

# **PENERAPAN DATA MINING DALAM PENGELOMPOKAN KATEGORI JURNAL MENGGUNAKAN METODE NAÏVE BAYES**

Muhammad Dimas Prasetya Himawan<sup>1</sup>, Muhammad Fatih 'Ad-Li<sup>2</sup>

<sup>1</sup>Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang

<sup>2</sup>Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang

[himawan.muhammad@student.unmer.ac.id](mailto:himawan.muhammad@student.unmer.ac.id) , [32602200017@std.nissula.ac.id](mailto:32602200017@std.nissula.ac.id)

## **Abstrak**

Pada penelitian ini, fokus utama adalah menerapkan metode naïve bayes dalam mengklasifikasi jurnal ilmiah. Menemukan akurasi metode naïve bayes dalam mengklasifikasi jurnal ilmiah. Dan dampak dari penerapan metode naïve bayes dalam mengklasifikasi jurnal ilmiah. Hasil penelitian ini, penerapan klaifikasi Naïve Bayes dengan pengujian menggunakan Algoritma Multinomial Naïve Bayes mendapatkan akurasi sebesar 84,62%, ada pula F1 score senilai 0,78. Hasil penerapan algoritma Multinomial Naïve Bayes menghasilkan 1 data loss. Berdasarkan algoritma tersebut, menunjukan bahwa jurnal dengan tema teknologi, lebih mendominasi daripada jurnal dengan kategori lainnya.

## **Abstract**

*In this research, the main focus is applying the Naïve Bayes method in classifying scientific journals. Finding the accuracy of the naïve Bayes method in classifying scientific journals. And the impact of applying the naïve Bayes method in classifying scientific journals. The results of this research, the application of Naïve Bayes classification by testing using the Multinomial Naïve Bayes Algorithm, obtained an accuracy of 84.62%, there was also an F1 score of 0.78. The results of applying the Multinomial Naïve Bayes algorithm produce 1 data loss. Based on this algorithm, it shows that journals with a technology theme dominate more than journals with other categories.*

## **1. Pendahuluan**

### **1.1 Latar Belakang**

Pada era digital, pertumbuhan informasi yang disimpan dalam bentuk digital semakin bertambah. Khususnya di bidang akademik, jurnal-jurnal ilmiah menjadi salah satu sumber informasi yang sangat penting. Dalam mengelola banyaknya jurnal yang semakin meningkat, maka diperlukan sebuah sistem pengelola basis data yang efisien dan cerdas. Salah satu cara yang dapat digunakan untuk mengoptimalkan database jurnal adalah dengan menerapkan teknik Data Mining.

Data Mining adalah proses menemukan pola-pola menarik dan berharga dari sebuah data yang banyak untuk mendukung pengambilan keputusan. Salah satu metode Data Mining yang umum digunakan adalah Naïve Bayes. Metode ini merupakan metode klasifikasi yang didasarkan pada teorema Bayes dengan asumsi bahwa semua variabel prediktif adalah independen satu sama lain.[1]

Dalam konteks pengelolaan jurnal ilmiah, aplikasi Data Mining dengan metode Naive Bayes mampu melakukan klasifikasi jurnal-jurnal tersebut berdasarkan berbagai kriteria. Kriteria ini dapat mencakup topik, bidang ilmu, penulis, institusi, atau bahkan popularitas jurnal tersebut. Dengan menerapkan teknik Data Mining, maka sistem pengelolaan jurnal dapat secara otomatis mengkategorikan jurnal-jurnal yang baru ditambahkan ke dalam basis data, memastikan bahwa informasi-informasi yang disimpan tersusun dengan baik dan mudah diakses.[2]

Dengan menyatukan konsep Data Mining dan metode Naive Bayes ke dalam pengelolaan database jurnal ilmiah, diharapkan penelitian ini dapat memberikan dampak yang signifikan dalam pengembangan sistem informasi akademik dan memfasilitasi perkembangan pengetahuan di kalangan ilmiah.

### **1.2 Permasalahan**

Pada penelitian ini, fokus utama adalah menerapkan metode naïve bayes dalam mengklasifikasi jurnal ilmiah. Menemukan akurasi metode naïve bayes dalam mengklasifikasi jurnal ilmiah. Dan dampak dari penerapan metode naïve bayes dalam mengklasifikasi jurnal ilmiah.

### **1.3 Tujuan Penelitian**

Tujuan utama penelitian ini adalah menerapkan metode Naïve Bayes Clustering pada data jurnal ilmiah untuk mengidentifikasi kategori yang sesuai. Dengan demikian, Meningkatkan aksesibilitas informasi bagi peneliti, akademisi, dan masyarakat umum dengan menyediakan antarmuka yang memungkinkan pencarian dan akses jurnal-jurnal ilmiah secara efisien dan akurat.

## **2. Landasan Teori**

### **2.1 Data Mining**

Data mining adalah suatu proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi atau data yang bermanfaat dan pengetahuan yang terdiri m,k/dari berbagai database besar. Teknologi data mining merupakan salah satu alat bantu untuk menggali data pada basis data berukuran besar dengan spesifikasi kerumitan tinggi dan banyak digunakan pada lingkungan aplikasi bisnis seperti perbankan, provider telekomunikasi, perusahaan pertambangan, perminyakan, dan lain-lain[1].

Data mining merupakan suatu pengetahuan yang dapat dibagi menjadi karakteristik, diskriminasi, asosiasi, klasifikasi, clustering, trend, dan outlier. Pada proses data mining melibatkan berbagai teknik seperti database, machine learning, statistik, dan visualisasi, dan dapat digunakan untuk analisis data dalam jumlah besar untuk mengetahui hubungan antar data yang disajikan dalam bentuk yang mudah dipahami[3]. Dengan bantuan software, data mining melakukan proses analisis data untuk menemukan pola atau aturan tersembunyi dalam lingkup himpunan data tersebut[1].

### **2.2 Naïve Bayes**

Naive Bayes adalah sebuah metode klasifikasi yang didasarkan pada teorema Bayes dengan asumsi independensi antar fitur. Metode ini telah diterapkan dalam berbagai bidang, termasuk analisis data gen ekspresi mikroarray[2], manajemen risiko lalu lintas, klasifikasi sentimen[4], klasifikasi kuartil jurnal, dan pembelajaran mesin.

Meskipun memiliki asumsi sederhana tentang independensi fitur, metode ini tetap efektif dan sering digunakan karena algoritmanya yang sederhana dan akurasi yang tinggi[2]. Salah satu penelitian juga memperbaiki kelemahan asumsi independensi dengan menggabungkan pembobotan fitur dan kalibrasi Laplace[4]. Dengan demikian, Naive Bayes telah terbukti menjadi metode klasifikasi yang kuat dan dapat diandalkan dalam berbagai aplikasi.

### **2.3 Sistem Pengelolaan Jurnal Ilmiah**

Sistem pengelolaan jurnal ilmiah melibatkan berbagai tahapan, mulai dari tata kelola editorial, manajemen, hingga akreditasi jurnal. Proses awal penerbitan jurnal meliputi mendefinisikan nama jurnal dan menyusun anggota dewan redaksi yang terdiri dari para ahli di bidang yang sesuai. Dalam pengelolaan jurnal ilmiah, terdapat beberapa hal yang perlu diperhatikan, seperti tata kelola editorial, manajemen, dan akreditasi jurnal[5].

### **2.4 Multinomial Naïve Bayes**

Algoritma ini berguna untuk masalah yang melibatkan data teks dengan fitur diskrit seperti frekuensi kata. Algoritma ini bekerja berdasarkan prinsip teorema Bayes dan mengasumsikan bahwa fitur-fitur bersifat independen secara kondisional terhadap variabel kelas. Model multinomial Naive Bayes memberikan kemampuan untuk mengklasifikasikan data yang tidak dapat direpresentasikan secara numerik, sehingga cocok untuk tugas klasifikasi teks. Model ini umumnya digunakan untuk klasifikasi dokumen, di mana setiap fitur mewakili jumlah atau frekuensi suatu istilah dalam dokumen yang ditugaskan ke berbagai kelas.[6]

Dengan demikian, Multinomial Naive Bayes adalah algoritma yang kuat untuk tugas klasifikasi teks, terutama dalam NLP, karena kemampuannya untuk menangani fitur diskrit seperti frekuensi kata dan asumsinya tentang independensi kondisional di antara fitur-fitur yang diberikan variabel kelas.[7]

### 3. Metode Penelitian

#### 3.1 Pengumpulan Data

Mengumpulkan data berupa jurnal ilmiah beserta metadata mereka dari sumber terpercaya.

#### 3.2 Pemrosesan Data

Melakukan preprocessing data, termasuk membersihkan data, transformasi data ke dalam format yang sesuai, dan pemilihan atribut atribut yang relevan untuk analisis.

#### 3.3 Implementasi Naive Bayes

Mengimplementasikan metode Naive Bayes dalam klasifikasi jurnal ilmiah berdasarkan atribut atribut yang telah ditentukan.

#### 3.4 Validasi dan Evaluasi

Melakukan uji validitas terhadap hasil klasifikasi menggunakan metode Naive Bayes, serta mengevaluasi akurasi sistem dengan membandingkannya dengan hasil klasifikasi manual atau acuan lainnya.

#### 3.5 Analisis Hasil

Menganalisis hasil eksperimen dan evaluasi sistem, termasuk mengidentifikasi keunggulan, potensi perbaikan, dan rekomendasi untuk pengembangan lebih lanjut..

## 4 Hasil dan Pembahasan

#### 4.1 Pengumpulan Data

Pada penelitian ini, data yang digunakan adalah jurnal jurnal acak yang tersebar pada internet sebanyak. Kemudian dikategorikan dengan 4 tema atau 4 kategori yaitu teknologi, kesehatan, humaniora, pertanian. Setelah itu

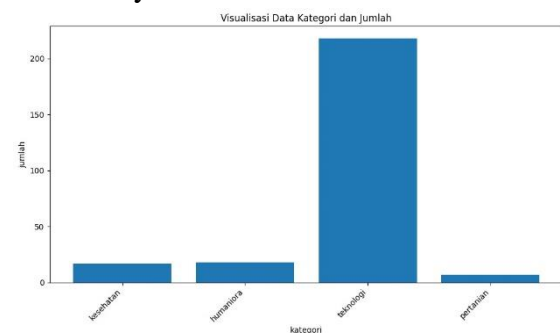
menghitung jumlah kategori yang ada. Data yang akan diproses berjumlah 262.

#### 4.2 Data Preprocessing

Pada data mentah, kami menambahkan kolom label. Kolom tersebut berisi tentang kategori yang ditulis manual dengan membaca abstrak jurnal yang tersedia. Guna menjadi tolak ukur klasifikasi menggunakan metode Naïve Bayes.

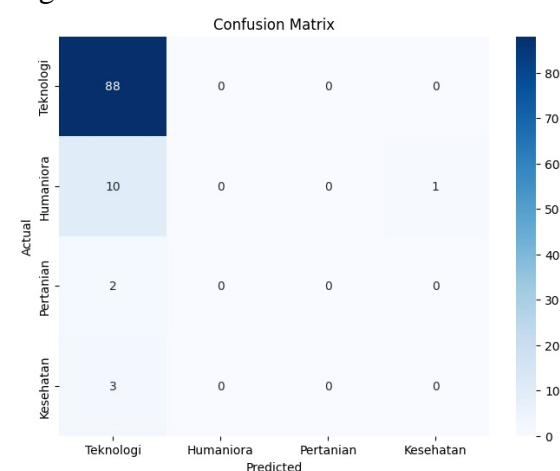
#### 4.3 Penerapan Klasifikasi Naïve Bayes

Setelah proses preprocessing dan penambahan data dilakukan, langkah selanjutnya adalah melakukan proses klasifikasi dengan penerapan metode Naïve Bayes



Berdasarkan diagram diatas, dapat dilihat bahwa kategori tertinggi adalah teknologi.

Peneliti juga menggunakan algoritma Multinomial Naïve Bayes. Berikut hasil confusioin matrix Multinomial Naïve Bayes dari dataset yang digunakan.



Akurasi Model	84,62%
Weigthed F1 Score	0,78
Weigthed Recall	0,85

Dari tabel diatas, didapatkan bahwa terdapat 1 klasifikasi data loss. hasil evaluasi akurasi yang didapat adalah 84,62%. Adapun hasil F1 Score senilai 0,78 dan nilai recall sebesar 0,85.

## 5 Kesimpulan dan Saran

### 5.1 Kesimpulan

Dari hasil penelitian ini, penerapan klaifikasi Naïve Bayes dengan pengujian menggunakan Algoritma Multinomial Naïve Bayes mendapatkan akurasi sebesar 84,62%, ada pula F1 score senilai 0,78. Hasil penerapan algoritma Multinomial Naïve Bayes menghasilkan 1 data loss. Berdasarkan algoritma tersebut, menunjukan bahwa jurnal dengan tema teknologi, lebih mendominasi daripada jurnal dengan kategori lainnya.

### 5.2 Saran

Pada penelitian selanjutnya dapat dikembangkan dengan menggunakan algoritma klasifikasi yang lain atau mengkombinasikan beberapa algoritma untuk mendapatkan hasil penelitian yang lebih baik.

### Daftar Pustaka

- [1] K. Latifah, "ANALISIS DAN PENERAPAN ALGORITHM C45 DALAM DATA MINING UNTUK MENUNJANG STRATEGI PROMOSI PRODI INFORMATIKA UPGRIS," *JURNAL TEKNIK INFORMATIKA*, vol. 11, no. 2, pp. 109–120, Nov. 2018, doi: 10.15408/jti.v11i2.6706.
- [2] A. P. Wibawa *et al.*, "Naïve Bayes Classifier for Journal Quartile Classification," *International Journal of Recent Contributions from Engineering, Science & IT (iJES)*, vol. 7, no. 2, p. 91, Jun. 2019, doi: 10.3991/ijes.v7i2.10659.
- [3] A. Pangestu and D. T. Ridwan, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS PENGELOMPOKAN PELANGGAN BERDASARKAN KUBIKASI AIR TERJUAL MENGGUNAKAN WEKA," 2021. [Online]. Available: <https://jurnal.umj.ac.id/index.php/jus-t-it/index>
- [4] H. Chen, S. Hu, R. Hua, and X. Zhao, "Improved naive Bayes classification algorithm for traffic risk management," *EURASIP J Adv Signal Process*, vol. 2021, no. 1, Dec. 2021, doi: 10.1186/s13634-021-00742-6.
- [5] "Final Panduan Jurnal LPPM UT 14122021".
- [6] M. T. H. Bunga *et al.*, "MULTINOMIAL NAIVE BAYES UNTUK KLASIFIKASI STATUS KREDIT MITRA BINAAN DI PT. ANGKASA PURA I PROGRAM KEMITRAAN," *J-ICON*, vol. 6, no. 2, pp. 30–34, 2018.
- [7] A. Sabrani, I. W. Gede Putu Wirarama Wedashwara, and F. Bimantoro, "METODE MULTINOMIAL NAÏVE BAYES UNTUK KLASIFIKASI ARTIKEL ONLINE TENTANG GEMPA DI INDONESIA (Multinomial Naïve Bayes Method for Classification of Online Article About Earthquake in Indonesia)." [Online]. Available: <http://jtika.if.unram.ac.id/index.php/JTIKA/>