

## Perbandingan Metode Backpropagation dan Naïve Bayes dalam Memprediksi Tingkat Pencemaran Air Sungai

Betran Roivaldo Sitinjak

Program Studi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

E-Mail : betransitnajak12@gmail.com

### Article Info

#### Article history:

Received Jun 09, 2025

Revised Jul 11, 2025

Accepted Jul 30, 2025

#### Kata Kunci:

Pencemaran  
Air  
Sungai  
Prediksi  
Perbandingan

#### Keywords:

Pollution  
Water  
River  
Prediction  
Comparasion

### ABSTRAK

Backpropagation dan Naïve Bayes merupakan dua algoritma yang umum digunakan dalam pengembangan model klasifikasi, meskipun menerapkan pendekatan pemrosesan data yang berbeda. Penelitian ini bertujuan untuk mengevaluasi dan membandingkan performa kedua algoritma dalam mengklasifikasikan tingkat pencemaran air sungai. Model klasifikasi dikonstruksi menggunakan 150 sampel data dengan 22 neuron input dan 3 neuron output. Evaluasi performa dilakukan dengan menerapkan validation matrix melalui empat skenario pengujian yang mengombinasikan variasi pembagian data (split) dan jumlah lipatan (fold). Berdasarkan hasil eksperimen, algoritma Naïve Bayes menunjukkan akurasi rata-rata sebesar 72,79%, lebih tinggi dibandingkan Backpropagation yang memperoleh akurasi rata-rata sebesar 64,75%. Temuan ini menunjukkan bahwa dalam konteks data yang digunakan, Naïve Bayes lebih efektif dalam memprediksi tingkat pencemaran air sungai dibandingkan Backpropagation.

### ABSTRACT

*Backpropagation and Naïve Bayes are two commonly used algorithms for developing classification models, although they employ different data processing approaches. This study aims to evaluate and compare the performance of both algorithms in classifying river water pollution levels. The classification models were built using 150 data samples with 22 input neurons and 3 output neurons. Performance evaluation was conducted using a validation matrix across four testing scenarios, combining variations in data splitting and fold numbers. The experimental results show that the Naïve Bayes algorithm achieved an average accuracy of 72.79%, which is higher than the Backpropagation algorithm, which reached an average accuracy of 64.75%. These findings indicate that, within the context of the dataset used, Naïve Bayes is more effective in predicting river water pollution levels compared to Backpropagation.*

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license.



#### Corresponding Author:

Betran Roivaldo Sitinjak

Program Studi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar

Jl. Jend. Sudirman Blok A No. 1,2&3, Pematangsiantar, Indonesia

Email: betransitnajak12@gmail.com

## 1. PENDAHULUAN

Sebelumnya, penulis telah melakukan penelitian tentang klasifikasi kualitas air sungai menggunakan algoritma Learning Vector Quantization (LVQ) dan Perceptron. Namun, kedua metode ini belum mampu memberikan tingkat akurasi yang optimal. Pada pengujian dengan 100 data latih dan 50 data uji, menggunakan parameter learning rate 0,01, penurunan learning rate 0,001, serta maksimal 20.000 iterasi, algoritma LVQ hanya menghasilkan akurasi sebesar 76 persen. Penelitian lanjutan dengan algoritma Perceptron pun belum menunjukkan hasil yang lebih baik. Dengan jumlah data latih dan uji yang sama, serta pengaturan learning rate 0,05 dan jumlah epoch tertentu, performa yang diperoleh masih belum memuaskan. 50.000, dan threshold

bernilai 0, algoritma Perceptron hanya memperoleh akurasi sebesar 75 persen (Rahimi, 2016)(Susetyoko et al., 2022).

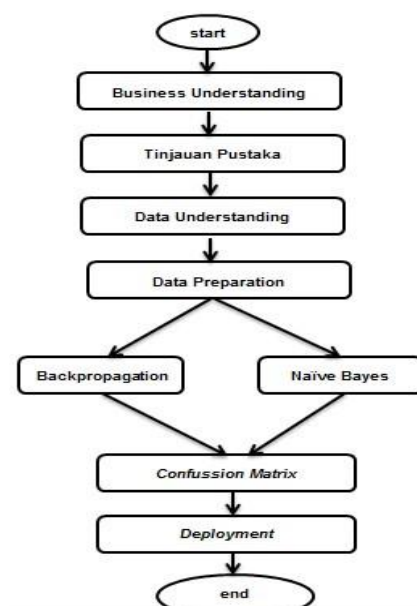
Penelitian yang dilakukan oleh Kim dan timnya membandingkan kinerja tiga algoritma klasifikasi yang berbasis jaringan saraf buatan, yaitu Neuro Fuzzy Network, Backpropagation, dan Support Vector Machine (SVM). Dalam percobaan pertama yang menggunakan dataset Iris dari Fisher-Anderson dengan lima neuron input, akurasi rata-rata yang diperoleh adalah 94,45 persen untuk SVM, 97,335 persen untuk Neuro Fuzzy Network, dan 95,33 persen untuk Backpropagation (Arifin & Ariesta, 2019). Pada percobaan kedua, dengan menggunakan dataset kanker payudara dari Wisconsin dan sembilan neuron input, akurasi yang diperoleh adalah 95,79 persen untuk SVM, 96,185 persen untuk Neuro Fuzzy Network, dan 96,28 persen untuk Backpropagation. Dalam percobaan ketiga yang menggunakan dataset wine dengan 13 neuron input, Neuro Fuzzy Network mencapai akurasi tertinggi sebesar 98,876 persen, diikuti oleh SVM dengan 97,183 persen, dan Backpropagation dengan 90,44 persen. Dari ketiga percobaan tersebut, dapat disimpulkan bahwa Backpropagation menunjukkan kinerja yang cukup baik, meskipun tidak selalu menjadi yang terbaik.

Penelitian lain yang dilakukan oleh Chaurasia dan rekan-rekannya berjudul "Pendekatan Data Mining untuk Mendeteksi Penyakit Jantung" membandingkan tiga algoritma klasifikasi, yaitu Naïve Bayes, J48 Decision Tree, dan Bagging (Ardyanti et al., 2020). Dengan menggunakan sebelas atribut untuk klasifikasi penyakit jantung dan dataset yang terdiri dari 294 data latih, serta metode pengujian fold validation, ditemukan bahwa algoritma Naïve Bayes memiliki akurasi terendah sebesar 82,31 persen, diikuti oleh J48 dengan akurasi 84,35 persen, dan algoritma Bagging yang memiliki akurasi tertinggi yaitu 94,50 persen (Yasar & Saritas, 2019).

Berdasarkan hasil dari berbagai penelitian sebelumnya, studi ini bertujuan untuk membandingkan dua algoritma klasifikasi, yaitu Naïve Bayes dan Backpropagation, dalam mengklasifikasikan tingkat pencemaran air sungai. Data yang digunakan dalam penelitian ini mencakup 22 parameter yang merepresentasikan kandungan fisika, kimia, dan biologi dari air sungai (Kusumadewi, 2009).

## 2. METODE PENELITIAN

Dalam menyelesaikan masalah penelitian ini, langkah-langkah yang perlu dilakukan sebagai landasan berfikir digambarkan pada Gambar 1 :



Gambar 1. Kerangka Penelitian

Adapun penjelasan dari kerangka yang tergambar di Gambar 1 adalah :

### 1. Memahami Kebutuhan Bisnis

Pertama-tama ditetapkan apa yang sebenarnya ingin dicapai, menilai kondisi saat ini, menetapkan target, lalu menyusun rencana kerja. Tujuannya di sini jelas: membuat model yang mampu



kedua metode ini, yang pada dasarnya melibatkan pertukaran peran antara data latih dan data uji. Pada percobaan pertama, sebagian data digunakan untuk pelatihan, sementara sisanya digunakan untuk pengujian. Pada percobaan kedua, data yang sebelumnya berfungsi sebagai data uji akan digunakan sebagai data latih, dan sebaliknya.

Dalam metode cross validation, pengujian dilakukan dengan dua skema lipatan, yaitu 10 fold dan 40 fold. Sementara itu, pada metode data split, data dibagi menjadi dua proporsi: 80% untuk pelatihan dan 20% untuk pengujian, serta 50% untuk pelatihan dan 50% untuk pengujian. Hasil akurasi dari pengujian algoritma Naïve Bayes dengan kedua metode tersebut, baik cross validation maupun data split—dapat dilihat pada Tabel 1 :

Tabel 1. Nilai akurasi Algoritma Naïve Bayes

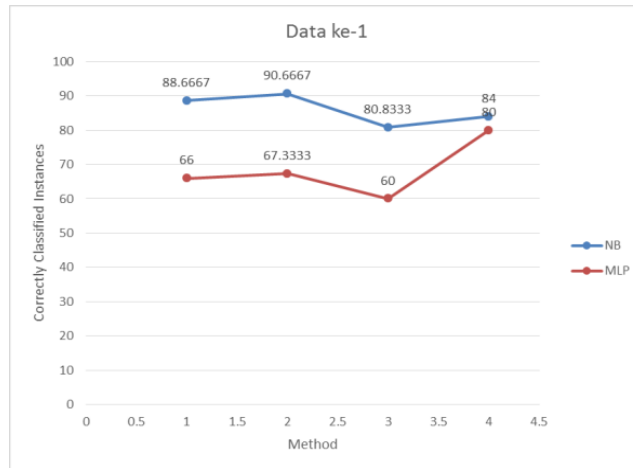
Data ke-	Cross Validation fold	Percentage Split (%)	Correctly Classified Instances(%)	Incorrectly Classified Instance (10%)
1	10	-	88.6667	11.3333
	40	-	90.6667	9.3333
	-	20	80.8333	19.1667
	-	50	84	16
2	10		84	18
	40		82	18.6667
		20	81.33333	30
		50	70	17.33333
.....	.....	.....	.....	.....

Sementara itu, hasil akurasi dari pengujian cross validation fold dan persentase split untuk algoritma Backpropagation dapat dilihat pada Tabel 2 :

Tabel 2. Nilai akurasi Algoritma Backpropagation

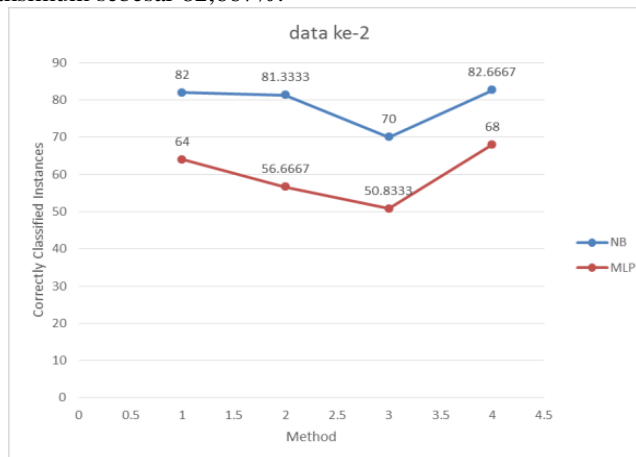
Data ke-	Cross Validation fold	Percentage Split (%)	Correctly Classified Instances(%)	Incorrectly Classified Instance (10%)
1	10	-	66	11.3333
	40	-	67.3333	9.3333
	-	20	60	19.1667
	-	50	80	16
2	10		56.6667	18
	40		50.8333	18.6667
		20	68	30
		50	58	17.33333
.....	.....	.....	.....	.....

Perbandingan tingkat akurasi dalam mengklasifikasikan tingkat pencemaran air sungai menggunakan algoritma Naïve Bayes dan Backpropagation untuk keempat kelas dapat dilihat pada Gambar 3 hingga Gambar 7. Pada pengujian pertama, yang ditampilkan pada Gambar 5, algoritma Naïve Bayes menunjukkan performa yang lebih unggul dibandingkan Backpropagation, dengan nilai akurasi tertinggi yang berhasil dicapai sebesar 90,67%.



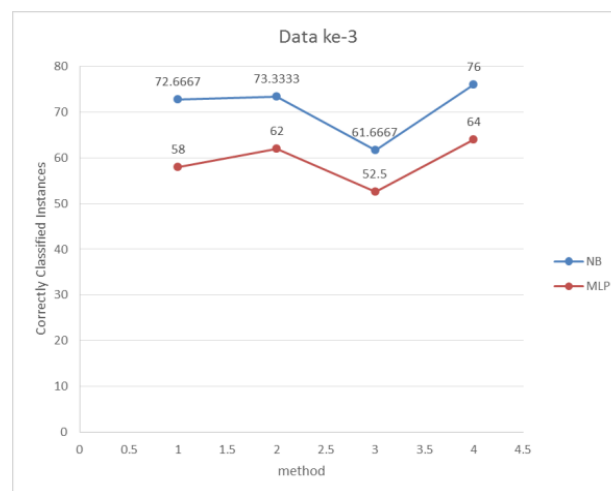
Gambar 3. Perbandingan Akurasi Backpropagation dan Naive Bayes pada Dataset Kelas Pertama

Pada pengujian terhadap dataset kedua, algoritma Naive Bayes kembali menunjukkan performa yang baik, dengan akurasi maksimum sebesar 82,667%.



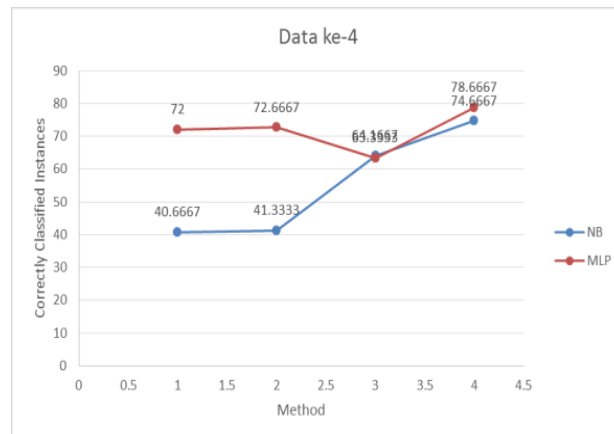
Gambar 4. Perbandingan Akurasi Backpropagation dan Naive Bayes pada Dataset Kelas Kedua

Pada pengujian dataset ketiga, algoritma Naive Bayes kembali mencatat akurasi tertinggi, yaitu sebesar 76%.

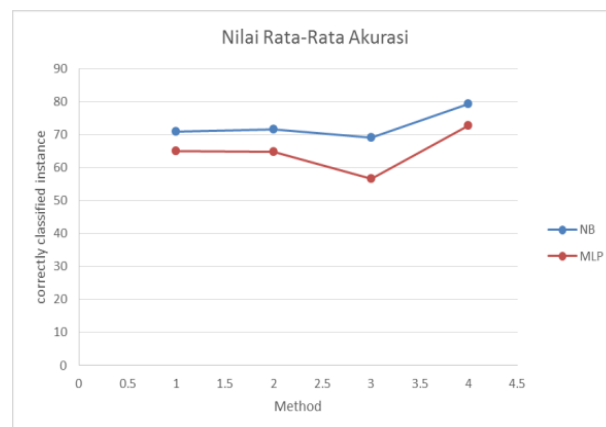


Gambar 5. Perbandingan Tingkat Akurasi Backpropagation dan Naive Bayes pada Dataset Kelas Ketiga

Pada dataset keempat, akurasi yang dicapai oleh Backpropagation lebih tinggi dibandingkan dengan Naïve Bayes..



Gambar 6. Perbandingan Hasil Akurasi Backpropagation dan Naïve Bayes pada Dataset Kelas Keempat



Gambar 7. Perbandingan Akurasi Backpropagation dan Naïve Bayes pada Keseluruhan Dataset

Hasil pengujian menunjukkan bahwa algoritma Naïve Bayes mampu menghasilkan akurasi yang lebih tinggi dibandingkan Backpropagation, dengan akurasi tertinggi mencapai 90,67%. Pengujian dilakukan dengan metode *cross validation* dan *data split*, serta membagi data ke dalam empat kelas pencemaran: A, B, C, dan D.

#### 4. KESIMPULAN

Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma Naïve Bayes lebih efektif dalam melakukan klasifikasi data pencemaran air sungai pada dataset yang digunakan. Ke depan, penelitian ini dapat dikembangkan lebih lanjut dengan menambah jumlah data, mencoba algoritma klasifikasi lain seperti Random Forest atau XGBoost, serta mengintegrasikan metode optimasi untuk meningkatkan akurasi dan efisiensi sistem klasifikasi. Selain itu, sistem ini berpotensi untuk diterapkan dalam pemantauan kualitas air secara real-time sebagai bagian dari pengelolaan lingkungan berbasis data.

#### ACKNOWLEDGEMENTS

Terima kasih disampaikan kepada semua pihak yang telah mendukung terlaksananya penelitian ini, baik secara langsung maupun tidak langsung, serta kepada semua yang telah memberikan semangat dan bantuan selama proses penelitian berlangsung.

#### REFERENCES

- Ardyanti, H., Goejantoro, R., Deny, F., & Amijaya, T. (2020). Perbandingan Metode Klasifikasi Naïve Bayes Dan Jaringan Saraf Tiruan (Studi Kasus: Pt Asuransi Jiwa Bersama Bumiputera Tahun 2018) Comparison Of Naïve Bayes And Artificial Neural Networks Classification Methods (Case Study: Pt Asuransi Jiwa Bersama Bumi. *Jurnal EKSPONENSIAL*, 11(2), 145–152.
- Arifin, T., & Ariesta, D. (2019). Prediksi Penyakit Ginjal Kronis Menggunakan Algoritma Naive Bayes Classifier Berbasis

- Particle Swarm Optimization. *Jurnal Tekno Insentif*, 13(1), 26–30. <https://doi.org/10.36787/jti.v13i1.97>
- Kusumadewi, S. (2009). Klasifikasi Status Gizi Menggunakan Naive Bayesian Classification. *CommIT (Communication and Information Technology) Journal*, 3(1), 6. <https://doi.org/10.21512/commit.v3i1.506>
- Rahimi, M. R. (2016). Penerapan Algoritma Learning Vector Quantization Pencemaran Air Sungai. *Semnasteknomedia Online*, 1(4), 1–6.
- Susetyoko, R., Yuwono, W., Purwantini, E., & Ramadijanti, N. (2022). Perbandingan Metode Random Forest, Regresi Logistik, Naive Bayes, dan Multilayer Perceptron Pada Klasifikasi Uang Kuliah Tunggal (UKT). *Jurnal Infomedia*, 7(1), 8. <https://doi.org/10.30811/jim.v7i1.2916>
- Yasar, A., & Saritas, M. M. (2019). Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification. *International Journal of Intelligent Systems and Applications in Engineering*, 7(2), 88–91. <https://doi.org/10.18201/ijisae.2019252786>