

## Penerapan Algoritma C4.5 Dalam Pemilihan Siswa Berprestasi di SMPN 10 Medan

Amir Mahmud Husein<sup>1\*</sup>, Ricki Eko Hasudungan Hutauruk<sup>2</sup>

<sup>1,2</sup> Universitas Prima Indonesia, Medan

Correspondence : [amirmahmud@unprimdn.ac.id](mailto:amirmahmud@unprimdn.ac.id)



### Histori Artikel:

Diajukan: xx xx 2022

Disetujui: xx xx 2022

Dipublikasi: xx xx 2022

### Kata Kunci:

C4.5, Data Mining, Decision Tree, Klasifikasi, Siswa Berprestasi

*Digital Transformation Technology (Digitech) is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).*

### Abstrak

Pada penelitian ini maka dibuat suatu sistem klasifikasi siswa berprestasi menggunakan teknik data mining dengan algoritma C4.5. Analisis dilakukan menggunakan data nilai tahun 2020/2021 untuk membangun model C4.5. Algoritma C4.5 merupakan teknik data mining yang dapat melakukan klasifikasi dan prediksi dengan mengolah beberapa variable. Algoritma C4.5 dengan metode pohon keputusan dapat memberikan informasi rule prediksi untuk menggambarkan proses yang terkait dengan prediksi pemilihan siswa berprestasi. Dalam melakukan penelitian ini dilakukan beberapa metode pengumpulan data. Setelah melakukan pengumpulan data, tahap selanjutnya adalah melakukan perhitungan untuk mengklasifikasikan hasil probabilitas dari sample yang digunakan. Apabila perhitungan probabilitas telah selesai, maka tahap selanjutnya dilakukan pengujian dengan menggunakan *tools data mining* yaitu *RapidMiner* untuk mengetahui hasil klasifikasi dan memvalidasi hasil perhitungan sebelumnya. Pengujian data ini berhasil mendapatkan akurasi yang cukup. Dengan akurasi yang didapat adalah 56,17%, *precision* 55,28% dan *recall* 31,25%.

## PENDAHULUAN

Pendidikan merupakan salah satu hal yang paling penting bagi pertumbuhan bangsa, dikarenakan melalui pendidikan maka dapat tercipta generasi muda bangsa yang cerdas. SMP Negeri 10 Medan adalah unit pelaksana teknis dinas pendidikan di kota Medan yang memiliki tujuan untuk meningkatkan kecerdasan, pengetahuan, akhlak mulia, serta berupaya untuk mencetak lulusan yang berprestasi agar para lulusan dapat melanjutkan ke lembaga pendidikan yang terbaik. Namun, setiap siswa membutuhkan bimbingan yang berbeda agar mendapatkan nilai yang bagus, karena nilai merupakan dataset utama dalam pemilihan siswa berprestasi. Pemilihan siswa berprestasi merupakan proses penilaian yang dilakukan SMP Negeri 10 Medan untuk mengetahui siswa yang cerdas dan layak mendapatkan prestasi. Permasalahannya saat ini yaitu sulitnya mengevaluasi (dalam hal ini memberikan penilaian) karena jumlah siswa yang banyak, kesulitan yang dialami akan memungkinkan terjadinya proses seleksi secara subjektif. Sehingga perlu dilakukan pengolahan data menggunakan metode data mining.

Pada penelitian ini maka dibuat suatu sistem klasifikasi siswa berprestasi menggunakan teknik data mining dengan algoritma C4.5. Analisis dilakukan menggunakan data nilai tahun 2020/2021 untuk membangun model C4.5. Algoritma C4.5 merupakan teknik data mining yang dapat melakukan klasifikasi dan prediksi dengan mengolah beberapa variable. Algoritma C4.5 dengan metode pohon keputusan dapat memberikan informasi rule prediksi untuk menggambarkan proses yang terkait dengan prediksi pemilihan siswa berprestasi.

Proses pemilihan siswa berprestasi dengan algoritma c4.5 tersebut dapat membentuk suatu pola keputusan bagi siswa tersebut dinyatakan "Layak" atau "Tidak Layak" sebagai siswa yang mendapatkan prestasi. Dengan metode ini maka akan didapatkan dua hasil yaitu siswa yang layak mendapatkan prestasi dan siswa yang memiliki potensi hambatan ataupun kurang berprestasi dalam belajarnya, selai itu pihak sekolah juga dapat langsung mengevaluasi metode belajar yang digunakan saat ini, dan lebih meningkatkan kualitas mengajar agar dapat menciptakan siswa yang berprestasi dalam jumlah yang lebih banyak lagi, dan tentunya hal ini dapat bermanfaat bagi tingkat kualitas profit SMP Negeri 10 Medan di mata masyarakat.

## STUDI LITERATUR

Algoritma C4.5 merupakan algoritma yang dikembangkan dari algoritma ID3. C4.5 ini merupakan algoritma turunan dari algoritma ID3 dengan beragam peningkatan. Beberapa peningkatan ini diantaranya adalah, penanganan atribut-atribut numerik, missing value dan noise pada dataset, dan aturan-aturan yang dihasilkan dari model pohon yang terbentuk (Wijaya, Y.A. *dkk*, 2021). Algoritma C4.5 yang merupakan salah satu algoritma modern untuk melakukan Data Mining.

Data mining merupakan proses *iterative* dan *interaktif* untuk menemukan pola atau model baru yang

sempurna, bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (*massive database*) (Sikumbang, E.D, 2018). Menurut (Putria, N.E, 2018) hal penting yang terkait dengan data mining adalah :

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan data mining adalah mendapatkan hubungan atau pola yang akan mungkin memberikan indikasi yang bermanfaat.

### METODE

Dalam melakukan penelitian ini dilakukan beberapa metode pengumpulan data, yaitu :

1. Studi Kepustakaan  
Mengumpulkan dan mempelajari jurnal penelitian nasional dan internasional yang berkaitan dengan objek penelitian.
2. Studi Lapangan
  - a. Observasi  
Penulis mengamati langsung terhadap kegiatan-kegiatan yang dilakukan di SMP Negeri 10 Medan agar mengetahui setiap proses yang dikerjakan pada bagian penilaian siswa berprestasi.
  - b. Wawancara  
Penulis langsung melakukan tanya jawab (*interview*) dengan pihak-pihak SMP Negeri 10 Medan yang berhubungan langsung dengan masalah yang diteliti pada perusahaan.
  - c. Sampling  
Pengumpulan data yang dilakukan dengan cara menerima data melalui whatshap atau email yang berkaitan dengan objek penelitian.

Setelah melakukan pengumpulan data, tahap selanjutnya adalah melakukan perhitungan untuk mengklasifikasikan hasil probabilitas dari sample yang digunakan. Apabila perhitungan probabilitas telah selesai, maka tahap selanjutnya dilakukan pengujian dengan menggunakan *tools data mining* yaitu *RapidMiner* untuk mengetahui hasil klasifikasi dan memvalidasi hasil perhitungan sebelumnya.

### HASIL

Hasil dari penelitian ini adalah sebesar 56.17% ini, dapat dilihat pada gambar 1.

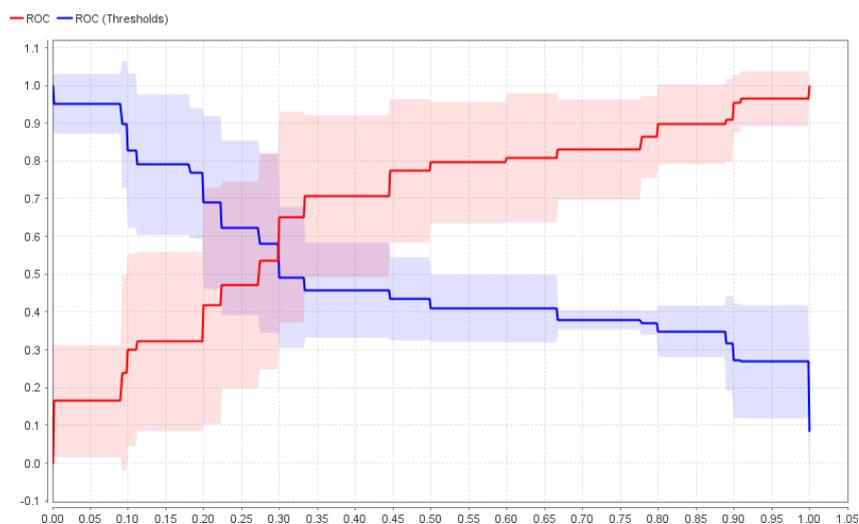
accuracy: 56.17% +/- 8.06% (micro average: 56.22%)

	true Tidak	true Layak	class precision
pred. Tidak	77	59	56.62%
pred. Layak	22	27	55.10%
class recall	77.78%	31.40%	

Gambar 1. Hasil

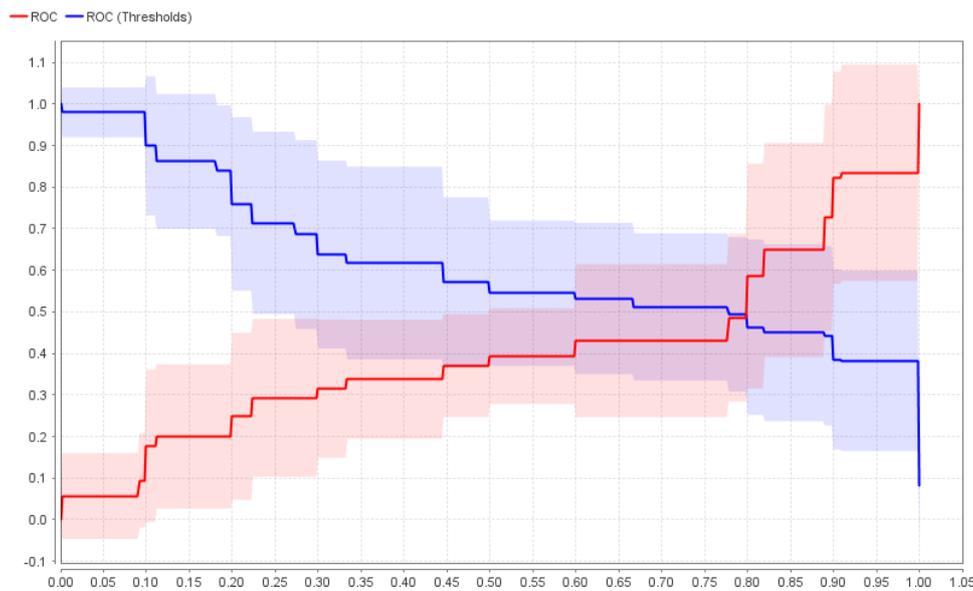
Kemudian hasil AUC (Optimistic) dan AUC (Pessimistic) dapat dilihat pada gambar 2 dan gambar 3.

AUC (optimistic): 0.671 +/- 0.135 (micro average: 0.671) (positive class: Layak)



Gambar 2. Hasil AUC (Optimistic)

AUC (pessimistic): 0.397 +/- 0.140 (micro average: 0.397) (positive class: Layak)



Gambar 3. Hasil AUC Pessimistic

Hasil AUC (*Optimistic*) yang ada pada *Rapid Miner* yaitu 0,671, dengan *positive class* = Layak dan Untuk AUC (*Pessimistic*) pada penelitian ini yang diperoleh yaitu 0,397.

### PEMBAHASAN

Dalam pembahasan ini akan membahas mengenai penerapan algoritma C4.5 pada pemilihan siswa yang layak mendapatkan prestasi. Pada penelitian ini terdapat 345 populasi data yang didapat, dengan jumlah sampel yang digunakan adalah 185 data dan diklasifikasikan menggunakan algoritma C4.5 sehingga membentuk model klasifikasi dan pohon keputusan dengan beberapa atribut yang dipakai yaitu: nilai agama, nilai pkn, nilai b.indonesia, nilai matematika, nilai ipa, nilai ips, nilai b.inggris, nilai seni budaya, nilai penjas, nilai prakarya, nilai rata-rata ijazah. Dari pengujian data ini berhasil mendapatkan akurasi yang cukup., dengan akurasi yang didapat adalah 56,17%, *precision* 55,28% dan *recall* 31.25%. Data ini dilihat dari hasil akurasi yang diuji di *Rapid Miner* yang menunjukkan bahwa siswa yang tidak layak mendapat prestasi lebih banyak dari pada siswa yang layak mendapat prestasi. Hasil ini juga dapat menjadi alternatif dalam pemilihan siswa berprestasi.

### KESIMPULAN

Setelah melalui beberapa tahapan dan pembahasan penelitian ini memiliki kesimpulan yaitu:

1. Pada *Data Mining* dengan metode C4.5 ini dapat diimplementasikan ke dalam pemilihan siswa berprestasi.
2. Pengujian data ini berhasil mendapatkan akurasi yang cukup. Dengan akurasi yang didapat adalah 56,17%, *precision* 55,28% dan *recall* 31.25%.
3. Hasil pengujian data pemilihan siswa berprestasi dapat disimpulkan bahwa lebih banyak siswa yang tidak layak mendapatkan prestasi dibanding siswa yang layak mendapatkan prestasi. Data ini dilihat dari hasil akurasi yang diuji di *Rapid Miner* yang menunjukkan bahwa siswa yang tidak layak mendapat prestasi lebih banyak dari pada siswa yang layak. Hasil ini juga dapat menjadi alternatif dalam pemilihan siswa berprestasi.

### REFERENSI

- Ardiansyah, D. (2019). Algoritma c4. 5 untuk klasifikasi calon peserta lomba cerdas cermat siswa smp dengan menggunakan aplikasi rapid miner. *Jurnal Inkofar*, 1(2).
- Avrizal, R., & Dwiasnati, S. (2019). Implementasi Data Mining dengan Algoritma C4.5 pada Penjualan Obat. *Format : Jurnal Ilmiah Teknik Informatika*, 8(1), 77. <https://doi.org/10.22441/format.2019.v8.i1/010>
- Elisa, E. (2017). Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti. *Jurnal Online Informatika*, 2(1), 36. <https://doi.org/10.15575/join.v2i1.71>
- Hardi. (2014). Pengaruh Minat Membaca Buku Terhadap Hasil Belajar Mahasiswa BKK Akutansi. *Analysis of Micro-Earthquakes in the San Gabriel Mountains Foothills Region and the Greater Pomona Area As Recorded By a Temporary Seismic Deployment*, 1(hal 140), 43. <http://www.springer.com/series/15440%0Apapers://ae99785b-2213-416d-aa7e-3a12880cc9b9/Paper/p18311>

- 
- Irnanda, K. F., Hartama, D., & Windarto, A. P. (2021). Analisa Klasifikasi C4. 5 Terhadap Faktor Penyebab Menurunnya Prestasi Belajar Mahasiswa Pada Masa Pandemi. *Jurnal Media Informatika Budidarma*, 5(1), 327-331.
- Kasih, P. (2019). Pemodelan Data Mining Decision Tree Dengan Classification Error Untuk Seleksi Calon Anggota Tim Paduan Suara. *Innovation in Research of Informatics (INNOVATICS)*, 1(2), 63–69. <https://doi.org/10.37058/innovatics.v1i2.918>
- Laila Qadrini, Andi Seppewali, & Asra Aina. (2021). Decision Treedan Adaboost Pada Klasifikasi Penerima Program Bantuan Sosial. *Jurnal Inovasi Penelitian*, 2(7), 1959–1965.
- Lubis, M. R. (2019). Analisa Prediksi Penjualan Produk Dengan Menggunakan Metode C4.5 (Studi Kasus : PT. Kawan Lama Ace Hardware). *Jurnal Riset Komputer*, 6(5), 545–549. <https://ejournal.stmik-budidarma.ac.id/index.php/jurikom/article/view/1656/1253>
- Putria, N. E. (2018). Data Mining Penjualan Tiket Pesawat Menggunakan Algoritma Apriori Pada Terminal Tiket Batam Tour & Travel. *Computer Based Information System Journal (CBIS)*, 6(1), 29–39. <http://ejournal.upbatam.ac.id/index.php/cbis/article/download/643/410>
- Ramadhani, S. (2019). Hubungan Antara Dukungan Sosial Dengan Motivasi Belajar Siswa Di Sma Negeri 1 Pantai Labu.
- Sikumbang, E. D. (2018). Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori. *Jurnal Teknik Komputer AMIK BSI (JTK)*, Vol 4, No.(September), 1–4.
- Suntoro, J. (2019). DATA MINING Algoritma dan Implementasi Menggunakan Bahasa Pemrograman PHP. *DATA MINING Algoritma Dan Implementasi Menggunakan Bahasa Pemrograman PHP*, 9(9), 259–278.
- Thahir, A., & Hindriyanti, B. (2014). Pengaruh Bimbingan Belajar terhadap Prestasi Belajar Siswa Pondok Pesantren Madrasah Aliyah Al-Utrujiyah Kota Karang. *Konseli : Jurnal Bimbingan Dan Konseling Fakultas Tarbiyah Dan Keguruan, IAIN Raden Intan Lampung*, Vol 1,(2), 55–66. <http://ejournal.radenintan.ac.id/index.php/konseli/article/view/671>
- Thi Bi Dan, T., Widya Sihwi, S., & Anggrainingsih, R. (2016). Implementasi Iterative Dichotomiser 3 Pada Data Kelulusan Mahasiswa S1 Di Universitas Sebelas Maret. *Jurnal Teknologi & Informasi ITSmart*, 4(2), 84. <https://doi.org/10.20961/its.v4i2.1770>
- Wijaya, Y. A., Bahtiar, A., Kaslani, & R, N. (2021). Analisa Klasifikasi menggunakan Algoritma Decision Tree pada Data Log Firewall. *Jurnal Sistem Informasi Dan Manajemen*, 9(3), 256–264. <https://ejournal.stmikgici.ac.id/>