

# **Prediksi Calon Mahasiswa Penerima KIP Pada Universitas Indo Global Mandiri menggunakan Algoritma Decision Tree**

**Zakaria Saputra<sup>\*</sup>, Dewi Sartika, Muhammad Haviz Irfani**

Ilmu Komputer, Teknik Informatika, Universitas Indo Global Mandiri, Palembang, Indonesia

Email: <sup>1,\*</sup>zakaria.zrs14@gmail.com, <sup>2</sup>dewi.sartika@uigm.ac.id, <sup>3</sup>m.haviz@uigm.ac.id

Email Penulis Korespondensi: zakaria.zrs14@gmail.com

**Abstrak**—Universitas Indo Global Mandiri telah menerima mahasiswa Kartu Indonesia Pintar sejak tahun 2020. Pembiayaan mahasiswa Kartu Indonesia Pintar diperoleh dari subsidi pemerintah dengan syarat mahasiswa tersebut wajib menyelesaikan perkuliahan selama kurang atau sama dengan empat tahun dan mempertahankan prestasi akademik dengan indeks prestasi kumulatif minimal 3,00. Seperti pada program studi Teknik Informatika harus menyeleksi 31 calon mahasiswa penerima Kartu Indonesia Pintar terlebih dahulu untuk mengisi kuota yang tersedia. Mahasiswa yang telah lulus seleksi akan dirangking untuk mengetahui jumlah calon mahasiswa yang berhak menerima bantuan yang dilakukan oleh Biro Kerjasama universitas. Tujuan penelitian yaitu melakukan pemrograman yang dapat memprediksi jumlah mahasiswa penerima bantuan tersebut menggunakan algoritma Decision Tree. Decision Tree (versi C4.5) salah satu metode klasifikasi yang menggunakan representasi struktur pohon dimana setiap node merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas. Adapun atribut yang digunakan yaitu program keahlian, nilai rata-rata, status KK, penghasilan, dan kartu bantuan. Berdasarkan perhitungan yang dilakukan menggunakan metode decision tree didapatkan nilai dari presentase tingkat akurasi cukup baik dengan nilai 73% untuk prediksi calon mahasiswa penerima KIP Universitas Indo Global Mandiri Palembang, sedangkan untuk kinerja dari model klasifikasinya masih memiliki kinerja yang cukup rendah dengan nilai precision sebesar 40%, nilai recall sebesar 56% dan nilai F1-score sebesar 47%.

**Kata Kunci:** Kartu Indonesia Pintar; Klasifikasi; Data Mining; Decision Tree C4.5

**Abstract**—Indo Global Mandiri University has accepted Smart Indonesia Card students since 2020. Financing for Smart Indonesia Card students is obtained from government subsidies on the condition that students are required to complete lectures for less than or equal to four years and maintain academic achievement with a minimum cumulative grade point average of 3.00. As in the Informatics Engineering study program, 31 prospective Smart Indonesia Card recipient students must be selected first to fill the available quota. Students who have passed the selection will be ranked to determine the number of prospective students who are entitled to receive assistance carried out by the university's Cooperation Bureau. The purpose of the research is to do programming that can predict the number of students receiving this assistance using the Decision Tree algorithm. Decision Tree (C4.5 version) is one of the classification methods that uses a tree structure representation where each node represents an attribute, the branch represents the value of the attribute, and the leaf represents the class. The attributes used are skill program, average score, KK status, income, and aid card. Based on calculations carried out using the decision tree method, the value of the percentage of accuracy is quite good with a value of 73% for predicting prospective KIP recipient students at Indo Global Mandiri University Palembang, while the performance of the classification model still has a fairly low performance with a precision value of 40%, a recall value of 56% and an F1-score value of 47%.

**Keywords:** Indonesia Smart Card; Classification; Data Mining; Decision Tree C4.5

## **1. PENDAHULUAN**

Kartu Indonesia Pintar (KIP) merupakan program pemerintah yang dirancang untuk membantu pendidikan para pelajar di Indonesia. KIP merupakan bagian dari Program Indonesia Pintar (PIP) yang dikembangkan oleh tiga instansi kementerian Indonesia yaitu Kementerian Pendidikan dan Kebudayaan (Kemendikbud), Kementerian Sosial (Kemensos), dan Kementerian Agama (Kemenag). Melalui program ini pemerintah mengharapkan berkurangnya jumlah anak usia pelajar yang tidak bisa melanjutkan pendidikan karena terkendala oleh biaya. Program ini dapat meringankan biaya pendidikan para pelajar yang ingin tetap melanjutkan pendidikannya, baik itu biaya secara langsung maupun biaya secara tidak langsung.

Universitas Indo Global Mandiri telah menerima mahasiswa Kartu Indonesia Pintar (KIP) sejak tahun 2020, dengan pembiayaan mahasiswa KIP diperoleh dari subsidi pemerintah dan sebagian lagi dibantu oleh pendanaan Yayasan Universitas Indo Global Mandiri. Mahasiswa KIP wajib menyelesaikan perkuliahan selama kurang atau sama dengan empat tahun dan mempertahankan prestasi akademik (IPK > 2,75) dan prestasi non akademik. Penerimaan mahasiswa KIP melalui berapa tahapan yaitu: pendaftaran, seleksi administratif, dan wawancara. Pada tahun 2021 masing-masing program studi di Universitas Indo Global Mandiri memperoleh kuota yang berbeda-beda kuota. Seperti pada program studi Teknik Informatika harus menyeleksi 31 calon mahasiswa penerima KIP terlebih dahulu untuk mengisi kuota 16 mahasiswa yang tersedia. Mahasiswa yang telah lulus seleksi berhak menerima KIP, untuk prediksi calon mahasiswa penerima KIP akan digunakan algoritma decision tree.

Beberapa indikasi yang ingin diketahui seperti kelayakan mahasiswa yang diseleksi sudah diperoleh dengan informasi yang ada dan nyata, birokrasi internal dalam proses pengumpulan data calon peserta KIP, dan pelaporan internal yang saat ini sudah dilakukan untuk memberikan hasil kepada pihak eksternal, keberlanjutan program KIP yang dilakukan agar mendapat manfaat dan memberikan jaminan bagi keluarga yang perekonomiannya di bawah rata-rata.

Beberapa penelitian algoritma decision tree sering disebut juga sebagai pohon keputusan yang dijadikan sebagai alat pendukung keputusan dengan struktur seperti pohon yang memodelkan kemungkinan hasil, biaya sumber daya, utilitas, dan kemungkinan konsekuensi [1], [2]. Algoritma decision tree pada umumnya sering digunakan untuk

pengambilan keputusan. Decision tree memiliki kemungkinan untuk menampilkan hasil dari setiap pemilihan secara terorganisir [1]–[4]. Konsep dari decision tree mengubah data menjadi tree dan aturan-aturan keputusan yang menggunakan model klasifikasi sistem pohon sebagai nilai akhirnya [5]–[7]. Algoritma decision tree merupakan metode pemecahan masalah dengan menjadikan kriteria sebagai node (akar) yang saling berhubungan dan akan membentuk struktur seperti pohon [5], [8]–[12].

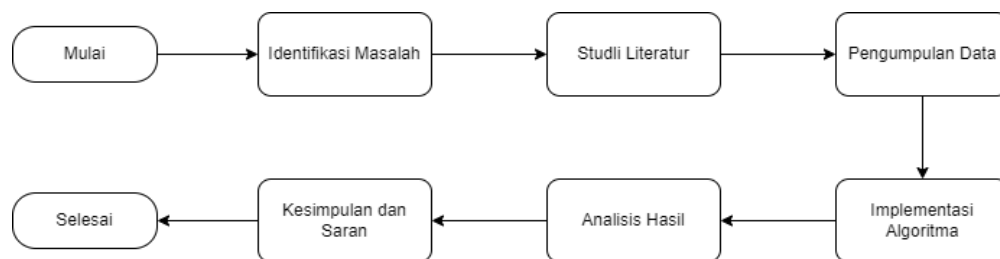
Khusnul Khotimah memprediksi seleksi KIP pada Universitas Muhammadiyah Kotabumi memperoleh hasil dari uji dua buah algoritma naïve bayes dan decision tree masing-masing tingkat akurasi 90,16% dan 100% [12], [13]. Perbandingan lainnya pada kasus penentuan penerimaan KIP memberikan hasil C.45 dan naïve bayes sama-sama efektif digunakan, sedangkan kasus penentuan usia kelahiran maka C.45 lebih baik, sementara kasus penentuan kelayakan calon anggota kredik di koperasi naïve bayes lebih baik pada presisi namun untuk recall dan accuracy C.45 lebih baik [3], [14].

Bagaimana hasil prediksi calon mahasiswa penerima bantuan Kartu Indonesia Pintar pada Universitas Indo Global Mandiri menggunakan algoritma decision tree nantinya, agar memberikan hasil dalam seleksi dan membantu penentuan data calon mahasiswa serta mempercepat proses yang dilakukan secara internal dalam lingkungan universitas

## 2. METODOLOGI PENELITIAN

### 2.1 Kerangka Dasar Penelitian

Penelitian ini bersifat kualitatif dan kuantitatif dengan data mahasiswa yang berjumlah 1058 buah data pada Universitas Indo Global Mandiri pada tahun 2020 dan tahun 2021. Data diperoleh dari Biro Kemahasiswaan Alumni dan Bursa Kerja (BKABK) yang ada di Universitas Indo Global Mandiri Palembang. Adapun variabel bebas (independent) yang digunakan diantaranya program keahlian, nilai rata-rata, prestasi, status Kartu Keluarga, kartu bantuan, dan penghasilan orang tua, sedangkan variabel terikat (dependent) yaitu keterangan pernyataan Lulus atau Tidak Lulus. Berikut ini di tampilkan tahapan penelitian dalam Prediksi Calon Mahasiswa Penerima KIP Pada Universitas Indo Global Mandiri menggunakan Algoritma Decision Tree.



**Gambar 1.** Tahapan Penelitian

Gambar 1 di atas adalah mengenai tahapan penelitian sebagai berikut: (1). Tahap identifikasi masalah merupakan tahapan pencarian dan penentuan masalah untuk diteliti, (2). Tahap studi literatur merupakan tahapan mencari referensi penelitian-penelitian terdahulu dan teori-teori dasar yang berhubungan dengan penelitian penulis, (3). Tahap pengumpulan data merupakan tahapan pengambilan data sekunder sebagai data untuk diteliti, (4). Tahap implementasi algoritma merupakan perancangan perangkat lunak yang akan dibangun mulai dari perhitungan manual (5). Tahap analisa hasil merupakan tahapan analisis dari pada hasil dari algoritma, visualisasi dari hasil prediksi, dan gambar decision tree, (6). Tahap kesimpulan dan saran merupakan tahapan kesimpulan dari penelitian yang telah dilakukan dan saran untuk penelitian selanjutnya.

### 2.2 Data Mining

Data Mining merupakan bagian dari proses Knowledge Discovery in Database (KDD), yang terdiri dari beberapa langkah: pemilihan data, preprocessing, transformasi, data mining, dan evaluasi hasil. Selain itu, data mining disebut juga langkah menuju penemuan pengetahuan dalam database (KDD), proses terdiri dari pembersihan data, integrasi data, pemilihan data, transformasi data, penambahan data, evaluasi pola, dan penyajian pengetahuan [3], [15]–[17]. Data mining adalah teknik yang digunakan organisasi untuk mengekstrak informasi yang berguna dari data mentah. Perangkat lunak ini diimplementasikan untuk menemukan pola yang diperlukan dalam volume besar data (gudang data) yang dapat membantu bisnis mempelajari pelanggan, memprediksi perilaku pelanggan, dan meningkatkan strategi pemasaran [16], [18]–[20].

### 2.3 Python

Python adalah bahasa pemrograman interpretatif, berorientasi objek dan semantik yang dinamis, high-level struktur data, dynamic typing dan dynamic binding. Python sudah ada sejak tahun 1991, pada tahun itu python pertama kali dikemukakan dan bertahan hingga sampai dengan sekarang, bahkan python masih populer. Python merupakan bahasa pemrograman yang telah banyak digunakan oleh perusahaan-perusahaan besar baik nasional maupun internasional seperti Gojek, Tokopedia, Grab, Facebook, dan Google (<https://www.kdnuggets.com/2012/11/best-python-modules-for-data-mining.html>). Python memiliki sintaks sederhana dan mudah dipelajari untuk penekanan pada kemudahan membaca dan mengurangi biaya perbaikan program (<https://docs.python.org/3.13/tutorial/index.html>).

### 2.3.1 TensorFlow

TensorFlow dirilis pada Oktober 2015, mengubah kerangka kerja dengan banyak cara berdasarkan masukan pengguna, untuk membuatnya lebih mudah digunakan (misalnya, dengan menggunakan API Keras yang relatif sederhana untuk pelatihan model) dan lebih berkinerja. Pelatihan terdistribusi lebih mudah dijalankan berkat API baru, dan dukungan untuk TensorFlow Lite memungkinkan penerapan model di berbagai platform yang lebih besar. Namun, kode yang ditulis untuk versi TensorFlow sebelumnya harus ditulis ulang terkadang hanya sedikit, terkadang secara signifikan untuk memanfaatkan fitur TensorFlow 2.0 baru secara maksimal. TensorFlow memungkinkan developer membuat grafik aliran data struktur yang mendeskripsikan bagaimana data bergerak melalui grafik, atau serangkaian node pemrosesan. Setiap node dalam grafik mewakili operasi matematika, dan setiap koneksi atau tepi antar node adalah larik data multidimensi, atau tensor [21]–[25]. TensorFlow menggabungkan banyak model dan algoritma machine learning yakni deep learning (neural network). Framework ini disusun menggunakan Python front-end API untuk membuat suatu aplikasi penggunaannya, selain itu TensorFlow dapat melatih dan menjalankan neural network untuk keperluan mengklasifikasikan tulisan tangan, pengenalan gambar/objek, serta menggabungkan suatu kata. Selanjutnya re-current neural network yang merupakan model sequential dapat digunakan untuk Natural Language Processing (NLP). Selain itu, TensorFlow digunakan pada skala yang besar untuk produksi dengan menggunakan model yang sama ketika proses training data [1], [26].

### 2.3.2 Numpy

Numpy merupakan salah satu library terbaik yang dimiliki Python. Library ini dapat digunakan untuk banyak case dalam data science. Numpy merupakan library open source, artinya, library numpy dapat digunakan dengan gratis bahkan untuk tujuan komersil (<https://www.w3schools.com/python/numpy/default.asp>). Numpy sendiri merupakan singkatan dari Numerical Python. Biasanya library ini digunakan untuk menghitung operasi matematika pada array [21]. NumPy (Numerical Python) adalah library python yang digunakan untuk bekerja dengan array dan juga memiliki fungsi yang bekerja dalam domain aljabar linier, transformasi fourier, dan matriks. Library yang dibuat pada 2005 oleh Travis Oliphant ini merupakan proyek open source sehingga Anda dapat menggunakannya secara bebas. Meski python memiliki daftar yang melayani tujuan array, prosesnya begitu lambat sehingga memerlukan NumPy yang bisa menyediakan objek array hingga lima puluh kali lebih cepat dari pada daftar python tradisional [22].

### 2.3.3 Scipy

SciPy (Scientific Python) adalah perpustakaan open-source yang digunakan untuk perhitungan ilmiah tingkat tinggi. Jenis library ini dibangun di atas ekstensi NumPy dan bekerja bersama untuk menangani komputasi yang kompleks. NumPy memungkinkan pengurutan dan pengindeksan data array, sementara kode data numerik disimpan di SciPy. Library python ini juga banyak digunakan oleh para developer dan engineer [21].

### 2.3.4 Pandas

Pandas adalah paket Python open source yang paling sering dipakai untuk menganalisis data serta membangun sebuah machine learning. Pandas dibuat berdasarkan satu package lain bernama Numpy, yang mendukung arrays multi dimensi. Pandas dikembangkan oleh Wes McKinney pada 2008, penggunaan Pandas dalam Python sendiri sering kali muncul pada domain akademis hingga komersial, termasuk keuangan, neurosciences, ekonomi, statistik, periklanan, analisis web, dan lain sebagainya [21], [23]. Pandas DataFrame sedikit berbeda karena merupakan struktur dua dimensi. Data yang diproses akan disusun dalam bentuk tabel dengan rows dan column. Umumnya, DataFrame terdiri atas tiga komponen penting: data, rows, dan column untuk membuat DataFrame, pengguna akan memuat datasets dari storage yang sudah dibuat sebelumnya, berupa database SQL, CSV atau bahkan Excel. Kegunaan pandas, terutama dalam bidang data analysis yang membutuhkan task repetitif dan memakan banyak waktu. Fungsi library Pandas antara lain: Data cleansing, Data fill, Normalisasi data, Penggabungan dan penyatuan data, Visualisasi data, Analisis statistik, Inspeksi data, Memuat dan menyimpan data [21].

### 2.3.5 Matplotlib

Matplotlib disusun oleh John Hunter di tahun 2002, dan di desain agar dapat digunakan selayaknya menggunakan MATLAB. Matplotlib dapat digunakan untuk memvisualisasikan data secara 2D maupun 3D dan menghasilkan gambar berkualitas yang bahkan dapat kamu simpan dalam berbagai format gambar, seperti JPEG dan PNG. Jika, kamu menggunakan python script maka kamu perlu menginstall matplotlib terlebih dahulu. Tapi, jika kebetulan kamu menggunakan Python melalui anaconda meliputi spyder dan jupyter notebook maka, kamu tidak perlu menginstallnya lagi karena sudah menjadi built-in library [27]. Matplotlib memiliki banyak contoh grafik yang bisa digunakan untuk memvisualisasikan data agar menjadi lebih menarik. Bar plot yang merupakan jenis grafik yang paling sering digunakan untuk merepresentasikan data numerik dan kategorik dalam bentuk bar. Lalu ada histogram yang digunakan untuk merepresentasikan distribusi frekuensi dan data numerik dengan batang. Selain itu, line plot yaitu jenis grafik untuk menampilkan informasi dengan menggunakan banyak titik yang saling terhubung dan membentuk garis lurus atau lengkung. Ada pula box plot untuk membuat bentuk visualisasi data secara statistik melalui lima dimensi utama yaitu nilai minimum, kuartil 1, kuartil 2, kuartil 3 dan nilai maksimum, box plot sering digunakan untuk memeriksa keberadaan outlier, dan masih banyak ada banyak lagi [28].

### 2.3.6 Keras

Library Keras termasuk API tingkat tinggi yang ditulis dengan Python dan dikembangkan dengan untuk memungkinkan eksperimen cepat, serta sudah open source sejak rilis awal pada Maret 2015. Keras adalah interface library yang bertujuan menyederhanakan implementasi algoritma-algoritma deep learning di atas TensorFlow. TensorFlow sendiri adalah platform high performance computing berbasis alur graph. Keras dikembangkan dan dikelola oleh Francois Chollet dan dirancang untuk manusia, bukan mesin. Termasuk pustaka Python yang kuat dan mudah, Keras digunakan untuk tujuan pengembangan dan pengevaluasian model deep learning. Mengikuti praktik terbaik untuk mengurangi beban kognitif, Keras menawarkan API yang konsisten dan sederhana. Beberapa kelebihan Keras dari model deep learning lain mencakup [27]:

- 1) Memungkinkan pengguna langsung menggunakan model atau mengekstrak fitur-fiturnya tanpa membuat model sendiri.
- 2) Keras adalah API yang dibuat agar mudah dipelajari orang sehingga dibuat sesederhana mungkin.
- 3) Durasi pembuatan prototipe dalam model Keras lebih sedikit sehingga ide-ide Anda dapat diimplementasikan dan disebarkan dalam waktu yang relatif lebih singkat.
- 4) Keras juga menyediakan berbagai opsi penerapan yang tergantung pada kebutuhan pengguna.

### 2.4 Sistem Pengambilan Keputusan

Sistem pengambilan keputusan merupakan sistem pendukung untuk menentukan keputusan pada situasi terstruktur dan tidak terstruktur. Sistem Pendukung Keputusan adalah suatu sistem informasi spesifik yang ditunjukan untuk membantu manajemen dalam mengambil keputusan yang berkaitan dengan persoalan yang bersifat semi terstruktur (Jawad & Mughal, 2018). Sedangkan menurut Little sistem pendukung keputusan sebagai suatu informasi berbasis komputer yang menghasilkan berbagai alternatif keputusan untuk membantu manajemen dalam menangani berbagai permasalahan yang terstruktur maupun tidak terstruktur dengan menggunakan data dan model [5].

### 2.5 Decision Tree

Decision tree merupakan metode klasifikasi yang menggunakan sistem pohon sebagai penentu nilai akhirnya. Konsep dari pohon keputusan adalah mengubah data menjadi decision tree dan aturan-aturan keputusan. Decision tree merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (tree) di mana setiap node merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas [21]. Algoritma decision tree merupakan algoritma yang umum digunakan untuk pengambilan keputusan. Decision tree akan mencari solusi permasalahan dengan menjadikan kriteria sebagai node yang saling berhubungan membentuk seperti struktur pohon. Decision tree adalah model prediksi terhadap suatu keputusan menggunakan struktur hirarki atau pohon. Setiap pohon memiliki cabang, cabang mewakili suatu atribut yang harus dipenuhi untuk menuju cabang selanjutnya hingga berakhir di daun (tidak ada cabang lagi). Konsep data dalam decision tree adalah data dinyatakan dalam bentuk tabel yang terdiri dari atribut dan record. Atribut digunakan sebagai parameter yang dibuat sebagai kriteria dalam pembuatan pohon [5]. Sebuah model pohon keputusan mungkin dibangun dengan seksama secara manual atau bisa tumbuh secara otomatis dengan menerapkan salah satu atau beberapa algoritma pohon keputusan. Banyak algoritma yang dapat dipakai dalam anatara lain ID3, CART, dan C4.5. Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Terdapat empat langkah dalam proses pembuatan pohon keputusan pada algoritma C4.5, yaitu [5]:

- a. Memilih atribut sebagai akar.
- b. Membuat cabang untuk masing-masing nilai.
- c. Membagi setiap kasus dalam cabang.
- d. Mengulangi proses dalam setiap cabang sehingga semua kasus dalam cabang memiliki kelas yang sama.

Kelebihan dan Kekurangan algoritma decision tree C45 [8]:

- a. Dapat Menemukan kombinasi data yang tidak terduga;
- b. Integrasi yang mudah ke dalam sistem basis data;
- c. Memiliki akurasi yang baik;
- d. Lokasi keputusan yang sebelumnya sangat kompleks dan sangat global dapat dibuat dengan lebih sederhana dan lebih spesifik;
- e. Dapat menghilangkan perhitungan yang tidak diperlukan. Dikarenakan dengan metode ini sampel hanya akan diuji berdasarkan kriteria ataupun kelas tertentu; dan
- f. Terdapat pemilihan fitur yang fleksibel dari node internal yang berbeda, sehingga fitur yang dipilih membedakan kriteria lain di node yang sama.

Sedangkan kekurangannya, yaitu:

- a. Tumpang tindih yang terjadi terutama ketika sangat banyak kelas dan kriteria yang digunakan. Ini dapat menyebabkan waktu keputusan yang lebih lama dan memori yang dibutuhkan lebih banyak;
- b. Perhitungan jumlah kesalahan dari setiap level dalam pohon keputusan lumayan besar;
- c. Kesulitan merancang pohon keputusan yang optimal; dan
- d. Hasil kualitas keputusan yang didapat dengan menggunakan metode pohon keputusan sangat tergantung kepada bagaimana pohon itu dirancang.

Persamaan C4.5 [14]:

Secara umum algoritma C4.5 membangun pohon keputusan adalah sebagai berikut [3]:

- Pilih atribut sebagai akar
- Buat cabang untuk tiap-tiap nilai
- Bagi kasus dan cabang.
- Ulangi proses untuk setiap cabang sampai semua kasus pada cabang cabang untuk memilih kelas yang sama.

Rumus Gain:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

Keterangan :

S: Himpunan Kasus

A: Atribut

n : Jumlah atribut A

|S<sub>i</sub>| : Jumlah kasus Pada Partisi ke-1

|S| : Jumlah kasus dalam S

Sementara itu, untuk menghitung nilai entropy sebagai berikut:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

Keterangan :

S : Himpunan Kasus

A: Fitur

N: Jumlah Partisi S

Pi : Proporsi dari S<sub>i</sub> terhadap S

## 2.7 Penelitian Terdahulu

Berikut ini adalah tabel 1 yang berisi penelitian yang terkait Prediksi Calon Mahasiswa Penerima KIP Pada Universitas Indo Global Mandiri menggunakan Algoritma Decision Tree.

**Tabel 1.** Penelitian Sebelumnya

Judul penelitian, penulis dan tahun	Focus penelitian	Hasil Penelitian
“Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji [1].	Metode decision tree algoritma C.45 dalam mengklasifikasi data penjualan	Dari hasil percobaan pencarian pohon hasil keputusan dari data penjualan gerai makanan cepat saji menggunakan algoritma C4.5 dihasilkan nilai entropy dan gain tertinggi yaitu 1,501991 pada atribut-atribut Menu Makanan pada perhitungan manual.
“Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree Dan Artificial Neural Network” [29].	Implementasi Decision Tree dan Artificial Neural Network dalam penentuan kelulusan.	Berdasarkan hasil pengujian metode decision tree memiliki akurasi sebesar 74,51% dan artificial neural network sebesar 79,74%.
“Implementasi Data Mining Dengan Metode C4.5 Untuk Prediksi Mahasiswa Penerima Beasiswa”[14].	Implementasi Data Mining Dengan Metode C4.5	Berdasarkan hasil dapat diketahui bahwa diantara total 100 data mahasiswa yang diolah didapatkan sebanyak ada sebanyak 9 orang bernilai YA menerima beasiswa dan 91 orang bernilai TIDAK.
“A Review on Data Mining and Machine Learning Methods for Student Scholarship Prediction”[30].	Penggunaan machine Learning (ML) dalam prediksi beasiswa mahasiswa	Menggunakan data minig dan ML dapat mempermudah prediksi penerima beasiswa mahasiswa

Pada tabel 1 di atas adalah daftar penelitian sebelumnya yang berisi Judul penelitian, penulis dan tahun serta focus penelitian serta hasil penelitian terkait dengan Prediksi Calon Mahasiswa Penerima KIP Pada Universitas Indo Global Mandiri menggunakan Algoritma Decision Tree.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Preprocessing

Preprocessing bertujuan untuk mengilangkan data-data yang tidak diperlukan, serta membersihkan data-data dari atribut



atau simbol-simbol yang mengganggu. Proses preprocessing ini akan digunakan untuk mengambil data-data yang akan digunakan saja dan akan menghilangkan data yang tidak digunakan. Preprocessing data mempunyai beberapa tahapan yaitu :

### 3.2 Data Cleaning

Tahapan kerja dari cleaning ini yaitu, data dibersihkan melalui beberapa proses seperti mengisi nilai yang hilang, menghilangkan noisy data, dan menyelesaikan inkonsistensi yang ditemukan. Data juga dibersihkan dengan cara dibagi menjadi segmen-segmen yang memiliki ukuran serupa lalu dihaluskan (binning), mengelompokkannya ke dalam kelompok-kelompok data yang serupa (grouping).

**Tabel 2.** Data Hasil Proses Cleaning

Program Keahlian	Nilai Rata- rata	Prestasi	Status KK	Penghasilan	Kartu Bantuan	Keterangan
Manajemen	78	2	LK	416667	PKH	Tidak Lulus
IPA	73	0	LK	333333	PKH	Tidak Lulus
IPA	78	0	LK	250000	PKH	Tidak Lulus
IPA	77	0	LK	416667	SKTM	Lulus
Adm	81	1	CH	250000	PKH	Tidak Lulus
Akuntansi	93	0	LK	300000	PKH	Tidak Lulus
IPA	87	4	LK	687500	PKH	Tidak Lulus
Adm	81	0	CM	200000	PKH	Lulus
IPA	78	4	LK	250000	SKTM	Lulus
TKJ	62	1	LK	225000	PKH	Lulus
IPS	71	3	LK	400000	SKTM	Tidak Lulus
Akuntansi	94	0	LK	250000	PKH	Tidak Lulus
Akuntansi	87	5	LK	200000	SKTM	Tidak Lulus

Keterangan :

LK : Lengkap; CM: Cerai Mati; CH: Cerai Hidup; YT: Yatim; SKTM: Surat Keterangan Tidak Mampu PKH: Program Keluarga Harapan

Pada tabel 2 di atas adalah data hasil proses cleaning terkait penelitian terkait dengan Prediksi Calon Mahasiswa Penerima KIP Pada Universitas Indo Global Mandiri menggunakan Algoritma Decision Tree. Pada tabel tersebut ditampilkan data Program Keahlian, Nilai Rata- rata, Prestasi, Status KK, Penghasilan, Kartu Bantuan, dan Keterangan.

**Tabel 3.** Data Transformasi

<b>Nilai</b>	Sangat Baik	Nilai > 90
	Baik	Nilai > 60
	Cukup Baik	Nilai ≤ 60
<b>Prestasi</b>	Sangat Berprestasi	Prestasi > 10
	Berprestasi	Prestasi > 5
	Kurang	Prestasi ≤ 5
	Berprestasi	
<b>Penghasilan</b>	Cukup	Penghasilan > 750.000
	Sedang	Penghasilan > 500.000
	Rendah	Penghasilan ≤ 500.000

Pada tabel 3 di atas adalah data hasil transformasi terkait penelitian terkait dengan Prediksi Calon Mahasiswa Penerima KIP Pada Universitas Indo Global Mandiri menggunakan Algoritma Decision Tree. Pada tabel tersebut ditampilkan data Nilai dengan rentang Sangat Baik, Baik, Cukup Baik. Untuk Data Prestasi dengan rentang Sangat Berprestasi, Berprestasi dan Kurang Berprestasi. Dan data Penghasilan antara lain Cukup, Sedang dan Rendah.

### 3.3 Proses Data Mining

Tahapan Data Mining untuk memprediksi Calon Mahasiswa Penerima KIP pada Universitas Indo Global Mandiri Palembang menggunakan algoritma Decision Tree (C4.5). Hasil pengolahan data Decision Tree selanjutnya akan dibagi menjadi 2 bagian yaitu, data training dan data testing. Proses dalam pembentukan tree data yang digunakan sebagai data training, sedangkan data testing digunakan untuk menghitung tingkat keakuratan dari prediksi calon mahasiswa penerima KIP. Langkah awal dari proses pembangunan tree yaitu menghitung setiap kejadian yang muncul pada tiap atribut menggunakan persamaan entropy pada decision tree C4.5 (Tabel 4).

**Tabel 4.** Data Hasil Jumlah Kemunculan Kejadian

Node	Kategori	Jumlah	Lulus	Tidak Lulus
<b>Total</b>		418	147	271
<b>Program keahlian</b>	IPA	154	55	99

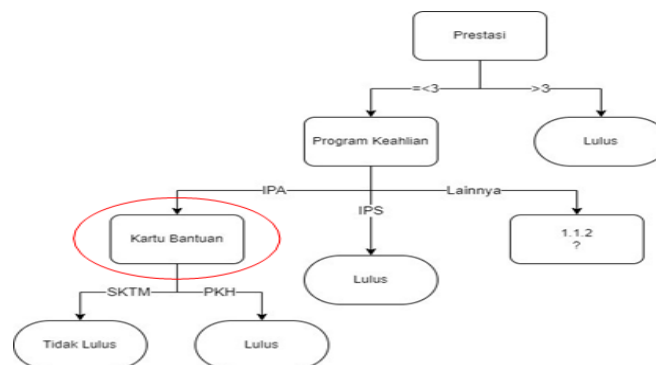
Node	Kategori	Jumlah	Lulus	Tidak Lulus
Nilai Rata-rata	IPS	140	37	103
	Lainnya	124	55	69
	Sangat Baik	15	6	9
	Baik	402	141	261
	Cukup Baik	1	0	1
Prestasi	Sangat Berprestasi	15	13	2
	Berprestasi	22	11	11
	Kurang Berprestasi	381	123	258
Status KK	LK	333	110	223
	CM	56	25	31
	CH	23	10	13
	YT	6	2	4
Penghasilan	Cukup	9	0	9
	Sedang	62	24	38
	Rendah	347	123	224
Kartu Bantuan	SKTM	211	69	142
	PKH	207	78	129

Pada tabel 4 di atas adalah Data Hasil Jumlah Kemunculan Kejadian terkait penelitian dengan Prediksi Calon Mahasiswa Penerima KIP Pada Universitas Indo Global Mandiri menggunakan Algoritma Decision Tree. Pada tabel tersebut ditampilkan data Program keahlian antara lain IPA, IPS dan Lainnya. Untuk data Nilai Rata-rata memiliki kategori Sangat Baik, Baik, Cukup Baik dan Sangat Berprestasi. Selanjutnya, data Prestasi antara lain Berprestasi dan Kurang Berprestasi. Untuk data Status KK antara lain LK, CM, CH, YT serta Cukup. Data Penghasilan antara lain Sedang, Rendah serta untuk data Kartu Bantuan memiliki status SKTM, dan PKH.

**Tabel 5.** Hasil Perhitungan Node Level 2 (Node 1.1.1)

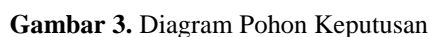
Node 1.1.1	Kategori	Jumlah/Total	Keterangan		Nilai	
			Tidak Lulus	Lulus	Entropy	Gain
PK (IPA)		4	2	2	1	
Nilai	=<85	1	0	1	0	0.311278
	>85	3	2	1	0.918296	
	LK	4	2	2	1	
Status KK	CM	0	0	0	0	0
	CH	0	0	0	0	
	YT	0	0	0	0	
Peghasilan	=<750000	4	2	2	1	0
	>750000	0	0	0	0	
Kartu Bantuan	SKTM	3	2	0	0	1
	PKH	2	0	2	0	

Pada tabel 5 di atas mengenai Hasil Perhitungan Node Level 2 (Node 1.1.1) terkait penelitian dengan Prediksi Calon Mahasiswa Penerima KIP Pada Universitas Indo Global Mandiri menggunakan Algoritma Decision Tree. Nilai informations gain terbesar (Tabel 5) didapatkan pada atribut Kartu Bantuan dengan nilai informations gain =1, maka atribut Kartu Bantuan akan menjadi bagian dari tree sebagai node level 2 (Node 1.1.1) (Gambar 2).



**Gambar 2.** Decision Tree Node Level 2 (Node 1.1.1)

Training merupakan query yang membentuk tree (pohon). Pembentukan tree dilakukan dengan cara menghitung nilai dari entropy dan nilai dari informations gain (Gambar 3). Tahapan inilah pola prediksi calon mahasiswa penerima KIP yang akan terbentuk.



### 3.4 Perhitungan Tingkat Akurasi

A heatmap visualization of the confusion matrix for the 'vhead' variable. The x-axis is labeled 'vhead' and has categories 0 and 1. The y-axis is labeled 'ytrue' and has categories 0 and 1. The color scale ranges from 10 (dark purple) to 50 (light orange). The values in the cells are: (0,0) = 10, (0,1) = 15, (1,0) = 8, and (1,1) = 51.

	0	1
0	10	15
1	8	51

### Gambar 2. Grafik Confusion Matrix

$$F1 - Score = \frac{2*Recall*Percision}{Recall+Precision} = \frac{2*0,556*0,4}{0,566+0,4} = \frac{0,445}{0,956} = 0,47$$



Perhitungan Confusion Matrix seperti di atas digunakan untuk mendapatkan hasil dari persentase pada Algoritma C4.5 untuk menentukan prediksi yang akan dilakukan (Tabel 6).

**Tabel 6.** Hasil Perhitungan Confusion Matrix

Perhitungan	Hasil	Persentase
Accuracy	0,73	73%
Precision	0,40	40%
Recall	0,56	56%
F1-Score	0,47	47%

Pada tabel 6 di atas adalah mengenai Hasil Perhitungan Confusion Matrix terkait penelitian dengan Prediksi Calon Mahasiswa Penerima KIP Pada Universitas Indo Global Mandiri menggunakan Algoritma Decision Tree. Dari hasil tabel 6 di atas dihasilkan nilai Accuracy dengan hasil 0,73 atau sebesar 73%. Nilai Precision dengan hasil 0,40 atau sebesar 40%, Nilai Recall dengan hasil 0,56 atau sebesar 56% serta nilai F1-Score dengan hasil 0,47 atau sebesar 47%.

## 4 KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang sudah dilakukan, maka didapatkan beberapa kesimpulan bahwa tingkat akurasi yang didapatkan dari prediksi menggunakan algoritma decision tree memiliki presentase yang cukup tinggi yaitu bernilai 73%, untuk prediksi calon mahasiswa penerima KIP Universitas Indo Global Mandiri Palembang. Berdasarkan prediksi yang dilakukan menggunakan algoritma decision tree C4.5 didapatkan hasil kinerja yang cukup rendah dengan besaran nilai precision sebesar 40%, nilai recall sebesar 56% dan nilai F1-score sebesar 47%, sebagai nilai prediksi calon mahasiswa penerima Universitas Indo Global Mandiri Palembang.

## REFERENCES

- [1] W. K. Tsang and D. F. Benoit, "Gaussian processes for daily demand prediction in tourism planning," *J. Forecast.*, vol. 39, no. 3, pp. 551–568, 2020, doi: 10.1002/for.2644.
- [2] A. W. N. Ulfy and P. A. R. Devi, "Penentuan Kenaikan Jabatan Menggunakan Pembobotan Metode AHP dan Didukung Metode Complex Proportional Assessment," *J. Sist. Komput. dan Inform.*, vol. 3, no. 3, p. 232, 2022, doi: 10.30865/json.v3i3.3867.
- [3] Yuli Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database ( KDD ) . Jurnal Edik Informatika," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2019.
- [4] M. Arfan et al., "Sistem Penunjang Keputusan Penentuan Prioritas Pembangunan Desa Dengan Metode AHP-TOPSIS," *J. Syst. Inf. Technol.*, vol. 3, no. 1, 2023.
- [5] Dr. V. Suma, "Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics," *J. Soft Comput. Paradig.*, vol. 2, no. 3, pp. 153–159, 2020, doi: 10.36548/jscp.2020.3.002.
- [6] G. Ouddai, I. Hamdi, and H. Ben Ghezala, "A Comparative Study of BRISK, ORB and DAISY Features for Breast Cancer Classification," *Int. Conf. Pattern Recognit. Appl. Methods*, vol. 1, no. Icpam, pp. 964–970, 2023, doi: 10.5220/0011902200003411.
- [7] H. March, "The Smart City and other ICT-led techno-imaginaries: Any room for dialogue with Degrowth?," *J. Clean. Prod.*, vol. 197, pp. 1694–1703, 2018, doi: 10.1016/j.jclepro.2016.09.154.
- [8] D. S. Prasvita, D. Chahyati, and A. M. Arymurthy, "Automatic Detection of Oil Palm Growth Rate Status with YOLOv5," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 3, pp. 529–537, 2023, doi: 10.14569/IJACSA.2023.0140361.
- [9] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *Int. J. Remote Sens.*, vol. 39, no. 9, pp. 2784–2817, 2018, doi: 10.1080/01431161.2018.1433343.
- [10] Y. Zeng, Y. Lan, Y. Hao, C. Li, and Q. Zheng, "Leveraging Multi-grained Sentiment Lexicon Information for Neural Sequence Models," *arxiv Comput. Lang.*, 2018, doi: <https://doi.org/10.48550/arXiv.1812.01527>.
- [11] A. P. Rodrigues et al., "Real-Time Twitter Spam Detection and Sentiment Analysis using Machine Learning and Deep Learning Techniques," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/5211949.
- [12] N. A. Husin, S. Khairunniza-Bejo, A. F. Abdullah, M. S. M. Kassim, D. Ahmad, and M. H. A. Aziz, "Classification of basal stem rot disease in oil palm plantations using terrestrial laser scanning data and machine learning," *Agronomy*, vol. 10, no. 11, 2020, doi: 10.3390/agronomy10111624.
- [13] K. Park, J. S. Hong, and W. Kim, "A Methodology Combining Cosine Similarity with Classifier for Text Classification," *Appl. Artif. Intell.*, vol. 34, no. 5, pp. 396–411, 2020, doi: 10.1080/08839514.2020.1723868.
- [14] P. Alam Jusia, F. Muhammad Irfan, and S. Dinamika Bangsa Jambi Jl Jend Sudirman Thehok Jambi, "Clustering Data Untuk Rekomendasi Penentuan Jurusan Perguruan Tinggi Menggunakan Metode K-Means," *J. IKRA-ITH Inform.*, vol. 3, no. 3, p. 75, 2019.
- [15] B. Bogdándy, "WiFi RSSI preprocessing library for Android," *Proc. 2018 19th Int. Carpathian Control Conf. ICCCC 2018*, pp. 649–654, 2018, doi: 10.1109/CarpathianCC.2018.8399710.
- [16] V. N. Sari, L. Y. Astri, and E. Rasywir, "Analisis Dan Penerapan Algoritma Naive Bayes Untuk Evaluasi," *J. Ilm. Mhs. Tek. Inform.*, vol. 2, no. 1, pp. 53–68, 2020.
- [17] Y. D. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2017, doi: 10.37676/jmi.v12i2.418.
- [18] A. Rohmah, F. Sembiring, and ..., "Implementasi Algoritma K-Means Clustering Analysis Untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: Smk Yaspin ...)," *... Sist. Inf. dan ...*, pp. 290–298, 2021.
- [19] D. Zaenal Abidin and E. Rasywir, "Penerapan Data Mining Klasifikasi Untuk Memprediksi Potensi Mahasiswa Berprestasi Di

- Stikom Dinamika Bangsa Jambi Dengan Metode Naive Bayes,” *J. Ilm. Mhs. Tek. Inform.*, vol. 3, no. 2, 2021.
- [20] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [21] K. Munir, H. Elahi, A. Ayub, F. Frezza, and A. Rizzi, “Cancer diagnosis using deep learning: A bibliographic review,” *Cancers (Basel)*, vol. 11, no. 9, pp. 1–36, 2019, doi: 10.3390/cancers11091235.
- [22] S. K. Yoki Donzia and H. K. Kim, “Implementation of recurrent neural network with sequence to sequence model to translate language based on tensorflow,” *Lect. Notes Eng. Comput. Sci.*, vol. 2237, pp. 375–379, 2018.
- [23] I. K. S. Buana, “DETEKSI GERAKAN KEPALA DAN KEDIPAN MATA DENGAN HAAR CASCADE CLASSIFIER CONTOUR DAN MORFOLOGI DALAM PENGOPERASIAN KOMPUTER UNTUK KAUM DIFABLE,” *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 5, no. 1, pp. 29–36, 2019, doi: 10.33330/jurteks.v5i1.273.
- [24] F. Utami, S. Suhendri, and M. Abdul Mujib, “Implementasi Algoritma Haar Cascade pada Aplikasi Pengenalan Wajah,” *J. Inf. Technol.*, vol. 3, no. 1, pp. 33–38, 2021, doi: 10.47292/joint.v3i1.45.
- [25] A. B. Adege, H. P. Lin, G. B. Tareegn, Y. Y. Munaye, and L. Yen, “An indoor and outdoor positioning using a hybrid of support vector machine and deep neural network algorithms,” *J. Sensors*, vol. 2018, 2018, doi: 10.1155/2018/1253752.
- [26] Izzudin, Hamzah, Nisrafariza, and Idris, “ANALYSIS OF MULTISPECTRAL IMAGERY FROM UNMANNED AERIAL VEHICLE (UAV) USING OBJECT-BASED IMAGE ANALYSIS FOR DETECTION OF GANODERMA DISEASE IN OIL PALM,” *J. Oil Palm Res.*, vol. 5, no. 5, pp. 543–546, 2020, [Online]. Available: <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52008DC0800&from=EL%0Ahttp://search.jamas.or.jp/link/ui/2005280140>.
- [27] F. S. Khan, M. I. Abbasi, M. Khurram, M. N. Haji Mohd, and M. D. Khan, “Breast cancer histological images nuclei segmentation and optimized classification with deep learning,” *Int. J. Electr. Comput. Eng.*, vol. 12, no. 4, pp. 4099–4110, 2022, doi: 10.11591/ijece.v12i4.pp4099-4110.
- [28] J. You, W. Liu, and J. Lee, “A DNN-based semantic segmentation for detecting weed and crop,” *Comput. Electron. Agric.*, vol. 178, no. September, p. 105750, 2020, doi: 10.1016/j.compag.2020.105750.
- [29] H. Priyatman, F. Sajid, and D. Haldivany, “Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan,” *JEPIN*, vol. 5, no. 1, pp. 62–66, 2019.
- [30] T. E. Salais-Fierro, J. A. Saucedo-Martinez, R. Rodriguez-Aguilar, and J. M. Vela-Haro, “Demand prediction using a soft-computing approach: A case study of automotive industry,” *Appl. Sci.*, vol. 10, no. 3, 2020, doi: 10.3390/app10030829.