OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 11, November 2024
ISSN 2828-2442 (media online)
Hal 2873-2879

Analisis Faktor-Faktor yang Mempengaruhi Kinerja Siswa Menggunakan Algoritma Naïve Bayes

Heliota Agung Christiesta[1], Muhammad Ali Sodikin[1], Nazira Apriliani Rosyadi[1], Siti Andiani[1], Devi Yulianti[1*]

[1]Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia
Email: [1]heliotaagung456@gmail.com, [2]mhmmdali2811@gmail.com, [3]naziraapriliani1313@gmail.com, [4]andianisiti21@gmail.com, [5*]dosen00846@unpam.ac.id
(* : coressponding author)

Abstrak− Peningkatan kualitas pendidikan sering kali membutuhkan pemahaman yang mendalam terhadap faktor-faktor yang memengaruhi kinerja siswa. Penelitian ini menggunakan algoritma Naïve Bayes untuk menganalisis data dari dataset "Student Performance Factors" yang tersedia di Kaggle guna mengidentifikasi pengaruh variabel seperti waktu belajar, kehadiran, dan dukungan keluarga terhadap hasil akademik siswa. Proses analisis dilakukan melalui tahapan preprocessing, termasuk mengidentifikasi nilai-nilai yang hilang dan mengubah variable numerik menjadi variable kategori. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes memiliki kemampuan yang baik dalam memprediksi kinerja siswa, dengan tingkat akurasi yang cukup tinggi. Faktor-faktor seperti dukungan keluarga dan intensitas waktu belajar menjadi indikator utama yang memengaruhi hasil prediksi. Evaluasi model dilakukan menggunakan metrik akurasi, precision dan recall, untuk memastikan keandalan hasil. Kesimpulan dari penelitian ini adalah algoritma Naïve Bayes dapat digunakan secara efektif untuk mendukung analisis kinerja siswa dalam konteks pendidikan. Dengan memahami faktor-faktor utama yang memengaruhi hasil belajar, temuan ini dapat dimanfaatkan untuk merancang strategi yang lebih tepat dalam meningkatkan mutu pendidikan.

Kata Kunci: Naïve Bayes, Kinerja Siswa, Analisis Data, Faktor Pendidikan, Data Mining, Student Performance Factors, Prediksi Akademik, Preprocessing Data, Evaluasi Model, Rapidminer

Abstract−*Improving the quality of education often requires a deep understanding of the factors that influence student performance. This research uses the Naïve Bayes algorithm to analyse data from the 'Student Performance Factors' dataset available on Kaggle to identify the influence of variables such as study time, attendance, and family support on student academic outcomes. The analysis process went through a preprocessing stage, including identifying missing values and converting numerical variables into categorical variables. The results show that the Naïve Bayes algorithm has a good ability to predict student performance, with a fairly high accuracy rate. Factors such as family support and study time intensity are the main indicators that influence the prediction results. Model evaluation was conducted using accuracy, precision and recall metrics, to ensure the reliability of the results. The conclusion of this research is that the Naïve Bayes algorithm can be effectively used to support the analysis of student performance in an educational context. By understanding the key factors that influence learning outcomes, the findings can be utilised to design more appropriate strategies to improve the quality of education.*

Keywords: *Naïve Bayes, Student Performance, Data Analysis, Educational Factors, Data Mining, Student Performance Factors, Academic Prediction, Data Preprocessing, Model Evaluation, Rapidminer*

1. PENDAHULUAN

Dalam era modern saat ini, pendidikan memainkan peranan penting dalam menentukan kualitas sumber daya manusia. Namun, tantangan yang dihadapi dunia pendidikan semakin kompleks, terutama dalam memahami faktor-faktor yang memengaruhi kinerja siswa. Waktu belajar, tingkat kehadiran, dukungan keluarga, serta lingkungan sekolah menjadi beberapa indikator utama yang memiliki pengaruh signifikan terhadap hasil akademik siswa. Menganalisis hubungan antara faktor-faktor ini dapat memberikan wawasan yang berharga bagi para pendidik dan pembuat kebijakan.

Data mining merupakan salah satu solusi teknologi yang efektif dalam membantu pengolahan data pendidikan yang kompleks. Dalam hal ini, algoritma Naïve Bayes menjadi salah satu metode yang banyak digunakan karena kesederhanaan dan akurasi yang dimilikinya. Algoritma ini memungkinkan analisis pola data dengan asumsi independensi antar variabel, yang menjadikannya cocok untuk analisis kinerja siswa. Melalui penelitian ini, diharapkan dapat ditemukan faktor-faktor

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 11, November 2024
ISSN 2828-2442 (media online)
Hal 2873-2879

utama yang berkontribusi pada keberhasilan akademik siswa, sehingga dapat mendukung pengambilan keputusan yang lebih baik di bidang pendidikan.

Penelitian ini bertujuan untuk menganlisis kinerja siswa menggunakan algoritma Naïve Bayes dan mengidentifikasi faktor-faktor yang mempengaruhi kinerja tersebut. Penelitian ini terbatas pada analisis dataset "Student Performance Factors" yang tersedia di Kaggle dengan menggunakan algoritma Naïve Bayes untuk memprediksi kinerja siswa.

2. METODOLOGI PENELITIAN

2.1 Metode Pengumpulan Data

Dataset yang digunakan dalam penelitian ini adalah dataset "Student Performance Factors". Dataset ini memberikan gambaran menyeluruh mengenai berbagai faktor yang mempengaruhi kinerja siswa dalam ujian. Data ini berisi 100 baris data dengan 20 kolom mencakup informasi mengenai kebiasaan belajar, kehadiran, keterlibatan orang tua, dan aspek-aspek lain yang mempengaruhi keberhasilan akademik. Berikut ini penjelasan singkat mengenai kolom-kolom yang terdapat dalam database tersebut:

Tabel 1. Deskripsi Database

Atribut	Deskripsi
Hours_Studied	Jumlah jam yang dihabiskan untuk belajar per minggu.
Attendance	Persentase kelas yang dihadiri.
Parental_Involvement	Tingkat keterlibatan orang tua dalam pendidikan siswa (Rendah, Sedang, Tinggi).
Access_to_Resources	Ketersediaan sumber daya pendidikan (Rendah, Sedang, Tinggi).
Extracurricular_Activities	Partisipasi dalam kegiatan ekstrakurikuler (Ya, Tidak).
Sleep_Hours	Jumlah rata-rata jam tidur per malam.
Previous_Scores	Nilai dari ujian sebelumnya.
Motivation_Level	Tingkat motivasi siswa (Rendah, Sedang, Tinggi).
Internet_Access	Ketersediaan akses internet (Ya, Tidak).
Tutoring_Sessions	Jumlah sesi bimbingan belajar yang dihadiri per bulan.
Family_Income	Tingkat pendapatan keluarga (Rendah, Sedang, Tinggi).
Teacher_Quality	Kualitas guru (Rendah, Sedang, Tinggi).
School_Type	Jenis sekolah yang diikuti (Negeri, Swasta).
Peer_Influence	Pengaruh teman sebaya terhadap prestasi akademik (Positif, Netral, Negatif).
Physical_Activity	Jumlah rata-rata jam aktivitas fisik per minggu.
Learning_Disabilities	Adanya ketidakmampuan belajar (Ya, Tidak).
Parental_Education_Level	Tingkat pendidikan tertinggi orang tua (SMA, Perguruan Tinggi, Pascasarjana).
Distance_from_Home	Jarak dari rumah ke sekolah (Dekat, Sedang, Jauh).
Gender	Jenis kelamin siswa (Laki-laki, Perempuan).
Exam_Score	Nilai ujian akhir.

2.2 Algoritma Naïve Bayes

Algoritma Naïve Bayes adalah salah satu metode klasifikasi berbasis probabilitas yang mengasumsikan independensi antar variabel. Proses perhitungan algoritma ini dimulai dengan menghitung probabilitas prior untuk setiap kelas target. Selanjutnya, probabilitas likelihood dari setiap fitur terhadap kelas dihitung. Hasil kombinasi dari probabilitas ini digunakan untuk menentukan prediksi kelas. Algoritma ini dipilih karena mampu menangani dataset pendidikan dengan efisien dan memberikan hasil yang interpretatif untuk analisis faktor.

Rumus Teori Bayes:

$$P(A|B) = \frac{P(B|A)\ .\ P(A)}{P(B)}$$

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 11, November 2024
ISSN 2828-2442 (media online)
Hal 2873-2879

Keterangan:
P(A|B) = Peluang terjadinya peristiwa A, jika peristiwa B telah terjadi
P(B|A) = Probabilitas terjadinya peristiwa B, jika peristiwa A telah terjadi
P(A) = Probabilitas kejadian A
P(B) = Probabilitas kejadian B

2.3 Tahapan *Preprocessing* Data

Preprocessing data merupakan serangkaian langkah untuk mengolah data mentah menjadi data yang siap digunakan dalam penelitian. Proses ini bertujuan memastikan bahwa data yang digunakan memiliki kualitas yang baik. Tahapan ini melibatkan berbagai langkah penting untuk mempersiapkan data, sehingga hasil analisis dapat dilakukan secara optimal. Pada penelitian ini, preprocessing data dilakukan melalui beberapa langkah berikut:

1. Mengidentifikasi nilai-nilai yang hilang dalam dataset menggunakan operator Replace Missing Values
2. Data numerik pada dataset diubah menjadi kategori menggunakan operator Discretize

2.4 RapidMiner

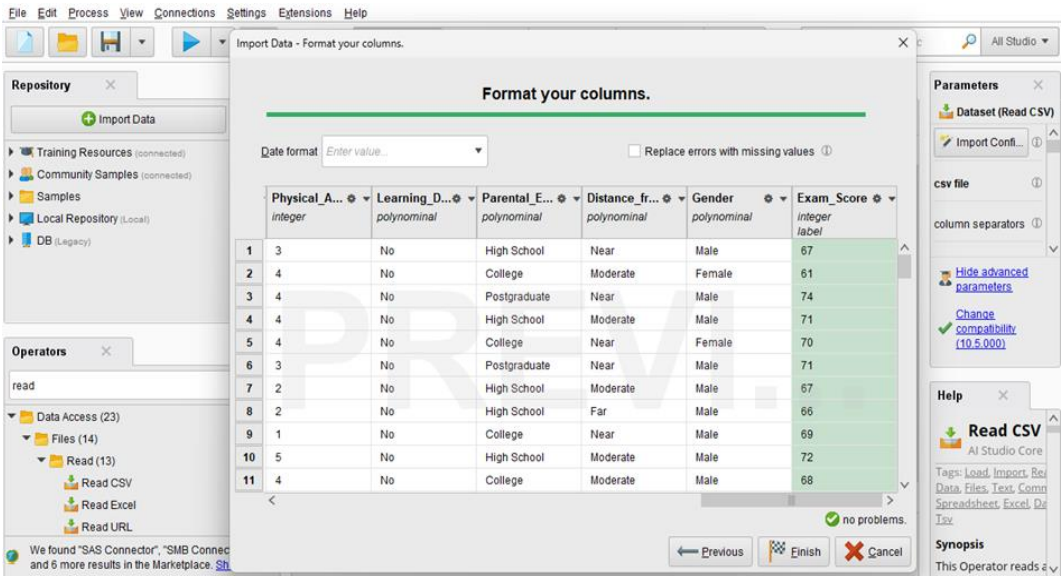
RapidMiner digunakan sebagai alat utama dalam penelitian ini. RapidMiner menawarkan berbagai modul untuk preprocessing data, penerapan algoritma Naïve Bayes, serta evaluasi model menggunakan metrik akurasi, precision, dan recall. Penggunaan RapidMiner dipilih karena antarmuka yang user-friendly dan kemampuan pengolahan data yang komprehensif.

3. ANALISA DAN PEMBAHASAN

3.1 Hasil Analisis

Penelitian ini bertujuan untuk menganalisis kinerja siswa menggunakan algoritma Naïve Bayes dengan dataset "Student Performance Factors" yang tersedia di Kaggle. Setelah melalui tahapan preprocessing, termasuk mengidentifikasi nilai-nilai yang hilang dan mengubah data numerik pada dataset menjadi kategori, model Naïve Bayes diterapkan untuk memprediksi nilai ujian akhir siswa (Exam_Score) berdasarkan faktor-faktor yang mempengaruhinya.

3.2 Transformasi Data



Gambar 1. Dataset

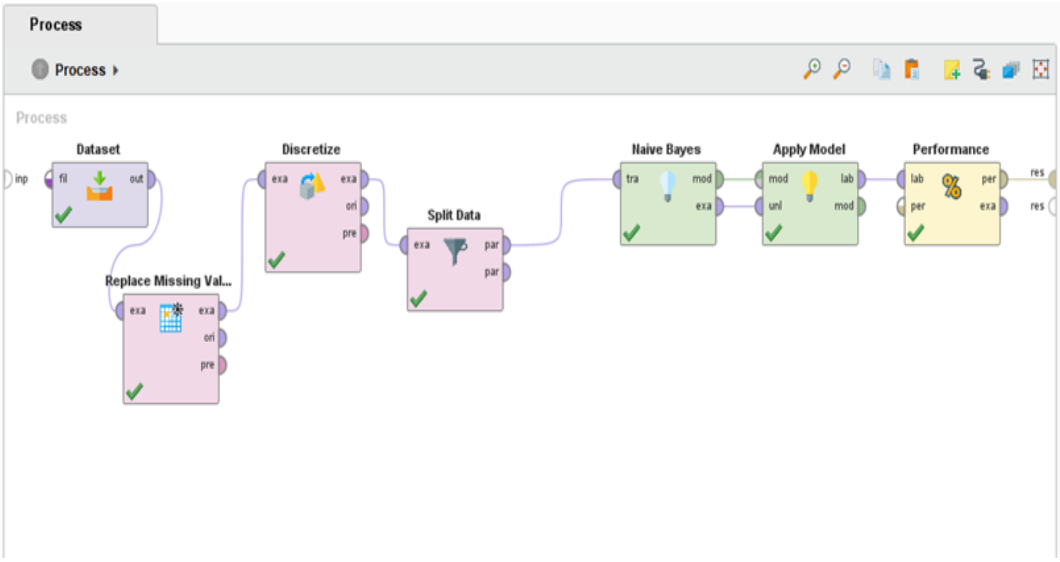
Pada Gambar 1 merupakan proses transformasi data yang akan diberikan label yang dibutuhkan algoritma Naïve Bayes.

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 11, November 2024
ISSN 2828-2442 (media online)
Hal 2873-2879

4. IMPLEMENTASI

4.1 Implementasi Algoritma Naïve Bayes

Data yang telah dikumpulkan dari dataset "Student Performance Factors" dibagi menjadi dua bagian, yaitu 80% untuk data pelatihan dan 20% untuk data pengujian. Pembagian ini dilakukan menggunakan operator Split Data di RapidMiner, yang bertujuan untuk melatih model dengan sebagian besar data dan menguji model dengan sisa data yang tidak digunakan dalam pelatihan. Pada tahap ini, algoritma Naïve Bayes diimplementasikan untuk menganalisis dan memprediksi kinerja siswa berdasarkan faktor-faktor yang mempengaruhi seperti waktu belajar, kehadiran, dukungan keluarga, dan lainnya.



Gambar 2. *Design*

Operator Naïve Bayes digunakan dalam RapidMiner untuk mbangun model klasifikasi, yang memungkinkan prediksi kelas berdasarkan probabilitas yang dihitung dari data pelatihan. Dengan asumsi independensi antar fitur, algoritma ini menghitung probabilitas untuk setiap kelas target (Exam_Score) dan memprediksi kelas untuk setiap data pengujian berdasarkan fitur-fitur yang ada. Proses ini menghasilkan model yang dapat digunakan untuk mengidentifikasi faktor-faktor utama yang mempengaruhi kinerja akademik siswa.

Evaluasi model dilakukan dengan menggunakan metrik akurasi, precision, dan recall untuk memastikan keandalan dan validitas hasil prediksi yang dihasilkan oleh model Naïve Bayes.

4.2 Hasil Penelitian

Setelah mengimplementasikan algoritma Naïve Bayes, evaluasi model dilakukan untuk mengukur kinerja model dalam memprediksi kinerja siswa. Tiga metrik yang digunakan untuk evaluasi adalah akurasi, precision, dan recall, yang masing-masing memberikan gambaran yang berbeda mengenai kualitas prediksi yang dihasilkan oleh model. Berikut ini hasil dari tiga metrik yang digunakan untuk evaluasi model:

accuracy: 90.69%

	true range1 [-∞ - 67.500]	true range2 [67.500 - ∞]	class precision
pred. range1 [-∞ - 67.500]	2669	282	90.44%
pred. range2 [67.500 - ∞]	210	2124	91.00%
class recall	92.71%	88.28%	

Gambar 3. Hasil Penelitian

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 11, November 2024
ISSN 2828-2442 (media online)
Hal 2873-2879

Dari hasil gambar diatas menunjukkan akrasi model tersebut sebesar 90.69% dapat dikatakan algoritma Naïve Bayes dianggap model yang cukup baik. Adapun angka presisi yang diberikan untuk setiap kelas menunjukan seberapa akurat model dalam memprediksi data untuk masing masing kategori:

1. Presisi kelas range1 sebesar 90.44% berarti bahwan 90.44% dari prediksi yang dilakukan untuk kelas ini benar, yaitu model memprediksi dengan tepat bahwa data termasuk dalam kelas kurang dari atau sama dengan 67.500
2. Presisi kelas range2 sebesar 91.00% berarti bahwa 90.00% dari prediksi yang dilakukan untuk kelas ini benar, yaitu model memprediksi dengan tepat bahwa data termasuk dalam kelas lebih dari 67.500
3. Recall kelas range1 92.71% angka ini menunjukkan kemampuan model dalam mengenali kelas range1 dengan baik dan recal kelas range2 menunjukkan bahwa 88.28% dari data yang sebenarnya termasuk dalam kelas range2 telah berhasil di prediksi dengan benar oleh model sebagai kelas range2.

4.3 Faktor-Faktor yang Mempengaruhi Kinerja Siswa

Berdasarkan hasil analisis menggunakan algoritma Naïve Bayes, beberapa faktor utama yang mempengaruhi kinerja siswa, khususnya nilai ujian akhir, antara lain:

1. Faktor dukungan keluarga (termasuk tingkat keterlibatan orang tua dalam pendidikan) terbukti menjadi indikator yang sangat berpengaruh terhadap kinerja siswa. Siswa yang mendapat dukungan tinggi dari orang tua cenderung memiliki kinerja akademik yang lebih baik. Hal ini dapat dijelaskan dengan adanya perhatian dan motivasi yang lebih besar dari orang tua, yang mendorong siswa untuk lebih giat belajar.
2. Intensitas waktu belajar siswa juga memainkan peran penting dalam prediksi kinerja akademik. Siswa yang menghabiskan lebih banyak waktu untuk belajar cenderung memiliki nilai ujian yang lebih baik. Temuan ini sejalan dengan hipotesis bahwa waktu yang lebih banyak dihabiskan untuk mempelajari materi ujian memberikan pemahaman yang lebih baik dan persiapan yang lebih matang.
3. Tingkat kehadiran di kelas menunjukkan hubungan positif dengan nilai ujian siswa. Siswa yang memiliki tingkat kehadiran tinggi lebih cenderung untuk memperoleh nilai ujian yang lebih baik. Hal ini menunjukkan bahwa partisipasi aktif dalam proses pembelajaran di kelas sangat penting untuk keberhasilan akademik.
4. Motivasi siswa juga memainkan peran penting dalam kinerja akademik. Siswa dengan tingkat motivasi yang tinggi cenderung lebih aktif dalam belajar dan memperoleh hasil yang lebih baik. Motivasi siswa dapat dipengaruhi oleh faktor eksternal seperti dukungan keluarga atau faktor internal seperti minat terhadap mata pelajaran.
5. Keterlibatan orang tua dalam pendidikan siswa adalah faktor yang sangat signifikan dalam memprediksi kinerja siswa. Orang tua yang terlibat dalam pendidikan anak-anak mereka, baik melalui bimbingan langsung maupun mendukung kegiatan akademik di rumah, dapat meningkatkan kinerja akademik anak mereka.

4.4 Evaluasi Model

Model Naïve Bayes berhasil menunjukkan hasil yang sangat baik berdasarkan metrik yang digunakan dalam evaluasi

- Akurasi: 90.69%
- Precision: 91.00%
- Recall: 88.28%

Evaluasi ini menunjukkan bahwa model Naïve Bayes memiliki kemampuan yang cukup baik untuk memprediksi hasil ujian siswa dengan tingkat keandalan yang tinggi. Hasil yang diperoleh memberikan gambaran yang jelas bahwa faktor-faktor seperti dukungan keluarga, waktu belajar, dan kehadiran berperan penting dalam memprediksi kinerja akademik siswa.

Untuk melakukan perhitungan manual dari tabel confusion matrix seperti yang ditunjukkan pada gambar, berikut adalah langkah-langkah dan rumus yang bisa kamu gunakan:

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 11, November 2024
ISSN 2828-2442 (media online)
Hal 2873-2879

4.4.1 Komponen:

- **True Positive (TP)**: Prediksi benar untuk kelas tertentu.
- **True Negative (TN)**: Prediksi benar untuk kelas lawan.
- **False Positive (FP)**: Prediksi salah untuk kelas tertentu.
- **False Negative (FN)**: Prediksi salah untuk kelas lawan.

Dari tabel gambar:

- **Kelas 1 (range1: [-∞ - 67.500])**
  - TP (Prediksi benar range1): 2669
  - FP (Prediksi salah range1): 282
  - FN (Prediksi salah range2): 210
  - TN (Prediksi benar range2): 2124

4.4.2 Perhitungan Manual:

Metode yang digunakan pada pengumpulan data dalam program aplikasi ini adalah sebagai berikut:

1. ***Accuracy***:

$$Accuracy = \frac{TP + TN}{Total\ Instances}$$

Metode yang digunakan pada pengumpulan data dalam program aplikasi ini adalah sebagai berikut:

Total Instances:

$$Total = 2669 + 282 + 210 + 2124 = 5285$$

Accuracy:

$$Accuracy = \frac{2669 + 2124}{5285} = 0.9069\ (90.44\%)$$

2. ***Precision* (*Class Precision*)**:

   - Untuk range1:

$$Precision\ (range1) = \frac{TP}{TP + FP} = \frac{2669}{2669 + 282} = 0.9044\ (90.44\%)$$

   - Untuk range2:

$$Precision\ (range2) = \frac{TP}{TP + FP} = \frac{2124}{2124 + 210} = 0.9100\ (91.00\%)$$

3. ***Recall* (*Class Recall*)**:

   - Untuk range1:

$$Recall\ (range1) = \frac{TP}{TP + FN} = \frac{2669}{2669 + 210} = 0.9271\ (92{,}71\%)$$

   - Untuk range2:

$$Precision\ (range2) = \frac{TP}{TP + FN} = \frac{2124}{2124 + 282} = 0.8828\ (88{,}28\%)$$

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 11, November 2024
ISSN 2828-2442 (media online)
Hal 2873-2879

5. KESIMPULAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa algoritma Naïve Bayes merupakan alat yang efektif untuk menganalisis kinerja siswa dalam konteks pendidikan. Faktor-faktor utama yang mempengaruhi kinerja siswa, seperti dukungan keluarga, waktu belajar, dan kehadiran, terbukti berpengaruh signifikan terhadap hasil ujian akhir. Dengan menggunakan model ini, pendidik dan pembuat kebijakan dapat merancang strategi yang lebih tepat untuk meningkatkan kualitas pendidikan dan membantu siswa mencapai potensi terbaik mereka.

Penelitian ini juga menunjukkan pentingnya penggunaan teknologi dalam pendidikan, terutama data mining dan algoritma pembelajaran mesin, untuk mendukung pengambilan keputusan yang lebih baik dalam konteks pendidikan. Algoritma Naïve Bayes, dengan kemampuannya dalam menangani data pendidikan yang kompleks, terbukti efektif dalam memberikan wawasan yang berguna bagi para pendidik dan pemangku kebijakan.

REFERENCES

Ridwan, M., Suyono, H., & Sarosa, M. (2013). Systematic Literature Review Prediksi Kinerja Siswa: Tren Penelitian, Metode, dan Atribut. *Jurnal Bisnis Digital dan Sistem Informasi, 1*(1), 1-8.

Shamsi, M. S., & Lakshmi, J. (2016). *A Comparative Analysis of Classification Data Mining Techniques: Deriving Key Factors Useful for Predicting Students Performance.* arXiv preprint arXiv:1606.05735.

Ramesh, V., Parkavi, R., & Ramar, K. (2013). Predicting Student Performance: A Statistical and Data Mining Approach. *International Journal of Computer Applications, 63*(8), 35-39.

Kumar, B., & Pal, S. (2011). Mining Educational Data to Analyze Students' Performance. *International Journal of Advanced Computer Science and Applications, 2*(6), 63-69.

Baradwaj, B. K., & Pal, S. (2012). *Mining Educational Data to Analyze Students' Performance.* arXiv preprint arXiv:1201.3417.

Yadav, S. K. and Pal, S. (2012) 'Data Mining : A Prediction for Performance Improvement of Engineering Students using Classification', *World of Computer Science and Information Technology Journal WCSIT, 2*(2), pp. 51–56. doi: 10.1142/9789812771728_0012.

Yadav, S., Bharadwaj, B. and Pal, S. (2012) 'Data mining applications: A comparative study for predicting student's performance', *International Journal of Innovative Technology & Creative Engineering, 1*(12), pp. 13–19. Available at: http://arxiv.org/abs/1202.4815.

Ridwan, M., Suyono, H. and Sarosa, M. (2013) 'Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier', *Eeccis, 7*(1), pp. 59–64.