

# ANALISIS POTENSI SISWA UNTUK KULIAH DENGAN CLASSIFICATION MENGGUNAKAN METODE DECISION TREE

Abdul Doahir <sup>1)</sup>, Annisa Nurul Qolbi <sup>2)</sup>

abduldoahir@kallabs.ac.id <sup>1)</sup>, annisanq@kallabs.ac.id <sup>2)</sup>

<sup>1, 2)</sup> Prodi Bisnis Digital, Institut Teknologi dan Bisnis Kalla

## Abstrak

Melanjutkan pendidikan ke jenjang perguruan tinggi memang menjadi mimpi besar bagi sebagian besar siswa lulusan menengah ke atas maupun menengah kejuruan. Namun menurut data statistika Kemendikbud tahun 2021 dari 3,6 juta siswa yang lulus hanya 1,3 juta dari mereka yang dapat melanjutkan pendidikannya ke jenjang perguruan tinggi. Maka tidak heran jika 2,3 juta dari mereka tidak dapat melanjutkan pendidikannya dan menjadi faktor rendahnya minat pendidikan di Indonesia. Semua itu dapat terjadi sesuai dengan potensi yang dimiliki dari masing-masing mereka baik itu potensi internal ataupun potensi eksternal. Penelitian ini mengajukan analisis potensi yang dimiliki oleh para siswa untuk melanjutkan dan meningkatkan jumlah pendidikan pada jenjang perguruan tinggi dengan teknik classification. Classification merupakan metode data mining untuk menemukan model data, salah satunya adalah metode Decision Tree. Tahapan dalam penelitian ini yaitu: data collection, decision tree, dan rule sebagai dasar perbandingan dengan data baru yang pada akhirnya akan menghasilkan prediksi apakah siswa akan melanjutkan pendidikan ke jenjang perguruan tinggi atau tidak dengan potensi yang mereka miliki. Adapun hasil penelitian ini adalah mengklasifikasi potensi yang dimiliki siswa untuk melanjutkan serta meningkatkan jumlah pendidikan pada jenjang perguruan tinggi.

**Kata Kunci :** *Rapidminer, Classification, Potensi Siswa, Decision Tree, Perguruan Tinggi*

## 1. PENDAHULUAN

Pendidikan merupakan usaha sadar yang dilakukan untuk menggali potensi yang ada di dalam diri seorang siswa atau mahasiswa untuk menjadikan dirinya menjadi seorang yang profesional. Melalui pendidikan yang baik, diperoleh hal-hal baru sehingga dapat digunakan untuk menciptakan sumber daya manusia yang berkualitas. Suatu bangsa apabila memiliki sumber daya manusia yang berkualitas, tentunya mampu membangun bangsanya menjadi lebih maju. Perguruan tinggi menjadi institusi yang memiliki peran penting dalam mencapai tujuan pendidikan.

Teknologi data mining merupakan salah satu cara untuk penggalian data pada basis data berukuran besar. Dengan data mining, kita dapat melakukan pengklasifikasian, memprediksi, memperkirakan dan mendapatkan informasi lain yang bermanfaat dari kumpulan data dalam jumlah yang besar[1].

Decision tree adalah algoritma machine learning yang menggunakan seperangkat aturan untuk membuat keputusan dengan struktur seperti pohon yang memodelkan kemungkinan hasil, biaya sumber daya, utilitas dan kemungkinan konsekuensi atau resiko. Konsepnya adalah dengan cara menyajikan

algoritma dengan pernyataan bersyarat, yang meliputi cabang untuk mewakili langkah-langkah pengambilan keputusan yang dapat mengarah pada hasil yang menguntungkan.

Dimana setiap cabang mewakili hasil untuk atribut, sedangkan jalur dari daun ke akar mewakili aturan untuk klasifikasi. Itulah mengapa algoritma ini disebut decision tree karena pilihannya bercabang, membentuk struktur yang terlihat seperti pohon. Sehingga bisa membuat pohon keputusan baik itu secara vertikal atau horizontal tergantung pada preferensi. Membaca pohon keputusan horizontal dari kiri ke kanan dan pohon keputusan vertikal dari atas ke bawah [2].

Penelitian ini analisa klasifikasi prediksi potensi yang dimiliki oleh seorang siswa dalam melanjutkan pendidikannya pada perguruan tinggi. Adapun manfaat dari penelitian ini adalah menganalisis potensi siswa, sehingga dari hasil analisis tersebut institusi dapat mengetahui kiat-kiat yang perlu diperhatikan untuk meningkatkan jumlah pendidikan pada perguruan tinggi, institusi dapat lebih memahami karakteristik masing-masing potensi kelompok siswa serta dapat memposisikan dirinya sebagai fasilitator peningkatan mutu pendidikan yang masih tergolong rendah.

## 2. METODE PENELITIAN

### 2.1 Studi Literatur

Data mining adalah proses mempekerjakan teknik pembelajaran untuk menganalisis dan mengekstraksi pengetahuan secara otomatis. Tahapan data mining adalah sebagai berikut: data selection, preprocessing, transformation, data mining, dan evaluation [3].

Tahap preprocessing dilakukan supaya bersih dari noise, memiliki dimensi yang lebih kecil, lebih terstruktur sehingga bisa diolah lebih lanjut. Pada preprocessing juga dilakukan normalisasi data.

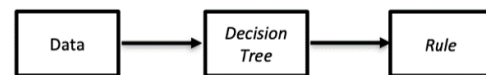
Transformation adalah mengubah format data yang disimpan menjadi standar format data yang sesuai dengan aplikasi yang

digunakan. Penelitian ini mengubah format data excel menjadi .csv supaya bisa diproses menggunakan Rapidminer 9.10.

Beberapa teknik yang ada dalam data mining yaitu classification, clustering, association rule discovery, sequential pattern, regression, dan deviation detection [4].

Algoritma Decision Tree merupakan salah satu algoritma untuk metode data mining yang sering diterapkan sebagai solusi untuk mengklasifikasikan sebuah masalah. Decision Tree juga merupakan salah satu klasifikasi pada data mining. Pembuatan Decision Tree sendiri menggunakan metode supervised machine learning yaitu proses pembelajaran dimana data baru diklasifikasikan berdasarkan training samples yang sudah ada. Decision Tree terdiri dari root, internal node dan leaf. Konsep yang sering digunakan untuk penentuan root, internal node dan juga leaf pada Decision Tree juga berupa konsep entropy dan konsep Gini [5].

Proses pada decision tree adalah mengubah bentuk data tabel menjadi sebuah model tree. Model tree akan menghasilkan rule dan disederhanakan [6].



Gambar 1. Konsep Decision Tree

Decision Tree dapat diterapkan untuk mempelajari klasifikasi dan memprediksi pola dari data dan menggambarkan relasi dari variabel atribut x dan variabel target y dalam bentuk pohon. Decision Tree merupakan struktur yang menyerupai flowchart dimana untuk setiap internal node merupakan pengujian terhadap variabel atribut, tiap cabangnya adalah hasil dari pengujian tersebut, sedangkan node terluar yaitu leaf menjadi labelnya[5].

### 2.2. Metoda

Metode penelitian yang digunakan dalam penelitian ini menggunakan deskriptif kuantitatif. Metode penelitian kuantitatif

merupakan metode penelitian yang berlandaskan pada filsafat positivisme, digunakan untuk meneliti populasi atau sampel tertentu. Teknik pengambilan sampel menggunakan dataset “go to collage” pada website kaggle.com

### 2.2.1 Dataset

Dataset diperoleh dari kaggle.com, memiliki 1000 record dan total atribut sebanyak 11, yang terdiri dari 10 atribut reguler dan 1 atribut spesial yaitu atribut will\_go\_to\_collage. Sampel dapat dilihat pada Tabel 1.

Tabel 1. Sampel Dataset

No	Atribut	Deskripsi	Tipe
1	Type_school	Type of school that student attends	string
2	School_accreditation	Quality of school. A is better than B.	string
3	Gender	Gender of student	string
4	Interest	How interested are students if they go to college	string
5	Residence	Type of residence	string
6	Parent_age	Parent age	integer
7	Parent_salary	Parent salary per month in IDR/Rupiah	integer
8	House_area	Parent house area in meter square	decimal
9	Average_grades	Average of grades in scale of 0-100	decimal
10	parent_was_in_college	Was parent ever in college?	boolean

11	will_go_to_college	Predictions about going to college	boolean
----	--------------------	------------------------------------	---------

### 2.2.2 Pengolahan Data

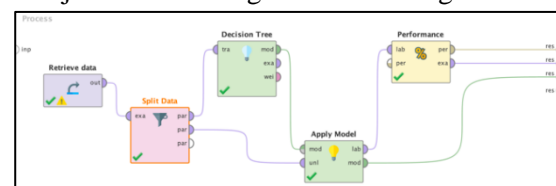
Pada tahap ini, pengklasifikasian data mining menggunakan software rapidminer 9.10. Pada tahap pre-processing ini digunakan 2 komponen operator yaitu Retrieve Data berfungsi untuk mengakses data yang tersimpan di repository atau menginput data ke process, serta pada operator ini telah ditentukan atribut yang menjadi Label yaitu atribut will\_go\_to\_collage. Operator kedua adalah Split Data berfungsi untuk menentukan ExampleSet yang mana pada operator ini dilakukan partisions atau membentuk parameter pembagi antara data training dengan data testing, pada proses ini pengujian mengambil perbandingan 70% data training dan 30% data testing dari 1000 record data sehingga dihasilkan 300 data testing.

### 2.2.3 Eksperimen dan Pengujian Metode

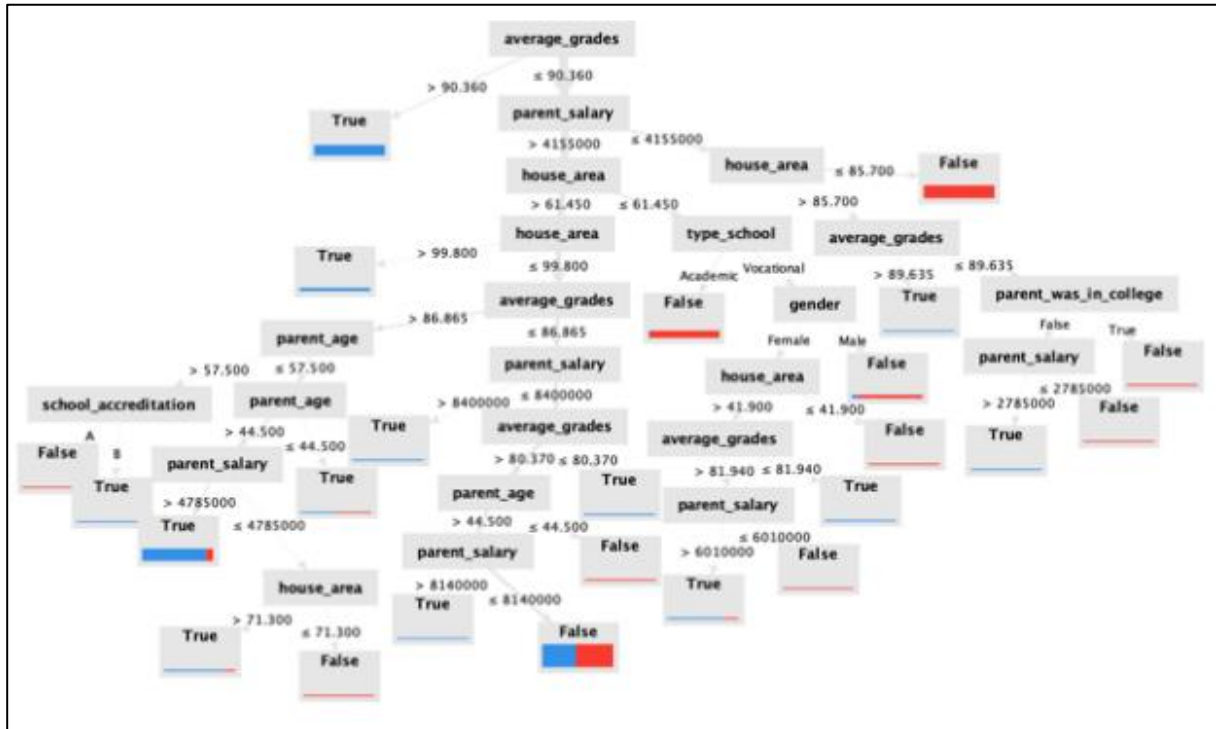
Pada tahap ini dilakukan eksperimen dalam mengimplementasikan pengujian algoritma Decision Tree. Ada 3 komponen operator pada pengujian ini yaitu; Decision Tree untuk menghasilkan pohon keputusan digunakan untuk klasifikasi data training, Apply Model untuk menampilkan model ExampleSet, Performance untuk menampilkan tingkat akurasi data testing. Tahap selanjutnya melakukan evaluasi dan validasi.

## 3. HASIL DAN PEMBAHASAN

Melakukan implementasi pengujian algoritma Decision Tree dan membagi dataset menjadi data training dan data testing.



Gambar 2. Model Proses



Gambar 3. Decision Tree

accuracy: 79.67%			
	true True	true False	class precision
pred. True	97	8	92.38%
pred. False	53	142	72.82%
class recall	64.67%	94.67%	

#### Gambar 4. Hasil Akurasi

Dataset dibagi menjadi 2 (dua) menggunakan operator Split Data dengan type sampling stratified dengan perbandingan 70% untuk data training, dan 30% untuk data testing.

Setelah diperoleh aturan keputusan dari pohon keputusan algoritma Decision Tree, proses uji akan mengukur sejauh mana keberhasilan model yang diperoleh dapat digunakan sebagai prediksi dengan evaluasi dan validasi melalui menggunakan matriks Confusion. Matriks Confusion yang diperoleh dari perhitungan algoritma Decision Tree.

Hasil dari Apply Model algoritma Decision Tree, diperoleh nilai accuracy klasifikasi true positif sebesar 92.38% dan false positif sebesar 72.82% class precision sedangkan true

negatif diperoleh sebesar 64.67% dan false negatif sebesar 94.67% class recall dengan keseluruhan nilai accuracy total sebesar 79.67% tingkat keakurasian antara prediksi decision tree terhadap data testing.

#### 4. KESIMPULAN

Tahapan dalam penelitian ini terdiri dari *data collection*, *Decision Tree*, *rule*, dan analisis hasil *klasifikasi*. Model klasifikasi dengan menggunakan metode algoritma decision tree dengan jumlah dataset 1000 yang dibagi menjadi 70% data training dan 30% data testing menghasilkan nilai recall tertinggi sebesar 94.67% dan tingkat akurasi sebesar 79.67%. Berdasarkan hasil nilai akurasi sebesar 79.67% membuktikan bahwa dengan menggunakan metode algoritma decision tree

ini menghasilkan nilai akurasi lebih baik dan tergolong *Good Classification* karena 79.67% setuju untuk melanjutkan pendidikannya ke perguruan tinggi dengan potensi yang mereka miliki.

Dari penelitian yang dilakukan dapat disimpulkan bahwa potensi yang dimiliki siswa untuk melanjutkan pendidikannya ke jenjang Perguruan Tinggi memiliki pengaruh besar untuk menentukan lanjut atau tidak. Metode *klasifikasi* menggunakan metode Decision Tree bisa digunakan untuk mendapatkan data atau rule baru sebagai parameter sehingga memungkinkan apabila suatu institusi atau perguruan tinggi untuk mengembangkan sistem informasi potensi siswa untuk melanjutkan kuliah dengan teknik *classification* menggunakan metode Decision Tree.

## 5. DAFTAR PUSTAKA

- [1] Mardi, Yuli. "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," Jurnal Edik Informatika, Vol.2, No.2, hal.213-219, 2017.
- [2] Rian Tineges "Pahami Metode Decision Tree Sebagai Algoritma Data Science" Dqlab.id Vol.03 Series. 28, 2018.
- [3] M.M. Benri dan L.S. Herlina "Analisis *Clustering* Menggunakan Metode K-Means dalam Pengelompokan Penjualan Produk pada Swalayan Fadhila," Jurnal Media Infotama, vol.11, No.2, hal.1101-1108, Sept, 2015.
- [4] A. Luh, "Seleksi Fitur dalam Klasifikasi Genre Music," Jurnal Ilmiah Ilmu Komputer, vol.X, No.1, hal 19-26, April, 2017.
- [5] Syaiful B, Akhyar L, "Metode kalsifikasi C4.5 prediksi juara premiere league" Jurnal Sintakis, Vol.2, No.1, April 2020
- [6] Yudhistira A.W., Agus B, Kaslani, Nining R, "Analisa Klasifikasi Menggunakan Algoritma Decision Tree Pada Data Log Firewall" Jurnal SIM, Vol.9 No.3, Desember 2021
- [7] W. A. Saputra, R. Fitri, A. S. B. Nugroho and S. Kustini, "Integration CLAHE and Seeded Region Growing for Segmentation of Rubber Tree in HSI Color Space," 2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 2021, pp. 334-338, doi: 10.1109/ISRITI54043.2021.9702812.
- [8] Gusti Jennie Febryza Indahsari, Annisa Kasiliyani, Wanvy Arifha Saputra, and Isna Wardiah, "SISTEM PENGAMBILAN KEPUTUSAN BEBAN KINERJA MENGGUNAKAN NAIVE BAYES STUDI KASUS PDAM BANDARMASIH", SENTRINOV, vol. 7, no. 1, pp. 571-581, Nov. 2021.
- [9] Saputra, Wanvy Arifha, and Evi Widiastuti. "Sistem Informasi Tanggap Darurat Bencana Kebakaran Hutan Berbasis Web Gis (Studi Kasus: Bnpb Kalimantan Selatan)." Seminar Nasional Riset Terapan. Vol. 4. 2019.
- [10] Riadi, Joni, et al. "Pengembangan Dan Pelatihan Penggunaan E-Learning Pada SMPN 7 Banjarmasin." Jurnal IMPACT: Implementation and Action 2.2 (2020): 112-118.