

## Prediksi Prestasi Siswa Berbasis Data Mining Menggunakan Algoritma *Decision Tree* (Studi Kasus: SMKN 2 Padang)

Rahmi Ramadhani<sup>1\*</sup>, Yeka Hendriyani<sup>2</sup>

<sup>1</sup>Prodi Pendidikan Teknik Informatika Fakultas Teknik Universitas Negeri Padang

<sup>2</sup>Jurusan Teknik Elektronika Universitas Negeri Padang

Jl. Prof. Hamka Kampus UNP Air Tawar Padang

\*Corresponding author e-mail:rahmiramadhani19@gmail.com

### ABSTRAK

Pengujian ini bertujuan untuk memperkirakan prestasi siswa tergantung pada status sosial ekonomi orang tua, disiplin, dan motivasi siswa menggunakan teknik penambangan informasi dengan algoritma J48 yang dibantu oleh aplikasi WEKA. Penelitian ini menggunakan 8 atribut yang dipartisi dengan 1 variabel terikat dan 7 variabel bebas. Pendekatan yang digunakan adalah kuantitatif. Subyek penelitian ini adalah siswa kelas XI SMK Negeri 2 Padang yang berjumlah 450 siswa. Strategi pengumpulan informasi yang digunakan adalah dokumentasi dan *polling*. Berdasarkan pengujian yang dilakukan pada siswa yang menggunakan perhitungan *decision tree*, algoritma J48 memiliki ketepatan 90,31% dan nilai ROC 0,987 termasuk ke dalam *Excellent Clasfication* Dengan cara ini pohon keputusan dapat meramalkan prestasi siswa dengan baik.

**Kata kunci** : Prestasi Belajar, *Data Mining*, *Decision Tree*

### ABSTRACT

*This examination plans to foresee understudy accomplishment dependent on the financial status of guardians, control, and understudy inspiration utilizing information mining techniques with the J48 calculation helped by the WEKA application. This investigation utilizes 8 partitioned ascribes with 1 ward variable and 7 autonomous factors. The exploration approach utilized is quantitative. The subjects of this investigation were understudies of class XI SMK Negeri 2 Padang, adding up to 450 understudies. Information assortment strategies utilized are documentation and polls. In light of the tests led on understudies utilizing the choice tree calculation J48 has an exactness of 90.31% and a ROC worth of 0.987 is remembered for the Great Order state. In this manner the choice tree can foresee understudy accomplishment well.*

**Keywords:** *Student Achievement, Data Mining, Decision Tree*

## I. PENDAHULUAN

Pendidikan adalah suatu usaha yang disadari dan terencana untuk menciptakan suasana belajar yang kondusif dan interaksi belajar siswa secara efektif menumbuhkan kemampuan mereka untuk memiliki kekuatan spiritual keagamaan, ketenangan, karakter, pengetahuan, dan kemampuan yang diperlukan untuk dirinya, masyarakat, bangsa dan

negara. Berdasarkan Sistem Pendidikan Nasional pada pasal 3 UU. 20 Tahun 2003 tentang, bahwa tujuan pendidikan nasional ialah membina kemampuan siswa agar menjadi pribadi yang beriman dan bertakwa kepada Tuhan Yang Maha Esa, dan berakhlak mulia. Mengingat tujuan pendidikan ini, kualitas dan pelaksanaan proses belajar mengajar di sekolah ataupun di lembaga pendidikan harus ditingkatkan. Diantaranya penanda

kualitas dan pelaksana suatu sekolah ataupun lembaga pendidikan bisa dilihat melalui prestasi siswa selama disekolah[1].

SMK (Sekolah Menengah Kejuruan) merupakan lembaga pendidikan yang berupaya mencetak lulusan yang bertalenta dan memiliki keahlian agar para lulusan dapat memasuki dunia kerja atau melanjutkan ke pendidikan selanjutnya. namun, setiap siswa membutuhkan bimbingan yang berbeda. Oleh karena itu, penting untuk diingat sebelumnya tentang kualitas dan kemampuan siswa ini. Sebagaimana ditunjukkan oleh Undang-undang No. 29 Tahun 1990 Pasal 3 ayat 3 ayat (2) menyatakan bahwa sekolah kejuruan merencanakan lulusannya jadi masyarakat yang produktif, adaptif serta normatif. [1].

Sebagian aspek yang pengaruhi prestasi belajar siswa SMK antara lain 1) faktor kedisiplinan merupakan salah satu komponen yang mempengaruhi prestasi belajar. disiplin memiliki dampak yang lebih penting daripada lingkungan belajar[2]. 2) faktor sosial ekonomi orang tua berpengaruh signifikan terhadap prestasi belajar siswa. faktor sosial ekonomi orang tua menunjukkan dalam kapasitas keuangan mereka, kapasitas keuangan yang berbeda akan mempengaruhi prestasi siswa [3]. 3) Besar kecilnya pengaruh motivasi terhadap hasil belajar siswa, dibuktikan dengan berbagai informasi melalui persepsi, dokumentasi, survei yang kemudian disusun secara bersamaan. [4].

Melihat uraian faktor-faktor yang mempengaruhi prestasi/kinerja belajar siswa di atas, penelitian ini bertujuan untuk memahami seberapa besar aspek-aspek tersebut pada prestasi siswa dan faktor-faktor apa yang paling besar pengaruhnya terhadap prestasi siswa. Diharapkan hasil tes ini merupakan hasil rinci dari faktor-faktor yang mempengaruhi prestasi belajar siswa. Mengingat aspek-aspek ini, diharapkan bahwa faktor mana yang memiliki hambatan yang kuat dalam prestasi belajar mereka, agar langkah antisipatif dapat dilakukan sejak dini kepada siswa tersebut.

Salah satu teknik yang dapat dimanfaatkan untuk meningkatkan informasi sehingga memiliki nilai yang bermanfaat dengan penggalian informasi. Pemanfaatan penggalian informasi dapat digunakan sebagai bahan pertimbangan dalam menentukan pilihan mengenai hal-hal yang mempengaruhi prestasi belajar.

Kehadiran data mining terinspirasi oleh masalah ledakan informasi akhir-akhir ini, yang banyak diantaranya telah mengumpulkan informasi sejak lama. Kaitannya dalam penelitian ini adalah adanya informasi, seperti informasi tentang status sosial ekonomi orang tua, daftar hadir siswa (disiplin), data identitas siswa, dan lain

sebagainya. Informasi tersebut dapat diperoleh di sekolah pada saat PPDB (Penguatan Siswa Baru).

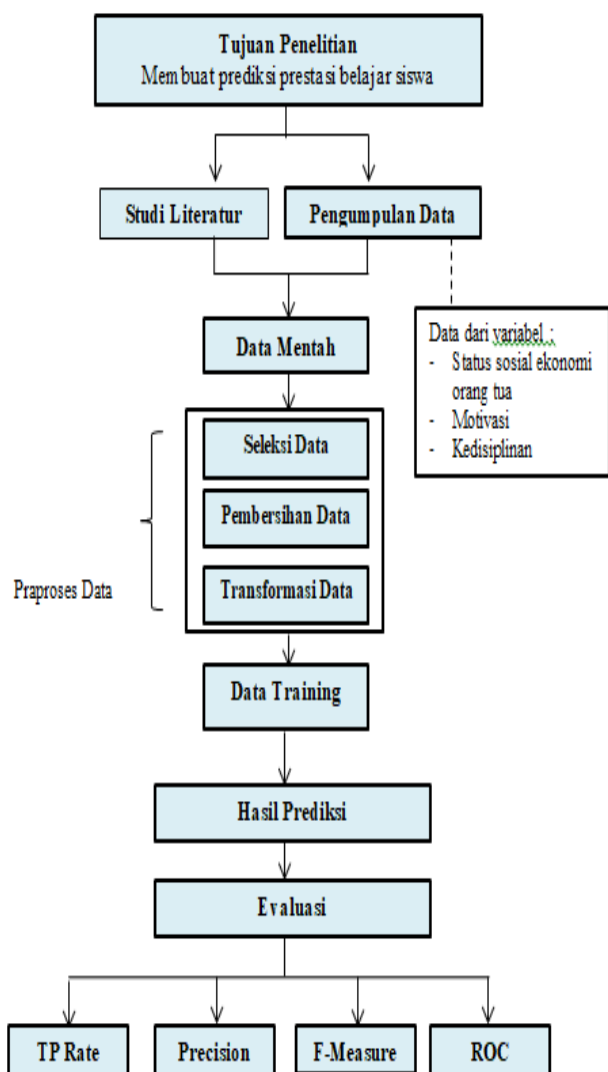
Strategi dalam *data mining* meliputi prediksi, klasifikasi, dan asosiasi. Teknik data mining yang akan digunakan dalam pengujian ini adalah Algoritma *Decision tree* yang digunakan untuk menguji karakterisasi dan ekspektasi desain informasi serta menggambarkan hubungan antara karakteristik faktor x dan faktor y dalam bentuk pohon.[5]. *Decision Tree* merupakan struktur menyerupai *flowchart* dimana setiap node terluar (node yang bukan leaf atau bukan node internal) merupakan pengujian dari variabel atribut, setiap cabangnya merupakan hasil dari pengujian tersebut, sedangkan hub terluar yaitu leaf menjadi labelnya[5].

## II. METODE

Penelitian ini menerapkan metodologi kuantitatif. Adapun tempat dilakukan di SMK Negeri 2 Padang. Sasaran/subjek eksplorasi adalah siswa kelas XI SMK 2 Padang tahun ajaran 2019/2020 yang menghimpun 450 siswa

Sistem menggunakan tahapan KDD (Information Revelation). Tahapan-tahapan diantaranya sebagai berikut: (1) *Data Selection*, tahap ini dimaksudkan untuk memilih data yang tepat dengan faktor-faktor yang dibutuhkan dalam pemeriksaan; (2) *Cleaning /Premeasure*, sebaiknya dilakukan *cleaning* sebelum dilakukan data mining pada informasi yang menjadi acuan KDD. Menggabungkan informasi yang diperlukan, memeriksa informasi yang bertentangan, dan merevisi kesalahan informasi, seperti kesalahan ketik (tipografi); (3) *Transformation*, pengkodean (*coding*) merupakan cara untuk mengubah data yang sudah dipilih, sehingga data itu layak untuk data mining; (4) *Analaisis information*; (5) *Pattern Evaluation*, adalah tingkat untuk menemukan contoh yang benar-benar menarik yang membahas informasi tergantung drarimana data berasal. Pada Gambar 1, dijelaskan bahwa diagram alur yang digunakan dalam penyelidikan ini.

Penelitian ini mendapatkan data melalui angket dan dokumentasi. Teknik dokumentasi yang dipakai untuk mengumpulkan informasi tentang status sosial ekonomi orang tua adapun data motivasi didapka dari angket.



Gambar 1. Diagram Alur Penelitian Teknik Analisis Data

Penelitian ini dilakukan di SMK Negeri 2 Padang untuk memprediksi prestasi siswa berdasarkan status sosial ekonomi orang tua, kedisiplin, dan motivasi siswateknik *Decision Tree*. Teknik ini untuk memprediksi prestasi siswa di SMK Negeri 2 Padang menurut status sosial ekonomi orang tua, kedisiplinan dan motivasi siswa.

### Decision Tree

Pohon keputusan akan menampilkan variabel potensial (kemungkinan) yang akan mempengaruhi kinerja siswa, dikombinasikan dengan perkiraan hasil akhir yang akan diperoleh jika komponen dalam pohon keputusan erpenuhi. Pohon keputusan mengubah data menjadi jenis visualisasi dalam bentuk diagram pohon dan aturan keputusan.

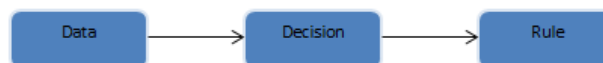
Informasi dalam pohon keputusan direpresentasikan sebagai tabel dengan atribut dan catatan. *Atribut* menggambarkan parameter yang dibuat sebagai kriteria dalam pembentukan *tree*. Salah satu *atribut* dari data solusi setiap item data

disebut atribut target. Atribut memiliki nilai yang disebut *instance*.

Alur proses analisis dalam pohon keputusan adalah mengubah bentuk data (tabel) menjadi model *tree*, dan model pohon menjadi aturan dan menyederhanakan aturan (pemangkasan). Informasi yang diperoleh dalam pengujian ini adalah 450 siswa yang akan digunakan untuk menentukan model prediksi pohon keputusan. Model yang dibuat akan ditentukan ketepatan ekspektasinya.

Pohon keputusan merupakan teknik klasifikasi yang paling imajinatif yang mencakup pengembangan pohon keputusan yang terdiri atas *node* pilihan yang diasosiasikan oleh cabang dari *node* akar menuju *node* daun (ujung). Dalam kualitas pilihan *node* akan dicoba, dan setiap hasil akan membuat cabang. Selanjutnya cabang tersebut akan dikoordinasikan ke *node* lainnya ataupun ke *node* akhir untuk membuat keputusan [6].

Pohon keputusan biasanya digunakan untuk memperoleh informasi yang bertujuan untuk pengambilan keputusan. Pohon keputusan diawali dengan *root node* (titik awal) yang akan dipakai oleh user untuk mengambil keputusan. Berdasarkan *root node* ini, *user* memecahkan *leaf node* sesuai dengan algoritma *decision tree*.



Gambar 2. Alur *Decision Tree*

Hasil akhir dari susunan *root node* dan *leaf node* tersebut adalah sebuah pohon keputusan dengan setiap cabang menunjukkan situasi potensial dari keputusan yang dibuat. Konsep pohon keputusan adalah mengubah data menjadi sebuah pohon keputusan (*decision tree*) dan aturan-aturan keputusan (*rule*).

TP Rate Class	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.904	0.022	0.944	0.904	0.924	0.894	0.954	0.897	C
0.882	0.031	0.921	0.882	0.901	0.862	0.947	0.873	B
0.913	0.032	0.931	0.913	0.922	0.886	0.978	0.922	D
0.944	0.010	0.850	0.944	0.895	0.890	0.987	0.812	A
0.909	0.029	0.526	0.909	0.667	0.678	0.970	0.487	E
W.Avg 0.903	0.028	0.914	0.903	0.906	0.874	0.962	0.880	

Tabel 1. *Detailed accuracy by class*

### III. HASIL DAN PEMBAHASAN

Dampak berkelanjutan dari pemeriksaan tergantung pada tahapan pengukuran *Knowledge Data Discovery* (KDD) sebagai berikut: (1)

Pemilihan seleksi data, sesi ini mencoba memilah data yang ditunjukkan oleh faktor-faktor yang diperlukan dalam proses pengujian. Hal ini dilakukan dengan mengkategorikan atau mencari tahu informasi yang mana yang akan digunakan untuk mengeksplorasi kumpulan informasi operasional yang ada. Salah satunya untuk mengetahui karakteristik faktor status sosial ekonomi orang tua dari informasi operasional, khususnya Informasi Pribadi Siswa di sekolah TU; (2) Pembersihan data, proses cleaning mencoba semua data dengan total 450 data siswa. Setelah pembersihan, berubah dari 450 data siswa menjadi 321 catatan data untuk analisis; (3) Transformasi, sesi ini membuat kumpulan catatan informasi untuk analisis; (4) Analisis Data

Analisis data menggunakan metode pohon keputusan algoritma J48. Aplikasi yang digunakan adalah WEKA 3.8.5. Hasilnya adalah kecepatan presisi 90,31%, seperti yang ditunjukkan pada tabel 1.

### IV. KESIMPULAN

Bersumber pada analisis data menggunakan decision tree data mining untuk memprediksi prestasi belajar siswa bersumber pada status sosial ekonomi orang tua, kedisiplinan serta motivasi siswa menggunakan metode data mining diperoleh hasil selaku berikut: (1) variabel motivasi merupakan variabel yang memastikan kemampuan seseorang siswa sukses ataupun tidak prestasi belajarnya di waktu yang akan datang. Hal ini dibuktikan dengan terdapatnya variabel motivasi yang menjadi simpul akar dari pohon yang terbentuk. Algoritma J48 memiliki tingkat keberhasilan rata-rata 90,31% untuk klasifikasi data. Hal ini menunjukkan bahwa

perhitungan ini memiliki kinerja profesional dalam klasifikasi.

### V. SARAN

Sehubungan dengan perihal tersebut, penyusun merekomendasikan beberapa saran, yaitu: (1) Selain data pendukung yang diklaim oleh pihak sekolah, masih terdapat kekurangan informasi yang memadai. Aksesibilitas informasi yang dimiliki sangat mempengaruhi keakuratan informasi dalam pengambilan keputusan. Salah satunya adalah informasi kedisiplinan, tidak hanya menggunakan rekap kehadiran siswa (2) Penyimpanan data jangka panjang perlu disimpan untuk memastikan bahwa informasi tersedia setiap saat bila diperlukan. (3) Kebijakan yang jelas dan tepat sasaran diperlukan untuk menangani siswa yang mungkin menghadapi kesulitan dalam kinerja akademik. Misalnya, dengan membimbing para siswa tersebut

### DAFTAR PUSTAKA

- [1] A. Noviriandini and N. Nurajijah, "Analisis Kinerja Algoritma C4.5 Dan Naïve Bayes Untuk Memprediksi Prestasi Siswa Sekolah Menengah Kejuruan," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 1, pp. 23–28, 2019, doi: 10.33480/jitk.v5i1.607.
- [2] P. Belajar and M. Pelajaran, "Pengaruh Disiplin Dan Lingkungan Belajar Terhadap Prestasi Belajar Mata Pelajaran Ekonomi," *Din. Pendidik.*, vol. 1, no. 2, pp. 206–218, 2005, doi: 10.15294/dp.v1i2.477.
- [3] A. Wulansari, "Pengaruh Motivasi Belajar dan Kondisi Sosial Ekonomi Orang Tua Terhadap Prestasi Belajar Mata Pelajaran Ekonomi (Studi Kasus Siswa Kelas VIII SMPN 2 Pringapus Tahun 2008/2009)," 2009. [Online]. Available: <http://lib.unnes.ac.id/2116/>.

- [4] Iswahyuni, “Pengaruh Motivasi Belajar Terhadap Prestasi Belajar IPS Siswa SMP Negeri 4 Sungguminasa Kabupaten Gowa,” 2017.
- [5] I. Sutoyo, “Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik,” *J. Pilar Nusa Mandiri*, vol. 14, no. 2, p. 218, 2018, doi: 10.33480/pilar.v14i2.926.