

PENERAPAN ALGORITMA C4.5 DALAM MEMPREDIKSI CALON SISWA BARU

Erpan Gustiarto

Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer
Universitas Indraprasta PGRI
Jl. Raya Tengah No. 80, Kel. Gedong Kec. Pasar Rebo, Jakarta Timur
erpangustiarto@gmail.com

Abstrak

Permasalahan yang dihadapi adalah pihak sekolah kesulitan dalam memprediksi calon siswa baru karena masih dilakukan secara manual, kemudian belum ada nya suatu sistem aplikasi dalam memprediksi calon siswa baru membuat kinerja bagian akademik terhambat setiap tahunnya. Tujuan dirancangnya suatu sistem aplikasi prediksi calon siswa baru agar memudahkan pimpinan dalam menerima siswa baru pada setiap tahunnya dari hasil kinerja dan proses perhitungannya terkomputerisasi dengan menggunakan algoritma C4.5. Metode yang digunakan dalam penelitian ini adalah C4.5 dalam penyelesaian permasalahan yang sudah diteliti. Hasil yang didapat dalam penelitian ini yaitu menghasilkan laporan data hasil calon siswa baru pada Sekolah Internasional Tutor Time Indonesia yang akurat sehingga dapat membantu dalam membuat strategi di masa yang akan datang. Sistem aplikasi yang dirancang sudah layak digunakan untuk proses penilaian prediksi calon siswa baru pada Sekolah Internasional Tutor Time karena sudah sesuai dengan kebutuhan, sehingga dapat mempermudah pimpinan dalam proses *inputting* data serta pembuatan laporan.

Kata Kunci: Sistem, Data Mining, C4.5

Abstract

*The problem faced is that the school has difficulty predicting prospective new students because it is still done manually, then there is no application system for predicting prospective new students which hampers the performance of the academic section every year. The purpose of designing an application system for predicting new prospective students is to make it easier for leaders to accept new students every year from the performance results and computerized calculation processes using the C4.5 algorithm. The method used in this study is C4.5 in solving the problems that have been studied. The results obtained in this study are to produce an accurate report on the results of prospective new students at the Tutor Time Indonesia International School so that it can assist in making strategies in the future. The designed application system is suitable for use in the process of assessing the predictions of prospective new students at the Tutor Time International School because it is in accordance with the needs, so that it can facilitate leaders in the process of *inputting* data and preparing reports.*

Keywords: System, Data Mining, C4.5

PENDAHULUAN

Seiring dengan semakin canggihnya teknologi yang berkembang, komputer telah mendorong terjadinya perubahan ilmu [1], baik dalam kajian ataupun implementasi dilapangan. Peran teknologi komputer sangat diperlukan oleh berbagai instansi pemerintah ataupun pendidikan [2]. Mengingat kebutuhan akan peningkatan efisiensi dan efektifitas dari setiap kegiatan dalam instansi tidak dapat diukur dan dilakukan secara cepat dan akurat tanpa adanya dukungan teknologi tersebut. Akurasi data, kecepatan waktu dan *relevansi* menjadi penentu kualitas informasi yang dihasilkan [3].

Penerimaan Peserta Didik Baru (PPDB) yang ada di Sekolah Internasional Tutor Time Indonesia dengan cara dilakukan secara mandiri yang dimana calon peserta didik baru bisa melakukan pendaftaran dengan datang langsung ke sekolah ataupun bisa melalui media email [4]. Dalam kegiatannya proses penerimaan PPDB masih dilakukan secara manual dimana calon siswa datang langsung ke sekolah dan pihak sekolah akan memberikan formulir pendaftaran dan kebijakan kesepakatan antara calon siswa dengan pihak sekolah. Setelah formulir tersebut di isi maka calon siswa dapat dianggap sebagai siswa dari Sekolah Internasional Tutor Time Indonesia. Sehingga bagian

akademik Sekolah Internasional Tutor Time Indonesia kesulitan dalam memberikan informasi data calon siswa baru pada setiap tahun ajaran baru. Kemudian pembuatan laporan calon siswa baru di Sekolah Internasional Tutor Time masih sering mengalami keterlambatan dikarenakan proses nya masih manual dengan di catat melalui Ms. Excel. Dari permasalahan tersebut, Sekolah Internasional Tutor Time kesulitan dalam memprediksi calon siswa baru karena masih dilakukan secara manual. Faktor utama belum ada nya suatu sistem aplikasi dalam memprediksi calon siswa baru membuat kinerja bagian akademik terhambat setiap tahunnya. Oleh karena itu, diperlukan sebuah metode yang dapat mengatasi masalah tersebut agar hasil yang diperoleh menjadi lebih objektif, akurat dan dapat mempersingkat waktu.

Suatu hal yang penting seharusnya dilakukan sekolah pada saat melakukan PPDB, yaitu dengan mengidentifikasi pola dari data PPDB yang sudah dilaksanakan, dengan melakukan klasifikasi dari variabel yang ditentukan [5]. Informasi yang dapat diterima sekolah dari penentuan tersebut dapat dijadikan dasar penentuan kebijakan sistem PPDB yang akan datang untuk kriteria penerimaan calon siswanya. Sistem PPDB dilakukan dengan teknik klasifikasi yaitu dengan menerapkan algoritma C4.5 [6]. Algoritma C4.5 adalah algoritma yang menghasilkan pohon keputusan, memiliki masukan berupa sampel klasifikasi. Algoritma C4.5 yaitu metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan .

Algoritma C4.5 digunakan pada penelitian ini untuk memprediksi penerimaan peserta didik baru pada Sekolah Internasional Tutor Time Indonesia. Algoritma ini menganalisa set data sampel dan membangun sebuah classifier yang harus mempunyai sebuah kapasitas untuk mengatur kasus latihan dan kasus percobaan dengan akurat.

Dengan permasalahan tersebut, perlu ada nya suatu sistem yang terkomputerisasi dalam penyelesaiannya. Sistem adalah sekelompok unsur yang erat hubungannya satu dengan yang lain, yang berfungsi bersama-sama untuk mencapai tujuan [7]. Suatu sistem yang dikembangkan merupakan sebagai instruksi yang dapat dikelola oleh programmer dalam menjalankan perintah script tersebut. Pembangunan sistem yang harus diketahui ini sekumpulan aktivitas yang biasa digambarkan bagaimana sistem ini dapat berjalan dengan semestinya sesuai dengan instruksi. Dengan ini tujuannya untuk menghasilkan suatu *software* yang dibutuhkan oleh calon *user* kita dalam penyelesaian masalahnya [8].

METODE PENELITIAN

Algoritma C4.5 merupakan jenis algoritma klasifikasi dan prediksi dengan menggunakan teknik decision tree dengan kelebihan-kelebihan yang dimilikinya yang terkenal dan sangat disukai di banding teknik klasifikasi dan prediksi yang lain, seperti naive bayes, random forest dan key nearest neighbour (KNN). Kelebihan – kelebihan yang dimiliki algoritma C4.5 dengan teknik decision tree ini, misalnya; nilai atribut yang hilang dapat ditangani, dapat mengolah data diskret dan numerik (kontinyu) pada komputer yang menggunakan memori utama dan mempunyai kemampuan paling cepat di antara algoritma-algoritma lainnya dengan menghasilkan adanya aturan-aturan (rules) yang akurat dan mudah untuk diimplementasikan [9].

Proses klasifikasi membagi dalam dua tahap pengolahan data, yaitu tahap belajar atau pelatihan yang dapat menghasilkan suatu model pelatihan dari data lampau dan tahap klasifikasi dari data testing (data uji) yang dapat menghasilkan suatu model untuk digunakan sebagai prediksi klasifikasi kelas. Pada tahap data training (pelatihan), dari data pelatihan yang ada berupa record – record data, kemudian dibangun metode decision tree C4.5 untuk menentukan decision. Nilai atribut – atribut untuk sebuah kelas berisikan dari kasus yang ada. Setiap atribut atau variabel dapat berupa data diskret atau numerik (kontinyu). Selain itu ketika ada kasus yang tidak mempunyai nilaiatribut atau lebih dari data yang ada dapat ditangani oleh algoritma C4.5 [10].

HASIL DAN PEMBAHASAN

Pemilihan Atribut Data Latih

Misalkan kita memiliki data pelatihan calon siswa baru yang mencakup atribut-atribut (C_i) dengan data latih (C_i) adalah Terima Perintah ($C1$), Umur ($C2$), Nilai Tes ($C3$), dan Calistung ($C4$). Berikut data-data nya terdapat di tabel 4.1

Tabel 1. Data Latih

Alternatif	Data Latih
C1	Terima Perintah
C2	Umur
C3	Nilai Tes
C4	Calistung

Berikut penilaian untuk 5 siswa, yang akan digunakan untuk masing-masing data latih.

Tabel 2. Penilaian Untuk Data Latih

No	Nama	Alternatif	C1	C2	C3	C4	Hasil
1	Andi	A1	Baik	70	70	Baik	Ya
2	Bagus	A2	Baik	70	70	Baik	Ya
3	Carli	A3	Baik	70	70	Baik	Ya
4	Denny	A4	Baik	70	70	Baik	Ya
5	Erika	A5	Kurang	70	70	Kurang	Tidak

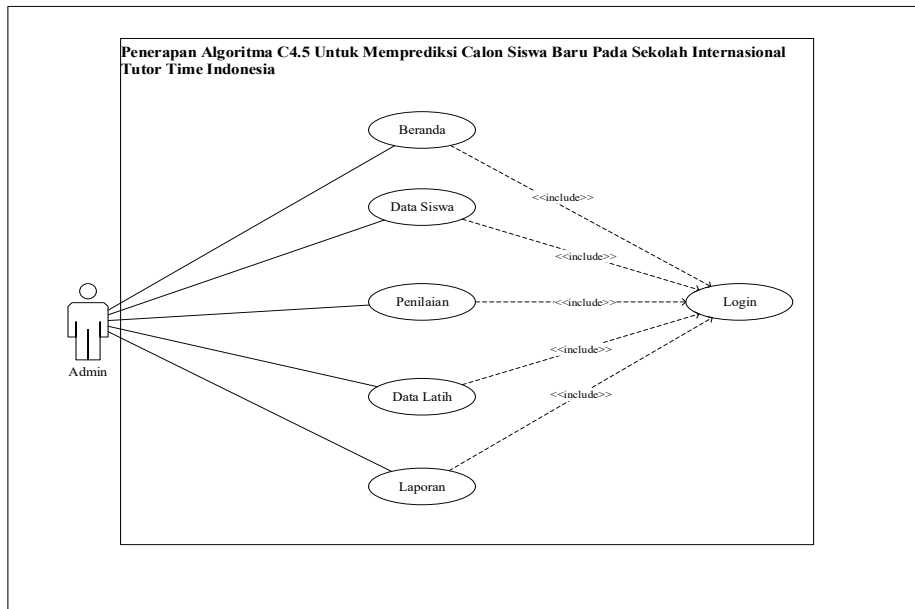
Entropy dan Information Gain

Tabel 3. Perhitungan Entropy dan Gain

Atribut	Alternatif	Jumlah	Ya	Tidak	Entropy	Gain
Total		5	4	1	1	0,72193
Siswa						0,72193
	A1	1	1	0	0	
	A2	1	1	0	0	
	A3	1	1	0	0	
	A4	1	1	0	0	
	A5	1	0	1	0	
C1						0,32193
	Baik	2	2	0	0	
	Cukup	2	1	1	1	
	Kurang	1	1	0	0	
C2						0,32193
	≤ 80	3	3	0	0	
	> 80	2	1	1	1	

C3						0,07291
	<=80	4	3	1	0,81127	
	>80	1	1	0	0	
C4						0,17095
	Cukup	3	2	1	0,91829	
	Kurang	2	2	0	0	

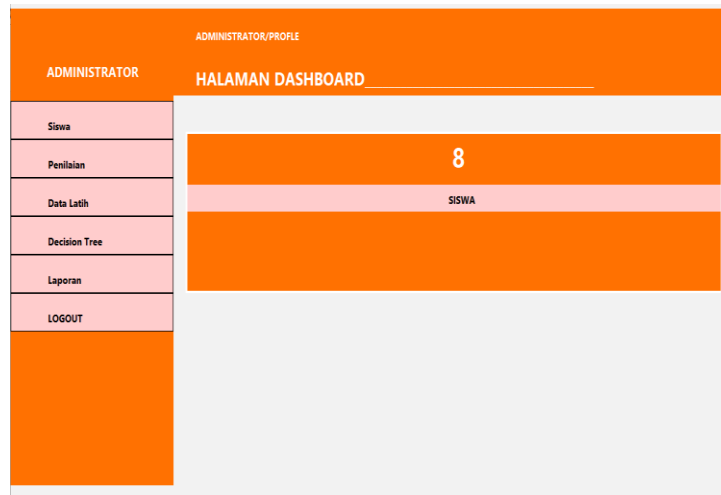
Use Case Diagram



Gambar 1. Use Case Diagram

Tampilan Layar Sistem

1. Tampilan Layar Menu Utama



Gambar 2. Tampilan Layar Menu Utama

Layar di atas menampilkan tampilan Menu Utama pada Sistem untuk memprediksi calon siswa baru. Pada layar utama tersedia *menu side* yang terdiri dari Siswa, Penilaian, Data Latih, *Decision Tree*, Laporan dan *Logout*.

2. Tampilan Layar Data Siswa

NIP	NAMA	ALAMAT	NO TELP
A2	Andi F Noya	Jalan	08xx
A3	Bagus Sianturi	Jalan	08xx
A4	Carli Caplin	Jalan	08xx
A5	Denny Sumargo	Jalan	08xx
A6	Erika Soebandono	Jalan	08xx

Gambar 3. Tampilan Layar Data Siswa

Layar di atas menampilkan tampilan menu data siswa. Pada layar menu siswa akan menampilkan inputan dari data kriteria seperti NIP, Nama, Alamat dan No Telp.

3. Tampilan Data Penilaian

ID	NAMA	C1	C2	C3	C4	Hasil
A1	Andi F ...	Enrolled	1	2	Ads	ya
A2	Bagus ...	Enrolled	1	1	Ads	ya
A3	Carli C...	Enrolled	1	2	Ads	ya
A4	Denny ...	Enrolled	1	1	Ads	ya
A5	Erika S...	Prospect	1	2	Referral	tidak

Gambar 4. Tampilan Data Penilaian

Layar di atas menampilkan tampilan menu penilaian. Pada layar menu penilaian akan menampilkan inputan seperti ID, Nama, Lead Status, Jarak, Follow up, Kualitas, dan Hasil.

4. Tampilan Menu Data Latih

ID	C1	C2	C3	C4	Class
1	Enrolled	1	1	Ads	ya
10	Prospect	2	2	Ads	ya
11	Enrolled	3	3	Referral	ya
12	Waiting	1	1	Referral	ya
13	Waiting	4	4	Ads	ya

Gambar 5. Tampilan Menu Data Latih

Layar di atas menampilkan tampilan menu data latih. Pada layar menu data latih akan menampilkan inputan seperti ID, Lead Status, Jarak, Follow up, Lead Source dan Hasil.

5. Tampilan Laporan Hasil

NO	Nama	C1	C2	C3	C4	NILAI
1	Andi F Noya	Enrolled	1.0	2.0	Ads	ya
2	Bagus Sianturi	Enrolled	1.0	1.0	Ads	ya
3	Carlil Caplin	Enrolled	1.0	2.0	Ads	ya
4	Denny Sumargo	Enrolled	1.0	1.0	Ads	ya
5	Erika Soebandono	Prospect	1.0	2.0	Referral	tidak

MENGETAHUI
JakartaMinggu, 20 Agustus 2023
NURKHOLIS

Gambar 6. Tampilan Laporan Data Hasil

Layar di atas menampilkan Report Hasil Perhitungan. Pada layar menu ditampilkan semua data dari hasil perhitungan berupa, ID, Nama, C1, C2, C3, C4 dan Nilai.

SIMPULAN

Terciptanya sebuah aplikasi yang memiliki kemampuan untuk memprediksi calon siswa baru di Sekolah Internasional Tutor Time Indonesia dengan menggunakan algoritma C4.5 sebagai landasan utama. Fokus utama dari rancangan aplikasi ini adalah untuk menciptakan solusi yang tidak hanya efektif namun juga efisien, sehingga dapat memenuhi kebutuhan spesifik dari Sekolah Internasional Tutor Time Indonesia dalam proses seleksi calon siswa baru. Melalui penerapan algoritma C4.5, diharapkan aplikasi ini mampu melakukan prediksi calon siswa baru dengan akurasi yang tinggi, serta memberikan kontribusi positif dalam meningkatkan efisiensi proses seleksi di sekolah tersebut. Dengan demikian, penelitian ini tidak hanya mengkaji aspek teknis dalam merancang aplikasi, tetapi juga menguji efektivitas algoritma C4.5 sebagai metode prediksi yang dapat mendukung pengambilan keputusan yang lebih baik dalam penerimaan calon siswa baru di Sekolah Internasional Tutor Time Indonesia.

DAFTAR PUSTAKA

- [1] R. Winanjaya, F. Amir, and R. Doni, "Penerapan Data Mining Untuk Memprediksi Penerimaan Peserta Didik Baru Menggunakan Algoritma C4.5," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 1, 2019, doi: 10.30645/senaris.v1i0.1.
- [2] P. P. Haryoto, H. Okprana, and I. S. Saragih, "Algoritma C4.5 Dalam Data Mining Untuk Menentukan Klasifikasi Penerimaan Calon Mahasiswa Baru," *TIN Terap. Inform. Nusant.*, vol. 2, no. 5, pp. 358–364, 2021, [Online]. Available: <https://ejournal.seminar-id.com/index.php/tin/article/view/919>
- [3] D. Made, A. Mertha, and A. A. I. I. Paramitha, "Penerapan Data Mining untuk Prediksi Mahasiswa Berpotensi Non- Aktif Menggunakan Algoritma C4 . 5 : Studi Kasus STMIK Primakara," *J. Ilm. Ilmu Terap.*, vol. 6, no. 1, pp. 84–97, 2022.
- [4] V. Anestiviya, A. Ferico, and O. Pasaribu, "Analisis Pola Menggunakan Metode C4.5 Untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum (Studi Kasus : Sman 1 Natar)," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 80–85, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTISI>
- [5] F. Ferdian Harryanto and S. Hansun, "Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE," *Jatiti*, vol. 3, no. 2, pp. 95–104, 2017.
- [6] M. dan F. D. Marleny, "Prediksi Calon Mahasiswa Baru Menggunakan Metode Klasifikasi Decision Tree," *CSRID J.*, vol. 7, no. 1, pp. 46–54, 2015.
- [7] Jogyanto, *Analisis & Desain Sistem Informasi*. Yogyakarta: Andi Offset, 2017.
- [8] Yanto, *Manajemen Basis Data Menggunakan MySQL*. Yogyakarta: DeePublish, 2020.
- [9] A. Purwanto and H. W. Nugroho, "Analisa Perbandingan Kinerja Algoritma C4.5 Dan Algoritma K-Nearest Neighbors Untuk Klasifikasi Penerima Beasiswa," *J. Teknoinfo*, vol. 17, no. 1, p. 236, 2023, doi: 10.33365/jti.v17i1.2370.
- [10] R. Haqmanullah Pambudi, B. Darma Setiawan, and Indriati, "Penerapan Algoritma C4.5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 7, pp. 2637–2643, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>