Implementasi Metode Naive Bayes Pada Pola Prestasi Siswa Dengan Menggunakan Rapidminer

Laila Devi Sari [1], Zaehol Fatah[2]
[1] Teknologi Informasi, Universitas Ibrahimy, Situbondo
[2] Sistem Informasi, Universitas Ibrahimy, Situbondo
Email : lailadevisari12@gmail.com

Abstrak

Penelitian ini membahas tentang penggunaan metode Naive Bayes untuk menganalisis pola prestasi siswa menggunakan RapidMiner. Tujuan utamanya adalah untuk mencari tahu faktor-faktor apa saja yang mempengaruhi prestasi belajar siswa dan seberapa akurat metode Naive Bayes dalam memprediksi prestasi siswa. Data yang digunakan mencakup informasi seperti jenis kelamin, usia, jurusan, jam belajar per minggu, tingkat kehadiran, status kerja paruh waktu, dan kegiatan ekstrakurikuler, dengan nilai GPA sebagai tolak ukur prestasi. Dalam prosesnya, data dibersihkan dan disiapkan terlebih dahulu melalui tahap preprocessing. Nilai GPA dibagi menjadi dua kategori: "Kurang_Baik" untuk nilai 2,00-2,99 dan "Baik" untuk nilai 3,00-3,99. Metode Naive Bayes kemudian diterapkan untuk menganalisis dan memprediksi kategori prestasi siswa berdasarkan faktor-faktor yang ada. Hasil penelitian menunjukkan bahwa model Naive Bayes mampu memprediksi prestasi siswa dengan tingkat akurasi 56,20%. Meskipun belum terlalu tinggi, hasil ini sudah cukup membantu dalam memahami pola prestasi siswa. Penelitian ini bisa menjadi dasar bagi pihak sekolah dalam mengambil keputusan untuk meningkatkan prestasi akademik siswa.

Kata Kunci: ***Naive Bayes, Prestasi Siswa, Data Mining, RapidMiner, Klasifikasi***

PENDAHULUAN

Di era digital seperti sekarang ini, penggunaan teknologi dalam dunia pendidikan sudah menjadi hal yang tidak bisa dihindari. Setiap sekolah menghasilkan data yang sangat banyak tentang siswanya, mulai dari data pribadi hingga catatan prestasi akademik. Data-data ini sebenarnya menyimpan informasi berharga yang bisa digunakan untuk meningkatkan kualitas pendidikan, namun sayangnya banyak sekolah yang belum memanfaatkan data tersebut dengan maksimal. Prestasi siswa menjadi salah satu indikator penting dalam menilai keberhasilan proses pembelajaran (Rasmi Daliana, 2020). Ada banyak faktor yang bisa mempengaruhi prestasi seorang siswa, seperti durasi waktu belajar, tingkat kehadiran di kelas, keaktifan dalam kegiatan ekstrakurikuler, atau bahkan status pekerjaan paruh waktu. Memahami hubungan antara faktor-faktor ini dengan prestasi siswa sangatlah penting bagi sekolah untuk bisa memberikan dukungan yang tepat kepada setiap siswa(Ilmu et al., 2020).

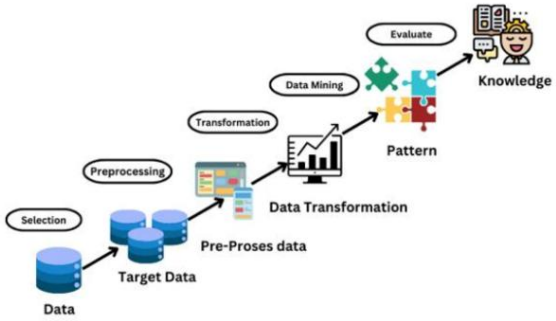
Untuk mengolah data siswa yang begitu banyak, diperlukan metode khusus yang bisa menemukan pola-pola tersembunyi dalam data tersebut. Salah satu metode yang cocok untuk hal ini adalah Naive Bayes, sebuah metode dalam data mining yang terkenal karena kesederhanaannya namun cukup efektif dalam melakukan klasifikasi. Metode ini bekerja dengan menghitung kemungkinan seorang siswa akan mendapat prestasi baik atau kurang baik berdasarkan karakteristik yang dimiliki. RapidMiner hadir sebagai solusi yang tepat untuk menerapkan metode Naive Bayes ini(Gandhi et al., 2021). Software ini menyediakan antarmuka yang mudah digunakan dan berbagai tools yang lengkap untuk mengolah data. Dalam penelitian ini, RapidMiner digunakan untuk menganalisis dataset yang berisi berbagai informasi tentang siswa, termasuk nilai GPA yang dibagi menjadi dua kategori:

"Kurang Baik" untuk nilai 2,00-2,99 dan "Baik" untuk nilai 3,00-3,99. Penelitian ini bertujuan untuk membantu sekolah dalam memahami faktor-faktor apa saja yang paling mempengaruhi prestasi siswa. Dengan memahami hal ini, sekolah bisa mengambil tindakan yang lebih tepat untuk membantu peningkatan prestasi akademik(Aslan, 2018). Misalnya, jika ditemukan bahwa jam belajar sangat mempengaruhi nilai siswa, sekolah bisa membuat program bimbingan belajar tambahan yang lebih terstruktur.

Hasil dari penelitian ini diharapkan bisa memberikan wawasan baru dalam dunia pendidikan. Dengan mengetahui pola-pola yang ada dalam prestasi siswa, guru dan pihak sekolah bisa merancang strategi pembelajaran yang lebih efektif. Selain itu, penelitian ini juga bisa menjadi contoh bagaimana teknologi data mining bisa dimanfaatkan untuk meningkatkan kualitas pendidikan di era digital ini. Dalam jangka panjang, penggunaan metode seperti ini dalam menganalisis data siswa bisa menjadi standar baru dalam dunia pendidikan(Gellysa Urva, Desyanti, 2023). Sekolah tidak lagi hanya mengumpulkan data, tapi juga bisa memanfaatkan data tersebut untuk membuat keputusan yang lebih baik dan memberikan pendidikan yang lebih berkualitas kepada para siswa. Hal ini tentunya sejalan dengan tujuan utama pendidikan yaitu menghasilkan lulusan yang berkualitas dan siap menghadapi tantangan di masa depan(Wanto et al., 2023).

METODE

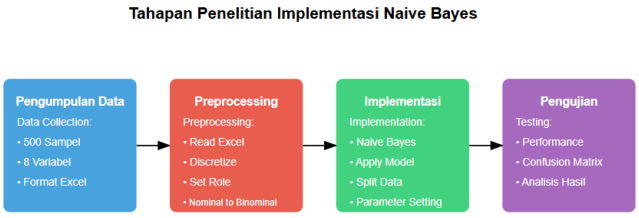
Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan metode Naive Bayes untuk menganalisis pola prestasi siswa. Data yang digunakan dalam penelitian ini berasal dari dataset prestasi akademik yang mencakup informasi seperti jenis kelamin, usia, jam belajar per minggu, dan nilai GPA siswa. Dataset ini terdiri dari catatan prestasi siswa yang mencakup berbagai karakteristik personal dan akademik siswa(Yuni Franata Sinurat, Masrizal, 2024).



Gambar 1. Ilustasi Proses Data Mining

Tahapan pengolahan data dalam penelitian ini dibagi menjadi beberapa langkah utama. Pertama, dilakukan preprocessing data menggunakan RapidMiner untuk membersihkan dan menyiapkan data. Dalam tahap ini, nilai GPA siswa dikelompokkan menjadi dua kategori: "Kurang Baik" untuk nilai 2.00-2.99 dan "Baik" untuk nilai 3.00-3.99. Pengelompokan ini penting untuk menyederhanakan proses klasifikasi menggunakan metode Naive Bayes. Implementasi metode Naive Bayes dilakukan menggunakan RapidMiner dengan beberapa tahapan penting(Alrasyid et al., 2024). Dimulai dari import dataset ke RapidMiner, pemilihan atribut yang relevan menggunakan operator Select Attributes, transformasi data menggunakan operator Discretize untuk mengkategorikan nilai GPA, dan penerapan algoritma Naive Bayes untuk proses klasifikasi. Untuk mengevaluasi keakuratan model, digunakan teknik cross-validation yang membagi data menjadi data training dan testing. Hasil analisis kemudian dievaluasi menggunakan beberapa metrik pengukuran seperti accuracy, precision, dan recall. Metrik-metrik ini penting untuk menilai seberapa baik model Naive Bayes dalam memprediksi prestasi siswa. Visualisasi hasil juga dilakukan untuk memudahkan interpretasi dan pemahaman terhadap pola-pola yang ditemukan dalam data prestasi

siswa(Jalil et al., 2024).

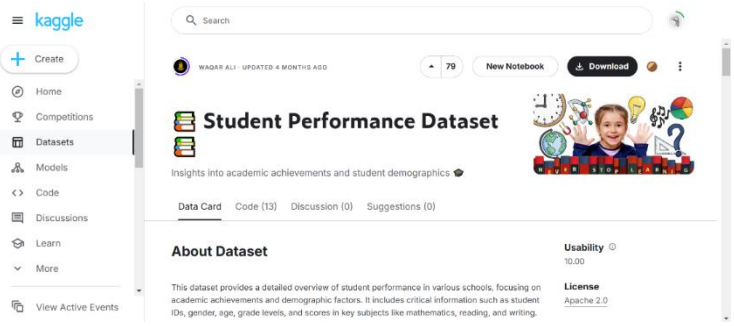


Gambar 2. Metodologi Penelitian

HASIL DAN PEMBAHASAN

Preprocessing data

Pada tahap preprocessing data, penelitian ini melakukan serangkaian proses untuk memastikan kualitas Penelitian ini menggunakan dataset prestasi siswa yang terdiri dari 500 record dengan 8 atribut, yang mencakup informasi demografis dan akademik siswa. Dataset ini diambil dari platform Kaggle dengan judul "Student Performance Dataset". Berikut rincian atribut yang digunakan dalam penelitian:



Gambar 3. Halaman Website Kaggle

a) Demograf
- Gender: Jenis kelamin siswa (Male/Female)
- Age: Usia siswa (dalam tahun)
- Major: Jurusan/bidang studi (Arts, Business, Education, Engineering, Science)

b) Akademik
- StudyHoursPerWeek: Jumlah jam belajar per minggu
- Attendance: Persentase kehadiran dalam pembelajaran
- GPA: Nilai rata-rata akademik (range 2.00-3.99)

c) Aktivitas
- PartTimeJob: Status kerja paruh waktu (Yes/No)
- ExtraCurric: Keikutsertaan dalam kegiatan ekstrakurikuler (Yes/No)

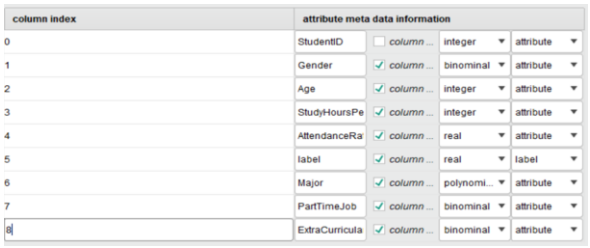
d) Label Klasifikasi:

Dalam penelitian ini, nilai GPA digunakan sebagai label dengan pembagian kategori:
- Kurang_Baik: GPA 2.00-2.99
- Baik: GPA 3.00-3.99

e) Karakteristik Dataset:
- Total Record: 500 data
- Tipe Data: Campuran (nominal, numerik)

f) Distribusi Label:
- Kurang_Baik: 247 record
- Baik: 253 record

Dataset ini dipilih karena memiliki variasi atribut yang relevan untuk menganalisis faktor-faktor yang mempengaruhi prestasi akademik siswa, serta memiliki distribusi kelas yang cukup seimbang antara kategori "Kurang Baik" dan "Baik".



Gambar 4. Pemilihan Atribut Dataset

Implementasi *K-Means Clustering*

Sebelum melakukan analisis menggunakan metode Naive Bayes, data yang ada perlu disiapkan terlebih dahulu melalui tahap preprocessing. Tahap ini penting untuk memastikan data bisa diolah dengan baik dan menghasilkan analisis yang akurat. Dalam penelitian ini, dataset prestasi siswa yang digunakan sudah cukup bersih, tidak ada data yang hilang atau rusak, sehingga bisa langsung diproses ke tahap berikutnya. Dalam proses preprocessing, langkah pertama yang dilakukan adalah mengubah format data kategori seperti jenis kelamin, jurusan, status kerja paruh waktu, dan kegiatan ekstrakurikuler menggunakan operator

'Nominal to Binominal' di RapidMiner(Ismai, 2017). Hal ini dilakukan agar data kategori tersebut bisa diproses dengan baik oleh metode Naive Bayes. Untuk data angka seperti usia, jam belajar per minggu, dan persentase kehadiran dibiarkan dalam format aslinya karena sudah sesuai untuk dianalisis.

Gambar 5. GPA sebagai label dataset

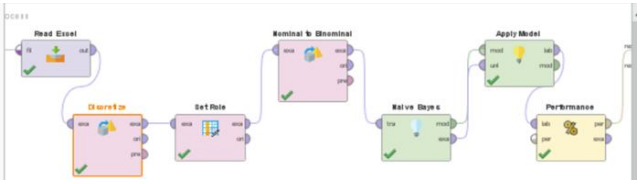
Langkah penting selanjutnya adalah mengubah nilai GPA menjadi dua kategori

class names	upper limit
Kurang_Baik	2.999
Baik	3.999

menggunakan operator 'Discretize by User Specification'. Nilai GPA antara 2,00-2,99 dikategorikan sebagai "Kurang Baik", sedangkan nilai 3,00-3,99 dikategorikan sebagai "Baik". Setelah itu, menggunakan operator 'Set Role', nilai GPA yang sudah dikategorikan ditetapkan sebagai label atau target yang akan diprediksi, sementara atribut lainnya ditetapkan sebagai predictor atau penentu dalam proses klasifikasi. Dengan selesainya tahap preprocessing ini, data sudah siap untuk dianalisis menggunakan metode Naive Bayes. Data dibagi menjadi dua bagian menggunakan operator 'Split Data', yaitu data training dan data testing dengan perbandingan 70:30. Data training digunakan untuk melatih model Naive Bayes agar dapat mengenali pola-pola yang ada, sedangkan data testing digunakan untuk menguji seberapa akurat model dalam memprediksi kategori prestasi siswa. Pembagian data ini penting untuk memastikan model yang dihasilkan bisa bekerja dengan baik pada data baru.

Implementasi Naïve Bayes

Proses klasifikasi menggunakan Naive Bayes di RapidMiner diimplementasikan dengan menghubungkan beberapa operator secara berurutan, yaitu operator Naive Bayes yang terhubung dengan data training, kemudian operator Apply Model yang menghubungkan hasil model dengan data testing, dan terakhir operator Performance untuk mengukur tingkat akurasi model.



Gambar 6. Proses Rapid Miner Naïve Bayes

Dalam prosesnya, metode Naive Bayes akan menghitung probabilitas setiap atribut (seperti jam belajar, kehadiran, jurusan, dll) terhadap kategori prestasi siswa (Kurang_Baik atau Baik). Parameter pada operator Naive Bayes diatur dengan mengaktifkan opsi 'laplace correction' untuk menghindari probabilitas nol yang bisa mengganggu perhitungan. Hasil dari implementasi ini kemudian dievaluasi menggunakan confusion matrix yang menunjukkan perbandingan antara hasil prediksi model dengan data sebenarnya, serta menghasilkan nilai accuracy, precision, dan recall yang menjadi ukuran keberhasilan model dalam mengklasifikasikan prestasi siswa.

Evaluasi Model

Setelah menerapkan metode Naive Bayes, kita perlu mengevaluasi seberapa baik model ini dalam memprediksi prestasi siswa. Berdasarkan hasil pengujian menggunakan operator Performance di RapidMiner, model menghasilkan beberapa nilai pengukuran penting yang bisa menunjukkan kualitas prediksi.

PerformanceVector

```
PerformanceVector:
accuracy: 56.20%
ConfusionMatrix:
True:   Kurang_Baik     Baik
Kurang_Baik:    126     98
Baik:   121     155
classification_error: 43.80%
ConfusionMatrix:
True:   Kurang_Baik     Baik
Kurang_Baik:    126     98
Baik:   121     155
weighted_mean_recall: 56.14%, weights: 1, 1
ConfusionMatrix:
True:   Kurang_Baik     Baik
Kurang_Baik:    126     98
Baik:   121     155
weighted_mean_precision: 56.20%, weights: 1, 1
ConfusionMatrix:
True:   Kurang_Baik     Baik
Kurang_Baik:    126     98
Baik:   121     155
```

Gambar 7. Hasil Evaluasi Model

Hasil evaluasi menunjukkan bahwa model Naive Bayes mencapai tingkat

akurasi sebesar 56.20%. Artinya, dari seluruh prediksi yang dilakukan, model berhasil memprediksi dengan benar sebanyak 56.20% kasus. Meskipun nilai ini tidak terlalu tinggi, namun masih bisa dianggap cukup mengingat kompleksitas faktor-faktor yang mempengaruhi prestasi siswa. Dari confusion matrix yang dihasilkan, dapat dilihat detail prediksi sebagai berikut:

a) Untuk kategori "Kurang_Baik": 126 prediksi benar dan 98 prediksi salah
b) Untuk kategori "Baik": 155 prediksi benar dan 121 prediksi salah

Selain itu, model juga menghasilkan nilai weighted mean precision sebesar 56.20% dan weighted mean recall sebesar 56.14%, yang menunjukkan keseimbangan antara kemampuan model dalam mengidentifikasi kasus positif dan negatif. Classification error sebesar 43.80% menunjukkan bahwa masih ada ruang untuk peningkatan model, misalnya dengan menambah jumlah data training atau mempertimbangkan faktor-faktor lain yang mungkin mempengaruhi prestasi siswa. Meski begitu, model ini sudah cukup baik untuk memberikan gambaran awal tentang faktor-faktor yang mempengaruhi prestasi akademik siswa.

KESIMPULAN

Penelitian tentang implementasi metode Naive Bayes untuk menganalisis pola prestasi siswa ini telah berhasil dilakukan dengan hasil yang cukup informatif. Berdasarkan analisis yang telah dilakukan menggunakan RapidMiner, model Naive Bayes mampu melakukan klasifikasi prestasi siswa dengan tingkat akurasi sebesar 56.20%. Model ini berhasil mengidentifikasi pola-pola dalam data siswa yang meliputi faktor demografis seperti jenis kelamin dan usia, serta faktor akademis seperti jam belajar dan tingkat kehadiran. Meskipun tingkat akurasi belum terlalu tinggi, model ini sudah mampu memberikan gambaran awal tentang hubungan antara berbagai faktor dengan prestasi akademik siswa. Untuk pengembangan penelitian selanjutnya, ada beberapa hal yang bisa dilakukan untuk meningkatkan kualitas analisis. Pertama, penambahan jumlah dataset yang lebih besar sangat disarankan untuk meningkatkan akurasi prediksi. Kedua, bisa ditambahkan variabel-variabel baru yang relevan dengan prestasi siswa, seperti metode pembelajaran, kondisi psikologis, atau fasilitas belajar yang tersedia. Selain itu, penggunaan metode klasifikasi lain seperti Decision Tree atau SVM bisa dicoba sebagai pembanding untuk mendapatkan hasil yang lebih optimal. Penerapan teknik feature selection juga bisa dilakukan untuk mengidentifikasi atribut-atribut yang paling berpengaruh, sehingga model yang dihasilkan bisa lebih efisien dan akurat dalam memprediksi prestasi siswa.

DAFTAR PUSTAKA

Alrasyid, H., Homaidi, A., Kom, M., Fatah, Z., & Kom, M. (2024). *Comparison Support Vector Machine and Random Forest Algorithms in Detect Diabetes*. *1*(1), 447–453.

Aslan, A. (2018). Kurikulum Pendidikan Islam di Amerika. *Journal of Al-Adzka: Jurnal Ilmiah Pendidikan Guru Madrasah Ibtidaiyah*, *8*(2), 117. https://doi.org/10.18592/aladzkapgmi.v8i2.2361

Gandhi, B. S., Megawaty, D. A., & Alita, D. (2021). Aplikasi Monitoring dan Penentuan Peringkat Kelas Menggunakan Naive Bayes Classifier. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, *2*(1), 54–63. https://doi.org/10.33365/jatika.v2i1.722

Gellysa Urva, Desyanti, I. A. (2023). PENERAPAN DATA MINING DI BERBAGAI BIDANG : Konsep, Metode, dan Studi Kasus. In *Google Buku*.

Ilmu, J., Muhammadiyah, P., Jati, K., Nurida, J. S., & Arianda, R. S. (2020). *JIPMuktj:Jurnal Ilmu Pendidikan Muhammadiyah Kramat Jati Volume 1 No 2 2020 https://jipmukjt.com/index.php/JIPMu*

Kjt Soraya Nurida dan Raka Sapto Arianda. *I*(2), 62–74.

Ismai. (2017). *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*.

Jalil, A., Homaidi, A., & Fatah, Z. (2024). Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita. *G-Tech: Jurnal Teknologi Terapan*, *8*(3), 2070–2079. https://doi.org/10.33379/gtech.v8i3.4811

Rasmi Daliana. (2020). Pentingnya Pendidikan Bagi Anak. In *Jurnal Manajemen, Kepemimpinan dan Supervisi Pendidikan* (Vol. 3, Issue 1).

Wanto, A., Siregar, M. N. H., Windarto, A. P., Hartama, D., Ginantra, N. L. W. S. R., Napitupulu, D., Negara, E. S., Lubis, M. R., Dewi, S. V., & Prianto, C. (2023). Data Mining : Algoritma dan Implementasi. In *Yayasan Kita Menulis*.

Yuni Franata Sinurat, Masrizal, I. (2024). *Data Mining Pengelompokan Siswa Berprestasi Menggunakan Metode Clustering*.