

Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indonesia Pintar (KIP)

Muhamad Yunus¹, Hanandriya Ramadhan², Dimas Rizki Aji³, Agus Yulianto⁴

¹Program Studi Sistem Informasi, Universitas Nusa Mandiri Jakarta
e-mail: 1muhamadyunus908@gmail.com, 2dimasrizkiaji@gmail.com, 3hanandriyaramadhan@gmail.com,
⁴agus.aag@nusamandiri.ac.id

Abstraksi - Pendidikan di Indonesia selalu berkonflik dengan masalah kemiskinan dan prestasi siswa dievaluasi setiap akhir semester untuk mengetahui hasil belajar yang telah dicapai. Bantuan KIP (Kartu Indonesia Pintar) dapat berupa uang sekolah, bantuan belajar siswa dan penghargaan bagi siswa yang berprestasi atau bagi siswa yang kurang mampu. Maka penelitian ini untuk memudahkan sekolah SMPN 38 Jakarta dalam menyeleksi dan juga lebih tepat sasaran guna mengoptimalkan sekolah untuk lebih menyeleksi calon KIP (Kartu Indonesia Pintar), dengan ini mampu meringankan tanggungan orang tua dalam biaya sekolah dan juga meningkatkan keinginan belajar siswa. Metode penelitian yang digunakan adalah metode klasifikasi Decision Tree data mining dengan algoritma C4.5. Berdasarkan hasil uji coba dan pembahasan yang telah dilakukan untuk pemilihan penerima Kartu Indonesia Pintar (KIP) dapat diselesaikan. Data siswa yang diperoleh dari SMPN 35 Jakarta dapat diprediksi dengan akurasi yang cukup tinggi, hasil ini diperoleh dengan akurasi mencapai 90,24% dengan *errorate* 9,75% sehingga dengan hasil tersebut algoritma C4.5 sangat baik dalam menentukan pemilihan penerima Kartu Indonesia Pintar (KIP).

Kata Kunci : KIP (Kartu Indonesia Pintar), Data Mining, Klasifikasi, *Decision Tree*, C4.5.

Abstracts - *Education in Indonesia is always in conflict with the problem of poverty and student achievement is evaluated at the end of each semester to determine the learning outcomes that have been achieved. KIP (Smart Indonesia Card) assistance can be in the form of tuition fees, student study assistance and awards for students who excel or for students who are less fortunate. So this research is to make it easier for SMPN 35 Jakarta schools in selecting and also being more targeted in order to optimize schools to better select KIP (Indonesia Cardcandidates Smart), with this it is able to ease the burden of parents in school fees and also increase students' desire to learn. The research method used is theclassification method Decision Tree data mining with the C4.5 algorithm. Based on the results of trials and discussions that have been carried out, the selection of KIP recipients (Kartu Indonesia Pintar) can be completed. Student data obtained from SMPN 35 Jakarta can be predicted with accuracy, high the results obtained with the accuracy of 90.24% with errorate 9,75% so with these results the C4.5 algoring is very good in determining the selection of KIP (Smart Indonesia Card).*

Keywords : KIP (Indonesia Cardcandidates Smart), Data Mining, Classification, Decision Tree, C4.5.

PENDAHULUAN

Pendidikan di Indonesia selalu bertentangan dengan masalah kemiskinan dan prestasi anak didik dievaluasi setiap akhir semester untuk mendeteksi hasil belajar yang sudah dicapai. Kemiskinan ini nanti menjadi alasan seorang mengajukan permintaan KIP (Kartu Indonesia Pintar) untuk lembaga pendidikan. Berbagai upaya telah dilakukan, antara lain pemberian dana bantuan sekolah melalui program Surat Keterangan Tidak Mampu (SKTM) yang dikeluarkan oleh kecamatan setempat. Agar penerima KIP ini terus terpacu meningkatkan prestasi akademisnya, tidak salah apabila lembaga pendidikan tidak serta merta menyetujui permintaan KIP yang diajukan (Pauziah, 2016).

SMPN 38 Jakarta merupakan sekolah menengah pertama yang beralamat di Jl. Karet Pasar Baru Barat I No.2, RT.2/RW.6, Karet Tengsin, Kecamatan Tanah Abang, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10220. Sekolah adalah suatu lembaga penyelenggara pendidikan bagi siswa (Ardiansyah, 2018). Banyak fasilitas – fasilitas yang kita dapatkan di sekolah seperti KIP (Kartu Indonesia Pintar) merupakan program pengembangan agar siswa terus dapat melanjutkan studinya, bantuan KIP (Kartu Indonesia Pintar) dapat berupa biaya pendidikan, bantuan belajar siswa dan suatu penghargaan untuk siswa yang berprestasi atau untuk siswa yang kurang mampu dengan adanya penerapan metode data mining untuk menyeleksi pemilihan penerima KIP (Kartu Indonesia Pintar) dapat mempermudah mendapatkan kriteria calon penerima. Dalam



penelitian ini permasalahan yang dihadapi untuk mendapatkan KIP (Kartu Indonesia Pintar) ada beberapa kriteria - kriteria yang sudah ditetapkan pada program studi ini seperti mata pelajaran, penghasilan orang tua, tanggungan orang tua, dan lain - lain. oleh sebab itu tidak semua yang mendaftarkan diri sebagai calon penerima KIP (Kartu Indonesia Pintar) dapat di terima, hanya yang memenuhi kriteria dan syarat tersebut yang dapat menerima KIP (Kartu Indonesia Pintar), maka perlu adanya penerapan metode data mining dengan algoritma C4.5 untuk membantu menyeleksi pemilihan siapa yang berhak menerima KIP (Kartu Indonesia Pintar) tersebut. Data mining adalah serangkaian proses untuk menggali nilai tambahan berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Algoritma C4.5 menurut (Vulandari, 2017) merupakan salah satu metode yang banyak digunakan untuk menghasilkan decision tree (Adinungroho, 2018). Algoritma yang dapat digunakan pada penelitian disini yaitu algoritma C4.5 untuk mempermudah menyeleksi pemilihan penerima KIP (Kartu Indonesia Pintar) di SMPN 38 Jakarta dengan tingkat akurasi yang tinggi. Dengan jumlah peserta yang mengajukan KIP (Kartu Indonesia Pintar) banyak serta indikator kriteria yang banyak, maka dibutuhkanlah suatu sistem yang dapat membantu untuk menentukan siapa yang berhak untuk mendapatkan KIP (Kartu Indonesia Pintar) dari SMPN 38 Jakarta.

METODOLOGI PENELITIAN

Metode penelitian merupakan suatu faktor yang sangat penting untuk memecahkan suatu masalah dan turut menentukan keberhasilan suatu penelitian. Metode penelitian pada dasarnya merupakan *cara ilmiah* untuk mendapatkan *data* dengan *tujuan* dan *kegunaan* tertentu (Prof.Dr.Sugiono, 2017). Penelitian yang peneliti gunakan peneliti ambil sesuai dengan judul yang merupakan metode klasifikasi *Decision Tree* data mining dengan algoritma C4.5, klasifikasi *Decision Tree* merupakan suatu metode pengelompokan atau classification data yang sangat mudah dimengerti (Prasetyowati, 2017). Ada banyak metode data mining yang digunakan untuk tujuan yang berbeda. Metode klasifikasi digunakan untuk membantu memahami pengelompokan data. Pada penelitian ini peneliti juga menggunakan metode algoritma C4.5, Algoritma C4.5 merupakan salah satu algoritma yang secara luas digunakan, khususnya di area *machine learning* yang memiliki beberapa perbaikan dari algoritma sebelumnya yaitu ID3. Algoritma C4.5 dan ID3 merupakan model yang tak terpisahkan, karena membangun sebuah pohon keputusan, dibutuhkan algoritma C4.5. Menjelaskan kronologis penelitian, termasuk desain penelitian, prosedur penelitian (dalam bentuk algoritma, Pseudocode atau lainnya), bagaimana untuk menguji dan akuisisi data. Deskripsi dari program penelitian harus didukung referensi, sehingga penjelasan

tersebut dapat diterima secara ilmiah.

1. Data Mining

Data *mining* merupakan langkah analisis terhadap proses pengetahuan didalam basis *knowledge discovery in database* (KDD). KDD berhubungan dengan integrasi dengan penemuan *ilmiah*, *Interpretasi* dan *visualisasi* dari pola-pola sejumlah data (Dr. Suyanto, 2017). Data mining menurut (Vulandari, 2017) adalah serangkaian proses untuk menggali nilai tambahan berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Berdasarkan dari sejumlah pengertian diatas, meyakini pandangan peneliti data mining adalah suatu tempat untuk mengolah data yang jumlahnya cukup banyak bahkan sampai ribuan untuk di olah dengan metode-metode serta adanya perhitungan data dan penyusunan data secara spesifik, kemudian membentuk sebuah informasi.

2. Algoritma C4.5

Algoritma C4.5 adalah algoritma yang digunakan untuk membentuk pohon keputusan, yang merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan-aturan dapat dengan mudah dimengerti dalam bahasa alami. (Sugiyarti, n.d.).

Dengan penjelasan di atas algoritma C4.5 kita gunakan untuk mempermudah peneliti memprediksi sebuah nilai menjadi informasi, dan menurut peneliti algoritma C4.5 merupakan pengembangan dari ID3 dengan membentuk pohon keputusan untuk memprediksi atau memperkirakan sebuah kejadian mendapatkan informasi.

Hasil dari algoritma C4.5 adalah pohon keputusan, sehingga mudah untuk dipahami dan dipahami proses penggunaan algoritma ini untuk menyelesaikan kasus adalah sebagai berikut (Azahari, 2021):

1) Mencari nilai Gain

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

Dimana keterangan dari rumus tersebut adalah:

- S = Himpunan Kasus
- A = atribut
- n = jumlah partisi
- pi = porsi dari si kepada s
- |si| = jumlah kasus pada nilai atribut/kriteria
- |s| = jumlah kasus atau total kasus

2) Mencari nilai Entropy

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

Keterangan:

S : Himpunan Kasus

N: Jumlah Partisi S

Pi: Proporsi dari Si terhadap S.

3. Decision Tree

Decision tree dalam (Prasetyowati, 2017) adalah suatu metode pengelompokan atau classification data yang sangat mudah dimengerti. Dalam perkembangan decision tree dijadikan sebagai acuan penalaran untuk mempermudah memahami dari permasalahan yang dihadapi. Metode ini sangat membantu untuk menggambarkan aturan yang dapat dengan mudah dipahami dengan bahasa alami dan juga dapat diekspresikan dalam bentuk bahasa basis data seperti Structured Query Language untuk mencari record pada kategori tertentu.

Decision tree menurut peneliti merupakan metode pengelompokan atau klasifikasi untuk mempermudah model klasifikasi yang tidak relevan dan mudah dimengerti. Salah satu metode decision tree yang peneliti ambil yaitu C4.5 yang termasuk dari bagian metode decision tree terpopuler.

Dalam referensi (Prasetyowati, 2017) Pohon Keputusan adalah pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukkan. Pohon yang dibentuk tidak selalu berupa pohon biner. Jika semua fitur dalam data set menggunakan 2 macam nilai kategorikal maka bentuk pohon yang didapatkan berupa pohon biner. Jika dalam fitur berisi lebih dari 2 macam nilai kategorikal atau menggunakan tipe numerik maka bentuk pohon yang didapatkan biasanya tidak berupa pohon biner.

Teknik pengumpulan data

Teknik pengumpulan data di penelitian ini untuk mendapatkan sebuah informasi yang dibutuhkan dalam mencapai sebuah penelitian. Teknik pengumpulan data dalam penelitian ini menggunakan observasi, studi Pustaka, dan wawancara.

1) Observasi

Dalam penelitian ini, peneliti terlibat dengan orang yang sedang diamati atau digunakan sebagai sumber data penelitian (Prof.Dr.Sugiono, 2017). Metode pengumpulan data yang digunakan untuk mengamati langsung, melihat dan mengambil suatu data di sekolah SMPN 38 JAKARTA.

2) Studi Pustaka

Data dalam penelitian ini kebanyakan diperoleh dari sumber manusia atau *human resources*, melalui observasi. Metode ini di gunakan untuk

mengumpulkan data berupa sejarah singkat sekolah, data-data siswa, sarana dan prasarana sekolah. Teknik ini untuk mendapatkan data yang berkenan dengan judul penelitian ini.

3) Wawancara

Untuk memahami aturan yang perlu ditetapkan dalam mengidentifikasi calon penerima Kartu Indonesia Pintar (KIP), penelitian ini mewawancarai wakil kepala sekolah bidang kesiswaan di SMPN 38 Jakarta.

Tahapan Penelitian

Dalam penelitian ini analisis data di lakukan dengan tahapan proses *Discovery in Databases* (KDD), tahapan tersebut dijelaskan sebagai berikut:

1) *Data Selection*

Tahapan ini data yang didapat dari SMPN 38 Jakarta adalah data calon penerima KIP (Kartu Indonesia Pintar), dengan jumlah 31 data dan atribut variabel yang ada sebanyak 9 atribut data yang digunakan sebagai berikut:

Tabel 1
keterangan atribut

Atribut	Detail Penggunaan
Nomer Induk	✓ Id
Nama Siswa	× No
Jenis Kelamin	✓ Variabel Bebas
Absensi	✓ Variabel Bebas
Nilai	✓ Variabel Bebas
Tanggungans Orang Tua	✓ Variabel Bebas
Bangunan Rumah	✓ Variabel Bebas
Penghasilann Orang Tua	✓ Variabel Bebas
Diterima	✓ Variabel Target

Sumber : (Yunus, Ramdhan, & Aji, 2021)

tabel diatas menjelaskan mengenai atribut-atribut yang akan digunakan dalam penelitian, indikator *yes* (✓) menandakan bahwa atribut bersangkutan akan digunakan dalam penelitian, sedangkan indikator *no* (×) menandakan atribut tersebut tidak digunakan pada tahap *data mining*.

2) *Processing (Cleaning)*

Dengan 9 atribut *predictor* dan 1 atribut target dan ada perubahan nama atribut yang bertujuan mempersingkat penyebutannya data terbut di urakan pada tabel dibawah ini:

Tabel 2
Absensi

0 – 2 hari	Rajin
3> hari	Tidak Rajin

Sumber : (Yunus et al., 2021)

Tabel 3
Nilai Rata Rata

<60	Kurang
60-70	Cukup
71-80	Baik
>80	Sangat Baik

Sumber : (Yunus et al., 2021)

Tabel 4
Tanggung Orang Tua

1	Sedikit
2-3	Cukup
>4	Banyak

Sumber : (Yunus et al., 2021)

Tabel 5
Pendapatan Orang Tua

<1.500.000	Sedikit
1.500.000 – 2.000.000	Rata - Rata
>2.000.000	Tinggi

Sumber : (Yunus et al., 2021)

Dari pengelompokan data yang telah dilakukan, maka atribut-atribut dapat dilihat sebagai berikut:

Tabel 6
Atribut dan Value

Atribut	Value
Absensi	Rajin
	Tidak Rajin
Nilai	Sangat Baik
	Baik
Tanggung Orang Tua	Banyak
	Cukup
	Sedikit
Bangunan Rumah	Sewa
	Permanen
Pendapatan Orang Tua	Tinggi
	Rata-Rata
	Rendah

Sumber : (Yunus et al., 2021)

3) Transformation

Dalam tahap ini data tidak ada yang di bersihkan, setelah dilakukan proses tersebut dilakukan pembagian data set dengan menggunakan process split data untuk menghasilkan data training dan data testing dengan rasio 60:40. Data training berjumlah 41 dan data testing berjumlah 31 dari total keseluruhan yaitu 102 data.

4) data Mining

Pada tahap ini, data diproses, diklasifikasikan menurut model, dan kemudian Buat beberapa aturan. Penelitian ini menggunakan algoritma C4.5

menggunakan langkah-langkah dalam algoritma C4.5.

5) Interpretation / Evaluation

Tahap ini digunakan untuk mengevaluasi empat hasil prediksi yang dihasilkan Algoritma dan metode algoritma yang dipilih Menghasilkan nilai yang mendekati klasifikasi data Sebenarnya.

HASIL DAN PEMBASAN

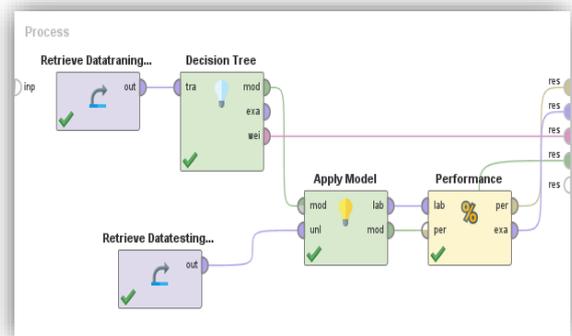
Hasil penelitian ini diharapkan untuk menentukan nilai akurasi Algoritma C4.5 dalam penyekleksian calon penerima KIP (Kartu Indonesia Pintar) yang memenuhi kriteria, dengan mengolah data dan memilih atribut yang diperlukan, kemudian menguji data tersebut dengan perhitungan manual dan perangkat lunak Rapidminer. Hasil dari penelitian ini berupa proses perhitungan berdasarkan algoritma C4.5.

1. Pengujian algoritma C4.5

Pembuatan algoritma C4.5 dilakukan pada kumpulan dataset, dengan data yang sudah di jelaskan sebelumnya pengujian data dapat dihitung dengan beberapa langkah-langkah perhitungan *Entropy* dan *Gain* pada setiap atribut yang dimiliki diterima yang sebagai variabel target. Dengan menghitung nilai gain disetiap atribut, lalu dengan membandingkan nilai gain tersebut yang nilainya tertinggi akan digunakan sebagai simpul akar atau atribut akar.

2. Pengujian Data Dengan RapidMiner

Pada proses pengujian data yang sudah di *import*, peneliti menggunakan data *training* dan data *testing* untuk menentukan sebuah model *decision tree* dengan menggunakan perbandingan 60% untuk data *training* dan 40% *testing*.



Gambar 1
Proses pengujian data

3. Hasil Pemodelan Pohon Keputusan



Gambar 2
Hasil Pohon Keputusan

Pohon keputusan C4.5 yang sudah di lihat pada gambar 2 dibaca dari atas ke bawah atau dari akar (simpul pertama paling atas) sampai ke daun (simpul terluar yang tidak lagi memiliki cabang). Berikut cara membacanya dengan mengacu pada tiap simpulnya. Jika Pendapatan Orang Tua rata-rata dan tinggi maka tidak diterima tanpa perlu lagi melihat atribut lainnya. Jika Pendapatan Orang Tua rendah maka terdapat dua kemungkinan, pertama jika Nilai baik maka tidak diterima. Kedua jika Nilai Sangat Baik maka terdapat dua kemungkinan lagi. Pertama, jika Tanggungan Orang Tua Sedikit maka tidak diterima. Kedua jika Tanggungan Orang tua Banyak maka diterima.

4. Hasil Accuracy

accuracy: 90,24%

	true tidak	true Diterima	class precision
pred. tidak	31	3	91.18%
pred. Diterima	1	6	85.71%
class recall	96.88%	66.67%	

Gambar 3
Hasil Accuracy

Hasil pengukuran hasil *accuracy* data yang diperoleh dari data *training* dapat dilihat pada gambar 4 dengan nilai *accuracy* mencapai 90,24%. Dari tabel dapat diketahui prediksi data siswa yang diterima dengan *true* diterima mencapai 6 siswa dan *true* tidak sebanyak 1 data, dengan hasil nilai presisi sebesar 85,71%. Sedangkan untuk prediksi data siswa yang tidak diterima untuk *true* diterima mencapai 3 siswa dan *true* tidak sebanyak 31 siswa, dengan hasil nilai presisi 91,18%. Untuk *class recall* data siswa yang sangat rendah terdapat pada *true* diterima mencapai 66,67%, sedangkan untuk *class recall true* tidak mencapai 96,88%.

KESIMPULAN

Berdasarkan hasil uji coba dan pembahasan yang sudah dilakukan untuk pemilihan penerima Kartu Indonesia Pintar (KIP) di SMPN 38 Jakarta dapat diambil kesimpulan bahwa menggunakan metode klasifikasi dengan algoritma C4.5 bisa selesai. Data siswa dapat diprediksi dengan *accuracy* yang cukup tinggi, hasil tersebut dapat diperoleh dari data *training* dan data *testing* dengan *accuracy* mencapai 90,24% sehingga dengan hasil tersebut algoritma C4.5 sangat baik dalam penentuan pemilihan penerima Kartu Indonesia Pintar (KIP). Dengan hasil *accuracy* sebesar 90,24% dapat dijadikan pendukung untuk digunakan oleh pihak SMPN 38 Jakarta dalam pemilihan penerima Kartu Indonesia Pintar (KIP) dengan algoritma C4.5.

REFERENSI

- Adinungroho, S. (2018). *Implementasi Data Mining Menggunakan Weka*. Malang.
- Azahari, N. (2021). *Rekomendasi Penerimaan Beasiswa Yayasan Untuk Siswa Baru SMK TI Airlangga dengan Algoritma C4 . 5*. 5(April), 609–614.
<https://doi.org/10.30865/mib.v5i2.2943>
- Dr. Suyanto. (2017). *Data Mining Untuk Klasifikasi Dan Klusterisasi Data*. Bandung.
- Pauziah, U. (2016). *Jurnal String Vol . 1 No . 1 Tahun 2016 ISSN : 2527 – 9661 KAJIAN KOMPARASI ALGORITMA C4 . 5 , NAÏVE BAYES DAN NEURAL NETWORK DALAM PEMILIHAN PENERIMA BEASISWA (Studi Kasus pada SMA Muhammadiyah 4 Jakarta) Pendahuluan Tinjauan Pustaka ISSN : 2527 – 9661 Metodologi Penelitian Hasil dan Pembahasan*. 1(1), 47–56.
- Prasetyowati, E. (2017). *DATA MINING Pengelompokan Data untuk Informasi dan Evaluasi*. Retrieved from https://www.google.co.id/books/edition/DAT_A_MINING_Pengelompokan_Data_untuk_Inf/rEH2DwAAQBAJ?hl=id&gbpv=0
- Prof.Dr.Sugiono. (2017). *METODE PENELITIAN KUANTITATIF, KUALITATIF, DAN R&D*. BANDUNG: ALFABETA.
- Sugiyarti, E. (n.d.). *SISTEM PENDUKUNG KEPUTUSAN (DSS) PENYELEKSIAN PEMILIHAN PENERIMA BEASISWA SMA N 1 ULUBELU TANGGAMUS MENGGUNAKAN DATA MINING* Eka Sugiyarti , Andino Maseleno Prodi Manajemen informatika STMIK Pringsewu Lampung. *File:///C:/Users/Muham/Downloads/Referensi Skripsi/29-176-1-PB.Pdf*, 62–69.
- Vulandari, R. T. (2017). *Data Mining Teori Dan Aplikasi Rapidminer*. Yogyakarta.

Yunus, M., Ramdhan, H., & Aji, D. R. (2021).
*PENERAPAN METODE DATA MINING C4.5
UNTUK PEMILIHAN PENERIMA KARTU
INDONESIA PINTAR (KIP) PADA SISWA
KELASS VIII DI SMPN 38 JAKARTA.
JAKARTA.*