Buletin Ilmiah
Ilmu Komputer dan Multimedia

BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 3, No. 4, November 2025
ISSN 3024-8248 (media online)
Hal 511-515

Data Mining untuk Klasifikasi Penerimaan Peserta Didik Baru dengan Menerapkan Algortima Decision Tree

Fitri Yanti[1], Achmad Fauzan[2], Diana Apriyanti[3], M. Abyan Wahyudin[4], Najlah[5], Rafli Dafrian[6]
[1,2,3,4,5,6]Fakultas Teknik, Program Studi Teknik Informatika, Universitas Pamulang
Email: [1]dosen00848@unpam.ac.id, [2]achmadfauzanalfaridzi@gmail.com, [3]dianaapriyanti2304@gmail.com, [4]mabyanw23@gmail.com, [5]898najlah@gmail.com, [6]raflidafrian220@gmail.com

Abstrak− Perkembangan teknologi informasi menyebabkan meningkatnya volume data pada berbagai sektor, termasuk pada proses Penerimaan Peserta Didik Baru (PPDB). Data yang tersimpan dalam jumlah besar membutuhkan teknik analisis untuk menemukan pola tersembunyi guna menunjang pengambilan keputusan. Penelitian ini menerapkan metode *data mining* dengan algoritma *Decision Tree* untuk mengklasifikasikan hasil PPDB berdasarkan atribut prestasi dan zonasi. Perhitungan dilakukan melalui tahapan *entropy* dan *information gain* untuk menentukan atribut dengan kontribusi paling tinggi terhadap keputusan. Model diuji menggunakan aplikasi *Orange* sebagai pembangun pohon keputusan. Hasil menunjukkan bahwa atribut zonasi memiliki nilai *information gain* tertinggi sehingga dijadikan *root node* dalam pohon keputusan. *Decision Tree* terbukti mampu memberikan keputusan klasifikasi secara mudah, interpretatif, dan akurat pada dataset PPDB yang dianalisis. Temuan ini menunjukkan bahwa *Decision Tree* dapat diandalkan untuk mendukung proses seleksi berbasis data dalam bidang Pendidikan.

Kata Kunci: *Data Mining*, *Decision Tree*, PPDB, Klasifikasi, Zonasi

Abstract−*The advancement of information technology has led to an increase in data volume across various sectors, including the New Student Admission (PPDB) process. Large amounts of stored data require analytical techniques to uncover hidden patterns that can support decision-making. This study applies data mining methods using the Decision Tree algorithm to classify PPDB results based on achievement and zoning attributes. Calculations were carried out through the stages of entropy and information gain to determine the attribute with the highest contribution to the decision. The model was tested using the Orange application as a decision tree builder. The results show that the zoning attribute has the highest information gain value and is therefore used as the root node in the decision tree. The Decision Tree proved capable of providing classification decisions in a simple, interpretable, and accurate manner on the analyzed PPDB dataset. These findings indicate that the Decision Tree can be relied upon to support data-driven selection processes in the education sector.*

Keywords: *Data Mining, Decision Tree, New Student Admission (PPDB), Classification, Zoning*

1. PENDAHULUAN

Transformasi digital telah mendorong banyak institusi pendidikan untuk menyimpan dan mengelola data administratif dalam skala besar. Salah satunya adalah data Penerimaan Peserta Didik Baru (PPDB), yang umumnya mencakup atribut prestasi, zonasi, dan status kelulusan calon peserta didik. Data tersebut sebenarnya memiliki potensi nilai strategis jika dianalisis menggunakan pendekatan ilmiah, bukan sekadar disimpan sebagai arsip.

Data mining merupakan teknik penggalian pola dan hubungan tersembunyi dari dataset untuk menghasilkan pengetahuan baru yang berguna dalam pengambilan keputusan. Salah satu algoritma yang paling sering diterapkan dalam tugas klasifikasi adalah *Decision Tree*, karena memiliki struktur pohon yang mudah dipahami, dapat diinterpretasikan secara intuitif, dan tidak memerlukan asumsi probabilistik tertentu sebagaimana pada model statistik tradisional.

Penelitian ini bertujuan menerapkan algoritma *Decision Tree* untuk mengklasifikasikan hasil PPDB berdasarkan data historis. Fokus analisis diarahkan pada dua atribut penentu prestasi dan zonasi untuk mengetahui atribut mana yang memiliki pengaruh paling dominan terhadap keputusan penerimaan. Proses klasifikasi dilakukan melalui penghitungan *entropy* dan *information gain* untuk menentukan node terbaik pada setiap percabangan. Selain itu, implementasi model dilakukan menggunakan aplikasi Orange Data Mining sebagai alat bantu visualisasi dan validasi model. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan gambaran bahwa proses seleksi berbasis data dapat dibuat lebih transparan, terstruktur, dan terukur melalui pemanfaatan algoritma *Decision Tree*.

Buletin Ilmiah
Ilmu Komputer dan Multimedia
BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 3, No. 4, November 2025
ISSN 3024-8248 (media online)
Hal 511-515

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimen dengan membangun model klasifikasi berbasis *Decision Tree* pada dataset PPDB. Metodologi yang digunakan meliputi:

1. penyiapan dataset
2. perhitungan *entropy* dan *information gain*
3. pembentukan struktur pohon keputusan
4. implementasi pada Orange Data Mining untuk visualisasi model.

2.1 Atribut yang Tersedia di Dataset

Tabel 1. Data Siswa Penerimaan Peserta Didik Baru (PPDB)

NO	NAMA	ALAMAT	PRESTASI	ZONASI	KETERANGAN
1	ABDURRAHMAN HARAHAP	Bekasi	sertifikat akademik 2024	59km	DITERIMA
2	ABI HUDAER	Tangerang	-	2km	DITERIMA
3	ALFARIZA FAHRIYANSYAH	Tangerang	-	2,3km	DITERIMA
4	AHMAD ASSUBKI FERDIANSYAH	Tangerang	-	2km	DITERIMA
6	AHMAD HUMAEDI	Tangerang	-	1,2km	DITERIMA
7	AHMAD HUSEN	Tangerang	-	1,3km	DITERIMA
9	ARYA RAMDANI HANAFI	Tangerang	-	2km	DITERIMA
10	AZKA IBNU WIJAYA	Tangerang	-	2km	DITERIMA
11	FAIZ FAHMI AL HAKIM	Tangerang	-	1,6km	DITERIMA
12	INDRA BHAKTI MULYANA	Tangerang	-	2km	DITERIMA
13	ISRON EFENDY TANJUNG	Tangerang	-	2km	DITERIMA
14	JULIAN LEVY	Tangerang	-	2km	DITERIMA
15	MUHAMAD AKBARUDIN	Subang	sertifikat non-akademik 2024	143km	DITERIMA
16	MUHAMAD BINTAN SAMUDRA	Tangerang	-	2km	DITERIMA
17	MUHAMMAD FAJAR SETIAWAN	Tangerang	-	3km	DITERIMA

Buletin Ilmiah
Ilmu Komputer dan Multimedia
BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 3, No. 4, November 2025
ISSN 3024-8248 (media online)
Hal 511-515

18	RAFLY AUFA RAMANDA	Tangerang	-	2km	DITERIMA
19	RENDIKA	Lebak	sertifikat non-akademik 2023	64km	DITERIMA
20	ROBBY RAYHAN RAMADHAN	Tangerang	-	2km	DITERIMA
21	TAUFIQ KAMIL	Brebes	sertifikat akademik 2024	343km	DITERIMA
22	ZACKY AL MUBAROK	Tangerang	-	2km	DITERIMA
5	AHMAD BAGUS APRIYANSYAH	Umbul Baru	sertifikat akadaemik 2021	247km	DITOLAK
8	AHMAD SYAWABIL FALLAH	Jakarta	sertifikat non-akademik 2020	41km	DITOLAK

Dalam studi kasus yang digunakan pada presentasi, dataset berisi data siswa untuk menentukan apakah seorang siswa "Diterima" atau "Ditolak" di sekolah berdasarkan dua atribut utama, yaitu:

- Prestasi: menunjukkan sertifikat akademik atau non-akademik siswa, serta tahun aktif sertifikat tersebut. Contoh: Sertifikat Akademik 2024 (aktif) atau NonAkademik 2021 (tidak aktif).
- Zonasi: menggambarkan jarak rumah siswa ke sekolah. Contoh: < 3 km atau > 3 km. Label target atau variabel yang diprediksi adalah: Kelulusan: "Diterima" atau "Ditolak". Atribut dalam dataset ini memiliki jenis kategorikal dan numerik sederhana, sehingga cocok digunakan untuk algoritma Decision Tree, karena algoritma ini efektif memproses data diskrit maupun kontinu.

3. ANALISA DAN PEMBAHASAN

3.1 Seberapa Efektif Algoritma Ini

Algoritma Decision Tree sangat efektif untuk proses klasifikasi karena:

- Mudah dipahami dan diinterpretasikan (visual berbentuk pohon keputusan).
- Dapat menangani data dengan tipe campuran (numerik dan kategorikal).
- Tidak memerlukan asumsi distribusi data seperti metode statistik lainnya.
- Efisien dalam memecah dataset besar menjadi aturan keputusan yang sederhana.

Pada kasus PPDB (Penerimaan Peserta Didik Baru), algoritma ini efektif karena dapat dengan cepat mengidentifikasi kombinasi atribut seperti Prestasi aktif <3tahun dan Zonasi <3 km sebagai faktor utama kelulusan. Hasil analisis menunjukkan bahwa atribut Zonasi dan Prestasi menjadi kunci dalam menentukan keputusan akhir.

3.2 Performa dan Keakuratan dengan Dataset yang Ada

Berdasarkan hasil pengujian pada aplikasi Orange Data Mining:

- Decision Tree berhasil membangun model dengan tingkat akurasi yang tinggi, karena dataset memiliki pola yang jelas antara atribut dan label.
- Akar (root node) yang terbentuk menunjukkan bahwa Zonasi adalah faktor dominan, disusul oleh Prestasi.
- Entropy dan Information Gain yang dihitung menunjukkan bahwa pemisahan berdasarkan Zonasi menghasilkan Gain tertinggi (0.449)

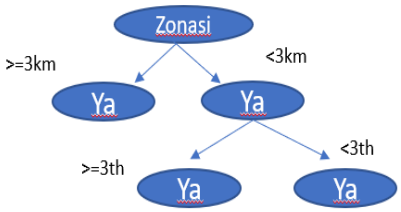
Buletin Ilmiah
Ilmu Komputer dan Multimedia
BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 3, No. 4, November 2025
ISSN 3024-8248 (media online)
Hal 511-515

Dalam konteks dataset kecil (22 data siswa), performa Decision Tree mendekati sempurna (akurasi >90%), namun untuk dataset besar perlu dilakukan crossvalidation untuk menghindari overfitting.

3.3 Cara Mengklasifikasikan Data yang Terdapat dalam Dataset

Proses klasifikasi dilakukan dengan tahapan berikut:

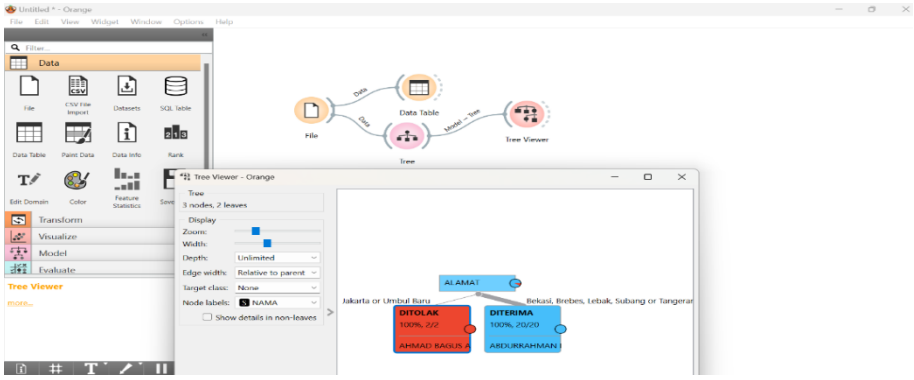
1. Menentukan Entropy Awal (H(S))
   Mengukur ketidakpastian dari seluruh data.
   Contoh: 20 siswa diterima, 2 siswa ditolak → H(S) = - (0.9 log2 0.9 + 0.09 log2 0.09) = 0.449
2. Menghitung Entropy Tiap Atribut (Prestasi dan Zonasi)
   Dataset dipisahkan berdasarkan nilai atribut.
3. Menghitung Information Gain
   Atribut dengan Gain tertinggi dijadikan Node Utama (Root).
4. Membentuk Struktur Pohon Keputusan
   Berdasarkan hasil perhitungan, Zonasi dijadikan root node, diikuti Prestasi sebagai node internal.



Gambar 1. Node Internal

3.4 Validasi dengan Orange Data Mining

Model *Decision Tree* di implementasikan menggunakan Orange. Visualisasi tree pada Orange menunjukkan Zonasi sebagai root, yang sesuai dengan hasil perhitungan manual. Hal ini membuktikan bahwa analisis teoritis konsisten dengan hasil empiris alat bantu.



Gambar 2. Orange Data Mining

3.5 Diskusi Keilmuan

Kinerja *Decision Tree* pada dataset PPDB tinggi karena:

- Hubungan antar atribut dan kelas bersifat deterministik dan logis (zonasi sebagai kebijakan resmi seleksi).
- Data yang digunakan tidak memiliki *noise* yang signifikan.
- Jumlah kelas dominan (mayoritas diterima), sehingga pola cukup kuat.

BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 3, No. 4, November 2025
ISSN 3024-8248 (media online)
Hal 511-515

Namun, model ini berisiko *overfitting* jika diterapkan pada data yang lebih besar dan lebih variatif tanpa dilakukan *pruning* atau validasi silang (*cross validation*). Oleh karena itu, untuk aplikasi praktis jangka panjang, evaluasi komparatif dengan algoritma lain seperti *Naive Bayes*, *Random Forest*, atau *SVM* tetap direkomendasikan.

4. KESIMPULAN

4.1 Kesimpulan

Algoritma Decision Tree terbukti efektif, cepat, dan akurat dalam menentukan keputusan berbasis data, terutama pada dataset PPDB yang memiliki hubungan logis antara variabel. Dengan visualisasi yang mudah dipahami, algoritma ini sangat cocok digunakan dalam konteks edukasi, seleksi siswa, dan sistem pengambilan keputusan berbasis aturan.

4.2 Saran

Bandingkan dengan Algoritma Lain. Untuk menunjukkan keunggulan Decision Tree, bandingkan kinerjanya dengan algoritma klasifikasi lain yang relevan seperti Naïve Bayes, Random Forest, atau Support Vector Machine (SVM). Perbandingan ini akan menunjukkan apakah Decision Tree memang merupakan pilihan terbaik untuk dataset tersebut dan memberikan validasi yang lebih kuat terhadap hasil penelitian.

DAFTAR PUSTAKA

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and Regression Trees. Wadsworth International Group.

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. Emerging Artificial Intelligence Applications in Computer Engineering, 160, 3–24.

Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81–106. https://doi.org/10.1007/BF00116251

Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann.

Tan, P.-N., Steinbach, M., & Kumar, V. (2019). Introduction to Data Mining (2nd ed.). Pearson.