



INNOVATIVE: Journal Of Social Science Research

Volume 4 Nomor 3 Tahun 2024 Page 10547-10555

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

Klasifikasi Kesiapan Anak Masuk Sekolah Dasar Menggunakan Algoritma Naïve Bayes dan Algoritma C4.5

M. Rudi Fanani^{1✉}, Dona Siska Sintia²

Institut Teknologi dan Sains Nahdlatul Ulama Pekalongan

Email: idurinanaf@gmail.com^{1✉}

Abstrak

Sekolah dasar merupakan dunia baru bagi anak-anak yang menuntut mereka untuk menjadi lebih serius dibandingkan saat mereka masih di taman kanak-kanak. Banyak faktor yang mempengaruhi prestasi belajar, salah satunya adalah kesiapan anak saat memasuki tingkat pendidikan tersebut. Penelitian ini bertujuan untuk menguji tingkat akurasi perbandingan antara algoritma Naive Bayes dan Algoritma C4.5 dalam memprediksi kesiapan anak saat memasuki tingkat pendidikan Sekolah Dasar. Dalam penelitian ini, dilakukan pengujian model dengan membandingkan dua metode data mining, yaitu algoritma C4.5 dan Naïve Bayes. Data yang digunakan adalah data siswa di TK Pratiwi di Desa Kemplong, Kecamatan Wiradesa. Data terdiri dari siswa yang siap dan belum siap, dengan total 135 data siswa. Model yang diuji akan menghasilkan confusion matrix seperti akurasi (accuracy), presisi (precision), recall, dan Area Under Curve (AUC) dari algoritma Naïve Bayes dan C4.5. Hasil evaluasi dan validasi model menunjukkan bahwa algoritma Naïve Bayes memiliki kinerja yang lebih baik dibandingkan dengan algoritma C4.5. Algoritma Naïve Bayes menghasilkan nilai akurasi sebesar 96,30% dan AUC sebesar 1,000, sementara algoritma C4.5 menghasilkan akurasi sebesar 92,59% dan AUC sebesar 0,974. Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa algoritma Naïve Bayes adalah metode yang efektif dalam memprediksi kesiapan siswa untuk memasuki sekolah dasar di TK Pratiwi.

Kata Kunci: *C4.5, Data Mining, Klasifikasi Anak Masuk Sekolah, Naive Bayes*

Abstract

Elementary school is a new world for children that requires them to be more serious than when they were in kindergarten. Many factors affect learning achievement, one of which is the readiness of children when they enter that level of education. This study aims to test the accuracy of the comparison between the C4.5 and Naïve Bayes algorithms in predicting children's readiness when they enter elementary school. In this study, model testing was carried out by comparing two data mining methods, namely the C4.5 and Naïve Bayes algorithms. The data used is student data at Pratiwi Kindergarten in Kemplong Village, Wiradesa District. The data consists of students who are ready and not ready, with a total of 135 student data. The tested model will produce a confusion matrix such as accuracy, precision, recall, and Area Under Curve (AUC) from the Naïve Bayes and C4.5 algorithms. The results of the model evaluation and validation show that the Naïve Bayes algorithm has better performance than the C4.5 algorithm. The Naïve Bayes algorithm produces an accuracy value of 96.30% and an AUC of 1.000, while the C4.5 algorithm produces an accuracy of 92.59% and an AUC of 0.974. Based on the results of this study, it can be concluded that the Naïve Bayes algorithm is an effective method in predicting students' readiness to enter elementary school in TK Pratiwi.

Keywords: *C4.5, Classification of Children Entering School, Data Mining, Naive Bayes*

PENDAHULUAN

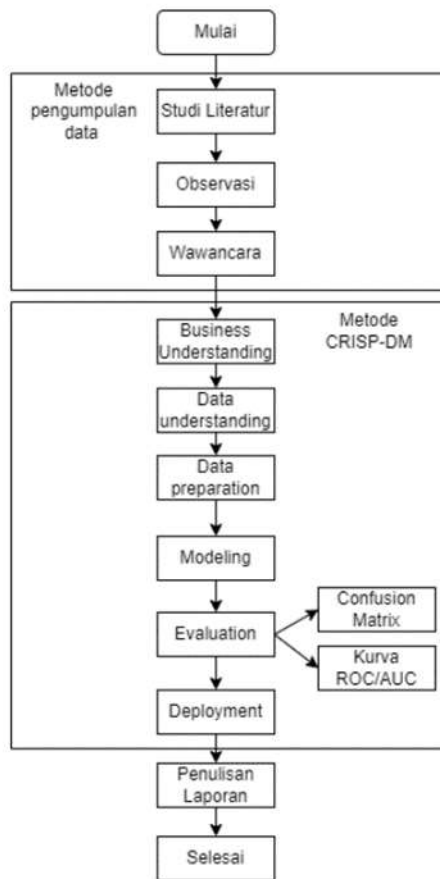
Sekolah Dasar (SD) merupakan masa peralihan yang penting bagi anak-anak setelah melewati masa Taman Kanak-Kanak (TK). Anak-anak mengalami kesulitan selama masa transisi ini karena banyaknya tuntutan, aturan, dan peraturan. Usia tujuh tahun ditetapkan sebagai usia wajib mengikuti pendidikan dasar. Pada usia ini, anak-anak secara umum sudah mengalami perkembangan yang memungkinkan mereka siap untuk belajar dan memiliki respons yang cepat terhadap lingkungan sekitar dan situasi baru. Pendidikan dasar pada usia ini menjadi langkah awal yang penting dalam proses pendidikan. Namun, dalam perjalanan menuju kemajuan, tentu ada tantangan yang harus dihadapi untuk memperkuat kemajuan yang telah dicapai oleh anak-anak (Ramadhini et al., 2022). TK Pratiwi merupakan salah satu TK yang berada di Kecamatan Wiradesa Kabupaten Pekalongan Jawa Tengah. Menurut Bu Nanda selaku Kepala TK Pratiwi menjelaskan bahwa permasalahan yang dihadapi oleh TK Pratiwi adalah ketidakseimbangan pengetahuan dan keterampilan anak-anak dalam kegiatan belajar mengajar. Beberapa kasus yang sering ditemui adalah anak-anak yang mampu menjawab pertanyaan secara lisan namun mengalami kesulitan dalam membaca, serta anak-anak yang memiliki kemampuan membaca dan berhitung tetapi mengalami kesulitan dalam memahami materi Pelajaran.

Diperlukan sebuah metode dalam membantu analisis anak TK Pratiwi masuk sekolah dasar dengan memprediksi siap atau belum siap anak untuk masuk sekolah dasar. Memprediksi kesiapan anak dapat dilakukan dengan mengekstraksi suatu informasi dari data-data dan berkas yang dikumpulkan saat penilaian anak dalam pembelajaran dan mengklasifikasikan data-data tersebut. Proses untuk mengekstraksi informasi tersebut disebut data mining. Data mining merupakan suatu proses untuk mengumpulkan informasi yang berharga dari sekumpulan data yang besar. Dalam proses ini, metode statistika dan matematika sering digunakan untuk menganalisis data tersebut (Nurmayanti, 2021). Klasifikasi dalam bidang data mining melibatkan berbagai metode perhitungan. Salah satu metode klasifikasi yang populer adalah algoritma decision tree C4.5 (Bayu Aji, 2023). Algoritma ini banyak digunakan dalam implementasi pengambilan keputusan karena memiliki banyak keunggulan. Salah satu keunggulan utamanya adalah kemampuan untuk menangani nilai atribut yang hilang. Selain itu, algoritma ini menghasilkan aturan-aturan yang mudah diinterpretasikan oleh manusia. Selain itu, performanya juga termasuk yang tercepat dibandingkan dengan algoritma klasifikasi lainnya (Fanani et al., 2023). Selain algoritma C4.5, dalam data mining terdapat juga algoritma naïve Bayes. Naïve Bayes adalah metode yang digunakan untuk mengembangkan klasifikasi dengan tujuan meningkatkan akurasi dalam menyelesaikan unit-unit permasalahan yang berbeda. Metode ini didasarkan pada Teorema Bayes dan mengasumsikan kemandirian (*independence*) antara setiap pasangan atribut dalam data. Meskipun memiliki asumsi sederhana, algoritma naïve Bayes sering kali memberikan kinerja yang baik dalam klasifikasi data, terutama ketika data yang digunakan memiliki ukuran besar (Liberti Duarte Tavares et al., 2023).

METODE PENELITIAN

Langkah langkah yang dilakukan dalam penelitian ini adalah Menyiapkan data mentah: Data mentah yang digunakan dalam penelitian diperoleh melalui studi literatur, observasi, dan wawancara. Data ini kemudian akan diolah untuk digunakan dalam analisis. Metode penelitian menggunakan CRISP-DM: Dalam penelitian ini, metode penelitian yang digunakan adalah model *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Model ini terdiri dari enam fase yang mencakup seluruh siklus penelitian data mining (Wurijanto et al., 2022). Fasenya adalah sebagai berikut :

- a. Business Understanding
Pada tahap ini, dilakukan survei lapangan di TK Pratiwi untuk secara langsung melihat kondisi dan permasalahan yang terjadi. Selain itu, faktor-faktor yang mempengaruhi kesiapan anak-anak untuk memasuki sekolah dasar juga dipelajari.
- b. Data Understanding
Dilakukan pengumpulan data dan analisis penyelidikan data mengenai siswa yang lulus masuk SD. Hal ini dilakukan untuk memperoleh pemahaman yang lebih dalam tentang data tersebut dan untuk mencari pengetahuan awal, serta mengevaluasi kualitas data tersebut
- c. Data Preparation
Data Preparation merupakan sebuah tahap persiapan data untuk pemodelan yang dimulai dengan data mentah awal dan berlanjut hingga tahap klasifikasi. Tahap persiapan data untuk pemodelan yang dimulai dengan data mentah awal dan berlanjut hingga tahap klasifikasi disebut persiapan data.
- d. Modelling
Dalam penelitian ini, model atau algoritma yang digunakan adalah Naive Bayes dan C4.5, yang diimplementasikan menggunakan RapidMiner. RapidMiner merupakan solusi yang digunakan untuk melakukan analisis data mining
- e. Evaluation
Evaluasi model dilakukan dengan menggunakan confusion matrix yang meliputi metrik-metrik seperti akurasi (accuracy), recall, dan presisi (precision). Selain itu, digunakan juga kurva ROC (Receiver Operating Characteristic) yang menghasilkan nilai AUC (Area Under Curve).
- f. Deployment
Pada tahap deployment, merupakan saat dimana hasil dari seluruh tahapan sebelumnya digunakan secara nyata.



Gambar 1 . Skema Alur Penelitian

Untuk menentukan root awal dalam perhitungan algoritma C4.5, perlu mencari nilai entropy dan gain untuk setiap atribut (Sukri & Handrianto, 2024):

$$\text{Entropy}(s) = \sum_{j=1}^m f(i,j) \log_2 f[(i,j)] \quad (1)$$

Dimana:

S : Himpunan Kasus

n : Jumlah Partisi Kasus

Pi : Proporsi dari Si

$$\text{Gain}(S,A) = \text{Entropy} - \sum |S_i| n_i - 1 * \text{Entropy}(S_i) \quad (2)$$

Dimana:

S : Himpunan Kasus

A : Atribut

N : Jumlah Partisi A

|Si| : Jumlah Kasus pada Partisi ke I

|S| : Jumlah Kasus dalam S

Tahapan dalam prediksi Naïve Bayes didasarkan pada teorema Bayes :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3)$$

Dimana:

B : data dengan class yang belum diketahui

A : Hipotesis data B

P(A|B) : Probabilitas A berdasarkan B

P(B|A) : Probabilitas B berdasarkan A

P(A) : Probabilitas dari A

P(B) : Probabilitas dari B.

HASIL DAN PEMBAHASAN

Algoritma C4.5

Data training diketahui ada 108 data, anak yang siap masuk Sekolah Dasar berjumlah 78 data dan belum siap ada 30 data. Untuk membuat algoritma C4.5 dihitung Nilai Entropy dan nilai gain

$$\text{Entropy} : \left(-\frac{78}{108} * \log_2 \left(\frac{78}{108} \right) \right) + \left(-\frac{30}{108} * \log_2 \left(\frac{30}{108} \right) \right) = 0,852405$$

$$\text{Gain} : 0,852405 - \left(\frac{53}{108} (0,94142) \right) + \left(\frac{55}{108} (0,72193) \right) = 0,022762$$

Confusion Matrix

Accuracy

Dari hasil pengujian accuracy nya adalah 92,59%

accuracy: 92.59%

	true Belum Siap	true Siap	class precision
pred. Belum Siap	7	1	87.50%
pred. Siap	1	18	94.74%
class recall	87.50%	94.74%	

Precision

Dari hasil pengujian nilai precision yaitu class true belum siap adalah 87,50% dan class true siap adalah 94,74%.

precision: 94.74% (positive class: Siap)

	true Belum Siap	true Siap	class precision
pred. Belum Siap	7	1	87.50%
pred. Siap	1	18	94.74%
class recall	87.50%	94.74%	

Recall

Dari hasil pengujian nilai recall yaitu class true belum siap adalah 87,50% dan class true siap adalah 94,74%.

recall: 94.74% (positive class: Siap)

	true Belum Siap	true Siap	class precision
pred. Belum Siap	7	1	87.50%
pred. Siap	1	18	94.74%
class recall	87.50%	94.74%	

Algoritma Naïve Bayes

Dalam membuat model naïve bayes terlebih dahulu kita mencari probabilitas hipotesis untuk masing-masing Kelas P(H). Hipotesis yang ada yaitu siswa yang siap masuk sekolah dasar dan siswa yang belum siap masuk sekolah dasar. Data *training* yang digunakan untuk metode naïve bayes menggunakan jumlah data 108, siswa yang Siap masuk sekolah dasar dalam kelas SIAP 78 *record* dan yang belum siap masuk sekolah dasar dalam kelas belum siap 30 *record*.

$$P(\text{SIAP}, n) = 78/108 = 0,72222222$$

$$P(\text{BELUM SIAP}, n) = 30/108 = 0,27777778$$

Confusion Matrix

Accuracy

Dari hasil pengujian *accuracy* adalah 96,30%

accuracy: 96.30%

	true Belum Siap	true Siap	class precision
pred. Belum Siap	7	0	100.00%
pred. Siap	1	19	95.00%
class recall	87.50%	100.00%	

Precision

Dari hasil pengujian nilai *precision* yaitu *class true* belum siap adalah 100% dan *class true* siap adalah 95%.

precision: 95.00% (positive class: Siap)

	true Belum Siap	true Siap	class precision
pred. Belum Siap	7	0	100.00%
pred. Siap	1	19	95.00%
class recall	87.50%	100.00%	

Recall

Dari hasil pengujian nilai *recall* yaitu *class true* belum siap adalah 87,50% dan *class true* siap adalah 100%.

recall: 100.00% (positive class: Siap)			
	true Belum Siap	true Siap	class precision
pred. Belum Siap	7	0	100.00%
pred. Siap	1	19	95.00%
class recall	87.50%	100.00%	

Komparasi Hasil Klasifikasi

Tabel 1 Hasil Komparasi

Algoritma	Accuracy	Recall		Precision	
		Siap	Belum siap	Siap	Belum siap
C4.5	92,59%	87,50%	94,74%	87,50%	94,74%
Naïve Bayes	96,30%	87,50%	100%	100%	95%

SIMPULAN

Pengujian model dengan membandingkan dua metode data mining yaitu algoritma C4.5 dan naïve bayes dengan menggunakan data siswa di TK Pratiwi di Desa Kemplong Kecamatan Wiradesa Kabupaten Pekalongan, yang terdiri dari siswa siap dan belum siap dengan total sebanyak 135 data siswa. Model yang diuji akan menghasilkan nilai *accuracy*, *precision*, dan *recall*, dari setiap algoritma. Hasil evaluasi dan validasi diketahui bahwa algoritma naïve bayes memiliki nilai *accuracy* 96,30 % lebih tinggi dibandingkan Algoritma C4.5 yang menghasilkan *accuracy* 92,59%. Dengan demikian algoritma naïve bayes merupakan metode yang cukup baik dalam memprediksi kesiapan anak masuk sekolah dasar.

DAFTAR PUSTAKA

- Bayu Aji, F. R. U. F. K. F. (2023). Klasifikasi Risiko Kematian Pasien Berdasarkan Penyakit Penyerta dan Usia Pasien Menggunakan Metode C4.5. *JIRE (Jurnal Informatika & Rekayasa Elektronik)*, 6, 9–17.
- Fanani, M. R., Fari, K., Fikriah, U., Azizah, M., & Ghufron, A. (2023). *Seleksi Fitur PSO untuk Klasifikasi Jenis Kekerasan dengan Algoritma C4.5*.
- Liberti Duarte Tavares, A., Nurraharjo, E., & Tri Lomba Juang No, J. (2023). Analisis Sentimen dan Klasifikasi Tweet Terkait Naiknya Kasus Omicron Menggunakan Naive

- Bayes Classifier. *Jurnal Informatika & Rekayasa Elektronika*, 6(1). <http://e-journal.stmiklombok.ac.id/index.php/jirelISSN.2620-6900>
- Nurmayanti, W. P. (2021). Penerapan Naive Bayes dalam Mengklasifikasikan Masyarakat Miskin di Desa Lepak. *Geodika: Jurnal Kajian Ilmu Dan Pendidikan Geografi*, 5(1), 123–132. <https://doi.org/10.29408/geodika.v5i1.3430>
- Ramadhini, F., Ahadiathul, L., Nasution, A., & Padangsidimpuan, I. (2022). *Pengukuran Kesiapan Sekolah : Analisis Empirik Berdasar Teori Tes Klasik*. 2(1).
- Sukri, M. H., & Handrianto, Y. (2024). Penerapan Algoritma C4.5 Dalam Menentukan Prediksi Prestasi Siswa Pada SMPN 51 Jakarta. In *Informatics and Computer Engineering Journal* (Vol. 4, Issue 1). <https://jurnal.bsi.ac.id/index.php/ijec/>
- Wurijanto, T., Setiawan, H. B., & Tjandrarini, A. B. (2022). Penerapan Model CRISP-DM pada Prediksi Nasabah Kredit yang Berisiko Menggunakan Algoritma Support Vector Machine. *Jurnal Ilmiah Scroll: Jendela Teknologi Informasi*, 10(1). <https://univ45sby.ac.id/ejournal/index.php/informatika>.