

PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE DAN NAIVE BAYES UNTUK KLASIFIKASI JENIS MOBIL TERLARIS BERDASARKAN PRODUKSI DI INDONESIA

Herdipaul, Anggri Sartika Wiguna, Heri Santoso

Program Studi Teknik Informatika S1, Fakultas.Sains.dan.Teknologi
Universitas PGRI Kanjuruhan Malang, Jalan S.Supriadi No.48 Malang, Indonesia
herdipaul2018425@gmail.com

ABSTRAK

Mobil merupakan salah satu alat transportasi yang paling banyak digunakan di Indonesia. Berdasarkan spesifikasinya, mobil memiliki berbagai varian atau jenis. Algoritma SVM yaitu untuk mencari *hyperline* yang berperan sebagai pemisah 2 buah *class* pada *input space* dan Naive Bayes merupakan salah satu metode klasifikasi yang menggunakan metode probabilitas dan statistik. Dari berbagai macam merek tersebut akan di bentuk sebuah *Class*, supaya produsen serta konsumen mampu untuk mengenali merk dari sebuah mobil yang sangat terlaris di produksi berdasarkan output ataupun kategorinya. GAIKINDO merupakan Gabungan Industri Kendaraan Bermotor Indonesia. Data yang diambil adalah data tahun 2021 dengan data berjumlah 574 data 25 atribut, untuk memudahkan riset sehingga atribut pada penggunaan 8 termasuk satu atribut kelas dengan penambahan peneliti supaya mempermudah untuk pencarian jumlah produksi mobil terbanyak. Temuan analisis dengan menggunakan metode SVM melalui *RapidMiner* mendapatkan nilai *accuracy* 96,52%, nilai *precision* 94,87% dan nilai *recall* 100%. Hasil analisis dengan menggunakan metode Naive Bayes melalui *RapidMiner* mendapatkan nilai akurasi 97,39%, nilai presisi 97,33% dan nilai *recall* 98,65%. Hasil *accuracy* Naive Bayes lebih tinggi dari SVM, yaitu 97,39%. Hasil *precision* Naive Bayes mendapatkan skor lebih tinggi dari SVM, yaitu 97,33%. Sedangkan hasil *recall* SVM lebih tinggi dari Naive Bayes, yaitu 100%.

Kata kunci : Merek Mobil, Klasifikasi, Support Vector Machine, Naive Bayes, Accuracy.

1. PENDAHULUAN

Indonesia adalah salah satu negara dengan kepemilikan penduduk yang berjumlah banyak, dalam mobilitas sehari-hari. Mobil adalah modal transportasi yang mayoritas dipakai oleh masyarakat. Seiring dengan meningkatnya jumlah pengguna mobil di Indonesia, banyak industri otomotif yang membangun pabrik di Indonesia sehingga mobil yang digunakan di Indonesia merupakan gabungan dari mobil buatan Indonesia dan mobil yang diimpor dari berbagai negara dan merupakan industri dimana banyak menampung angkatan kerja yang ada [1].

Berdasarkan spesifikasinya, mobil memiliki berbagai varian atau jenis. Namun GAIKINDO sendiri tidak memberikan peringkat pembeda di websitenya mengenai merek mobil mana yang paling diminati konsumen. GAIKINDO merupakan penggabungan industri motor yang bergerak pada sektor otomotif serta merupakan perusahaan nirlaba. Data yang diambil adalah data tahun 2021 dengan data berjumlah 574 data 25 atribut, untuk memudahkan riset dimana atribut dipergunakan yaitu 8 termasuk satu atribut kelas dengan peneliti menambahkannya supaya mempermudah dalam pencarian jumlah produksi mobil terbanyak. Maka penggolongan tersebut diperlukan supaya konsumen dapat menganalisa merk mobil terlaris dalam kaitannya dengan kategori yang sangat mereka minati. Banyaknya macam merk diperlukan pembentukan sebuah kelas yang tergolong terjual laris serta tidak, supaya produsen serta konsumen

mampu untuk menganalisa merk mobil tergolong terlaris di produksi berdasarkan output ataupun kategorinya [2].

Penggolongan merupakan sebuah cara dalam machine learning dimana dipakai dengan penggunaan mesin dalam memilih ataupun melakukan klasifikasi objek didasarkan pada kriteria tertentu. Machine Learning ataupun pembelajaran mesin adalah sebuah metode pada AI dimana telah banyak dipakai pada pengambil alihan maupun mengikut perilaku manusia dalam penyelesaian permasalahan dengan otomatis [3]. Salah satunya yaitu menggunakan *Support Vector Machine* (SVM) yang merupakan upaya guna supaya mendapatkan *hyperlane* yang memiliki nilai yang paling baik dimana memiliki fungsi pada pembeda dari 2 kelas di input ruang. Algoritma supervise dimana berbentuk penggolongan secara pembagian data yang menjadi 2 golongan *class* dengan penggunaan *hyperlane* [4]. Sedangkan pada penggunaan Naive Bayes adalah sebuah metode dimana memprediksi barisan kemungkinan secara sederhana dengan didasarkan diterapkannya aturan bayes menggunakan anggapan bebas serta kuat. Naive Bayes digunakan dengan dasar mempunyai kemampuan dalam penggolongan seragam serta jaringan netral. Metode ini teruji mempunyai tingkat keakuratan serta cepatnya dalam pengaplikasian pada data base yang besar [5].

2. TINJAUAN PUSTAKA

Pendukung penelitian ini diperlukan landasan teori yang akan digunakan pada penelitian. Landasan teori yang digunakan adalah Gaikindo, *Data mining, Machine Learning, Klasifikasi, RapidMiner, Support Vektor Machine, Naive Bayes Dan Confusion Matrix*

2.1. Gaikindo

GAIKINDO merupakan penggabungan industri motor yang merupakan salah satu sektor aspek pasar otomotif yang ada di Indonesia serta perusahaan dimana berperan kerjasama bersama Pemerintah untuk dapat melakukan pengembangan industri motor skala besar dan memiliki kesanggupan dalam kancah global [6].

2.2. Data Mining

Data Mining merupakan tahapan dalam penggunaan metode matematika, kecerdasan buatan, statistik, *machine learning* untuk mendapatkan serta identifikasi wawasan dimana berguna pada pengetahuan mengenai database yang berskala besar [7].

2.3. Machine Learning

Machine learning adalah paradigma yang dapat merujuk pada pembelajaran dari pengalaman masa lalu (yang dalam hal ini adalah data sebelumnya) untuk meningkatkan kinerja di masa mendatang. Pembelajaran mesin adalah pijakan kecerdasan buatan [8]. Satu-satunya fokus bidang ini adalah metode pembelajaran otomatis. Pembelajaran mengacu pada modifikasi atau peningkatan algoritma berdasarkan masa lalu pengalaman secara otomatis tanpa bantuan eksternal dari manusia.

2.4. Klasifikasi

Klasifikasi merupakan metode yang dipakai untuk mendapatkan sesuatu fungsi dengan tujuan untuk membedakan suatu jenis kategori atau kelas data. [9] Didalam data mining banyak mempunyai teknik yang dapat dicoba untuk mengolah data dalam jumlah yang besar jadi suatu informasi yang dapat berguna. Terdapat sebagian teknik data mining adalah, deskripsi, estimasi, prediksi, klasifikasi, clustering, dan asosiasi. Tujuan dari klasifikasi merupakan untuk memprediksi ataupun memperkirakan kelas dari suatu data terkini yang belum mempunyai merek. Untuk mencapai tujuan klasifikasi yaitu, perlu dibuat suatu bentuk untuk membedakan suatu *class* data berdasarkan metode tertentu.

2.5. RapidMiner

Rapidminer merupakan perangkat lunak dengan sifat yang terbuka, maka dijelaskan yaitu sesuatu pemecahan dalam rangka analisa pada data mining, analisa kemungkinan dan text mining. Rapidminer ini penggunaannya dengan metode deskripsi serta melakukan *forecast* didalam membagikan wawasan

pada konsumen mengenai ketetapan yang sangat bagus. Metode ini mempunyai kurang lebih 500 data mining operators, *preprocessing, output, input serta visualisasi*. Banyak metode dimana menyediakan metode ini yang bermula pada *clustering, klasifikasi, forecasting, association* serta lainnya [10].

2.6. Support Vector Machine

Support Vector Machine (SVM) merupakan awal kali dipublikasikan oleh Vapnik yang dimulai tahun 1992 selaku susunan yang beriringan pada teori-teori favorit didalam aspek pengenalan bertahap. Seperti contoh tata pengenalan tahapan, umur *Support Vektor Machines* terhitung sedang termasuk awal [11]. Sebagai salah satu tata cara *pattern recognition*, umur *Support Vektor Machine* terhitung sedang relatif muda. Algoritma *Support Vektor Machine* yaitu untuk mencari *hyperline* yang berperan sebagai pemisah 2 buah *class* pada *input space*. *Support Vector Machine* sudah jadi metode klasifikasi dan regresi yang kerap dipakai untuk permasalahan linear serta nonlinear.

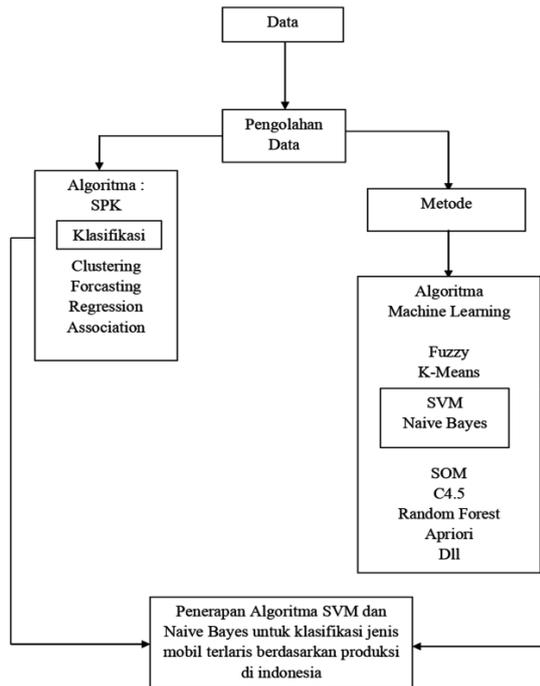
2.7. Naive Bayes

Naive Bayes adalah sebuah metode dimana memprediksi barisan kemungkinan secara sederhana dengan didasarkan diterapkannya aturan bayes menggunakan anggapan bebas serta kuat [12]. *Naive Bayes* digunakan dengan dasar mempunyai kemampuan dalam penggolongan seragam serta jaringan netral. Metode ini teruji mempunyai tingkat keakuratan serta cepatnya dalam pengaplikasian pada database yang besar. *Naive Bayes* merupakan metode perkiraan probabilitas yang beralasan pada penerapan *teorema Bayes* dengan karakteristik penting merupakan yang amat kuat akan independensi dari tiap-tiap situasi ataupun peristiwa [13]. Penggolongan *Naive Bayes* memiliki tingkat keakuratan yang bagus bila dibandingkan dengan bentuk klasifikasi yang lain.

2.8. Confusion Matrix

Berdasarkan pada mining data guna beberapa metode dalam pengukuran kinerja sebuah model maka digunakan matrix campuran dimana merupakan sebuah metode yang penggunaannya dalam rangka menghitung tingkat akurasi sebuah konsep mining data. Keakuratan dari metode ini diprediksi positif. *Sensitivity* atau *recall* merupakan bagian pada kasus positif dimana dilakukan *forecasting* dengan benar [14].

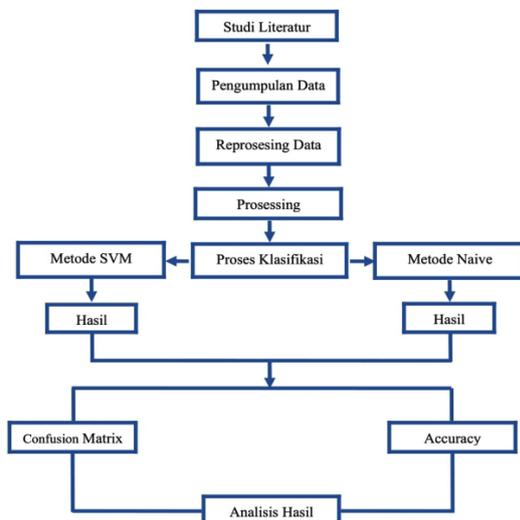
Gambar 1 adalah gambaran dari model penelitian yang dilakukan.



Gambar 1. Representasi Pengetahuan

3. METODE PENELITIAN

Pada riset yang dilakukan ini penggunaan dari data mining yaitu penggunaan *Support Vector Machine* serta *Naive Bayes* dalam penggolongan Jenis Mobil Terbanyak Berdasarkan Jumlah Produksi di Indonesia. Berikut merupakan proses pada penelitian ini.



Gambar 2. Tahapan Penelitian

3.1. Studi Literatur

Tahap ini merupakan tahapan pembelajaran yang berkaitan dengan topik yang diangkat sebagai penelitian, serta mengkaji dan menganalisis metode *Support Vector Machine* dan *Naive Bayes* dalam melakukan klasifikasi. Penelusuran referensi ini bersumber dari buku, jurnal maupun penelitian yang

telah ada sebelumnya mengenai hal - hal yang berhubungan dengan metode *Support Vector Machine* dan *Naive Bayes*. Hal ini dilakukan untuk mengetahui bagaimana metode yang digunakan dalam kajian teorinya sebelum digunakan dalam penelitian.

3.2. Pengumpulan Data

Proses didatakannya sebuah data adalah menganalisa jawaban berdasarkan rumusan masalah yang disusun. Data berupa 574 data dimana merupakan dataset yang diperoleh dari laman GAKINDO dengan fokus di tahun 2021. Dalam bentuk file pdf.

3.3. Reprosesing Data

Data dari website ini yang didapatkan dari website Gaikindo merupakan data mulai sejak tahun 2005 sampai dengan 2022, tapi menurut teori pengambilan data secara sampel. Namun pada penelitian ini menggunakan data 1 tahun terakhir yaitu tahun 2021, karena dengan berbagai alasan yaitu:

- a) Penggunaan mobil variasi hampir sama.
- b) Data relevan dan terbaru.

3.4. Prossesing

Setelah dilakukannya Reprosesing data selanjutnya akan dilakukan pengolahan data yang sudah didapatkan dari gaikindo, sehingga diperlukannya tahapan hitung data dengan tahapan penggunaan data mining adalah algoritma SVM serta metode algoritma *Naive Bayes* proses ini dalam rangka menganalisa temuan pada kedua algoritma yang digunakan yaitu tingkat akurasi dari hasil data yang digunakan.

3.5. Proses Klasifikasi Metode Algoritma SVM dan Naive Bayes

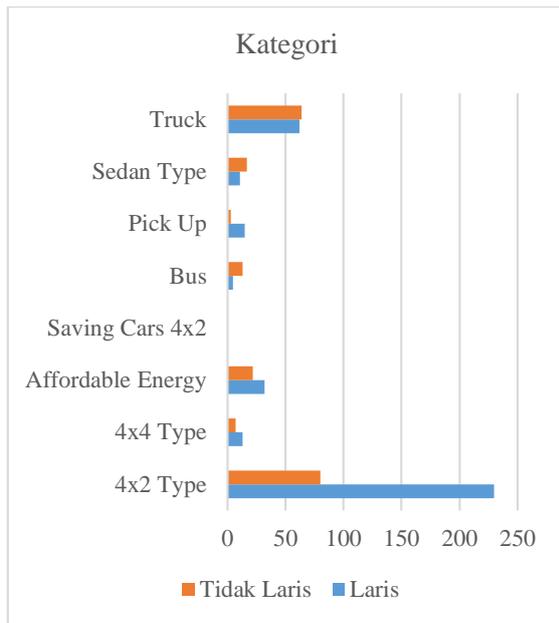
Pada tahap ini dilakukannya penerapan. Untuk mengklasifikasi jenis mobil terlaris berdasarkan kategori. Berdasarkan data dari gaikindo yang sudah dilakukan Reprocessing selanjutnya akan dilakukan perhitungan akurasi dengan menggunakan metode Algoritma *Support Vector Machine* dan *Naive Bayes*.

3.6. Hasil Analisis

Pada tahap ini setelah semua data yang dibutuhkan terkumpul, selanjutnya adalah menganalisis data yang akan diklasifikasi. Analisis data yang akan digunakan oleh peneliti adalah analisis kuantitatif dengan menggunakan *Confusion Matrix*.

4. HASIL DAN PEMBAHASAN

Jumlah data yang digunakan dalam penelitian ini adalah sebanyak 574 data, dengan menggunakan 7 atribut, yaitu *category*, *brand*, *type model*, *cc*, *trans*, *origin country*, *output*. Pada atribut *type* terdapat tujuh kategori yang didapatkan hasil sebagai berikut.



Gambar 3. Hasil Atribut Kategori

Berdasarkan pada gambar 3 didapatkan category mobil terlaris adalah 4X2 Type dengan total output sebesar 310 unit, disusul dengan category Truck dengan total output sebesar 126 unit, dan posisi ketiga adalah Affordable Energy Saving Cars 4X2 dengan total output sebesar 54 unit. Berikut merupakan hasil dari kriteria brand dengan menggunakan 19 brand mobil.

Tabel 1. Hasil Atribut Brand

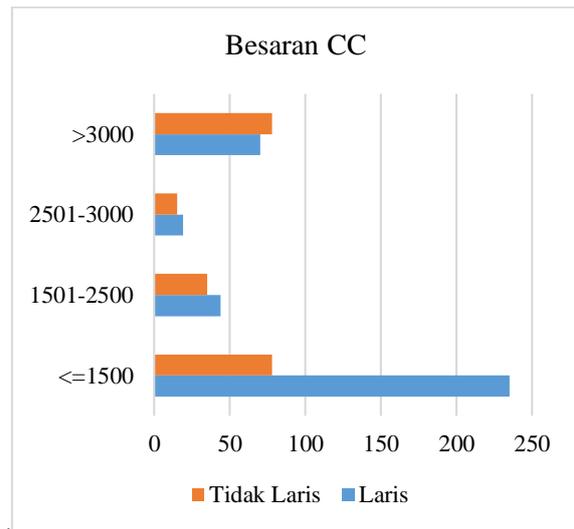
Kategori	Laris	Tidak Laris
BMW	6	4
Daihatsu	54	10
Datsun	0	12
DFSK	6	1
FAW	0	1
Hino	35	34
Honda	26	4
Hyundai	0	1
Hyundai Him	1	2
Isuzu	13	11
Mercedes-Benz CV	2	9
Mercedes-Benz PC	7	12
Mini	1	0
Mitsubishi Fuso	16	11
Mitsubishi Motors	64	23
Suzuki	23	10
Toyota	96	51
UD Truck	1	7
Wuling	17	3

Berdasarkan pada tabel 1, dapat dilihat bahwa brand mobil terlaris adalah Toyota dengan total output sebesar 147 unit, disusul dengan brand Mitsubishi Motors dengan total output sebesar 87 unit, dan posisi ketiga adalah Hino dengan total output sebesar 69 unit

Tabel 2. Hasil Atribut Type Model

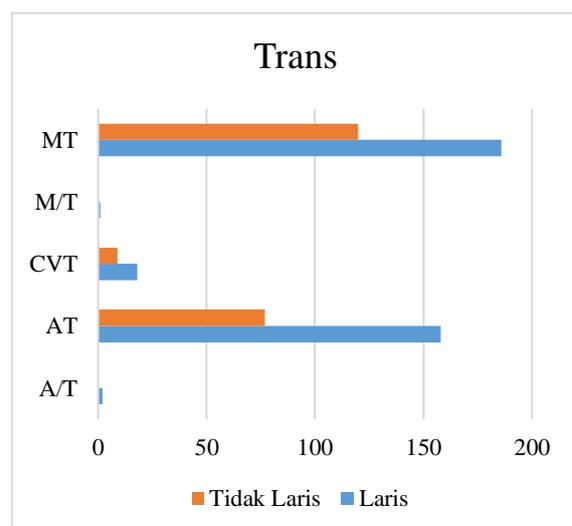
Atribut	Laris	Tidak Laris
Type Model	278	172
Jumlah Tertinggi	24	15

Berdasarkan pada tabel 2 didapatkan tipe model yang paling laris pada tipe Fortuner dengan 24 kali laris, sedangkan tidak laris pula didapatkan pada tipe Fortuner dengan jumlah nilai 15. Berdasarkan pada kriteria besaran CC disajikan pada gambar berikut



Gambar 4. Hasil Atribut Besaran CC

Berdasarkan pada gambar 4 didapatkan mobil terlaris adalah yang menggunakan 1499 cc dengan total output sebesar 76 unit, disusul dengan 1500 cc dengan total output sebesar 68 unit, dan 2400 cc dengan total output adalah 38 unit.. Pada atribut trans disajikan pada gambar berikut.



Gambar 5. Hasil Atribut Trans

Berdasarkan pada gambar 5 didapatkan mobil terlaris adalah yang menggunakan trans MT dengan total output sebesar 306 unit, disusul dengan AT dengan total output sebesar 27 unit, dan CVT dengan total output adalah 27 unit. Pada Atribut *Origins* disajikan dalam tabel berikut.

Tabel 3. Atribut Origins

Kategori	Laris	Tidak Laris
INA	364	196
IND	4	10

Berdasarkan pada tabel 3 menjelaskan produk asal Indonesia yang paling laris dan tidak laris dikarenakan penelitian ini berfokus pada produksi lokal. Sedangkan pada atribut Output disajikan dalam tabel berikut ini.

Tabel 4. Atribut Output

Kategori	Laris	Tidak Laris
>100	368	0
<=100	0	206

Berdasarkan pada tabel 4 menunjukkan perbandingan data mobil laris dan tidak laris. Jumlah mobil laris adalah 368 unit, sedangkan mobil tidak laris adalah 206 unit.

Implementasi klasifikasi yang menggunakan *support vector machine* (SVM) untuk melakukan 2 klasifikasi terhadap 20% dari data yang didapatkan. Cara pengujian yang dilakukan dalam penelitian ini adalah dengan menggunakan *Confusion Matrix* seperti yang digambarkan dalam tabel berikut.

Tabel 5. *Confusion Matrix* SVM

Kelas	True Tidak Laris	True Laris
Tidak Laris	37	0
Laris	4	74

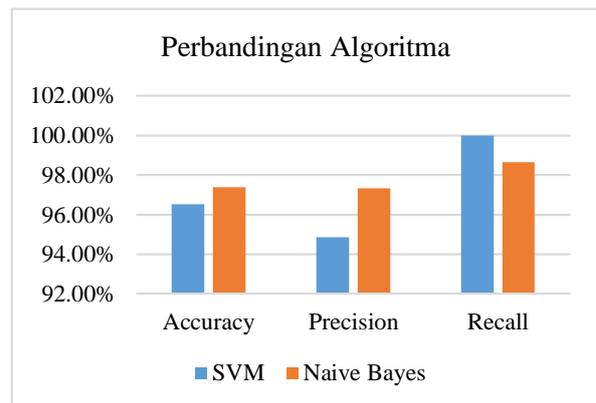
Berdasarkan pada tabel 5 menunjukkan bahwa data mobil tidak laris yang di klasifikasi dengan benar adalah sejumlah 37 data, data mobil laris yang diklasifikasi benar adalah 74 data, data mobil tidak laris yang diklasifikasi salah adalah 4, dan data mobil laris yang diklasifikasi salah adalah 0 data. Pada penggunaan klasifikasi SVM ini peneliti memperoleh tingkat *accuracy* klasifikasi SVM berada pada angka 96,52%, pada tingkat *precision* didapatkan nilai 94,87% dengan *positive class* adalah laris dan hasil *recall* pada klasifikasi SVM didapatkan nilai 100%.

Sedangkan pada implementasi klasifikasi menggunakan *Naive Bayes* menggunakan *confusin matrix* sebagai berikut.

Tabel 6. *Confusion Matrix* Naive Bayes

Kelas	True Tidak Laris	True Laris
Tidak Laris	39	1
Laris	2	73

Berdasarkan pada tabel 6 menunjukkan bahwa data mobil tidak laris yang di klasifikasi dengan benar adalah sejumlah 61 data, data mobil laris yang diklasifikasi benar adalah 47 data, data mobil tidak laris yang diklasifikasi salah adalah 0, dan data mobil laris yang diklasifikasi salah adalah 7 data. Pada penggunaan klasifikasi *Naive Bayes* ini peneliti memperoleh nilai *accuracy* sebesar 97,39%, tingkat *precision* menunjukkan besaran 97,33% dengan *positive class* adalah laris dan hasil *recall* pada besaran 98,65%. Berdasarkan hasil yang didapatkan pada kedua algoritma, maka dapat di bandingkan hasil yang disajikan dalam gambar berikut.



Gambar 6. Perbandingan Algoritma *Support Vector Machine* dan *Naive Bayes*

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, dapat ditarik kesimpulan sebagai berikut: Hasil analisis dengan menggunakan metode SVM melalui RapidMiner mendapatkan nilai *accuracy* 96,52%, nilai *precision* 94,87%, dan nilai *recall* 100%. Hasil analisis dengan menggunakan metode Naive Bayes melalui RapidMiner mendapatkan nilai *accuracy* 97,39%, nilai *precision* 97,33% dan *recall* 98,65%. Hasil *accuracy* Naive Bayes lebih tinggi dari SVM, yaitu 97,39%. Hasil *precision* Naive Bayes mendapatkan skor lebih tinggi dari SVM, yaitu 97,33%. Sedangkan hasil *recall* SVM lebih tinggi dari Naive Bayes yaitu 100%.

Adapun saran dari penelitian ini yaitu sebagai berikut: Jumlah data yang digunakan lebih banyak agar didapatkan hasil akurasi yang lebih baik. Penelitian ini tidak terimplementasi pada sebuah sistem aplikasi, sehingga perlu dibuat sistem aplikasi. Penelitian selanjutnya dapat dikembangkan lagi dengan menggunakan data ataupun metode yang berbeda.

DAFTAR PUSTAKA

[1] Retno, A., Hayati, T., Putri, I.K., Informasi, J.T., & Malang, P.N.(2021). *Peramalan Penjualan Mobil Menggunakan Holt–Winters*.

- [2] Susanto, & Suryadi. (2018). *Penerapan algoritma naive bayes dalam menentukan merek mobil terlaris*. 1–58.
- [3] Surahman, M., Abdillah, L.A., & Ferdiansyah. (2020). *Penerapan Metode SVM - Based Machine Learning Untuk Menganalisa Pengguna Data Trafik Internet (Studi Kasus Jaringan InternetWlan Mahasiswa Bina Darma)*. 196–206.
<http://arxiv.org/abs/2008.12099>
- [4] Maksun, A., Swanjaya, D., Informatika, T., Teknik, F., Nusantara, U., & Kediri, P. (2021). *Perbandingan Antara Metode Decision Tree Dan Support Vector Machine Pada Model Rekomendasi Mobil Bekas*.
- [5] Gaikindo. (2022). *GAIKINDO*. Gaikindo. gaikindo.co.id
- [6] Wicaksana, I. W. S. (2013). *Belajar Data Mining dengan Rapid Miner*. Wiga.
- [7] Imron, M. (2017). Penerapan Data Mining Algoritma Naives Bayes Dan Part Untuk Mengetahui Minat Baca Mahasiswa Di Perpustakaan Stmik Amikom Purwokerto. *Jurnal Telematika*, 10(2), 121–135.
- [8] Sani, Susanto, & Suryadi, D. (2010). *Pengantar Data Mining: Menggali Pengetahuan dari Bongkahan Data*. Andi.
- [9] Tran, H. (2019). *Survei Machine Learning dan Teknik Data Mining yang digunakan dalam Sistem* September. <https://doi.org/10.13140/RG.2.2.20395.49446/1>.
- [10] Fitriana, D., Dwiasnati, S., Hikmayanti, H., & Baihaqi, K.A. (2021). *Penerapan Metode Machine Learning untuk Prediksi Nasabah Potensial menggunakan Algoritma Klasifikasi Naïve Bayes*. 14(2), 1979–276.
- [11] Wibawa, A.P., Guntur, M., Purnama, A., Akbar, M.F., & Dwiyanto, F.A. (2018). *Metode-metode Klasifikasi*. 3(1), 134–138.
- [12] Sihombing, J. (2021). *Klasifikasi Data Antropometri Individu Menggunakan Algoritma Naïve Bayes Classifier*. *BIOS :Jurnal Teknologi Informasi Dan Rekayasa Komputer*, 2(1), 1–10.
<https://doi.org/10.37148/bios.v2i1.15>
- [13] Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Citee Journal*, 2(3), 73–81.
<https://doi.org/10.20895/inista.v1i2.73>
- [14] Sirli. (2016). *Penerapan Algoritma ID3 Untuk Prediksi Minat Studi Mahasiswa Teknik Informatika*. Universitas Dian Nuswantoro.