

Pemodelan Data Menggunakan Algoritma Decision Tree untuk Visualisasi Data Hubungan Nilai Siswa dengan Makan Siang, Partisipasi Kursus Persiapan Tes, dan Tingkat Pendidikan

Rangga Gelar Guntara¹, Muhammad Rizki Nugraha², Riska Aprilia³, Nur Azizyah Putri Dewinta⁴, Fiky Budi Dewanto⁵, Salsa Nurahma⁶

^{1,2,3,4,5,6} Universitas Pendidikan Indonesia, Jl. Dr. Setiabudhi No. 229, Bandung 40154, Indonesia
e-mail: ranggagelar@upi.edu¹, murinu@upi.edu²

INFORMASI ARTIKEL

A B S T R A K

Sejarah Artikel:

Diterima Redaksi : 21 Agustus 2023

Revisi Akhir : 01 November 2023

Diterbitkan Online : 30 November 2023

Kata Kunci:

Pendidikan, Sosio ekonomi, Visualisasi Data, Decision Tree

Korespondensi :

Telepon / Hp : +62

8112196116

E-mail: ranggagelar@upi.edu

Pendidikan memegang peranan vital dalam perkembangan baik individu maupun komunitas. Prestasi belajar siswa dianggap sebagai tolak ukur kunci untuk menilai keberhasilan dari suatu sistem pendidikan. Terdapat beragam faktor yang berpengaruh terhadap prestasi belajar siswa, termasuk kondisi sosio ekonomi mereka serta keikutsertaan dalam berbagai program pendukung, misalnya kelas persiapan ujian. Studi ini dirancang untuk mengeksplorasi dampak dari faktor-faktor tersebut terhadap prestasi akademik siswa. Metodologi yang digunakan melibatkan penggalian data dari sumber-sumber beragam, analisis statistik, dan evaluasi data. Sedangkan untuk pemodelan menggunakan algoritma decision tree. Informasi yang diperoleh kemudian diolah menjadi presentasi visual seperti grafik, tabel, dan diagram untuk memudahkan dalam pemahaman dan penyampaian temuan. Skor rata-rata untuk siswa yang mendapatkan makan siang standar adalah 212.51, sementara untuk mereka yang mendapatkan makan siang gratis atau dengan subsidi adalah 186.60. Skor rata-rata untuk siswa yang mengikuti kelas persiapan ujian berada di angka 218.01, dibandingkan dengan mereka yang tidak mengikutinya, yang skornya adalah 195.12. Terdapat juga korelasi antara tingkat pendidikan orang tua dengan skor siswa, dimana anak-anak dari orang tua dengan pendidikan lebih tinggi cenderung mendapatkan skor yang lebih baik dibandingkan dengan mereka yang orang tuanya memiliki latar belakang pendidikan yang lebih rendah, meskipun perbedaannya tidak terlalu signifikan.

1. PENDAHULUAN

Pendidikan adalah salah satu aspek kunci dalam pembangunan individu dan masyarakat. Kinerja akademik siswa menjadi indikator penting dalam mengukur efektivitas sistem Pendidikan [1]. Banyak faktor yang dapat mempengaruhi kinerja akademik siswa, termasuk lingkungan sosio ekonomi dan partisipasi dalam program pendukung, seperti kursus persiapan tes [2]. Penting untuk memahami peran faktor sosioekonomi dalam kinerja akademik siswa. Salah satu indikator sosio ekonomi yang dapat digunakan adalah tipe makan siang yang mereka terima. Makan siang standar dan makan siang gratis atau subsidi adalah dua tipe yang mencerminkan perbedaan dalam akses terhadap sumber daya pendidikan. Makan siang standar mungkin lebih umum diakses oleh siswa dengan latar belakang sosio ekonomi yang lebih stabil, sementara makan siang gratis atau subsidi diberikan kepada siswa yang memenuhi kriteria ekonomi tertentu. Oleh karena itu, perbandingan skor siswa berdasarkan tipe makan siang dapat memberikan wawasan tentang dampak lingkungan sosio ekonomi pada kinerja akademik.

Selain itu, kursus persiapan tes juga merupakan faktor yang dapat mempengaruhi kinerja akademik siswa. Kursus ini dirancang khusus untuk membantu siswa mempersiapkan diri untuk ujian dan meningkatkan pemahaman mereka tentang materi ujian [3]. Kursus persiapan tes dapat memberikan manfaat

tambahan dalam persiapan siswa untuk ujian, yang mungkin mencakup pengembangan keterampilan ujian, pemahaman yang lebih mendalam tentang materi pelajaran, dan peningkatan rasa percaya diri [4]. Oleh karena itu, perbandingan skor siswa yang mengikuti kursus persiapan tes dengan mereka yang tidak mengikuti dapat memberikan gambaran tentang dampak positif kursus ini pada kinerja akademik.

Untuk mengetahui detail mengenai hal tersebut, maka dilakukan proses pengolahan data hingga visualisasi data untuk memudahkan penjelasan. Proses pengolahan data mencakup pengumpulan data dari berbagai sumber, pemrosesan statistik, dan analisis data [5]. Data ini kemudian disajikan dalam bentuk visualisasi yang jelas, seperti grafik, tabel, dan diagram, untuk memfasilitasi pemahaman dan penyajian hasil dengan lebih mudah [6].

Melalui analisis perbandingan skor siswa berdasarkan tipe makan siang dan partisipasi dalam kursus persiapan tes, dapat diidentifikasi adanya perbedaan yang signifikan dalam kinerja akademik. Hasil dari analisis ini dapat membantu sekolah dan sistem pendidikan dalam mengembangkan strategi pendukung yang lebih efektif untuk siswa. Kesimpulan penelitian ini akan membahas implikasi dari perbedaan ini dan menyoroti pentingnya mempertimbangkan faktor sosio ekonomi serta partisipasi dalam program pendukung dalam upaya untuk meningkatkan kinerja akademik siswa. Dengan memahami faktor-faktor yang

mempengaruhi kinerja siswa, kita dapat merancang pendekatan pendidikan yang lebih inklusif dan berkelanjutan untuk mencapai kesetaraan dalam pendidikan.

2. METODE

Pada penelitian ini akan menggunakan metode pengembangan AI Life Cycle. Metode AI Life Cycle menggambarkan proses sistematis pengembangan dan penerapan solusi kecerdasan buatan (AI) dari konsepsi awal hingga penerapan dan pemeliharaan berkelanjutan [7]. Siklus hidup AI (Artificial Intelligence) mengacu pada proses komprehensif yang melibatkan berbagai tahapan dalam pengembangan dan penerapan solusi berbasis kecerdasan buatan [8]. Siklus hidup ini memastikan bahwa sistem AI dikembangkan, dites, dikerahkan, dan dikelola dengan efektif sepanjang masa pakainya. Pada gambar 1 disajikan tahapan pada AI Life Cycle.



Gambar 1. AI Life Cycle [9]

a. Perumusan Masalah

Tahap awal ini melibatkan pengidentifikasian masalah yang spesifik dan formulasi bagaimana AI dapat menjadi solusi. Ini termasuk pemahaman tentang apakah masalah tersebut memang sesuai untuk dipecahkan dengan AI dan penentuan tujuan yang jelas.

b. Pengumpulan dan Persiapan Data

Setelah masalah dirumuskan, langkah selanjutnya adalah mengumpulkan data yang relevan yang akan digunakan untuk melatih model AI. Data ini kemudian dibersihkan, dinormalisasi, dan diubah menjadi format yang sesuai untuk pemrosesan lebih lanjut. Tahap ini sangat kritis karena kualitas dan kuantitas data yang baik sangat penting untuk pengembangan model AI yang efektif.

c. Desain dan Pengembangan Model

Pada tahap ini, para pengembang dan data scientist memilih algoritma yang sesuai dan teknologi yang akan digunakan untuk membangun model AI. Mereka kemudian mulai mengkode dan melatih model dengan dataset yang telah disiapkan, menyesuaikan parameter untuk meningkatkan performa.

d. Pelatihan Model

Model AI dilatih menggunakan dataset yang telah disiapkan. Proses ini melibatkan penyesuaian

bobot dalam algoritma berdasarkan input dan output yang diharapkan. Proses pelatihan berlangsung hingga model mencapai tingkat akurasi yang diinginkan atau sampai tidak ada peningkatan yang signifikan lagi.

e. Evaluasi Model

Setelah pelatihan, model dievaluasi untuk menentukan seberapa baik ia melakukan tugas yang ditentukan. Ini biasanya melibatkan pengujian model terhadap set data yang tidak digunakan selama pelatihan (dikenal sebagai data pengujian) untuk mengukur performanya secara objektif.

f. Penyebaran

Jika model memenuhi kriteria evaluasi, selanjutnya ia dapat dikerahkan dalam lingkungan produksi. Ini berarti mengintegrasikan model ke dalam aplikasi atau sistem yang lebih besar untuk digunakan oleh pengguna akhir atau proses bisnis.

g. Pemantauan dan Pemeliharaan

Setelah dikerahkan, model harus terus dipantau untuk memastikan bahwa ia tetap efektif seiring berubahnya kondisi atau data. Model mungkin perlu diperbarui atau di-retuning secara berkala untuk mempertahankan atau meningkatkan kinerjanya.

h. Umpan Balik dan Iterasi

Siklus hidup AI bersifat iteratif. Berdasarkan umpan balik dari pengguna akhir dan pemantauan kinerja model, langkah-langkah dalam siklus hidup dapat diulang untuk meningkatkan, memperbarui, atau mengadaptasi model sesuai dengan kebutuhan yang berubah atau pemahaman yang lebih dalam tentang masalah yang dihadapi.

3. HASIL DAN PEMBAHASAN

3.1. Proses Akuisisi Data

Data yang kami dapatkan bersumber dari Web Scraping Kaggle yang merupakan website dengan berbagai kumpulan data publik yang berkualitas tinggi. Dari website tersebut, kami mengambil data yang terdiri dari data nilai yang diperoleh siswa dalam berbagai mata pelajaran. Nilai siswa didapatkan dari sekolah menengah yang ada di Amerika Serikat. Dengan memahami pengaruh latar belakang orang tua, persiapan ujian, serta makan siang terhadap kinerja siswa. Ada 1.000 data siswa dengan berbagai kategori seperti pendidikan orang tua, kelompok, gender, tipe makan siang, pelaksanaan ujian serta skor-skor yang diperoleh dari ujian tersebut yaitu skor matematika, skor membaca, dan skor menulis.

3.2. Data Preprocessing

Tahap ini merupakan proses pengolahan data supaya data dapat digunakan dan terhindar dari segala kerusakan seperti data kosong, duplikat, salah format, dan lain-lain [10].

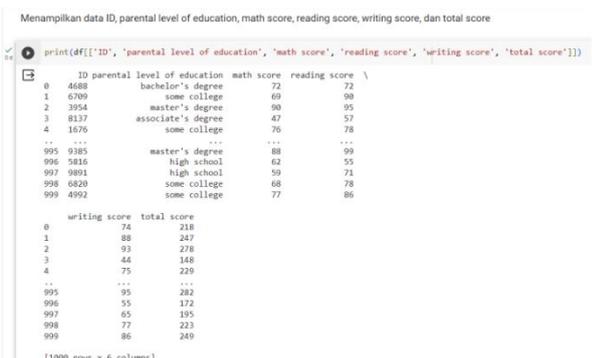
3.3. Pemodelan

Setelah melakukan data pre-processing dan menghasilkan data yang bersih serta siap digunakan, maka langkah selanjutnya adalah pemodelan. Pemodelan data dilakukan dengan menggunakan Google Collab dan diimplementasikan dengan bahasa pemrograman Python untuk menampilkan dan mengeksplorasi data yang dibutuhkan dalam menganalisis [11]. Selain itu, pemodelan dilakukan untuk menjawab beberapa pertanyaan berikut.

- 1) Tampilkan data yang menunjukkan kategori ID, lunch, dan total score!
- 2) Berapa rata-rata score siswa yang memiliki makan siang standar atau standard lunch?
- 3) Berapa rata-rata score siswa yang memiliki makan siang gratis/bersubsidi atau free/reduced lunch?
- 4) Tampilkan data yang menunjukkan kategori ID, test preparation course, dan total score!
- 5) Berapa rata-rata score siswa yang berpartisipasi dalam test preparation course?
- 6) Berapa rata-rata score siswa yang tidak berpartisipasi dalam test preparation course?
- 7) Tampilkan data yang menunjukkan kategori ID, parental level of education, dan total score!
- 8) Berapa rata-rata score siswa yang memiliki orang tua dengan tingkat pendidikan high school?
- 9) Berapa rata-rata score siswa yang memiliki orang tua dengan tingkat pendidikan some high school?
- 10) Berapa rata-rata score siswa yang memiliki orang tua dengan tingkat pendidikan bachelor's degree?
- 11) Berapa rata-rata score siswa yang memiliki orang tua dengan tingkat pendidikan associate's degree?
- 12) Berapa rata-rata score siswa yang memiliki orang tua dengan tingkat pendidikan some college?



Gambar 2. Menampilkan Seluruh Data serta Dilengkapi dengan Total Score



Gambar 1. Menampilkan Data ID, Parental Level of Education, Math Score, Reading Score, Writing Score, dan Total Score

3.4. Implementasi Algoritma Decision Tree

Berikut adalah penjelasan singkat tentang setiap cabang dalam decision tree:

- a. Skor Matematika adalah node akar yang menentukan cabang berikutnya yang akan diikuti.
- b. Jenis Kelamin adalah faktor yang digunakan untuk menentukan kelompok siswa jika skorma tematikamereka antara 0 dan 20.
- c. Level Edukasi Orang Tua adalah faktor yang digunakan untuk menentukan kelompok siswa jika skorma tematika mereka antara 0 dan 20 dan mereka perempuan.
- d. Test Preparation Course adalah faktor yang digunakan untuk menentukan kelompok siswa jika skorma tematika mereka antara 21 dan 40.
- e. Makan Siang adalah faktor yang digunakan untuk menentukan kelompok siswa jika skorma tematika mereka antara 41 dan 60.
- f. Level Edukasi Orang Tua adalah faktor yang digunakan untuk menentukan kelompok siswa jika skorma tematika mereka antara 61 dan 80.
- g. Skor Membaca adalah faktor yang digunakan untuk menentukan kelompok siswa jika skorma tematika mereka antara 81 dan 100 dan mereka perempuan.
- h. Skor Menulis adalah faktor yang digunakan untuk menentukan kelompok siswa jika skorma tematika mereka antara 81 dan 100 dan mereka perempuan.
- i. Decision tree ini dapat digunakan untuk memprediksi kelompok siswa berdasarkan skorma tematika, jenis kelamin, tingkat pendidikan orang tua, kursus persiapan tes, jenis makan siang, skor membaca, dan skor menulis mereka.

3.5. Pengujian dan Evaluasi

Matrix confusion adalah sistem yang digunakan dalam analisis statistik dan pembelajaran untuk mengevaluasi kinerja model klasifikasi, terutama dalam konteks masalah klasifikasi biner (dua kelas)[12].

Tabel 1. Siswa yang memiliki standard lunch diprediksi memiliki total score tinggi

Confusion Matrix		Prediksi	
		Yes	No
Aktual	Yes	349 <i>Standard lunch, total score di antara 210-300 (True Positive)</i>	245 <i>Free/reduced lunch, total score di antara 0-209 (False Positive)</i>
	No	296 <i>Standard lunch, total score 0-209(False Negative)</i>	110 <i>Free/reduced lunch, total score 210-300 (True Negative)</i>

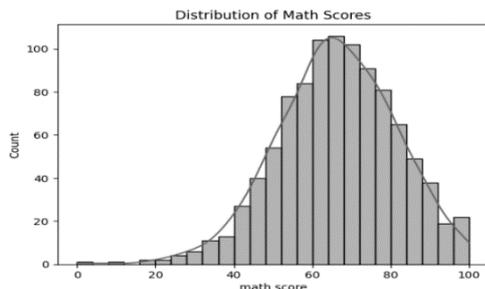
Tabel 2. Siswa dengan Latar Belakang Pendidikan Orangtua Tinggi Diprediksi Memiliki Total Score Tinggi

Confusion Matrix		Prediksi	
		Yes	No
Aktual	Yes	320 <i>Education Parents associate's</i>	236 <i>Education Parents high</i>

	<i>degree, bachelors degree, some college, total score di antara 210-300 (True Positive)</i>	<i>school, some high school, total score di antara 0-209 (False Positive)</i>
No	305	139
	<i>Education Parents associate's degree, bachelors degree, some college, total score 0-209(False Negative)</i>	<i>Education Parents high school, some high school, total score 210-300 (True Negative)</i>

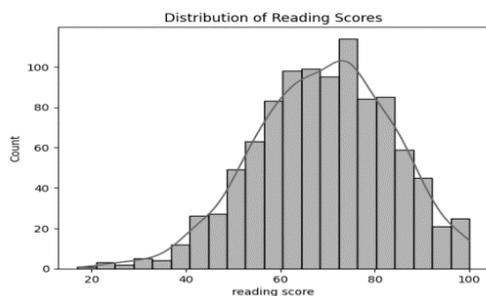
3.6. Deployment

Visualisasi data adalah proses representasi informasi atau data menggunakan elemen grafis seperti grafik, diagram, dan tabel untuk membantu orang memahami pola, tren, dan hubungan dalam data [13]. Tujuan utamanya adalah untuk membuat data lebih mudah dipahami, memungkinkan penemuan wawasan, dan membantu pengambilan keputusan yang lebih baik.



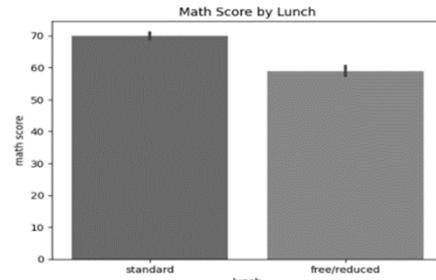
Gambar 4. Diagram PerolehanSkor Berhitung

Dilihat dari diagram gambar 4, rata-rata terbanyak skor matematika yang diperoleh adalah 63-65.

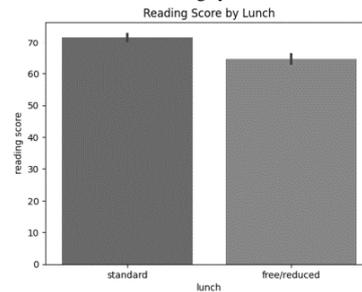


Gambar 5. Diagram perolehan skor membaca

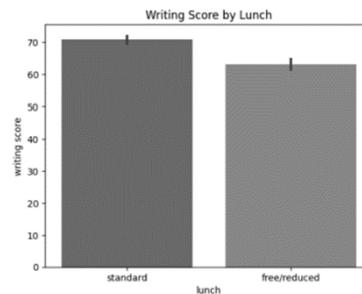
Dilihat dari diagram gambar 5, rata-rata terbanyak skor membaca yang diperoleh adalah 70.



Gambar 6. Diagram perolehan skor berhitung berdasarkan tipe makan siangnya



Gambar 7. Diagram perolehan skor membaca berdasarkan tipe makan siangnya



Gambar 8. Diagram perolehan skor menulis berdasarkan tipe makan siangnya

4. KESIMPULAN

Rata-rata skor siswa dengan tipe makan siang standar (standard lunch type) adalah 212.51, sedangkan rata-rata skor siswa dengan tipe makan siang gratis atau subsidi (free/reduced lunch type) adalah 186.60. Ini menunjukkan adanya perbedaan skor yang signifikan antara kedua kelompok ini. Siswa dengan akses makan siang standar cenderung memiliki skor yang lebih tinggi secara keseluruhan dibandingkan dengan siswa yang mendapatkan makan siang dengan harga yang lebih terjangkau atau gratis. Hal ini mungkin mencerminkan perbedaan dalam lingkungan sosio ekonomi dan akses terhadap sumber daya pendidikan yang mempengaruhi kinerja akademik.

Rata-rata skor siswa yang mengikuti kursus persiapan tes adalah 218.01, sedangkan rata-rata skor siswa yang tidak mengikuti kursus tersebut adalah 195.12. Ini menunjukkan bahwa siswa yang mengikuti kursus persiapan tes memiliki skor yang lebih tinggi secara keseluruhan dibandingkan dengan siswa yang tidak mengikuti kursus. Kursus persiapan tes dapat memberikan manfaat tambahan dalam persiapan siswa

untuk ujian dan membantu meningkatkan kinerja mereka dalam tes.

Hubungan nilai dengan tingkat pendidikan orang tua, rata-rata skor peserta yang didasarkan dengan latar belakang pendidikan orang tuanya menunjukkan bahwa semakin tinggi pendidikan orang tuanya maka serupa pula dengan anaknya, yaitu memiliki hasil yang lebih tinggi dibandingkan peserta dengan latar belakang pendidikan orang tuanya yang lebih rendah, namun dengan hasil perbandingan yang tidak terlalu jauh.

DAFTAR PUSTAKA

- [1] Adnan, A., Zohriah, A., & Muin, A. (2024). Evaluasi Kinerja Tenaga Pendidik. *JIP-Jurnal Ilmiah Ilmu Pendidikan*, 7(2), 1463-1468.
- [2] Zulkifli, E. (2022). Pengaruh Optimalisasi Pembelajaran Online, Partisipasi Mahasiswa Dan Gaya Mengajar Dosen Terhadap Motivasi Belajar Pada Pembelajaran Daring Ditengah Pandemi Covid-19 (Studi Pada Kampus STIE Indonesia Jakarta).
- [3] Manurung, S. M., Purba, I. P., Sitinjak, L. T. S., Aryani, N., Herman, H., Sinaga, Y. K., ... & Batubara, J. (2023). Sosialisasi Pengenalan TOEFL Sebagai Persiapan Studi Lanjut Pada Siswa-Siswi SMAN 4 Pematang Siantar. *IKHLAS: Jurnal Pengabdian Dosen*
- [4] Gs, A. D., Sormin, R. K., Siregar, Z. A., Winanjaya, R., & Wanto, A. (2023). Pelatihan Pembuatan Pin Press Digital Bagi Siswa untuk Meningkatkan Keterampilan dan Menumbuhkan Semangat Wirausaha. *JPM: Jurnal Pengabdian Masyarakat*, 4(2), 142-152.
- [5] Akbar, M., & Rahmanto, Y. (2020). Desain data warehouse penjualan menggunakan Nine Step Methodology untuk business intelegency pada PT Bangun Mitra Makmur. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 1(2), 137-146.
- [6] Kusuma, M. D. H., & Hidayat, S. (2024). Penerapan Model Regresi Linier dalam Prediksi Harga Mobil Bekas di India dan Visualisasi dengan Menggunakan Power BI. *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, 5(2), 1097-1110.
- [7] Rifai, A. I. (2021). Manajemen Perkerasan: Implementasi Data Mining & Artificial Intelligence.
- [8] Guntara, R. G. (2023). Pelatihan Sains Data Bagi Pelaku UMKM di Kota Tasikmalaya Menggunakan Google Colab. *Joong-Ki: Jurnal Pengabdian Masyarakat*, 2(2), 245-251.
- [9] Nguyen, D. C., Cheng, P., Ding, M., Lopez-Perez, D., Pathirana, P. N., Li, J., ... & Poor, H. V. (2020). Enabling AI in future wireless networks: A data life cycle perspective. *IEEE Communications Surveys & Tutorials*, 23(1), 553-595.
- [10] Amanda Hasna, C. (2024). Penilaian Kemampuan Pembayaran Kredit Dengan Menggunakan Machine Learning Logistic Regression Dan Random Forest Classifier Pada Home Credit.
- [11] Guntara, R. G. (2023). Visualisasi Data Laporan Penjualan Toko Online Melalui Pendekatan Data Science Menggunakan Google Colab. *ULIL ALBAB: Jurnal Ilmiah Multidisiplin*, 2(6), 2091-2100.
- [12] Ratnasari, R., Wahidin, A. J., Setiawan, A. E., & Bintoro, P. (2024). mechine Mechine Learning Untuk Klasifikasi Penyakit Jantung. *Aisyah Journal Of Informatics and Electrical Engineering (AJIEE)*, 6(1), 145-150.
- [13] Putri, A. K., & Nur, D. I. (2023). Penggunaan Bahasa Python Untuk Analisis Dan Visualisasi Data Penduduk Di Desa Sumberjo, Nganjuk. *Karya: Jurnal Pengabdian Kepada Masyarakat*, 3(3), 206-217.