

Gaussian Naïve Bayes dengan Harmonic Mean untuk Klasifikasi Hasil Produksi Gula Merah

Gaussian Naïve Bayes with Harmonic Mean for Brown Sugar Yield Classification

Jaoda Tifalliyu Haqqesda¹, Elindra Ambar Pambudi^{2*}

^{1,2}Teknik Informatika, Fakultas Teknik dan Sains, Universitas Muhammadiyah Purwokerto

Jl. K. H. Ahmad Dahlan, Kembaran 53182, Indonesia

Email: *elindraambarpambudi@ump.ac.id

ABSTRAK

Industri gula merah pada CV Muara Abadi termasuk salah satu industri makanan yang ada di daerah kabupaten Banyumas. Hasil produksi gula merah yang dimiliki oleh CV tersebut belum diklasifikasi menjadi kelas naik atau turun sehingga CV masih belum dapat mengetahui naik atau turunnya hasil produksi yang didapat untuk minggu berikutnya. Permasalahan tersebut diperlukan adanya teknik data mining dalam pengklasifikasian hasil produksi gula merah untuk membantu CV agar mengetahui naik atau turunnya hasil produksi gula merah pada minggu berikutnya dan agar dapat mengetahui fluktuatif hasil produksi per tahun. Algoritma klasifikasi yang digunakan dalam teknik data mining yaitu algoritma Naïve Bayes dengan Harmonic Mean. Algoritma Naïve Bayes memiliki keuntungan dalam klasifikasi kelas yaitu hanya membutuhkan jumlah data pelatihan (training data) yang sedikit pada saat proses perhitungan. Data set yang digunakan merupakan data hasil produksi dari bulan Juli 2019 hingga bulan April 2022 yang telah diakumulasi per minggu sebanyak 136 data set. Data tersebut nantinya akan dibagi (split) terlebih dahulu menjadi data training dan data testing. Persentase perbandingan data yang digunakan yaitu 90% data training dan 10% data testing. Hasil yang diperoleh dari proses perhitungan klasifikasi hasil produksi yang telah dilakukan menghasilkan nilai accuracy sebesar 71,43% dengan nilai precision sebesar 100,00% dan recall sebesar 33,33%.

Kata Kunci: Klasifikasi, Gaussian Naïve Bayes, Harmonic Mean.

(Dikirim: 25 November 2024, Direvisi: 30 November 2024, Diterima: 1 Desember 2024)

ABSTRACT

The brown sugar industry at CV Muara Abadi is one of the food industries in the Banyumas regency area. The brown sugar production results owned by the CV have not been classified into rising or falling classes so that the CV is still unable to know the rise or fall of the production results obtained for the next week. This problem requires a data mining technique in classifying brown sugar production results to help CVs to find out the rise or fall of brown sugar production results in the following week and to be able to find out the fluctuating production results per year. The classification algorithm used in data mining techniques is the x algorithm with Harmonic Mean. The Naïve Bayes algorithm has the advantage of class classification, which only requires a small amount of training data during the calculation process. The data set used is production data from July 2019 to April 2022 which has been accumulated per week as many as 136 data sets. The data will later be divided (split) first into training

data and testing data. The percentage of data comparison used is 90% training data and 10% testing data. The results obtained from the calculation process of the classification of production results that have been carried out produce an accuracy value of 71.43% with a precision value of 100.00% and recall of 33.33%.

Keywords Classification, Gaussian Naïve Bayes, Harmonic Mean

1. Pendahuluan

Berbagai macam bentuk-bentuk badan wirausaha yang mana dalam setiap bentuk badan wirausaha memiliki perbedaannya masing-masing, salah satu bentuk-bentuk badan wirausaha adalah badan usaha industri. Badan usaha industri adalah badan usaha yang kegiatan produksinya mengolah bahan mentah (bahan yang belum diolah sama sekali) atau bahan setengah jadi (bahan yang sudah diolah namun belum sampai bahan yang siap dikonsumsi) menjadi bahan jadi (bahan yang siap dikonsumsi oleh konsumen) (Mahkamah, 2019). Badan usaha industri yang bergerak pada bidang industri makanan dalam sektor pertanian contohnya adalah gula. Gula adalah produk pangan yang dikonsumsi oleh hampir semua orang dengan berbagai jenis ragam produk gula seperti gula pasir, gula merah, gula semut, dan masih banyak lagi (Prabawati et al., 2018).

Badan usaha industri makanan yang ada di daerah Kabupaten Banyumas salah satunya adalah industri gula merah pada CV Muara Abadi. Industri gula merah tersebut merupakan industri rumahan (*home industry*) dimana alat dan cara pembuatannya masih menggunakan alat dan cara pembuatan yang sederhana, sehingga proses pembuatannya membutuhkan waktu yang cukup lama. Hasil produksi gula merah sudah dipasarkan keluar daerah Kabupaten Banyumas.

Produksi gula merah pada CV Muara Abadi menunjukkan hasil produksi gula merah yang masih belum diklasifikasi menjadi kelas naik atau turun di setiap minggunya. Contohnya hasil produksi gula merah pada minggu sebelumnya termasuk ke dalam kelas naik, sedangkan hasil produksi gula merah pada minggu berikutnya termasuk ke dalam kelas turun. Permasalahan tersebut diperlukan adanya teknik data mining dalam pengklasifikasian hasil produksi gula merah setiap minggunya. Data mining merupakan suatu bentuk implementasi yang diterapkan untuk menemukan pola (Basit, 2020; Ermanto & Humaeroh, 2020) dan model berdasarkan data sebelumnya selama periode waktu tertentu (Basit, 2020).

Salah satu algoritma yang digunakan dalam teknik data mining untuk klasifikasi yaitu *Naïve Bayes*. *Naive bayes* adalah salah satu metode klasifikasi populer dan termasuk di antara sepuluh algoritma terbaik dalam data mining (Jahromi & Taheri, 2018; Mukminin & Riana, 2017; Prasetyo & Pratiwi, 2015). *Naïve bayes* merupakan pengklasifikasi probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari kumpulan data yang diberikan (Ananda et al., 2021; Devita et al., 2018; Rezekika, 2020; Saleh, 2015b, 2015a). Kinerja algoritma *naïve bayes* dalam kebanyakan situasi dunia nyata sering bekerja jauh lebih baik daripada yang diharapkan (Ermanto & Humaeroh, 2020; Saleh, 2015a).

Keuntungan penggunaan algoritma *naïve bayes* dalam proses klasifikasi yaitu untuk menentukan estimasi parameter hanya membutuhkan jumlah data pelatihan (*training data*) yang sedikit (Devita et al., 2018; Ermanto & Humaeroh, 2020; Harahap et al., 2018; Manalu et al., 2017; Saleh, 2015a). Klasifikasi *bayesian* juga memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network* (Ermanto & Humaeroh, 2020; Mukminin & Riana, 2017; Prasetyo & Pratiwi, 2015).

Pengklasifikasi *naive bayes* menggunakan *Gaussian* berkinerja sedikit lebih baik daripada implementasi *naïve bayes* lainnya (*Multinomial*, *Bernoulli*, dan melengkapi asumsi distribusi probabilitas) pada kumpulan data seperti *Gaussian* (Kelly & Johnson, 2021). Menurut (Shah et al., 2020) kinerja

Gaussian Naïve Bayes Classifier lebih baik daripada *Multinomial Naïve Bayes Classifier*. *Gaussian Naïve Bayes* digunakan untuk mengestimasi probabilitas di setiap kelas, distribusi estimasi *Gaussian* cukup baik bila datanya kontinu dan cukup besar (Jahromi & Taheri, 2018).

Perhitungan *Naïve Bayes* dalam data numerik diselesaikan dengan menghitung nilai *mean* terlebih dahulu menggunakan perhitungan *Harmoni Mean*. Penggunaan *harmonic mean* lebih baik digunakan daripada *arithmetic mean* (Nasution & Ginting, 2017; Oliva & Hassanien, 2018; Prabhakar et al., 2019; Ramya et al., 2018).

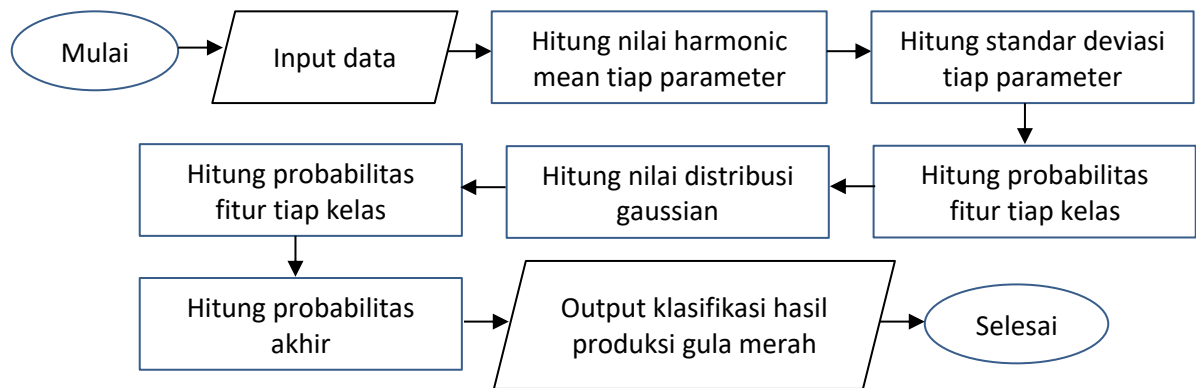
Berdasarkan permasalahan yang telah dijelaskan mengenai klasifikasi hasil produksi gula merah yang berubah-ubah setiap minggunya pada CV Muara Abadi, maka dapat diselesaikan dengan melakukan pengklasifikasian hasil produksi gula merah setiap minggunya menggunakan algoritma *Naïve Bayes* dengan *Harmonic Mean*. Penggunaan algoritma tersebut dilakukan dengan cara mengolah data hasil produksi gula merah pada minggu sebelumnya yang nantinya menghasilkan data klasifikasi kelas naik atau turun produksi gula merah pada minggu berikutnya. Pengklasifikasian tersebut nantinya dapat digunakan pada CV Muara Abadi agar CV tersebut mengetahui klasifikasi hasil produksi gula merah setiap minggunya.

2. Metode

2.1 Pengumpulan Data

Data set yang digunakan merupakan data hasil produksi gula merah bulan Juli 2019 sampai dengan bulan April 2022 yang dibagi menjadi data *training* dan data *testing* dengan persentase perbandingan 90%:10%. Perbandingan data *training* sebesar 90% dan data *testing* sebesar 10% menghasilkan nilai akurasi terbaik atau tertinggi dari perbandingan data lain (Gunawan et al., 2018);(Habib et al., 2021);(Nugroho et al., 2021). Data *training* adalah data untuk menyusun fungsi dari sebuah model sedangkan data *testing* adalah data untuk menguji keakuratan model (Ikhsan, 2020);(Maulana & Yahya, 2019).

2.2 Metode Usulan



- a. Membaca data input
- b. Hitung nilai harmonic mean dari masing-masing parameter dengan rumus yang dapat dilihat pada persamaan 1:

$$H = \frac{n}{\sum_{i=1}^n \left(\frac{1}{x_i}\right)} \quad \dots(1)$$

Keterangan:

H = Rata-rata harmonic

n = Jumlah data sampel

x_i = Nilai x ke-i

- c. Hitung nilai standar deviasi (simpangan baku) dari masing-masing parameter dengan rumus yang dapat dilihat pada persamaan 2:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}} \quad \dots(2)$$

- d. Hitung nilai probabilitas fitur setiap kelas dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
e. Hitung nilai distribusi Gaussian untuk probabilitas fitur data uji, rumus dapat dilihat pada persamaan 3.

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad \dots(3)$$

- f. Hitung nilai probabilitas akhir setiap kelas dengan memasukkan semua nilai distribusi gaussian yang ada ke dalam satu kelas yang sama 4:

$$P(X|Kelas) = P(V1|Kelas) * P(V2|Kelas) \quad \dots(4)$$

- g. Hitung nilai probabilitas akhir melalui perhitungan nilai probabilitas ke dalam rumus Naïve Bayes, rumus dapat dilihat pada persamaan 5:

$$P(Kelas|X) = P(Kelas) * P(X) \quad \dots(5)$$

- h. Hitung nilai normalisasi dengan cara membagikan nilai probabilitas satu kategori dengan jumlah nilai semua kategori, rumus dapat dilihat pada persamaan 6:

$$P(Kelas) = \frac{P(Kelas|X)}{(P(X|Kelas) + P(X|Kelas))} \quad \dots(6)$$

2.3 Metode Evaluasi

Evaluasi dilakukan untuk menguji dan menganalisis algoritma *naïve bayes* dalam mengklasifikasi data. Evaluasi menggunakan *Confusion Matrix* yang terdapat pada Tabel 1 untuk menghitung nilai *accuracy*, *precision*, dan *recall*. *Confusion Matrix* digunakan untuk menunjukkan perbandingan jumlah prediksi tepat dan tidak tepat yang dibuat oleh model dengan hasil klasifikasi sebenarnya pada data uji (Syahputra et al., 2018).

Tabel 1. *Confusion Matrix*

	Predicted	
Actual	Naik	Turun

Naik	TP	FP
Turun	FN	TN

Keterangan:

- TP (*True Positive*) = Jumlah objek naik yang benar diklasifikasikan
- TN (*True Negative*) = Jumlah objek turun yang benar diklasifikasikan
- FP (*False Positive*) = Jumlah objek naik yang salah diklasifikasikan
- FN (*False Negative*) = Jumlah objek turun yang salah diklasifikasikan

Hasil dan Pembahasan

Data yang digunakan dalam perhitungan manual menggunakan data hasil produksi gula merah bulan Juli 2019 sampai dengan bulan April 2022 yang diakumulasi per minggu.

A. Perhitungan Naive Bayes

Data training yang digunakan dimulai dari bulan Juli 2019 minggu 1 hingga bulan Januari 2022 minggu 2, terdapat pada Tabel 2.

Tabel 2. Data *Training*

Bulan	Minggu	Hasil Produksi	Kelas
Juli 2019	Minggu 1	8350 Kg	Naik
	Minggu 2	9360 Kg	Naik
	Minggu 3	9640 Kg	Naik
	Minggu 4	7540 Kg	Turun
Agustus 2019	Minggu 1	9060 Kg	Naik
	Minggu 2	7910 Kg	Turun
	Minggu 3	10750 Kg	Naik
	Minggu 4	10280 Kg	Naik
...
Januari 2022	Minggu 1	7975 Kg	Turun
	Minggu 2	7030 Kg	Turun

Data yang digunakan merupakan data numerik. Langkah pertama yang dilakukan yaitu mengelompokkan jumlah kelas Naik dan Turun berdasarkan data training yang digunakan sebelum menghitung nilai harmonic mean dan standar deviasi dari masing-masing klasifikasi.

Kelas Naik = 64

Kelas Turun = 58

B. Menghitung Nilai Harmonic Mean

Perhitungan nilai *harmonic mean* menggunakan persamaan (1).

$$H(\text{Naik}) = \frac{64}{\frac{1}{8350} + \frac{1}{9360} + \frac{1}{9640} + \dots + \frac{1}{2810} + \frac{1}{5940} + \frac{1}{9860}}$$

$$H(\text{Naik}) = \frac{64}{0,006405451}$$

$$H(\text{Naik}) = 9991,490322$$

$$H(\text{Turun}) = \frac{58}{\frac{1}{7540} + \frac{1}{7910} + \frac{1}{7120} + \dots + \frac{1}{6970} + \frac{1}{7975} + \frac{1}{7030}}$$

$$H(\text{Turun}) = \frac{58}{0,007708015}$$

$$H(\text{Turun}) = 7524,63476$$

C. Menghitung Nilai Standar Deviasi

Perhitungan nilai standar deviasi menggunakan persamaan (2) ketika dihubungkan dengan kasus maka akan menjadi.

$$\sigma(\text{Naik}) = \sqrt{\frac{(8350 - 9991,490322)^2 + (9360 - 9991,490322)^2 + (9640 - 9991,490322)^2 + \dots + (2810 - 9991,490322)^2 + (5940 - 9991,490322)^2 + (9860 - 9991,490322)^2}{64 - 1}}$$

$$\sigma(\text{Naik}) = \sqrt{\frac{2694490,476 + 398780,0263 + 123545,4462 + \dots + 51573803,24 + 16414573,83 + 17289,70467}{63}}$$

$$\sigma(\text{Naik}) = \sqrt{\frac{215931709,1}{63}}$$

$$\sigma(\text{Naik}) = \sqrt{3427487,4}$$

$$\sigma(\text{Naik}) = 1851,347468$$

$$\sigma(\text{Turun}) = \sqrt{\frac{(17880 - 7524,365754)^2 + (18760 - 7524,365754)^2 + (19770 - 7524,365754)^2 + \dots + (6970 - 7524,365754)^2 + (7975 - 7524,365754)^2 + (7030 - 7524,365754)^2}{58 - 1}}$$

$$\sigma(\text{Turun}) = \sqrt{\frac{236,0905913 + 148506,368 + 163729,2892 + \dots + 307619,7173 + 202828,8491 + 244663,5461}{57}}$$

$$\sigma(\text{Turun}) = \sqrt{\frac{83602141}{57}}$$

$$\sigma(\text{Turun}) = \sqrt{1466704,2}$$

$$\sigma(\text{Turun}) = 1211,075649$$

D. Menghitung Nilai Probabilitas Fitur Setiap Kelas

$$P(\text{Naik}) = \frac{64}{122}$$

$$P(\text{Naik}) = 0,524590164$$

$$P(\text{Turun}) = \frac{58}{122}$$

$$P(\text{Turun}) = 0,475409836$$

E. Menghitung nilai distribusi gaussian

Nilai distribusi gaussian dihitung untuk data testing yang datanya berupa numerik atau angka. Data testing yang digunakan dimulai dari bulan Februari 2022 minggu 4 hingga bulan April 2022 minggu 4, seperti pada Tabel 3.

Tabel 3. Data *Testing*

Bulan	Minggu	Hasil Produksi
Januari 2022	Minggu 3	9910 Kg
	Minggu 4	7760 Kg
Februari 2022	Minggu 1	9880 Kg
	Minggu 2	8430 Kg
	Minggu 3	9530 Kg

	Minggu 4	9250 Kg
Maret 2022	Minggu 1	8920 Kg
	Minggu 2	8090 Kg
	Minggu 3	8590 Kg
	Minggu 4	8890 Kg
April 2022	Minggu 1	8100 Kg
	Minggu 2	8330 Kg
	Minggu 3	8350 Kg
	Minggu 4	8500 Kg

Lakukan proses Distribusi *gaussian* dihitung menggunakan persamaan (3).

$$P(\text{Hasil Produksi} = 9910 | \text{Naik}) = \frac{1}{\sqrt{2 \times \pi \times 1851,347468}} \times e^{-\frac{(9910-9991,490322)^2}{2 \times (1851,347468)^2}}$$

$$P(\text{Hasil Produksi} = 9910 | \text{Naik}) = \frac{1}{107,853415} \times 2,718281828^{-0,00096874}$$

$$P(\text{Hasil Produksi} = 9910 | \text{Naik}) = \frac{1}{107,853415} \times 0,99903173$$

$$P(\text{Hasil Produksi} = 9910 | \text{Naik}) = 0,00926287$$

$$P(\text{Hasil Produksi} = 9910 | \text{Turun}) = \frac{1}{\sqrt{2 \times \pi \times 1211,075649}} \times e^{-\frac{(9910-7524,63476)^2}{2 \times (1211,075649)^2}}$$

$$P(\text{Hasil Produksi} = 9910 | \text{Turun}) = \frac{1}{87,2319478} \times 2,718281828^{-1,93971191}$$

$$P(\text{Hasil Produksi} = 9910 | \text{Turun}) = \frac{1}{87,2319478} \times 0,14374535$$

$$P(\text{Hasil Produksi} = 9910 | \text{Turun}) = 0,00164785$$

Berikut hasil keseluruhan dari nilai distribusi *gaussian* pada Tabel 4:

Tabel 4. Distribusi *Gaussian*

Bulan	Minggu	Naik	Turun
	Minggu 3	0,009262866	0,001647852
	Minggu 4	0,004484237	0,011249232
Februari 2022	Minggu 1	0,009255046	0,001729714
	Minggu 2	0,006496681	0,008668999
	Minggu 3	0,008988211	0,002910303
	Minggu 4	0,00855723	0,004155212
Maret 2022	Minggu 1	0,007842042	0,005902858
	Minggu 2	0,00547138	0,010280203
	Minggu 3	0,006961879	0,007785489
	Minggu 4	0,007767819	0,006071894
April 2022	Minggu 1	0,005501738	0,010240303
	Minggu 2	0,006198298	0,009189598
	Minggu 3	0,006258318	0,009087991
	Minggu 4	0,006702409	0,008288536

F. Menghitung Nilai Probabilitas Akhir Setiap Kelas

G. Menghitung Nilai Probabilitas Akhir

H. Menghitung nilai normalisasi

Perhitungan nilai normalisasi berdasarkan persamaan (6).

$$P(\text{Naik}) = \frac{0,004859208}{0,004859208 + 0,000783405}$$

$$P(\text{Naik}) = 0,861162716$$

$$P(\text{Turun}) = \frac{0,000783405}{0,004859208 + 0,000783405}$$

$$P(\text{Turun}) = 0,138837284$$

Tabel 1. Hasil Klasifikasi

Bulan	Minggu	Naik	Turun	Kelas
Januari 2022	Minggu 3	0,861162716	0,138837284	Naik
	Minggu 4	0,305489591	0,694510409	Turun
Februari 2022	Minggu 1	0,855159243	0,144840757	Naik
	Minggu 2	0,452636961	0,547363039	Turun
	Minggu 3	0,773134648	0,226865352	Naik
	Minggu 4	0,694417428	0,305582572	Naik
Maret 2022	Minggu 1	0,594476596	0,405523404	Naik
	Minggu 2	0,369992504	0,630007496	Turun
	Minggu 3	0,49665701	0,50334299	Turun
	Minggu 4	0,585346062	0,414653938	Naik
April 2022	Minggu 1	0,372191415	0,627808585	Turun
	Minggu 2	0,426692794	0,573307206	Turun
	Minggu 3	0,431777579	0,568222421	Turun
	Minggu 4	0,471539233	0,528460767	Turun

Berdasarkan tabel 5, penentuan kelas naik maupun turun hasil produksi dapat dilihat dari nilai normalisasi masing-masing kelas. Jika nilai normalisasi kelas naik lebih besar daripada kelas turun maka diperoleh hasil klasifikasi kelas naik, dan jika nilai normalisasi kelas turun lebih besar daripada kelas naik maka diperoleh hasil klasifikasi kelas turun. Kinerja penggunaan naïve bayes pada minggu-minggu di 4 bulan antara Januari 2022 hingga April 2022 mendapatkan hasil akurasi sebesar **71,43**, *recall* 33,33, dan presisi sebesar 100.

Kesimpulan

Berdasarkan hasil pengujian yang telah dilakukan mengenai klasifikasi hasil produksi gula merah dengan menerapkan algoritma Naive Bayes dengan Harmonic Mean, maka dapat disimpulkan sebagai berikut:

- A. Hasil klasifikasi algoritma Naive Bayes yang diperoleh dapat membantu perusahaan dalam mengklasifikasi kelas naik atau turun hasil produksi gula merah.
- B. Algoritma Naive Bayes memanfaatkan data training untuk mendapatkan hasil nilai probabilitas dari kelas yang berbeda untuk mengklasifikasi hasil produksi gula merah.
- C. Algoritma Naive Bayes dengan persentase perbandingan 10% data uji berhasil mengklasifikasi hasil produksi gula merah dengan nilai akurasi sebesar 71,43%, precision sebesar 100,00%, dan recall sebesar 33,33%.

DAFTAR PUSTAKA

- Ananda, W., Safii, M., & Fauzan, M. (2021). Prediksi Jumlah Hasil Panen Sawit Menggunakan Algoritma Naive Bayes. In *TIN: Terapan Informatika Nusantara* (Vol. 1, Issue 10, pp. 513–519).
- Basit, A. (2020). Implementasi Algoritma Naive Bayes Untuk Memprediksi Hasil Panen Padi. *Jurnal Teknik*

Informatika Kaputama (JTIK), 4(2), 208–213.

- Devita, R. N., Herwanto, H. W., & Wibawa, A. P. (2018). Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTII)*, 5(4), 427–434. <https://doi.org/10.25126/jtiik.201854773>
- Ermanto, & Humaeroh, E. (2020). Penerapan Data Mining menggunakan Metode Algoritma Naive Bayes untuk Menentukan Kelayakan Kredit Rumah Bersubsidi. *Jurnal Teknologi Pelita Bangsa*, 10(4), 117–134.
- Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 4(2), 113–118. <https://doi.org/10.26418/jp.v4i2.27526>
- Habib, S. M., Haerani, E., Gusti, S. K., & Ramadhani, S. (2021). Klasifikasi Berita Menggunakan Metode Naïve Bayes. *SCAN - Jurnal Teknologi Informasi Dan Komunikasi*, 16(3), 248–258. <https://doi.org/10.33005/scan.v16i3.2870>
- Harahap, F., Harahap, A. Y. N., Ekadiansyah, E., Sari, R. N., Adawiyah, R., & Harahap, C. B. (2018). Implementation of Naïve Bayes Classification Method for Predicting Purchase. *6th International Conference on Cyber and IT Service Management, CITSM 2018*, 1–6. <https://doi.org/10.1109/CITSM.2018.8674324>
- Ikhsan, M. (2020). *Peramalan Energi Photovoltaic dengan Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbors*.
- Jahromi, A. H., & Taheri, M. (2018). A Non-Parametric Mixture of Gaussian Naive Bayes Classifiers based on Local Independent Features. *19th CSI International Symposium on Artificial Intelligence and Signal Processing, AISP 2017*, 1, 209–212. <https://doi.org/10.1109/AISP.2017.8324083>
- Kelly, A., & Johnson, M. A. (2021). Investigating the Statistical Assumptions of Naïve Bayes Classifiers. *2021 55th Annual Conference on Information Sciences and Systems, CISS 2021*. <https://doi.org/10.1109/CISS50987.2021.9400215>
- Mahkamah, W. (2019). *Peningkatan Hasil Belajar Ips Materi Keragaman Kegiatan Ekonomi Masyarakat Melalui Metode Cooperative Script Dan Media Flat Opaque Picture Pada Siswa Kelas Iv Mi Ma'arif Global Blotongan Salatiga Tahun Pelajaran 2018/2019*.
- Manalu, E., Sianturi, F. A., & Manalu, M. R. (2017). Penerapan Algoritma Naive Bayes Untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan dan Jumlah Pemesanan Pada CV. Papa dan Mama Pastries. *Jurnal Mantik Penusa*, 1(2), 16–21.
- Maulana, D., & Yahya, R. (2019). Implementasi Algoritma Naive Bayes untuk Klasifikasi Penderita Penyakit Jantung di Indonesia menggunakan Rapid Miner. *SIGMA-Jurnal Teknologi Pelita Bangsa*, 10(2), 191–197.
- Mukminin, A., & Riana, D. (2017). Komparasi Algoritma C4.5, Naïve Bayes Dan Neural Network Untuk Klasifikasi Tanah. *Jurnal Informatika*, 4(1), 21–31.
- Nasution, A. Y., & Ginting, G. (2017). Implementasi Metode Harmonic Mean Filter Dan Canny. *Jurnal Pelita Informatika*, 6(1), 72–76.
- Nugroho, R. A., Cholissodin, I., & Indriati. (2021). Implementasi Naïve Bayes Classifier untuk Klasifikasi Emosi Tweet Berbahasa Indonesia pada Spark. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(1), 301–310.
- Oliva, D. A., & Hassanién, A. E. (2018). *Studies in Computational Intelligence: Advances in Soft Computing and Machine Learning in Image Processing* (A. E. Hassanién & D. A. Oliva, Eds.).
- Prabawati, E. S., Ayuningtyas, A. P., Putri, R., Pratama, P. K., Indarto, M., Erawati, E., Kimia, T., Teknik, F., Surakarta, U. M., Sipil, T., Teknik, F., & Muhammadiyah, U. (2018). Wirausaha Pembuatan Gula Semut Kesehatan Dari Ekstrak Buah Kersen. *Jurnal Teknologi Bahan Alam*, 2(2), 118–123.
- Prabhakar, S., Gowda, S., Rukmini, T., Khan, S. A., & Vikramadhithan. (2019). Evaluation of Digital Filters for

- Noise Present in Digital Images. *International Journal of Engineering Science and Computing*, 9(5), 22157–22162.
- Prasetio, R. T., & Pratiwi. (2015). Penerapan Teknik Bagging Pada Algoritma Klasifikasi Untuk Mengatasi Ketidakseimbangan Kelas Dataset Medis. *Jurnal Informatika*, 11(2), 395–403.
- Ramya, A., Murugan, D., & Manish, T. I. (2018). A Comparative Analysis on Various Noise Reduction Techniques Tested for Medical Images. *Journal of Network and Information Security*, 6(June), 18–23.
- Rezekika, D. (2020). Penerapan Algoritma Naïve Bayes Untuk Memprediksi Penjualan Spare Part Sepeda Motor. *Jurnal Pelita Informatika*, 8(3), 326–329.
- Saleh, A. (2015a). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Creative Information Technology Journal*, 2(3), 207–217.
- Saleh, A. (2015b). Penerapan Data Mining Dengan Metode Klasifikasi Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa Dalam Mengikuti English Proficiency Test. *Universitas Potensi Utama, June*, 1–6.
- Shah, K., Punjabi, R., Shah, P., & Rao, M. (2020). Real Time Diabetes Prediction using Naïve Bayes Classifier on Big Data of Healthcare. *International Research Journal of Engineering and Technology (IRJET)*, 07(05), 102–107.
- Syahputra, I. K., Bachtiar, F. A., & Wicaksono, S. A. (2018). Implementasi Data Mining untuk Prediksi Mahasiswa Pengambil Mata Kuliah dengan Algoritme Naive Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(11), 5902–5910.