



Comparison of Data Mining Methods for
Predictive Achievement of Vocational Middle
Education Students at State Vocational School
(SMKN 1) Gadingrejo Pringsewu Lampung

Jani Triwidianti, Sri Yanto and Margi Prasajo

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

July 16, 2021

PERBANDINGAN METODE DATA MINING UNTUK PREDIKSI PRESTASI SISWA TINGKAT PENDIDIKAN MENENGAH KEJURUAN PADA SEKOLAH MENENGAH KEJURUAN NEGERI (SMKN 1) GADINGREJO PRINGSEWU LAMPUNG

Jani Triwidianti^{1a*}, Sriyanto^{2b}, Margi Prasajo^{3c}

^aFakultas Ilmu Komputer, IBI Darmajaya

^bFakultas Ilmu Komputer, IBI Darmajaya

^cFakultas Ilmu Komputer, IBI Darmajaya

^a jani.triwidianti.2021211019@mail.darmajaya.ac.id

^b sriyanto@darmajaya.ac.id

^c margi.prasajo.2021211023@mail.darmajaya.ac.id

Abstrak

Penelitian ini memiliki tujuan untuk membuat prediksi prestasi siswa menggunakan data mining, Sekolah Menengah Kejuruan (SMK) merupakan salah satu lembaga pendidikan yang berupaya menciptakan sumber daya manusia yang memiliki kemampuan, keterampilan, dan keahlian sehingga lulusannya dapat mengembangkan keterampilannya agar dapat terjun dalam dunia kerja maupun untuk melanjutkan ke jenjang pendidikan yang lebih tinggi, sekolah kejuruan bertujuan menyiapkan siswa dalam memenuhi lapangan kerja, menyiapkan siswa agar mampu memiliki karir, dan menyiapkan tamatan agar menjadi warga Negara yang produktif, adaptif, dan normatif.

Berdasarkan latar belakang dan identifikasi maka peneliti melakukan penelitian dengan menggunakan 3 metode pada data mining untuk melakukan tingkat akurasi prestasi siswa dengan menggunakan data mining metode yang digunakan dalam penelitian ini adalah decision tree, naive bayes, K-NN

Kata kunci : Data mining, Prestasi

Abstract

This study aims to make predictions of student achievement using data mining, Vocational High School (SMK) is one of the educational institutions that seeks to create human resources who have the ability, skills, and expertise so that graduates can develop their skills so that they can enter the world of work and To continue higher education, vocational schools aim to prepare students for employment, prepare students to be able to have careers, and prepare graduates to become productive, adaptive, and normative citizens.

Based on the background and the search method in using the research method, the researchers conducted research using 3 data mining to perform the level of accuracy of data mining used in this study were decision tree, naive bayes, K-NN

Keywords: Data mining, Achievement

1. PENDAHULUAN

Sekolah Menengah Kejuruan Negeri 1 (SMKN 1) Gadingrejo Pringsewu Lampung merupakan sekolah kejuruan yang sangat mengedepankan prestasi siswa dan skill lulusan, prestasi belajar siswa dipengaruhi oleh banyak faktor. Salah satu faktor penting dalam pendidikan agar prestasi belajar siswa meningkat adalah manajemen pembelajaran di sekolah. Semakin baik manajemen pembelajaran yang ditawarkan sekolah kepada siswa semakin besar pula kemungkinan prestasi belajar siswa akan baik

Pendidikan merupakan aspek penting bagi setiap negara terutama bagi negara berkembang seperti Indonesia. Selain itu, pendidikan merupakan suatu kebutuhan yang penting bagi kehidupan setiap individu. Melalui pendidikan, individu dapat mengembangkan potensi diri dan kepribadian yang dimilikinya dan generasi penerus bangsa dibentuk kualitasnya

Jumlah peserta didik saat ini untuk angkatan 2019 pada SMKN 1 Gadingrejo pringsewu berjumlah 377 dan memiliki 5 paket keahlian yaitu Teknik Konstruksi dan Properti, Teknik Elektronika, Teknik Otomotif, Teknik Komputer dan Informatika, Senirupa dan 7 program keahlian. Desain Pemodelan Informasi Bangunan, Bisnis Konstruksi dan Properti, Teknik Elektronika Industri, Teknik Audio Video, Teknik Kendraan Ringan, Multi Media dan animasi

Tujuan diselenggarakan sekolah kejuruan adalah untuk membekali lulusan dengan kompetensi yang berguna bagi diri sendiri dalam karir dan kehidupan bermasyarakat, tujuan sekolah menengah kejuruan akan lebih terarah jika kurikulum yang digunakan tepat dan dilaksanakan dengan baik. Usaha untuk meningkatkan kualitas lulusan SMK, dibutuhkan kerjasama dan sinkronisasi yang baik antara pemerintah, manajemen sekolah, kemitraan dengan dunia

industri, kompetensi guru, orang tua dan siswa sendiri. Beberapa elemen tersebut sangat berpengaruh baik secara langsung maupun tidak langsung terhadap peningkatan mutu siswa, yang pada akhirnya akan berimbas pada peningkatan kualitas pendidikan

Dari latar belakang yang peneliti paparkan maka peneliti menggunakan beberapa metode melakukan tingkat akurasi prestasi siswa dengan menggunakan data mining metode yang digunakan dalam penelitian ini adalah decision tree, naive bayes, K-NN untuk memperkirakan kelulusan tepat waktu siswa dengan melihat pengaruh dari nilai

2. KERANGKA TEORI

2.1. Prediksi

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan diantaranya adalah prediksi yaitu (Larose dalam Leidiyana, 2013). Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Contoh dari prediksi dalam bisnis dan penelitian adalah:

- a. Prediksi harga beras dalam tiga bulan yang akan datang.
- b. Prediksi persentase kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikan. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

2.2. Data Mining

Data mining merupakan langkah dalam proses KDD yang terdiri dari penerapan analisis data dan algoritma penemuan yang menghasilkan penghitungan pola atau model tertentu melalui data. Fayyad et al. (1996) di dalam penelitiannya menjelaskan bahwa terdapat beberapa langkah di dalam proses KDD (Knowledge Discovery in Database) diantaranya secara berurutan selection, preprocessing, transformation, DATA MINING, dan Interpretation/evaluation.

2.3. Decision Tree

Decision Tree merupakan faktor-faktor kemungkinan (probabilitas) yang akan sangat mempengaruhi alternatif-alternatif prestasi belajar siswa, disertai dengan prediksi hasil akhir yang akan didapat bila faktor-faktor dalam *Decision Tree* terpenuhi. *Decision Tree* merubah data kedalam bentuk visual berupa diagram pohon serta aturan-aturan keputusan, data dalam Decision Tree dinyatakan dalam bentuk tabel dengan atribut dan record

2.4. Naïve Bayes

Naive Bayes adalah teknik prediksi berbasis probabilistic sederhana yang berdasar pada penerapan Teorema Bayes (aturan Bayes) dengan asumsi independensi (ketidakketergantungan) yang kuat. (Prasetyo, 2012).

2.5. Algoritma K-Nearest Neighbor (KNN)

Algoritma **K-Nearest Neighbor (KNN)** adalah sebuah metode untuk melakukan klasifikasi terhadap objek yang berdasarkan dari data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Alasan untuk mengapa memakai Metode KNN ini, karena memiliki keunggulan dapat mengklasifikasikan data prestasi siswa yang tidak diketahui dengan adanya data latih dan data uji. Metode ini dapat mengklasifikasikan data secara akurat dengan **memilih** terlebih dahulu nilai tetangga terdekat dengan tepat.

3. METODOLOGI

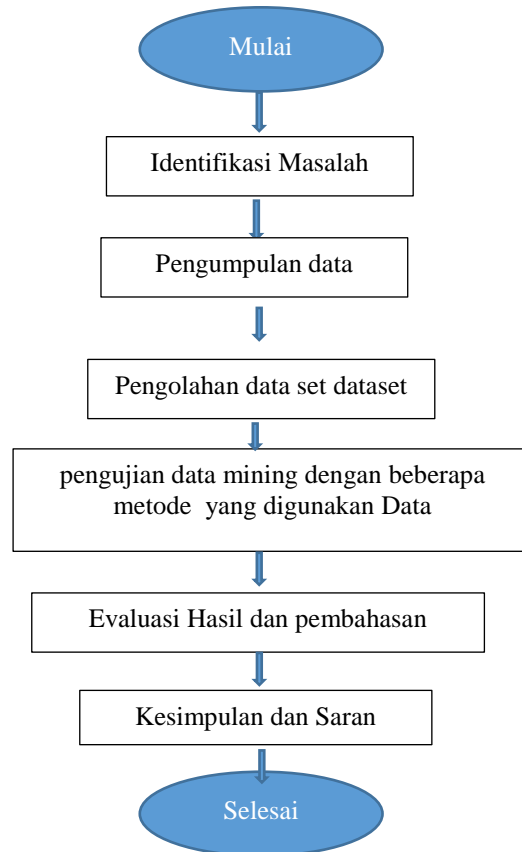
3.1. Metode Pengumpulan data

Metode pengumpulan data yang digunakan untuk memperoleh data-data yang dibutuhkan dalam penelitian ini adalah :

- a. Wawancara
Metode wawancara dilakukan dengan cara menyampaikan sejumlah pertanyaan dari pewawancara untuk di jawab oleh narasumber yaitu bagian kurikulum dan kesiswaan untuk mendapatkan suatu informasi.
- b. Studi Pustaka
Merupakan metode pengumpulan data yang diperoleh dari hasil olahan orang lain berupa dokumen, buku pustaka, jurnal, dengan membaca berbagai bahan penulisan, mengenai permasalahan yang berhubungan dengan penulisan dan khususnya penelitian yang berkaitan karya ilmiah.
- c. Observasi/survei
Metode ini digunakan dengan cara terjun langsung dan mengamati apa saja yang di perlukan untuk menemukan informasi dan pengetahuan yang diperlukan untuk bahan penelitian.

3.2 Tahapan Penelitian

Metodologi penelitian merupakan tahapan-tahapan yang sistematis dilakukan pada penelitian dan alur penelitian yang dilakukan adalah sebagai berikut :



Gambar 3.1 Alur Penelitian

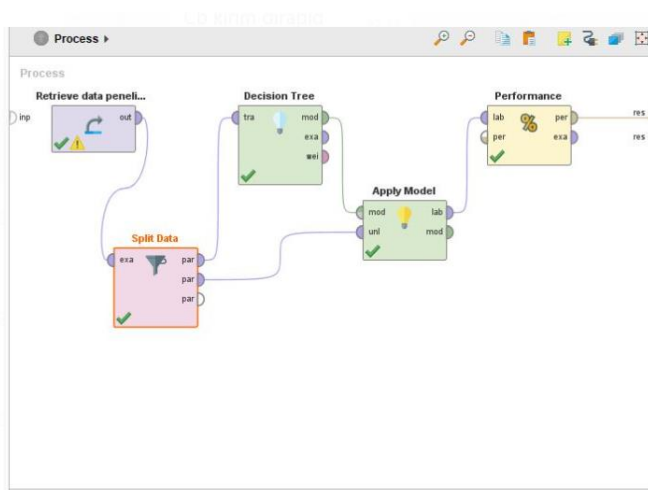
Alur penelitian dimulai dengan identifikasi masalah pencarian data set, data yang diperoleh tersebut diolah atau pre processing data menjadi data set, kemudian melakukan pengujian cross validation setiap model metode, setelah itu peneliti melakukan evaluasi dan pembahasan yang terakhir peneliti menyimpulkan hasil.

4. HASIL DAN PEMBAHASAN

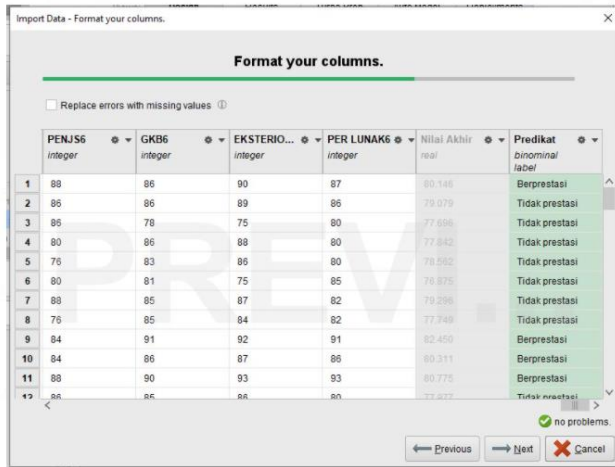
4.1 Hasil Accuracy

Pengolahan dataset yang dihasilkan 377 record yang terdiri dari beberapa atribut. Data tersebut bisa dilihat pada table dibawah ini.

1. Metode Decision Tree



Gambar 4.1 Hasil proses metode Decision tree



Gambar 4.2 data set menggunakan decision tree

accuracy: 97.37%

	true Berprestasi	true Tidak prestasi	class precision
pred. Berprestasi	2	0	100.00%
pred. Tidak prestasi	1	35	97.22%
class recall	66.67%	100.00%	

Gambar 4.3 Tingkat akurasi menggunakan decision tree

Adapun hasil Accuracy dari performance vector sebesar 97,37 % untuk proses yang di laksanakan pada model diatas dapat dilihat pada Gambar 4.3



Gambar 4.4 Grafik metode decision tree

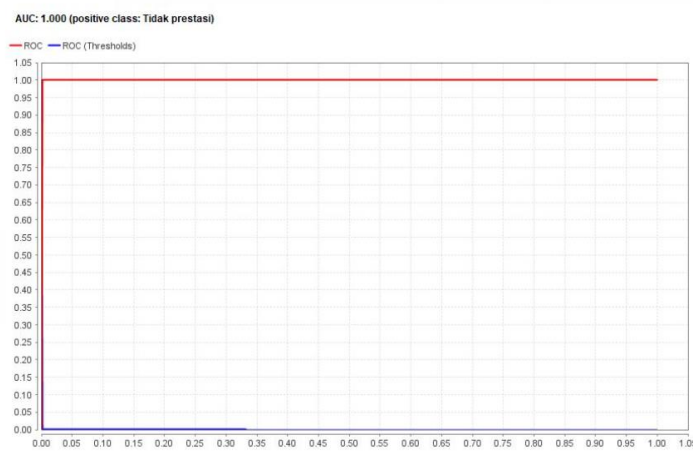
2. Metode Naive bayes

accuracy: 94.74%

	true Berprestasi	true Tidak prestasi	class precision
pred. Berprestasi	3	2	60.00%
pred. Tidak prestasi	0	33	100.00%
class recall	100.00%	94.29%	

Gambar 4.5 Tingkat akurasi menggunakan Naive Bayes

Adapun hasil Accuracy dari metode Naive bayes sebesar 94,74 % untuk proses yang di laksanakan pada model diatas dapat dilihat pada Gambar 4.5



Gambar 4.6 Grafik metode Naive bayes

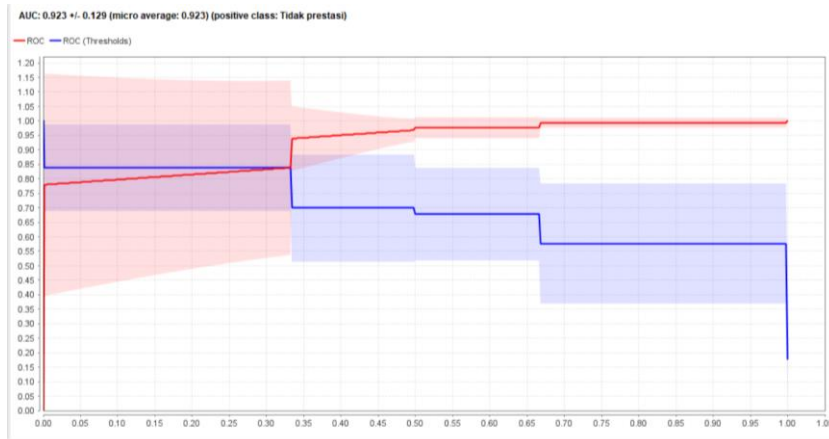
3. Metode K-Nearest Neighbor (KNN)

accuracy: 95.21% +/- 3.07% (micro average: 95.24%)

	true Berprestasi	true Tidak prestasi	class precision
pred. Berprestasi	13	3	81.25%
pred. Tidak prestasi	15	347	95.86%
class recall	46.43%	99.14%	

Gambar 4.7 Hasil Perhitungan Nilai akurasi metode KNN

Adapun hasil Accuracy dari metode Naive bayes sebesar 95,21 % untuk proses yang di laksanakan pada model diatas dapat dilihat pada Gambar 4.7



Gambar 4.8 Grafik metode KNN

Hasil dari ketiga metode yang kita uji Nama Metode Accuracy Decision Tree 97,37 % % Naive Bayes 94,74 % K-Nearest Neighbor 95,21 %

4.2. Pembahasan

Hasil penelitian berdasarkan tahapan tahapan proses dalam KDD (Knowledge Data Discovery) sebagai berikut:

1. Data Selection, tahapan ini dilakukan untuk memilih data yang sesuai dengan variabel yang dibutuhkan dalam penelitian. Caranya adalah dengan memilih atau menentukan atribut-atribut data mana yang akan digunakan dalam penelitian dari sekelompok data operasional yang ada. Salah satunya adalah menentukan atribut-atribut untuk variabel nilai PPDB, Nilai UN, Nilai US dan Nilai matapelajaran selama 3 tahun
2. Preprocessing/Cleaning, proses cleaning tersebut dilakukan terhadap keseluruhan data yang diteliti yang berjumlah 377 siswa. Setelah dilakukan proses cleaning data sejumlah 377, dihasilkan data bersih sebanyak 377 record data yang digunakan untuk proses analisis berikutnya;
3. Transformation, tahap ini menghasilkan satu recordset data yang siap untuk analisis data;
4. Analisis data. Analisis data yang pertama adalah dengan menggunakan 3 metode (Decision Tree, Naive Bayes, dan KNN) dan menggunakan aplikasi Rapidminer

5. KESIMPULAN

Berdasarkan hasil uji menggunakan RapidMiner maka didapat hasil untuk tingkat akurasi menggunakan 3 metode adalah sebagai berikut : Metode Accuracy Decision Tree 97,37 % % Naive Bayes 94,74 % K-Nearest Neighbor 95,21 %, berdasarkan hasil penelitian menunjukkan bahwa metode yang terbaik dari penelitian prediksi nilai adalah metode Decision Tree dengan tingkat akurasi 97,37%

UCAPAN TERIMA KASIH

Alhamdulillah puji syukur kepada Allah swt, karena kehendak -Nya peneliti dapat menyelesaikan penelitian ini. Pada kesempatan ini penulis mengucapkan rasa syukur dan terimakasih kepada :

1. Kelurgg tercinta
2. Bapak Sriyanto selaku pembimbing dan selalu mengingatkan dan memberi support sehingga terselesaikannya penelitian ini
3. Teman-teman seperjuangan MTI IIB Darmajaya

Penulis berharap semoga Allah SWT berkenan membalas segala kebaikan semua pihak yang telah membantu dan semoga penelitian ini membawa manfaat bagi pengembangan ilmu pengetahuan.

DAFTAR PUSTAKA

- Han, et al. *Data Mining : Concepts and Techniques* Third Edition New York : Morgan Kaufmann, 2012.
- Deshpande, S. P., dan V. M. Thakare. *Data Mining System and Applications: A Review*. *International Journal of Distributed and Parallel systems (IJDPS)*. Vol.1, No.1, hal. 32-44, 2010.
- Bellazzi, Riccardo, dan Blaz Zupan. *Predictive Data Mining In Clinical Medicine: Current Issues And Guidelines*. *International Journal Of Medical Informatics* 77, hal. 81-97, 2008.
- Larose, Daniel T. "Discovering Knowledge in Data : An Introduction to Data Mining." John Wiley & Sons, Inc., New Jersey. 2005
- Saefulloh, A. & Moedjiono. Penerapan Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Tepat Waktu. *InfoSys Journal*, Volume 2, pp. 42- 54, 2013.
- Pambudi, R., Setiawan, B., & Indriati, I. Penerapan Algoritma C4.5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal. **Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer**, vol. 2, no. 7, p. 2637-2643, sep. 2018
- Fayyad, Usama, et al. *From Data Mining to Knowledge Discovery in Databases*. *AI Magazine American Association for Artificial Intelligence*. Vol. 17, No. 3, hal. 37-54, 1996.
- Saefulloh, A. & Moedjiono. Penerapan Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Tepat Waktu. *InfoSys Journal*, Volume 2, pp. 42- 54, 2013.
-