Router: Jurnal Teknik Informatika dan Terapan
Volume. 2 Nomor. 4 Tahun 2024
e-ISSN: 3032-3312; dan p-ISSN: 3026-3611, Hal 177-186
DOI: https://doi.org/10.62951/router.v2i4.302
Available Online at: https://journal.aptii.or.id/index.php/Router

Implementasi Data Mining untuk Mengetahui Minat Baca Peserta Didik Menggunakan Naives Bayes pada Perpustakaan SMP Negeri 2 Palembang

Tiara Siti Nadira [1]*, Tata Sutabri [2]
[1,2] Universitas Bina Darma, Indonesia
Email: tiarasitinadira@gmail.com [1], tata.sutabri@gmail.com [2]*

Abstract. *Students reading interest is a crucial factor in enhancing the quality of education. However, the lack of structured data makes it challenging to identify specific patterns of reading interest. This study aims to implement a data mining method using the Naive Bayes algorithm to analyze students' reading interest at SMP Negeri 2 Palembang's library. The data used includes book borrowing history, types of books, and library visit frequency over one semester. The analysis results indicate that the Naive Bayes method achieves an accuracy rate of 80% in classifying reading interest based on predetermined categories. These findings are expected to assist the school in designing more effective literacy programs.*

Keyword: *Data mining, Reading Interest, Naive Bayes, Library, Education.*

Abstrak. Minat baca peserta didik merupakan faktor penting dalam meningkatkan kualitas pendidikan. Namun, kurangnya data yang terstruktur membuat sulit untuk mengetahui pola minat baca secara spesifik. Penelitian ini bertujuan untuk mengimplementasikan metode data mining menggunakan algoritma Naive Bayes dalam menganalisis minat baca peserta didik di perpustakaan SMP Negeri 2 Palembang. Data yang digunakan meliputi riwayat peminjaman buku, jenis buku, dan frekuensi kunjungan perpustakaan selama satu semester. Hasil analisis menunjukkan bahwa metode Naive Bayes memiliki tingkat akurasi sebesar 80% dalam mengklasifikasikan minat baca berdasarkan kategori yang telah ditentukan. Temuan ini diharapkan dapat membantu pihak sekolah dalam merancang program literasi yang lebih efektif.

Kata Kunci: *Data Mining*, Minat Baca, *Naive Bayes*, Perpustakaan, Pendidikan.

1. PENDAHULUAN

Membaca adalah proses berpikir yang dilakukan dengan sengaja dengan tujuan untuk memahami keseluruhan bahasa tulis. (R. D. Utami, D. C. Wibowo & Y. Susanti, 2018).Minat baca merupakan salah satu faktor penting dalam membentuk karakter dan pengetahuan peserta didik. Minatnya terhadap bacaan dapat didefinisikan sebagai keinginan seseorang untuk membaca dan keinginan untuk membaca lebih banyak dan lebih banyak sendiri tanpa tekanan dari pihak lain. (M. Asgari, M. Khan, & Fazal, 2019). Data menunjukkan bahwa minat baca masyarakat Indonesia hanya 0,001%, atau hanya 1 dari 1000 orang. (Rahmawati, 2020). Untuk meningkatkan kualitas pendidikan di SMP Negeri 2 Palembang, diperlukan strategi khusus untuk mengetahui pola minat baca peserta didik, sehingga program literasi dapat dirancang lebih efektif. Perpustakaan adalah institusi yang bertugas mengumpulkan, mengelola, melestarikan, dan mendistribusikan informasi kepada para pengguna melalui berbagai jenis media, baik cetak maupun digital. Penerapan Teknologi Informasi dan Komunikasi di bidang pendidikan saat ini menjadi kebutuhan yang tak bisa dielakkan (Tata Sutabri, 2012).

Received: November 01, 2024; Revised: November 15, 2024; Accepted: November 28, 2024; Online Available : Desember 02, 2024

Kemampuan membaca sangat erat kaitannya dengan minat serta kebiasaan membaca, di mana setiap peserta didik diharapkan memiliki ketertarikan dan keterampilan membaca yang baik.

Menurut (Abdurrahman, 2003) Membaca berkembang dalam lima tahap: kesiapan membaca, membaca permulaan, kemampuan membaca cepat, membaca luas, dan membaca yang sebenarnya. Perpustakaan sekolah membantu siswa mendapatkan bahan bacaan yang diminati. Apabila materi bacaan menarik bagi siswa, membaca menjadi menyenangkan dan mereka akan membaca dengan sungguh-sungguh, yang pada gilirannya meningkatkan pemahaman bacaan mereka. (R. Masri Sareb Putra, 2008)

Penelitian ini dilakukan untuk mengidentifikasi minat baca peserta didik dengan membandingkan tingkat akurasi algoritma Naive Bayes dan PART pada data mining, berdasarkan data peminjaman buku di perpustakaan. Tujuannya adalah untuk mengetahui seberapa akurat algoritma tersebut dalam mengukur minat baca peserta didik. Data mining sendiri merupakan proses dasar yang metode kecerdasan buatan untuk menemukan pola dari data memanfaatkan yang ada (Han, 2013) Selain itu, data mining atau Knowledge Discovery in Database (KDD) merupakan upaya menemukan pengetahuan yang menarik dari sejumlah besar data yang tersimpan dalam database, gudang warehouses, atau sumber informasi lainnya (Romario, 2013)

Penelitian ini menggunakan pendekatan data mining dengan metode Naive Bayes untuk menganalisis data peminjaman buku di perpustakaan. Algoritma Naive Bayes merupakan sebuah metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik (Yohanes B. W., Silvia A. & Tata S., 2021). Teknik ini dipilih karena kemampuannya dalam mengklasifikasikan data yang bersifat teks dan kategorikal.

2. METODE PENELITIAN

Data Mining

Data perlu diolah lebih lanjut karena bentuk mentah mereka tidak dapat menceritakan banyak hal. (Tata Sutabri, 2012). Istilah "data mining" mengacu pada proses penemuan pengetahuan di dalam basis data yang menggunakan statistik, matematika, AI, dan pembelajaran mesin untuk mengekstrak dan mengidentifikasi informasi penting dari berbagai kumpulan data yang sangat besar. (Turban, 2005). Data mining pada dasarnya adalah bidang ilmu yang bekerja untuk menemukan, menggali, dan mengekstrak pengetahuan dari data yang tersedia. Metode ini juga dikenal sebagai Knowledge Discovery in Database (KDD), yang terdiri dari beberapa langkah berurutan (Thomas, 2004). Tahapan dalam KDD meliputi:

1. Pembersihan Data: Menghilangkan data yang tidak relevan atau berisik.
2. Integrasi Data: Menggabungkan data dari berbagai sumber.
3. Seleksi Data: Memilih data yang relevan untuk dianalisis lebih lanjut.
4. Transformasi Data: Mengubah data ke dalam format yang sesuai untuk dieksplorasi.
5. Data Mining: Menerapkan metode tertentu untuk menemukan pola dalam data.
6. Evaluasi Pola: Mengidentifikasi pola yang signifikan sebagai pengetahuan baru.
7. Presentasi Pengetahuan: Menyajikan pengetahuan yang ditemukan dalam bentuk visualisasi untuk memudahkan pemahaman oleh pengguna.

Naive Bayes

Metode klasifikasi Naive Bayes didasarkan pada teori probabilitas dan statistik oleh ilmuwan Inggris Thomas Bayes. Algoritma ini menggunakan data historis untuk memprediksi kemungkinan kejadian di masa depan, yang dikenal sebagai Teorema Bayes. Keunggulan utama metode Naive Bayes adalah asumsi independen (naif) yang kuat, di mana setiap fitur dianggap tidak bergantung satu sama lain.

Menurut (Karthika, S., & Sairam, 2015), pendekatan klasifikasi ini sangat sederhana. Dalam penerapannya, metode Naive Bayes menentukan nilai probabilitas dan kemungkinan maksimum untuk setiap atribut pada setiap kategori sebelum melakukan klasifikasi.

Metode ini menggunakan model fitur independen, di mana setiap fitur dianggap tidak bergantung pada fitur lainnya. Naive Bayes berfungsi sebagai pengklasifikasi statistik yang mampu memprediksi probabilitas keanggotaan suatu kelas tertentu. Meskipun sederhana, metode ini memiliki kemampuan klasifikasi yang sebanding dengan algoritma seperti decision tree dan neural network.

Keunggulan dari Naive Bayes adalah kemampuannya untuk mencapai akurasi dan kecepatan yang tinggi, terutama ketika diterapkan pada basis data berukuran besar (Eko, 2014). Prediksi bayes didasarkan pada formula teorema bayes dengan formula umum sebagai berikut:

$$P(H|X) = \frac{P(X|H) x\ P(H)}{P(X)}$$

Persamaan dari Teorema Bayes (Salputral et al.,2018) :

Keterangaln :

X : data class yang belum diketahui

H : data hipotesis yaitu suatu class spesifik

P(H|X) : jumlah probabilitas hipotesis H berdasarkan kondisi X (posterior probabilitas)

P(H) :jumlah probabilitas hipotesis H (prior probabilitas)

P(X) : jumlah probabilitas X

P(X|H) : jumlah probabilitas X berdasarkan kondisi pada Hipotesis H.

Confusion Maltrix

Confusion Matrix digunakan untuk mengevaluasi kinerja model klasifikasi dengan mengukur kemampuan model dalam melakukan prediksi secara akurat. Kemampuan prediktif ini mencerminkan tingkat ketepatan aturan klasifikasi yang dihasilkan oleh model dalam mengelompokkan data pada test set ke dalam kelas yang sesuai berdasarkan atribut yang dimilikinya.

Dengan menggunakan persentase, akurasi seratus persen menunjukkan bahwa semua kasus yang dianalisis oleh aturan klasifikasi berhasil dimasukkan dengan tepat ke dalam kelas yang diprediksi. Untuk menghitung nilai akurasi prediktif, perbandingan antara jumlah kasus yang diklasifikasikan dengan benar dan jumlah kasus yang diklasifikasikan secara keliru diperlukan. Tabel yang disebut confusion matrix mengumpulkan perhitungan tersebut.

Table 1. Confusion Matrix

Actual Class	Predicted Class		
		Kelas=Ya	Kelas=Tidak
	Kelas=Ya	al (TP)	b (FN)
	Kelas=Tidak	c (FP)	d (TN)

Rumus yang digunakan untuk confusion matrix adalah :

$$\text{Akurasi}=(TP+TN)/(TP+TN+FP+FN) \times 100\ \%$$

Keteralngaln :

Dalam pengukuran kinerja dengan confusion matrix, empat kata digunakan untuk menggambarkan hasil proses klasifikasi, yaitu:

a. TP (Data Positif Asli): jumlah data positif yang berhasil dianggalp positif;
b. TN (Data Negatif Asli): jumlah data negatif yang berhasil dianggap negatif;
c. FP (Data Negatif Keliru): jumlah data negatif yang keliru dianggap positif;
d. FN (Data Negatif Keliru): jumlah data positif yang keliru dianggap negatif.

3. HASIL DAN PEMBAHASAN

Hasil

1. Pre-processing / Menentukan Variabel

Tahap ini mencakup pembuatan himpunan data yang dimaksud, memilih himpunan data, atau berkonsentrasi pada subset variabel atau sampel data yang akan digunakan untuk proses penemuan. (Erik P. & Tata S., 2024). Penelitian ini memprediksi minat baca berdasarkan usia dengan menggunakan data peminjaman buku, jenis buku, dan frekuensi kunjungan perpustakaan selama satu semester. Setelah dilakukan preprocessing, diperoleh 2 variabel yang digunakan untuk analisis. Variabel-variabel tersebut meliputi nama siswa, kategori buku, jumlah hari keterlambatan, dan tingkat minat baca. Variabel tingkat minat baca dibagi menjadi dua kategori, yaitu tinggi dan rendah. Data yang digunakan dalam penelitian ini terdiri dari 50 entri yang diperoleh dari catatan peminjaman di Perpustakaan SMP Negeri 2 Palembang. Berikut adalah Tabel yang menunjukkan data prediksi yang digunakan.emua simbol maupun kata dapat dibaca (readable).

Tabel 2. Data Peminjalmaln Buku

No	Nama	Kriteria 1	Kriteria 2	Tingkat Minat Baca
1	Intan	Ilmu Sosial	Tepat Waktu	Tinggi
2	Nyimas	Sejarah	Tepat Waktu	Tinggi
3	Satria	Olahraga	Tidak Tepat Waktu	Rendah
4	Hidayat	Agama	Tepat Waktu	Tinggi
5	Putri	Matematika	Tepat Waktu	Tinggi
6	Syawal	Ilmu Sosial	Tepat Waktu	Tinggi
7	Aira	Sejarah	Tepat Waktu	Tinggi

8	Dea	Buku Fiksi	Tepat Waktu	Tinggi
9	Amel	Agama	Tidak Tepat Waktu	Rendalh
10	Khaidir	Olahraga	Tidak Tepat Waktu	Rendah
....	...	....	....	....
50	Dwi	Teknologi Informasi	Tidak Tepat Waktu	Rendah

Tabel 3 menunjukkan keterangan untuk setiap variabel.

Tabel 3. Keteralngaln Valrialbel

Variabel	Kriteria
Kategori Buku	Kriteria 1
Waktu Pengembalian Buku	Kriteria 2

Berikut keterangan dari variabel jumlah hari telat pada tabel 4.

Tabel 4. Keterangan Variabel Waktu Pengembalian

Jumlah Hari Telat	Keteralngaln
Tepat Waktu	< 3 Hari
Tidak Tepat Waktu	> 3 Hari

Dengan menunjukkan Tinggi dan Rendah, minat baca digunakan untuk membuat keputusan.

Setelah tahapan pre-processing, atribut yang sebelumnya terdiri dari tiga atribut dikumpulkan, dan setelah tahap pre-processing selesai, atribut tersebut menjadi dua, yang digunakan dalam dataset untuk mengidentifikasi minat baca di SMP Negeri 2 Palembang, berikut tabel hasil pre-processing data.

Talbel 5. Hasil Data Uji

No	Nama	Kriteria 1	Kriteria 2	Tingkat Minat Baca
1	Intan	Ilmu Sosial	Tepat Waktu	Tinggi

2. Perhitungaln nilai peluang (Probabilitas) untuk setiap kelas (label) pada kasus baru, yaitu "P(XK|Ci)"

Pertama, probabilitas total untuk kelas kejadian masing-masing dihitung dari tiga variabel yang digunakan. Proses ini dimulai dengan membagi jumlah data pada masing-masing kelas kejadian dengan jumlah total data dalam tabel untuk menentukan peluang untuk variabel tingkat minat baca yang "tinggi" dan "rendah". Ini dilakukan dengan membagi jumlah data pada masing-masing kelas kejadian dengan jumlah total data dalam tabel. Maka perhitungannya dapat menjadi sebagai berikut :

a. P(Y = Tinggi) = 25/50: Probabilitas data peminjaman dengan tingkat minat baca "Tinggi" diperoleh dengan membagi jumlah data kategori "Tinggi" dengan total data.
b. P(Y = Rendah) = 25/50: Probabilitas data peminjaman dengan tingkat minat baca "Rendah" diperoleh dengan membagi jumlah data kategori "Rendah" dengan total data.
c. P(Kriteria 1 = "Ilmu Sosial" | Tingkat = "Tinggi") = 3/25 = 0,08: Probabilitas data dengan kriteria "Ilmu Sosial" pada tingkat minat baca "Tinggi".
d. P(Kriteria 2 = "Tepat Waktu" | Tingkat = "Tinggi") = 17/25 = 0,7: Probabilitas data dengan kriterial "Tepat Walktu" paldal tingkat minalt balcal "Tinggi".
e. P(Kriteria 1 = "Ilmu Sosial" | Tingkat = "Rendah") = 3/25 = 0,12: Probabilitas data dengan kriteria "Ilmu Sosial" pada tingkat minat baca "Rendah".
f. P(Kriteria 2 = "Tepat Waktu" | Tingkat = "Rendah") = 6/25 = 0,24: Probabilitas data dengaln kriteria"Tepat Waktu" pada tingkat minat baca "Rendah".

3. Perhitungaln nilai akumulasi peluang dari setiap kelas (label) “P(X|Ci)”

Langkah kedua adalah menghitung probabilitas untuk setiap kasus. Proses ini dilakukaln dengan menghitung jumlah kasus yang terjadi pada setiap variabel, sesuai dengan data tambahan yang relevaln, untuk masing-masing kelas kejadian. Perhitungan dilakukan sebagai berikut:

1. P(Tingkat Minat Baca = "Tinggi") x P(Y = "Tinggi") = 0,08 x 0,7 x 0,3 = 0,015232
2. P(Tingkat Minat Baca = "Rendah") x P(Y = "Rendah") = 0,12 x 0,24 x 0,24 = 0,006912

4. Perhitungan nilai Probabilitas akhir setiap kelas (label) "P(X|Ci) x P(Ci)"

Langkah ketiga adalah mengalikan semua hasil dari variabel pada setiap kelas kejadian. Perhitungannya adalah sebagai berikut:

1. Untuk kelas kejadian dengan tingkat minat baca "Tinggi":

P(X | Tingkat Minat Baca = "Tinggi") x (X | Tingkat Minat Baca = "Tinggi") = 0,015232 x 25 / 50 = 0,007616

2. Untuk kelas kejadian dengan tingkat minat baca "Rendah":

P(X | Tingkat Minat Baca = "Rendah") x (X | Tingkat Minat Baca = "Rendah") = 0,006912 x 25 / 50 = 0,003456

5. Menentukaln nilai Probabilitas akhir terbesar dari setiap kelas

Pada tahap terakhir, hasil dari setiap kelas dibandingkan. Kelas dengan hasil tertinggi yang dipilih berdasarkan data di atas:

Tabel 6. Hasil data Uji

No	Nalma	Kriteria 1	Kriteria 2	Tingkat Minat Baca
1	Intan	Ilmu Sosial	Tepat Waktu	Tinggi

Bisa disimpulkan bahwa hasilnya untuk data input ini adalah "Tinggi" berdasarkan estimasi probabilitas yang diperoleh dari data pelatihan menggunakan Naive Bayes. Kami juga dapat menghitung probabilitas kondisional untuk pilihan "Tingkat Minat Baca Sedang" dengan memberikan nilai atribut. Dalam kasus ini, probabilitasnya adalah:

$$Probabilitals = \frac{0{,}007616}{0{,}007616+0{,}003456} = 1{,}003456$$

Pembahasan

Pada langkah ini, pengukuran akurasi dilakukan dengan menggunakan matriks kekacauan. Pengukuran ini dilakukan dengan membandingkan hasil prediksi pada data latih, yang terdiri dari variabel yang telah ditentukan, dengan data asli yang seharusnya.

Tabel 7. Hasil Counfusion Matrix

Aktual Class	Predicted as	
	Tinggi	Rendah
Tinggi	4	1
	2	8

4. KESIMPULAN

Hasil penelitian menunjukkan bahwa pendekatan Naive Bayes dapat digunakan sebagai alat pendukung keputusan untuk memprediksi minat baca secara bertahap. Setelah diuji, metode ini menunjukkan akurasi sebesar 80%, yang menunjukkan bahwa itu layak dan layak digunakan.

DAFTAR PUSTAKA

Abdurrahman, M. (2003). *Pendidikan Bagi Anak Berkesulitan Belajar*.

Eko, P. (2014). *Data Mining Konsep dan Aplikasi Menggunakan Matlab. Yogyakarta*.

Han. (2013). Data Mining Concepts and Techniques 3rd Edition. Morgan Kaufmann, USA. *Morgan Kaufmann*.

Karthika, S., & Sairam, N. (2015). A Naïve Bayesian Classifier for Educational Qualification. ,. *Indian Journal of Science and Technology*, 8(16), 1–5. https://doi.org/http://doi.org/10.17485/ijst/2015/v8i16/62055

M Asgari, S Ketabi, & Z Amirian. (2019). Interest-Based Language Teaching: EnhancingStudents' Interest and Achievement in L2 Reading Iranian J. Language Teaching Research7(1) 61–75

Priansyah, E., & Sutabri, T. (2024). Analisis Sentimen Berbasis Naïve Bayes Pada Media Sosial Twitter Terhadap Hasil Pemilu Indonesia 2024. *IJM: Indonesian Journal of Multidisciplinary*, *2*(3), 128-138.

Rahmawati. (2020). Komunitas Baca Rumah Luwu Sebagai Inovasi Sosial Untuk Meningkatkan Minat BacaDi Kbupaten Luwu. Jurnal BENING, 1-5.

R D Utami, D C Wibowo, & Y Susanti. (2018). Analisis Minat Membaca Siswa Pada KelasTinggi di Sekolah Dasar Negeri 01 Belitang J. Pendidikan Dasar PerKhasa 4(1) 179–188

R. Masri Sareb Putra. (2008). *Menumbuhkan Minat Baca: Panduan Praktis bagi Pendidik, Orang Tua, dan Penerbit.* PT. Macanan Jaya Cemerlang.

Romario. (2013). *Penerapan Data Mining Pada Rsup Dr.Moh Hoesin Sumatera Selatan Untuk Pengelompokan Hasil Diagnosa Pasien Pengguna Asuransi Kesehatan Miskin (askin).*

Tata Sutabri. (2012). Analisis Sistem Informasi. Penerbit Andi.

Tata Sutabri. (2012). Konsep Sistem Informasi. Penerbit Andi.

Thomas. (2004). *Data Mining : Definition and Decision Tree Examples*. e-book.

Turban. (2005). *Decision Support System and Intelligent Systems - 7th ed. Pearson Education, Inc. Pearson Education, Inc. Dwi Prabantini (penterjemah)*. Sistem Pendukung Keputusan dan Sistem Cerdas. Penerbit ANDI.

Widodo, Y. B., Anggraeini, S. A., & Sutabri, T. (2021). Perancangan Sistem Pakar Diagnosis Penyakit Diabetes Berbasis Web Menggunakan Algoritma Naive Bayes. *J. Teknol. Inform. dan Komput*, *7*(1), 112-123.