa masih banyaknya mobil bekas diminati oleh masyarakat.

Kata Kunci— Mobil Bekas, Penjualan, Algoritma Naive Bayes, Forward Selection.

ABSTRACT

Cars are one of the vehicles that are the daily needs of the people, not only the use of new cars is in demand, but now used cars are also in great demand because the quality of used cars is still good and the many types of used cars are sold in the market. The aim of the researchers is to increase public interest in switching to buying used cars. This study uses data mining methods, one of which is prediction using the Naive Bayes algorithm as an algorithm that uses probabilistic and statistical methods to predict the future, besides that the data is also processed using forward selection feature selection which aims to reduce the level of complexity of a classification algorithm while increasing accuracy. The research data used were 2318 records, in this study an experiment was carried out with the accuracy results obtained using split validation on the naive Bayes algorithm of 96.98% and then another experiment was carried out to obtain accurate results using split validation on the naive bayes algorithm based on forward selection of 97.82 %. Thus the naive Bayes algorithm based on forward selection is suitable for predicting, as well as being used for handling in the future that there are still many used cars that are of interest to the public..

Keywords— Used Cars, Selling, Naïve Bayes Algorithm, Forward Selection.

1. Pendahuluan

Mobil merupakan salah satu kendaraan yang digunakan masyarakat dalam aktivitas sehari-hari. Penggunaan mobil sangatlah membantu aktivitas masyarakat terutama di perkotaan. Masalah dari penelitian ini terdapat beberapa masyarakat yang saat ini, tidak hanya berminat untuk membeli mobil baru tetapi ada juga yang berminat membeli mobil bekas dikarenakan dengan penawaran harga yang murah dan masih memiliki kualitas yang bagus.

Tingginya peminat mobil mengakibatkan banyak dari perusahaan otomotif seperti nissan, toyota, daihatsu, honda, dan lainnya, harus berkompetisi untuk memenuhi kebutuhan pasar, termasuk pembelian mobil bekas yang semakin banyak diminati oleh masyarakat. Maka dari itu penelitian ini bertujuan meningkatkan minat masyarakat untuk beralih membeli mobil bekas yang memiliki jenis dan harga bervariasi sesuai kualitas yang ada, selain itu juga penelitian ini bertujuan mengetahui peningkatan hasil akurasi algoritma *naïve bayes* menggunakan *forward selection* dengan algoritma *naïve bayes* tanpa fitur seleksi. Penjualan mobil bekas pada perusahaan otomotif masih saja berkurang, karena masih sering terjadi pembeli mengalami kerugian dikarenakan pembeli tidak memperhatikan beberapa faktor yang mendukung pada mobil bekas tersebut. Faktor yang mendukung mobil bekas digunakan penelitian ini mencakup harga, merk, model, tahun, jarak tempuh, kapasitas mesin mobil, bahan bakar. Dataset penelitian ini berasal dari website dataset <https://www.kaggle.com/datasets/dojoalsenani/used-cars-dataets> yang mencakup jumlah data 2318 record.

Data mining merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat terkait *database* yang besar [1], salah satu algoritma yang akan digunakan untuk memprediksi yaitu algoritma *naïve bayes*, data yang tersimpan akan diolah menggunakan algoritma *naïve bayes* yang merupakan salah satu algoritma yang terdapat pada klasifikasi dengan penggunaan metode probabilitas dan statistik untuk memprediksi peluang di masa depan berdasarkan pengalaman sebelumnya. [2].

Penggunaan algoritma *naïve bayes* dikarenakan algoritma ini merupakan teknik prediksi berbasis probabilitas sederhana dengan asumsi independensi yang kuat pada fitur, tetapi memiliki kelemahan dengan atribut atau fitur independent sering salah, dan hasil estimasi probabilitas tidak dapat berjalan optimal [3]. maka dari itu diperlukan *fitur selection*, dimana *fitur selection* merupakan salah satu cara untuk menentukan atribut yang paling berpengaruh dalam dataset, dan berperan memilih subset yang tepat dari set fitur asli [4]. salah satu *fitur selection* yaitu *forward selection* yang bertujuan untuk mengurangi tingkat kompleksitas dari sebuah algoritma klasifikasi, meningkatkan akurasi dari algoritma klasifikasi tersebut, dan mampu mengetahui fitur-fitur yang paling berpengaruh terhadap tingkat akurasi [5].

Penelitian terkait yang merupakan menjadi acuan yaitu penelitian yang dilakukan. penelitian yang dilakukan

oleh moh. fajarianditya n, setyoningisih w, dengan judul fitur seleksi *forward selection* untuk menentukan atribut yang berpengaruh pada klasifikasi kelulusan mahasiswa fakultas ilmu komputer unaki semarang menggunakan algoritma *naïve bayes* [6], reynaldi, wahyu s, m faris al hakim, dengan judul analisis perbandingan akurasi metode *fuzzy tsukamoto* dan *fuzzy sugeno* dalam prediksi penentuan harga mobil bekas [7] dan penelitian yang dilakukan oleh dodu guswandi, gushelmi, yang berjudul sistem pendukung keputusan pemilihan mobil bekas menggunakan metode *analytical hierarchy process* [8].

Dari beberapa penelitian terkait yang telah dijelaskan, maka penelitian ini bertujuan untuk memprediksi dan menghitung akurasi dengan menggunakan algoritma *naïve bayes* berbasis *forward selection*. sehingga dimasa akan datang penjualan mobil bekas lebih banyak diminati lagi oleh masyarakat.

2. TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

Penelitian yang dilakukan oleh Dodi Guswandi, Gushelmi (2021). Dengan judul “Sistem Pendukung Keputusan Pemilihan Mobil Bekas Menggunakan Metode Analytical Hierarchy Process”. Penelitian ini membahas tentang bagaimana perusahaan memenuhi target penjualan yang dilakukan, dan memasarkan mobil bekas tidak hanya melalui iklan surat kabar dan brosur sehingga hal tersebut belum memenuhi target penjualan yang diinginkan, sehingga dibutuhkan sistem pendukung keputusan dengan menggunakan metode AHP.

Penelitian yang dilakukan oleh Pratiwi Susanti, Kelik Susolaikah (2022). Dengan judul “ Penerapan Metode Regresi Linear Untuk Memprediksi Harga Jual Mobil Bekas Yaris dan Jazz pada Wilayah DKI Jakarta”, penelitian ini membahas tentang tingginya peminat mobil mengakibatkan banyak perusahaan otomotif harus berkompetisi untuk berinovasi memenuhi keinginan kebutuhan pasar.

Penelitian yang dilakukan oleh Sriani,dkk. (2024) dengan judul “ Implementasi Data Mining terhadap Evaluasi Kinerja Guru dalam Mengajar Metode Naïve Bayes Clasifier”, peneliti ini membahas tentang klasifikasi kinerja guru untuk menilai kompetensi kepribadian, kompetensi social dan kompetensi keprofesional dengan menggunakan perhitungan Naïve Bayes Clasifier.

2.2 Forward Selection

forward selection merupakan salah satu seleksi teknik untuk mereduksi dimensi dataset dengan menghilangkan atribut yang tidak relevan, *forward selection* juga sebagai pemodelan yang dimulai dari nol peubah (*empty model*), kemudian satu persatu peubah dimasukkan sampai kriteria tertentu dipenuhi [9]. Selain itu juga metode *forward selection* seleksi fitur sendiri, yang biasa digunakan *machine learning* yang mana sekumpulan dari fitur yang dimiliki oleh data digunakan untuk proses pembelajaran algoritma, dan merupakan pemodelan dari nol peubah

(empty model), lalu salah satu peubah dimasukkan sampai kriteria tertentu dipenuhi [10]

2.3 Algoritma Naïve Bayes

Algoritma *naïve bayes* merupakan salah satu algoritma yang digunakan untuk prediksi peluang atau kemungkinan suatu kelas, yang mempunyai pendapat yang kuat untuk tidak terkait pada kondisi masing-masing [11].

Persamaan dari algoritma *naïve bayes* sebagai berikut [12]:

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)} \dots \dots \dots (1)$$

keterangan :

x = sampel data yang dimiliki kelas (label yang tidak diketahui)

y = hipotesa bahwa x adalah data kelas (label)

p(y) = peluang dari hipotesa h

p(x) = peluang dari data sampel yang diamati

p(x|y) = peluang dari data sampel x bila diasumsikan bahwa hipotesa benar.

2.4 K-fold cross validation

K-fold cross validation dapat didefinisikan sebagai Teknik validasi silang yang menerapkan pemecahan data ke dalam k sub-set data dengan pembagian jumlah yang seimbang. Pengujian data dengan menggunakan *10-fold cross validation* [13].

Split validation adalah metode yang digunakan untuk menilai kemampuan generalisasi model untuk tune parameter regularisasi atau *hyper-parameter* lain dari suatu proses pembelajaran, selain itu juga sebagai metode validasi model untuk menilai dan mengetahui berapa hasil statistic analisis yang akan menggeneralisasi himpunan data independent [14].

3. METODE YANG DIUSULKAN

3.1 Tahapan Penelitian

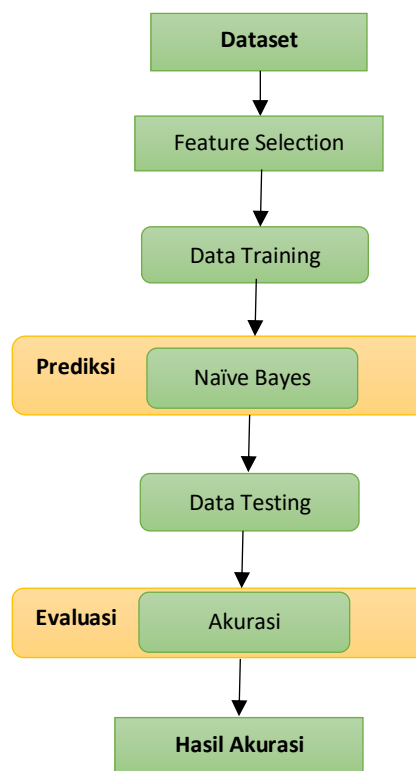
Penelitian ini dikategorikan sebagai penelitian eksperimen yang dilakukan terhadap data yang ada, dalam penelitian ini terdapat beberapa tahap yang dilakukan. Penelitian ini dimulai dengan pengumpulan data yang berasal dari dataset data publik <https://www.kaggle.com/datasets/doaaalsenani/used-cars-dataets> . kemudian dataset tersebut di olah menggunakan *tools rapid miner* untuk mendapatkan hasil prediksi. Jumlah dataset yang digunakan sebanyak 2318 record tentang data mobil bekas, contoh dataset terdapat pada gambar 1 di bawah ini :

	price	mark	model	year	mileage	engine_ca	transmissi	drive	hand_driv	fuel
0	80	nissan	march	2003	80000	1240	at	2wd	rhd	gasoline
1	110	nissan	march	2010	53000	1200	at	2wd	rhd	gasoline
2	165	nissan	lafesta	2005	47690	2000	at	2wd	rhd	gasoline
3	190	toyota	avensis	2008	130661	1990	at	2wd	rhd	gasoline
4	190	daihatsu	mira	2006	66300	660	at	2wd	rhd	gasoline
5	190	daihatsu	mira	2004	81400	660	at	2wd	rhd	gasoline
8	220	nissan	march	2010	117000	1200	at	2wd	rhd	gasoline
9	230	volkswage	passat	2008	127763	3190	at	4wd	rhd	gasoline
10	275	mazda	bongo var	2010	178218	1800	mt	2wd	rhd	gasoline
11	283	honda	step wgn	2005	121655	2000	at	2wd	rhd	gasoline
12	290	subaru	sambar	2004	121000	650	mt	2wd	rhd	gasoline
13	290	honda	inspire	2002	61000	2490	at	2wd	rhd	gasoline
14	290	mercedes	mercedes	2006	86000	2030	at	2wd	rhd	gasoline
15	290	nissan	march	2011	62000	1200	at	2wd	rhd	gasoline
16	290	nissan	march	2010	38000	1200	at	2wd	rhd	gasoline
17	290	nissan	march	2011	130000	1200	at	2wd	rhd	gasoline
18	290	nissan	march	2011	113000	1200	at	2wd	rhd	gasoline
19	290	nissan	march	2010	152300	1200	at	2wd	rhd	gasoline
20	290	nissan	note	2010	105000	1500	at	2wd	rhd	gasoline

Gambar 1. Contoh Dataset Mobil Bekas

3.2 Desain Eksperimen

Model yang diusulkan dapat dilihat pada gambar 2 di bawah ini :



Gambar 2. Desain Eksperimen

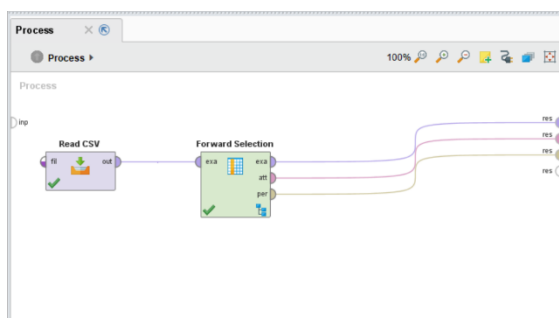
Keterangan :

Data yang digunakan merupakan dataset dari gudang data dari bank data. Dataset yang digunakan *extensi .csv* dan diolah menggunakan *rapidminer* kemudian data tersebut diimport menggunakan *read csv*. Setelah itu *forward selection* digunakan untuk menyeleksi fitur atribut dalam dataset dan algoritma *naïve bayes* digunakan untuk memprediksi penjualan mobil bekas, dalam penelitian ini menggunakan evaluasi dalam bentuk akurasi sehingga diharapkan akurasi yang didapatkan meningkat.

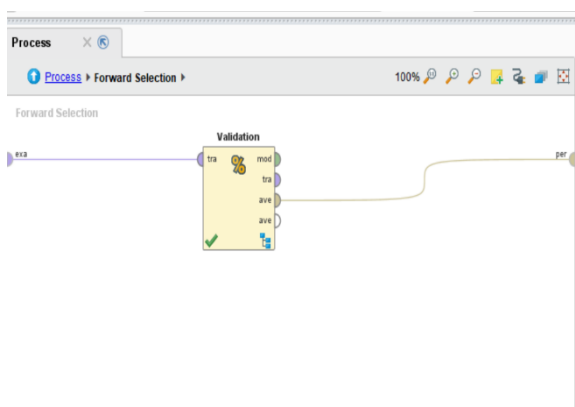
4 Hasil Penelitian

4.1 Skenario Pengujian

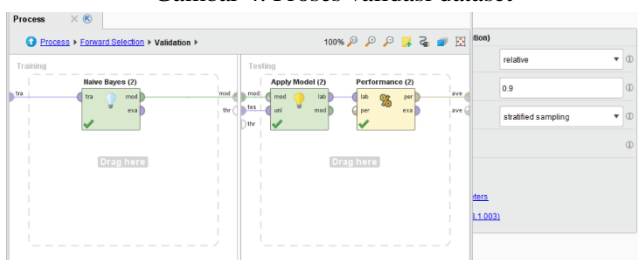
Memprediksi penjualan mobil bekas ini, peneliti mencoba melakukan pengolahan dataset dengan menggunakan *tools rapid miner* seperti di bawah ini :



Gambar 3. Proses *forward selection*



Gambar 4. Proses validasi dataset



Gambar 5. Proses Naïve bayes

Gambar 3, 4, dan 5 di atas menjelaskan tentang bagaimana proses prediksi menggunakan algoritma *naïve bayes* berbasis *forward selection*, dengan penggunaan dataset sebanyak 2318 *record*.

4.2 Pengujian Menggunakan Algoritma *Naive Bayes* berbasis *Feature Selection*

Pengujian model *split validation* menggunakan fitur seleksi, telah diuji menggunakan *tools rapid miner* dan mendapatkan hasil akurasi seperti di bawah ini :

Tabel 2. Nilai akurasi *split validation* pada data mobil bekas algoritma *naïve bayes* berbasis *forward selection*

Sampling Type	Split Ratio								
	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
Linear Sampling	92.67	92.89	93.24	92.77	92.58	88.64	87.92	86.25	76.08
Shuffled Sampling	96.12	95.91	95.68	96.22	95.34	93.31	92.73	90.35	86.19
Stratified Sampling	97.82	95.87	95.68	95.78	94.65	93.61	92.54	90.20	84.86

Tabel 2. Nilai akurasi *split validation* di atas dilakukan pembagian nilai presentase berdasarkan *sampling type* yang ada sehingga didapatkan nilai akurasi tertinggi dari setiap *sampling type*, diantaranya *sampling type* tertinggi yaitu *stratified sampling* dengan *split ratio* 0.9 yang mempunyai nilai akurasi 97.82 % disusul *shuffled sampling* 96.12% dan *linear sampling* 92.67%.

1. Hasil pengujian akurasi dataset mobil bekas menggunakan Algoritma *Naive Bayes* berbasis *Forward Selection*

accuracy: 97.82%

	true niss...	true toyota	true daih...	true volk...	true maz...	true honda	true sub...	true mer...	true kub...	true mits...	true suzu
pred. niss...	32	0	0	0	0	0	0	0	0	0	0
pred. toy...	0	102	0	0	0	0	0	0	0	0	0
pred. dai...	0	0	6	0	0	0	0	0	0	0	0
pred. vol...	0	0	0	5	0	0	0	0	0	0	0
pred. ma...	0	0	0	0	15	0	0	0	0	0	0
pred. ho...	0	0	0	0	0	37	0	0	0	0	0
pred. su...	0	0	0	0	0	0	5	0	0	0	0
pred. me...	0	0	0	0	0	0	0	0	0	0	0
pred. ku...	0	1	0	0	0	0	0	0	0	0	0
pred. mit...	0	0	0	0	0	0	0	0	0	5	0
pred. suz...	0	2	0	0	0	1	0	0	0	1	13
pred. bmw	0	0	0	0	0	0	0	0	0	0	0

Gambar 6. Hasil pengujian akurasi dataset mobil bekas menggunakan Algoritma *Naive Bayes* berbasis *Forward Selection*.

Gambar di atas merupakan hasil dari proses akurasi menggunakan *tools Rapid Miner* dengan menunjukkan bahwa hasil akurasi yang diperoleh menggunakan Algoritma *Naive Bayes* dengan *split ratio* 0.9 yang mempunyai nilai akurasi 97.82 %.

2. Hasil *Attribute Weight (Forward Selection)*

Tabel 3. Hasil *Attribute Weight*

Attribut	Weight
Att1	0
Price	0
model	1
Year	0
Mileage	0
Engine	0
Transmission	0
Drive	0
Hand_drive	0
Fuel	1

Forward Selection yang dihasilkan pada label jenis mobil menghasilkan dua atribut yang berpengaruh yakni model dengan nilai 1 dan bahan bakar menghasilkan nilai 1. Sedangkan atribut yang lain menghasilkan nilai 0 sehingga menandakan bahwa atribut yang lain tidak berpengaruh pada data yang telah diujikan.

4.3 Pengujian Menggunakan Algoritma *Naïve Bayes*

Pengujian model *split validation* tanpa menggunakan fitur seleksi, telah diuji menggunakan *tools rapid miner* dan mendapatkan hasil akurasi seperti di bawah ini :

Tabel 4. Nilai Akurasi *Split Validation* pada data Mobil Bekas Menggunakan Algoritma *Naïve Bayes*

Sampling Type	Split Ratio									
	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	
Linear Sampling	90.52	89.01	88.20	85.44	78.17	73.54	37.03	29.72	11.94	
Shuffled Sampling	96.98	94.61	94.10	93.96	93.18	92.31	90.94	82.04	80.35	
Stratified Sampling	94.76	94.35	94.67	93.19	92.67	89.88	88.78	86.87	79.54	

Tabel 4. Nilai akurasi *split validation* di atas dilakukan pembagian nilai presentase berdasarkan *sampling type* yang ada sehingga didapatkan nilai akurasi tertinggi dari setiap *sampling type*, diantaranya *sampling type* tertinggi yaitu *shuffled sampling* dengan *split ratio* 0.9 yang mempunyai nilai akurasi 96,98 % disusul *stratified sampling* 94,76% dan *linear sampling* 90,52%.

1. Hasil pengujian akurasi dataset mobil bekas menggunakan Algoritma *Naïve Bayes*.

accuracy: 96.98%

	true niss...	true toyota	true daih...	true volk...	true maz...	true honda	true sub...	true mer...	true i
pred. niss...	35	0	0	0	0	0	0	0	0
pred. toy...	0	108	0	0	0	0	0	0	0
pred. dai...	0	0	3	0	0	0	0	0	0
pred. vol...	0	0	0	5	0	0	0	0	0
pred. ma...	0	0	0	0	15	0	0	0	0
pred. ho...	0	0	0	0	0	34	0	0	0
pred. su...	0	0	0	0	0	0	5	0	0
pred. me...	0	1	0	0	0	0	0	0	0
pred. ku...	0	0	0	0	0	0	0	0	0
pred. mit...	0	0	0	0	0	0	0	0	0
pred. suz...	0	0	0	0	0	0	0	0	0
pred. bmw	0	1	0	0	0	0	0	0	0

Gambar 7. Hasil Pengujian Akurasi Dataset Mobil Bekas Menggunakan Algoritma *Naïve Bayes*

Gambar di atas merupakan hasil pengujian akurasi dataset mobil bekas menggunakan *tools rapid miner* dengan menunjukkan bahwa hasil akurasi yang diperoleh menggunakan algoritma *naïve bayes* dengan *split ratio* 0.9 yang mempunyai nilai akurasi 96.98%

2. Hasil probabilitas dataset mobil bekas menggunakan *tools rapid miner* seperti pada tabel 5.

Tabel 5. Hasil probabilitas dataset mobil bekas

Class	Distributions	Value Label
Nissan	10	0.138
Toyota	10	0.452
Daihatsu	10	0.024
Volkswagen	10	0.022
Mazda	10	0.064
Honda	10	0.165
Mercedes-Benz	10	0.002

- [11] H. Annur, “Penerapan Algoritma Naïve Bayes Berbasis Backward Elimination Untuk Prediksi Pemesanan Kamar Hotel,” *J. Ilm. Ilmu Komput. Banthayo Lo Komput.*, vol. 1, no. 1, pp. 1–5, 2022.
- [12] N. Nurajijah and D. Riana, “Algoritma Naïve Bayes, Decision Tree, dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah,” *J. Teknol. dan Sist. Komput.*, vol. 7, no. 2, pp. 77–82, 2019.
- [13] F. Andraini and E. Mahdiyah, “Analisis Sentimen Twitter Terhadap Peperangan Rusia Dan Ukraina Menggunakan Algoritma Support Vector Machine,” *J. Apl. Komput.*, vol. 2, no. 1, pp. 46–58, 2022.
- [14] R. Tuntun, “Analisis Perbandingan Kinerja Algoritma Klasifikasi dengan Menggunakan Metode K-Fold Cross Validation,” vol. 6, pp. 2111–2119, 2022.