

KLASIFIKASI JENIS MOBIL PALING DIMINATI DI INDONESIA MENGUNAKAN ALGORITMA NAÏVE BAYES

JAJAM HAERUL JAMAN
ADITYA RIZKY SANJAYA
CARUDIN

Program Studi Teknik Informatika
Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang
Jl. H.S. Ronggowaluyo Telukjambe Timur Karawang 41361
Email: jajam.haeruljaman@staff.unsika.ac.id, aditya.rizky15006@student.unsika.ac.id,
carudin@staff.unsika.ac.id

***Abstract.** Passenger car sales in Indonesia are quite high, in 2014-2018 there were 4,157,580 units sold according to GAIKINDO data, the highest car sales were in MPV cars as the most interested in Indonesia, meanwhile type of car Sedan, SUV, SUV 7 Seater, and City Car have not been classified yet to any category, this is why this study aims to classify the most popular types of cars in Indonesia using data mining techniques. The algorithm that will be used in this research is Naive Bayes with the help of WEKA tools, from the research that has been done by the Naive Bayes algorithm can be used to classify the most exciting types of cars in Indonesia with 96.7% accuracy using the Test Options Cross-validation feature 5. Results the research obtained was that the type of MPV car became the "Most Interested", while the type of Sedan, SUV, 7 Seater SUV, and City Car classified as "Interested" category.*

***Keywords:** Data Mining, Classification, Naive Bayes, Car*

Abstrak. Penjualan *Passenger car* di Indonesia mencatatkan angka cukup tinggi, dalam kurun waktu 2014 – 2018 tercatat sebanyak 4.157.580 unit terjual menurut data GAIKINDO, penjualan mobil tertinggi berada pada jenis mobil MPV sebagai paling diminati di Indonesia, ada banyak jenis mobil di Indonesia diantaranya adalah jenis Sedan, SUV, SUV 7 Seater, dan City Car dan peminatnnya pun tidak main-main sangat banyak sekali sehingga sangat menarik untuk dikaji, metode klasifikasi adalah salah satu cara untuk mendapatkan hasil bagai mana pola penjualan mobil tersebut di Indonesia, dan pada makalah ini kami akan melakukan klasifikasi berdasarkan jenis mobil yang ada dengan beberapa sekenario yang di buat, hasil menunjukkan bahwa sekenario 1 dengan akurasi 96,7% klasifikasi algoritma Naive Bayes, didapatkan jenis mobil MPV menjadi “Paling Diminati”, sedangkan jenis mobil Sedan, SUV, SUV 7 Seater, dan City Car diklasifikasikan ke dalam kategori “Diminati”.

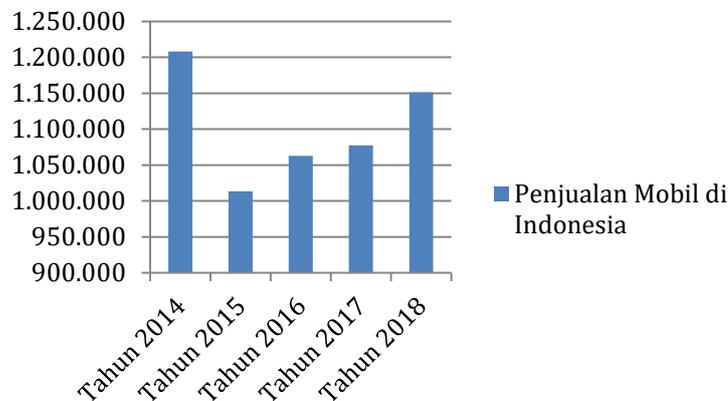
Kata Kunci: Data Mining, Klasifikasi, Naive Bayes, Mobil

PENDAHULUAN

Penjualan *Passenger car* di Indonesia mencatatkan angka yang cukup tinggi dalam kurun waktu 2014 – 2018 berdasarkan data yang diperoleh dari GAIKINDO (Gabungan Industri Kendaraan Bermotor Indonesia) yaitu sebanyak 4.157.580 unit yang terjual (Gaikindo, 2017), penjualan tertinggi yaitu ada jenis mobil MPV menjadi jenis mobil paling diminati, , ada banyak jenis mobil di Indonesia diantaranya adalah jenis Sedan, SUV, SUV 7 Seater, dan *City Car* dan peminatnnya pun tidak main-main sangat banyak

sekali sehingga sangat menarik untuk dikaji, maka dari itu penelitian ini dibuat untuk mengklasifikasi jenis mobil Sedan, SUV 7 Seater, SUV, dan City Car dapat dikategorikan sebagai paling diminati, diminati, atau kurang diminati. Selain itu juga dapat membuktikan apakah benar jenis mobil MPV masuk ke dalam kategori paling diminati atau bukan. Dengan adanya penelitian ini diharapkan bahwasanya jika mobil masuk ke dalam kategori yang peminatnya tinggi maka ketersediaan suku cadang mudah didapatkan dan jika akan dijual kembali maka harganya masih cukup tinggi dikarenakan banyak peminatnya (Prabowo & Muljono, 2018).

Semua anggota GAIKINDO adalah perusahaan agen pemegang merk (APM) yang terdiri dari produsen kendaraan bermotor, distributor kendaraan bermotor, serta pembuat komponen utama kendaraan bermotor (*manufacturer*) (GAIKINDO, t.thn.). Data penjualan mobil di Indonesia disajikan kedalam bentuk grafik pada gambar 1 di bawah ini.



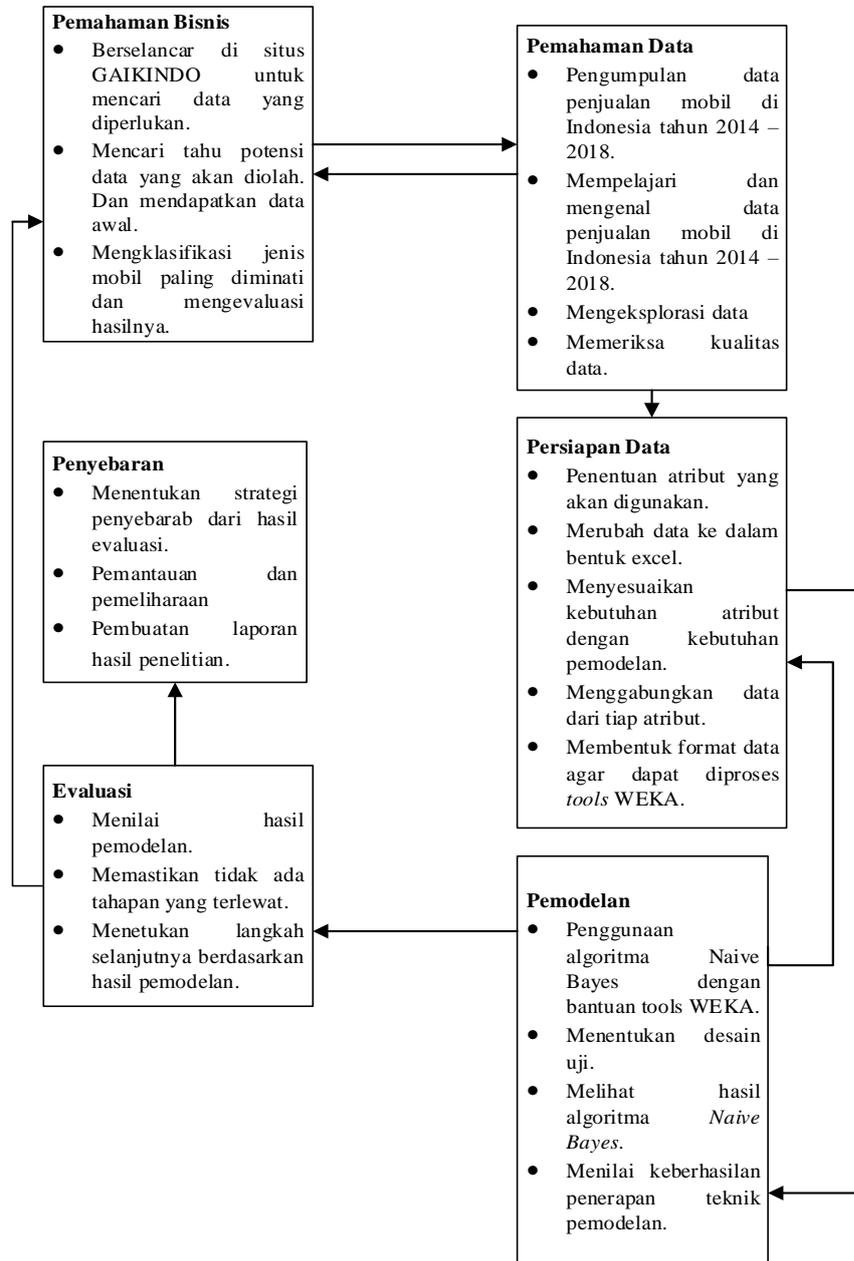
Gambar 1. Grafik Penjualan Mobil di Indonesia

Berdasarkan gambar 1 di atas bahwa penjualan tertinggi terdapat pada tahun 2014 yang mencatatkan angka sebanyak 1.208.028 unit mobil yang terjual, lalu ditahun berikutnya pada tahun 2015 penjualan mengalami penurunan dari tahun sebelumnya, tercatat hanya 1.013.518 unit mobil yang berhasil terjual pada tahun tersebut, lalu pada tahun selanjutnya yaitu tahun 2016 penjualan tercatat sebanyak 1.062.694 unit, mengalami peningkatan walau sedikit dari tahun sebelumnya, lalu 2 tahun setelahnya walau tidak banyak penjualan tetap mengalami kenaikan dari tahun sebelumnya, tercatat tahun 2017 sebanyak 1.077.365 unit terjual dan tahun 2018 sebanyak 1.151.284 unit berhasil terjual.

Data mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstrasi dan menggali pola yang penting atau menarik dari data yang terdapat dalam basis data (Fadlan, Ningsih, & Windarto, 2018) (Waluyo, 2017). Dalam penerapannya pun teknik data mining yang akan digunakan dalam penelitian ini menggunakan algoritma Naive Bayes. Algoritma Naive Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema bayes (aturan bayes) dengan asumsi indenpendensi (ketidaktergantungan) yang kuat (naif). Dalam bayes (terutama Naive Bayes), indenpendensi yang kuat pada fitur adalah sebuah fitur pada sebuah data tidak berkaitan dengan adanya atau tidaknya fitur lain dalam data yang sama (Laia, Buulolo, & Sirait, 2018).

METODE

Metodologi penelitian yang digunakan penelitian ini menggunakan metode yang ada dalam *data mining*, yaitu metode CRISP-DM (*Cross-Industry Standard Process Model for Data Mining*). Berikut merupakan bagian alur CRISP-DM yang ditunjukkan pada gambar di bawah ini.



Gambar 2. Alur CRISP-DM

Pemahaman Bisnis (*Business Understanding*)

Pada tahap ini peneliti akan melakukan observasi penelitian, karena GAIKINDO (Gabungan Industri Kendaraan Bermotor Indonesia) menyediakan data yang dapat diunduh secara bebas oleh publik. Data yang telah diunduh kemudian akan dicari potensi yang dapat diperoleh dari data tersebut setelah diolah dengan teknik data mining guna mengklasifikasikan jenis mobil paling diminati di Indonesia yang pastinya informasi yang dihasilkan dapat bermanfaat bagi produsen mobil, konsumen, hingga para penjual mobil bekas, adapun tahapan ini membutuhkan waktu selama 7 hari, dalam pemahaman bisnis ada beberapa tahap yang harus dilakukan antara lain: Menentukan Tujuan Bisnis (*Determine Business Objectives*), Menilai Situasi (*Assess Situation*), Menentukan Tujuan Data Mining (*Determine Data Mining Goals*).

Pemahaman Data (*Data Understanding*)

Pada tahap ini pengumpulan data akan dilakukan, di dalam penelitian ini data yang didapat dari GAIKINDO (Gabungan Industri Kendaraan Bermotor Indonesia) yaitu data *wholesales* atau penjualan keseluruhan mobil di Indonesia tahun 2014 – 2018, tahapan ini membutuhkan waktu selama 10, dengan tahapan sebagai berikut: Mengumpulkan Data Awal (*Collect Initial Data*), Mendeskripsikan Data (*Describe Data*), Mengeksplorasi Data (*Explore Data*), Verifikasi Kualitas Data (*Verify Data Quality*).

Persiapan Data (*Data Preparation*)

Pada tahap ini perlu dilaksanakan secara intensif. Memilih kasus atau variabel yang ingin dianalisis, melakukan perubahan pada beberapa variabel jika diperlukan sehingga data siap untuk dimodelkan, mengubah nilai angka dengan nilai ordinal dan memilih atribut yang dipakai, tahapan ini membutuhkan waktu selama 20 hari dengan tahapan sebagai berikut: Memilih Data (*Select Data*), Membersihkan Data (*Clean Data*), Membangun Data (*Construct Data*), Mengintegrasikan Data (*Integrate Data*), Format Data.

Pemodelan (*Modeling*)

Tahap ini merupakan gambaran dari rangkaian kegiatan yang dilakukan untuk pelatihan dan pengujian terhadap data yang sudah dikumpulkan dan diolah. Berikut merupakan teknik pemodelan dengan algoritma Naive Bayes dan menggunakan *tools* WEKA untuk membantu menyelesaikan permasalahan data mining, tahapan ini membutuhkan waktu selama 3 hari dengan tahapan sebagai berikut: Memilih Teknik Pemodelan (*Select Modeling Techniques*), Menghasilkan Desain Uji (*Generate Test Design*), Membangun Model (*Build Model*), Menilai Model (*Assess Model*).

Evaluasi

Pada tahap ini hasil dari tahap pemodelan diinterpretasikan ke dalam bentuk grafik menggunakan pengukuran tingkat akurasi yang merupakan tingkat kedekatan nilai klasifikasi dengan nilai aktual, *precision* yaitu tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem, *recall* tingkat keberhasilan sistem dalam menemukan kembali informasi dan *f-measure* yang merupakan salah satu perhitungan evaluasi dengan mengkombinasikan *recall* dan *precision*.

Penyebaran

Tahap terakhir dari metodologi CRISP-DM adalah tahap *Deployment*. Pada tahap ini akan dilakukan pembuatan laporan hasil kegiatan yang sudah dilakukan. Laporan akhir mengenai pengetahuan yang didapat tentang hasil analisis yang telah dilakukan dan yang telah dievaluasi.

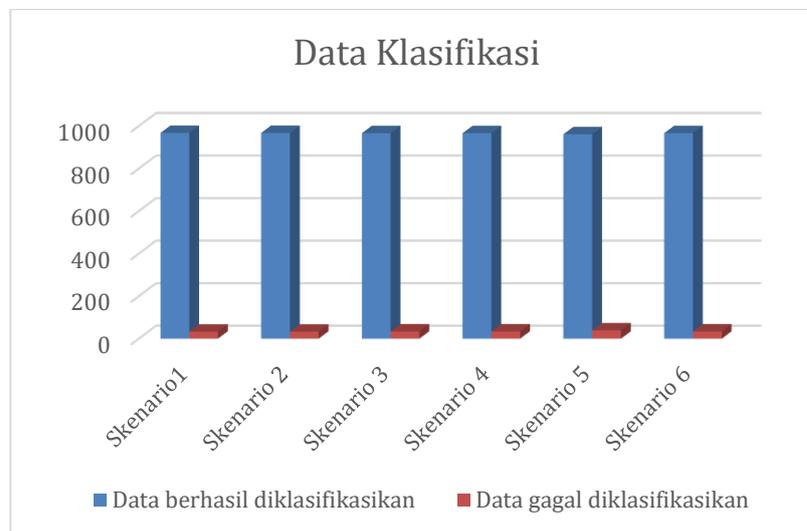
HASIL DAN PEMBAHASAN

Proses awal yang dilakukan pada penelitian ini yaitu mengumpulkan data dari situs GAIKINDO (Gabungan Industri Kendaraan Bermotor Indonesia) yang merupakan asosiasi yang dianggotai para APM (Agen Pemegang Merk). Setelah *dataset* terkumpul maka transformasi data dilakukan guna merubah tipe data numerik menjadi nominal. Selanjutnya melakukan klasifikasi dengan algoritma *Naive Bayes*, pengujian dilakukan dengan 6 skenario menggunakan *cross validation* dengan nilai *Folds* 5 sampai dengan 10 dan menggunakan *Confusion Matrix*.

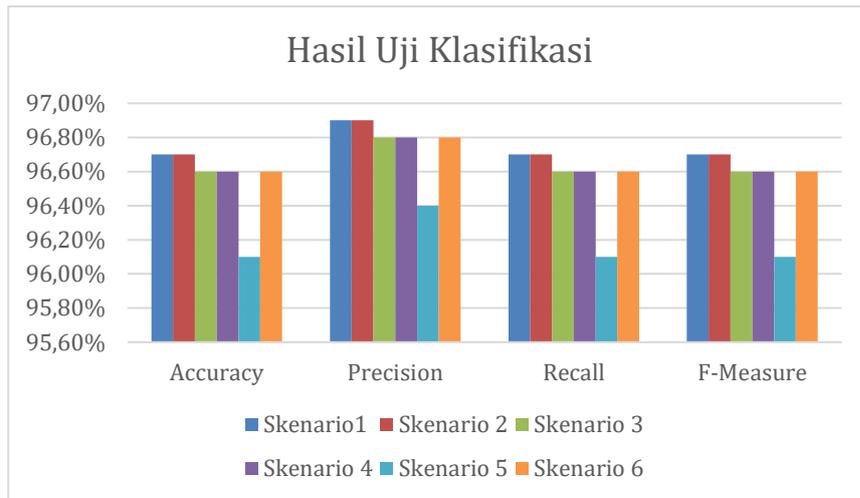
Tabel 1. Hasil Algoritma *Naive Bayes*

Klasifikasi Algoritma Naïve Bayes	Data berhasil diklasifikasi	Data gagal diklasifikasi	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
Skenario 1	967	33	96,7%	0,969	0,967	0,967
Skenario 2	967	33	96,7%	0,969	0,967	0,967
Skenario 3	966	34	96,6%	0,968	0,966	0,966
Skenario 4	966	34	96,6%	0,968	0,966	0,966
Skenario 5	961	39	96,1%	0,964	0,961	0,961
Skenario 6	966	34	96,6%	0,968	0,966	0,966

Tabel 1 menunjukkan hasil klasifikasi pada *dataset* dengan menggunakan algoritma *Naive Bayes* dan menunjukkan bahwa klasifikasi yang terbentuk dengan pengukuran *Cross Validation* menunjukan bahwa pada skenario 1 dan 2 menghasilkan hasil yang terbaik dengan *accuracy* 96,7%, *precision* 0,969, *recall* 0,967, dan *f-measure* 0,967. Dengan hasil tersebut klasifikasi dengan algoritma *Naive Bayes* dapat membentuk model yang baik.



Gambar 3. Data Kalsifikasi Berhasil dan Tidak Berhasil



Gambar 4. Hasil Uji Klasifikasi

Tabel 2. Klasifikasi Jenis Mobil Paling Diminati Skenario 1

Category	Data berhasil diklasifikasikan	Kategori Klasifikasi	Accuracy	Error Rate
City Car	245	Diminati	96,7%	3,3%
Sedan	14	Diminati		
SUV	54	Diminati		
SUV 7 Seater	7 125	Diminati		
MPV	529	Paling Diminati		

Tabel 2 menunjukkan hasil klasifikasi skenario 1 dimana hasilnya adalah jenis mobil *City Car*, *Sedan*, *SUV*, dan *SUV 7 Seater* diklasifikasikan ke dalam kategori “Diminati”, sedangkan jenis mobil *MPV* diklasifikasikan ke dalam kategori “Paling Diminati”.

Tabel 3. Klasifikasi Jenis Mobil Paling Diminati Skenario 2

Category	Data berhasil diklasifikasikan	Kategori Klasifikasi	Accuracy	Error Rate
City Car	245	Diminati	96,7%	3,3%
Sedan	15	Diminati		
SUV	54	Diminati		
SUV 7 Seater	7 125	Diminati		
MPV	528	Paling Diminati		

Tabel 3 menunjukkan hasil klasifikasi skenario 2 dimana hasilnya adalah jenis mobil *City Car*, Sedan, SUV, dan SUV 7 Seater diklasifikasikan ke dalam kategori “Diminati”, sedangkan jenis mobil MPV diklasifikasikan ke dalam kategori “Paling Diminati”.

PENUTUP

Simpulan

Algoritma Naive Bayes dengan metode CRISP-DM menggunakan tools WEKA 3.6.9 dapat digunakan untuk mengklasifikasi jenis mobil paling diminati di Indonesia. Pengujian algoritma Naive Bayes diukur berdasarkan nilai accuracy, precision, recall, dan f-measure yang dihasilkan. Pengujian kinerja dilakukan dengan 6 skenario menggunakan Cross-validation dengan nilai Folds 5 sampai dengan 10. Nilai tertinggi terdapat pada skenario 1 dan 2 dengan accuracy 96,7%, precision 0,969, recall 0,967, dan f-measure 0,967. Hasil yang klasifikasi didapatkan dari penelitian yang telah dilakukan adalah bahwa jenis mobil City Car mendapatkan predikat “Diminati”, selanjutnya jenis mobil Sedan mendapatkan predikat “Diminati”, jenis mobil SUV “Diminati”, lalu SUV 7 Seater “Diminati”, dan yang terakhir jenis mobil MPV menjadi “Paling Diminati”.

Saran

Untuk penelitian selanjutnya yaitu: Menambahkan variabel atau atribut seperti kabupaten/kota atau provinsi, lalu *brand* dari setiap ATPM, dan *Merk* mobilnya pada data penjualan mobil di Indonesia agar hasil klasifikasi dapat lebih terinci.

DAFTAR PUSTAKA

- Dwi, A., Prabowo, R., Studi, P., Informatika, T., Komputer, F. I., Dian, U., & Semarang, N. (2018). Prediksi Nasabah Yang Berpotensi Membuka Simpanan Deposito Menggunakan Naive Bayes Berbasis Particle Swarm Optimization. *Techno.COM*, 17(2), 208–219.
- Fadlan, C., Ningsih, S., & Windarto, A. P. (2018). PENERAPAN METODE NAÏVE BAYES DALAM KLASIFIKASI KELAYAKAN KELUARGA PENERIMA BERAS RASTRA. *JUTIM*, 3(1), 1–8.
- Fathor, R., Fauzi, M. A., & Rizal, S. P. (2017). Prediksi Rating Pada Review Produk Kecantikan Menggunakan Metode Naïve Prediksi Rating Pada Review Produk Kecantikan Menggunakan Metode Naïve Bayes dan Categorical Proportional Difference (CPD). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(August).
- Gaikindo. (2017). Indonesian Automobile Industry Wholesales. Retrieved from Gaikindo.or.id website: <https://files.gaikindo.or.id>
- Indraswari, N. R., & Kurniawan, Y. I. (2018). Aplikasi prediksi usia kelahiran dengan metode naive bayes. *SIMETRIS*, 9(1), 129–138.
- Jabez, J., Gowri, S., Vigneshwari, S., Mayan, J. A., & Srinivasulu, S. (2019). Anomaly Detection by Using CFS Subset and Neural Network with WEKA Tools. *Information and Communication Technology for Intelligent Systems, Smart Innovation, Systems and Technologies* 107, 675–682. <https://doi.org/10.1007/978-981-13-1747-7>
- Kurniawan, F., & Ivandari. (2017). Komparasi Algoritma Data Mining Untuk Klasifikasi Penyakit Kanker Payudara. *Jurnal Stmik*, XII(1), 1–8.
- Laia, D., Bulololo, E., & Sirait, M. J. F. (2018). Implementasi Data Mining Untuk Memprediksi Pemesanan Driver Go-Jek Online Dengan Menggunakan Metode Naive Bayes (Studi Kasus: Pt. Go-Jek Indonesia). *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 2(1), 434–439. <https://doi.org/10.30865/komik.v2i1.972>

- Muhammad, F., Asror, I., & T, I. L. S. S. (2018). Analisis Perbandingan CPU dan GPU (CUDA) Pada Klasifikasi Data Mining dengan Menggunakan Metode K-Nearest Neighbor Kernel Algorithm. *E-Proceeding of Engineering*, 5(3), 7494–7506.
- Pratiwi, R. W., & Nugroho, Y. S. (2017). Prediksi Rating Film Menggunakan Metode Naïve Bayes. *Jurnal Teknik Elektro*, 8(December 2016).
- Priyono. (2016). *Metode Penelitian Kuantitatif* (2016th ed.; T. Chan, Ed.). Sidoarjo: Zifatama.
- Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Citec Journal*, 2(3), 207–217.
- Trihandoko, K. (2017). *Analisis Algoritma Naïve Bayes dengan Test Options dalam Memprediksi Produktivitas Tanaman Padi*. Universitas Singaperbangsa Karawang.
- Waluyo, S. H. (2017). *Klasifikasi Pemanfaat Program Beras Sejahtera (RASTRA) Berdasarkan Tingkat Kemiskinan Dengan Menggunakan Algoritma Decision Tree C4 . 5 Berbasis Particle Swarm Optimization*. 7(2), 19–24.
- Wicaksono, T. P., Hidayat, N., & Rahayudi, B. (2019). Implementasi Metode Naive Bayes Pada Diagnosis Penyakit Lambung. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(1), 227–232.