



## Implementasi Data Mining Prediksi Kelulusan Siswa Menggunakan Metode Decision Tree Pada SMK IPTEK Tangsel

Suryaningrat<sup>1</sup>, Sartika Lina Mulani Sitio<sup>2</sup>, Ayni Suwarno Herry<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang

<sup>1</sup>[d02362@unpam.ac.id](mailto:d02362@unpam.ac.id), <sup>2</sup>[dosen00847@unpam.ac.id](mailto:dosen00847@unpam.ac.id), <sup>3</sup>[dosen00707@unpam.ac.id](mailto:dosen00707@unpam.ac.id)

### Abstract

Vocational High School (SMK) is a formal educational institution that prepares its graduates for the world of work. However, SMK Iptek Tangsel faces challenges in optimizing student data and overcoming labor shortages in the field of education administration. To overcome this problem, this study applies a prediction system using a data mining technique with the decision tree method. Aims to improve the accuracy of student graduation predictions. The methodology of this study adopts a quantitative approach with structured steps, including observation, interviews, data collection, and documentation. The results showed that the accuracy of Rapid miners in motorcycle and multimedia business engineering reached 98.49%, while in hospitality and accounting accommodation reached 99.05%. This graduation prediction system helps schools identify students at risk of not graduating and who are at potential to graduate, optimize data management, and enable appropriate interventions. This research makes a significant contribution to improving the decision-making process in the field of education through the use of data mining technology. Thus, the graduation prediction system can be an effective tool in supporting data management and increasing student graduation in vocational schools. Some explanations about how the intervention will be carried out and its impact on schools by doing; Identify students who are at risk of not graduating based on historical data and variables that affect graduation, such as academic grades, attendance, and participation in extracurricular activities. In addition, schools can design special guidance programs for students identified as at risk, then conduct periodic monitoring and evaluation of student progress and then also increase parental involvement, and use data for decision-making at the school level.

Keywords: *Data mining, Decision tree, Student Graduation Prediction, Rapid miner.*

### Abstrak

Sekolah Menengah Kejuruan (SMK) merupakan lembaga pendidikan formal yang mempersiapkan lulusannya untuk dunia kerja. Namun, SMK Iptek Tangsel menghadapi tantangan dalam mengoptimalkan data siswa dan mengatasi kekurangan tenaga kerja di bidang administrasi pendidikan. Untuk mengatasi masalah tersebut, penelitian ini menerapkan sistem prediksi menggunakan teknik data mining dengan metode decision tree. Bertujuan untuk meningkatkan akurasi prediksi kelulusan siswa. Metodologi penelitian ini mengadopsi pendekatan kuantitatif dengan langkah-langkah terstruktur, termasuk observasi, wawancara, pengumpulan data, dan dokumentasi. Hasil penelitian menunjukkan bahwa akurasi Rapid miner dalam teknik bisnis sepeda motor dan multimedia mencapai 98.49%, sementara dalam akomodasi perhotelan dan akuntansi mencapai 99.05%. Sistem prediksi kelulusan ini membantu sekolah mengidentifikasi siswa yang berisiko tidak lulus dan yang berpotensi lulus, mengoptimalkan pengelolaan data, serta memungkinkan intervensi yang tepat. Penelitian ini memberikan kontribusi signifikan dalam meningkatkan proses pengambilan keputusan di bidang pendidikan melalui pemanfaatan teknologi data mining. Dengan demikian, sistem prediksi kelulusan dapat menjadi alat yang efektif dalam mendukung pengelolaan data dan peningkatan kelulusan siswa di SMK. Beberapa penjelasan mengenai bagaimana intervensi akan dilakukan dan dampaknya pada sekolah dengan melakukan ; Identifikasi siswa yang berisiko tidak lulus berdasarkan data historis dan variabel yang mempengaruhi kelulusan, seperti nilai akademik, kehadiran, dan partisipasi dalam kegiatan ekstrakurikuler. selain itu Sekolah dapat merancang program bimbingan khusus untuk siswa yang teridentifikasi berisiko, kemudian melakukan monitoring dan evaluasi berkala terhadap kemajuan siswa kemudian juga melakukan peningkatan keterlibatan orang tua, dan penggunaan data untuk Pengambilan Keputusan di tingkat sekolah.



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

Kata kunci: Data mining, *Decision tree*, Prediksi Kelulusan Siswa, *Rapid miner*.

## 1. Pendahuluan

Data mining merupakan inovasi atau temuan data baru dengan mencari pola maupun aturan tertentu dari beberapa informasi dalam jumlah besar yang diharapkan bisa menanggulangi keadaan tersebut. Data mining sendiri mempunyai sebagian metode salah satunya ialah klasifikasi. Metode klasifikasi dengan membangun pohon keputusan berdasarkan aturan-aturan yang diperoleh dari data yang dikumpulkan (Sianturi, 2018). Dalam decision tree, setiap node pada pohon mewakili sebuah atribut, setiap cabang mewakili nilai atribut dan setiap daun digunakan untuk mewakili kelas (Rizmayanti et al., 2021). Algoritma C4.5 pohon keputusan merupakan model klasifikasi sangat terkenal sebab bisa dengan gampang ditafsirkan oleh manusia. Banyaknya algoritma bisa digunakan buat membangun pohon keputusan semacam ID3, C4.5, CART serta GUIDE. Algoritma C4.5 ialah salah satu algoritma klasifikasi informasi dengan metode pohon keputusan yang bisa mencerna informasi numerik serta diskrit, bisa menanggulangi nilai 1 2 atribut yang lenyap, menciptakan aturan- aturan yang gampang untuk diinterpretasikan.

Sekolah Menengah Kejuruan (SMK) merupakan Lembaga Pendidikan formal yang lulusannya dipersiapkan untuk dunia kerja, dari sekian rangkaian program pendidikan yang dijalani oleh siswa/i perlu adanya sebuah sistem prediksi kelulusan dengan tujuan membantu sekolah dalam mengetahui informasi kelulusan sejak dini. Ada 2 aspek permasalahan yang dihadapi, yaitu ;

Pertama, Data siswa tidak dioptimalkan dengan baik. Informasi yang tidak akurat dan terkini tentang siswa menyulitkan sekolah dalam memahami pola dan tren kelulusan siswa secara menyeluruh. Dalam proses pengambilan keputusan, terutama terkait dengan strategi pendidikan dan bimbingan konseling, kurangnya akurasi data menjadi hambatan. Siswa yang membutuhkan bantuan tambahan untuk meningkatkan peluang kelulusan sulit diidentifikasi dengan tepat dan tepat waktu.

Kedua, Kurangnya tenaga kerja di bidang administrasi pendidikan juga menjadi permasalahan serius. Proses manual dalam mengumpulkan, mengelola, dan menganalisis data memakan waktu yang cukup lama dan meningkatkan peluang 3 kesalahan dalam pengolahan data. Kurangnya efisiensi dalam pengelolaan data ini menghambat respons cepat terhadap kebutuhan siswa. Selain itu, dalam proses prediksi kelulusan, data yang tidak dioptimalkan membuat analisis pola perilaku siswa dan faktor-faktor yang berkontribusi pada tingkat

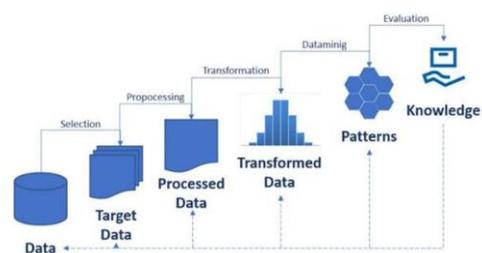
kelulusan yang rendah menjadi sulit. Tanpa sistem prediksi yang akurat dan efisien, sekolah kesulitan mengidentifikasi masalah secara dini dan menerapkan intervensi yang tepat.

Berikut ini adalah penjelasan mengenai bagaimana intervensi akan dilakukan dan dampaknya terhadap sekolah dengan sebagai: mengidentifikasi siswa yang berisiko tidak lulus berdasarkan data historis dan variabel yang mempengaruhi kelulusan, seperti nilai akademik, kehadiran, dan partisipasi dalam kegiatan ekstrakurikuler. juga meningkatkan keterlibatan orang tua, kemudian menggunakan Data untuk Pengambilan Keputusan di tingkat sekolah

Maka dari itu, perlunya adanya penerapan sistem prediksi dalam mengelola data untuk membantu sekolah dalam memprediksi kelulusan siswa. Sehingga, dengan adanya hal tersebut penulis memutuskan untuk mengambil judul penelitian “Implementasi Data Mining Prediksi Kelulusan Siswa Menggunakan Metode Decision Tree pada SMK Iptek Tangsel”

## 2. Metode Penelitian

Tahapan proses dalam Data Mining Prediksi Kelulusan Siswa Menggunakan Metode Decision Tree adalah sebagai berikut.



Gambar 2. Tahapan Data mining pada KDD

a. Data selection, Mempunyai tujuan utama membuat kumpulan data target dari data asli, yaitu memilih subset variable atau sampel data.

b. Pre-processing / cleaning, sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi focus KDD. Proses cleaning antara lain membuang duplikasi data, memeriksa data yang konsisten dan memperbaiki kesalahan pada data.

c. Transformasi, memiliki tugas untuk mengurangi dan memproyeksikan data untuk mendapatkan representasi yang sesuai untuk tugas tertentu yang akan dilakukan, biasanya dicapai dengan melibatkan Teknik

transformasi atau metode yang mampu menemukan representasi data yang tidak berubah.

d. Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan Teknik, metode atau algoritma tertentu.

e. Interpretation/evaluation ,pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak bekerpentingan. Dimanfaatkan oleh pengguna untuk menafsirkan dan mengekstrak pengetahuan dari pola yang dihasilkan, dengan memvisualisasikan pola.

Adapun langkah-langkah proses penelitian penulisan perlakuan tingkat kelulusan mahasiswa disajikan pada tabel berikut:



Gambar 3. Tahapan Proses Data Penelitian

- Data sampel yang sama dengan populasi akan diolah di Microsoft excel.
- Pengolahan data dengan Microsoft excel untuk mendapatkan data hasil pemeriksaan.
- Pemrosesan data menggunakan aplikasi Rapid miner untuk mengimpor data pengujian untuk pengujian.
- Pada tahap ini, data uji akan diproses dengan metode pohon keputusan, kemudian hasil data akan menghasilkan validasi data dengan hasil yang sama persis dengan data uji di aplikasi RapidMiner.
- Data telah diolah oleh aplikasi Rapid miner dan Decision tree akan menghasilkan hasil data yang akurat dari pengolahan data siswa dengan hasil yang akurat.

Beberapa tantangan spesifik yang dihadapi adalah volume data yang besar dan beragam, misalnya memiliki lebih dari 500 mahasiswa dengan berbagai latar belakang akademik dan sosial. Selain itu, ada keterbatasan sumber daya manusia, misalnya sekolah hanya memiliki beberapa tenaga pengajar dan staf administrasi yang bertanggung jawab untuk mengelola data siswa, kemudian kualitas datanya tidak konsisten, ada perbedaan catatan, selain kurangnya pemahaman teknik analisis data, misalnya banyak tenaga pengajar yang tidak memiliki latar belakang analisis data atau pemrograman sehingga sulit digunakan alat analisis yang lebih canggih seperti RapidMiner atau perangkat lunak penambangan data lainnya.

#### 4. Perhitungan Entropy dan Gain

Pada saat pembuatan pohon keputusan, kita perlu menghitung jumlah kelas yang lulus atau tidak lulus, entropi setiap kelas berdasarkan atribut yang telah ditentukan oleh data nilai.

Dari data tersebut diketahui jumlah kasus jurusan teknik bisnis sepeda motor dan multimedia sebanyak 331 record, jumlah siswa yang lulus sebanyak 318 record dan jumlah tidak lulus sebanyak 13 record, sedangkan kasus pada jurusan akomodasi perhotelan dan akuntansi sebanyak 207 record, jumlah siswa yang lulus sebanyak 205 record dan siswa yang tidak lulus sebanyak 2 record pada tahun ajaran 2021 s/d 2023, sehingga nilai entropy nya adalah:

$$1. \text{ Perhitungan entropy (s) untuk TBSM dan MM} \\ E(S) = ((-318/331)*\text{IMLOG}_2(318/331)+ \\ (-13/331)*\text{IMLOG}_2(13/331))\text{Entropy}(S) = 0,238957811$$

$$2. \text{ Perhitungan entropy (s) untuk APH dan AK} \\ E(S) = ((-205/207)*\text{IMLOG}_2(205/207)+(2/207)*\text{IMLOG}_2(2/207) \\ ) \\ \text{Entropy}(S) = 0,078542898$$

Untuk mendapatkan gain setiap atribut, entropy atribut harus dihitung tergantung pada kasusnya. Berikut hasil dari perhitungan gain:

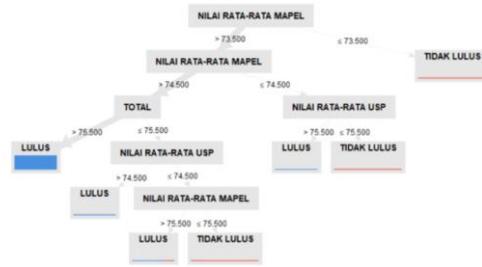
$$3. \text{ Perhitungan gain untuk TBSM dan MM} \\ (0,238957811)-(105/331)*0,388611749+ \\ ((226/331)*0,153205042)=0,220287436 \\ 4. \text{ Perhitungan gain untuk APH dan AK} \\ =(0,078542898)-((69/207)*0,109285715)+ \\ ((138/207)*0,061927389) =0,083399252$$

#### 3. Hasil dan Pembahasan

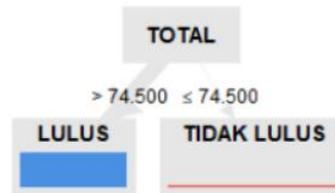
Berdasarkan hasil ini, tindakan konkret yang dapat diambil di tingkat sekolah meliputi:

- Siswa yang teridentifikasi berisiko tidak lulus akan dimasukkan dalam program intervensi.

- b. Menyediakan kelas tambahan untuk mata pelajaran yang menjadi kelemahan siswa.
- c. Menghubungkan siswa dengan mentor yang dapat memberikan dukungan moral dan akademik.
- d. Menginformasikan orang tua tentang status akademik anak dan langkah-langkah yang diambil oleh sekolah.
- e. Setelah intervensi dilakukan, penting untuk melakukan monitoring dan evaluasi berkala terhadap kemajuan siswa. Ini bisa dilakukan melalui:
- f. Mengadakan ujian untuk mengukur peningkatan pemahaman siswa.
- g. Mengumpulkan umpan balik dari guru mengenai perkembangan siswa.
- h. Monitoring yang konsisten memungkinkan sekolah untuk menyesuaikan intervensi jika diperlukan dan memastikan bahwa siswa mendapatkan dukungan yang tepat

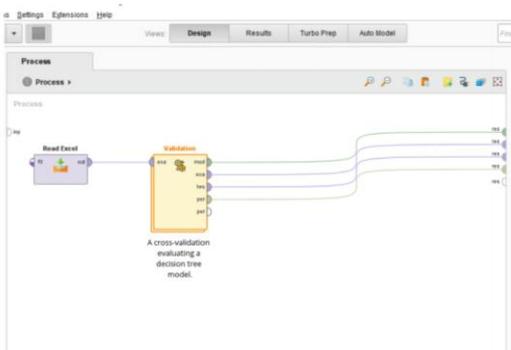


Gambar 4 3 Tree Jurusan Teknik Bisnis Sepeda Motor dan Multimedia

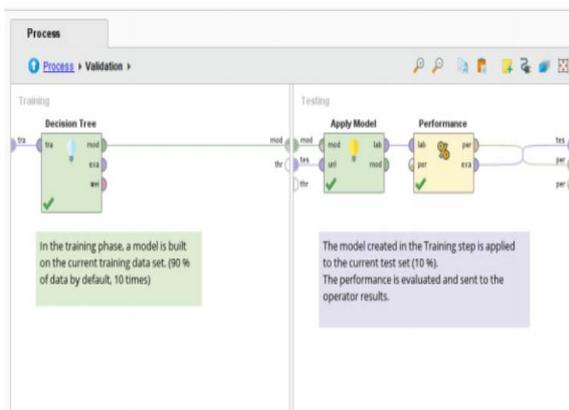


Gambar 4 4 Tree Jurusan Akomodasi Perhotelan dan Akuntansi

Untuk melakukan validasi, melibatkan analisis model yang berbeda dan memilih model dengan kinerja yang baik setelah membaca file data ke dalam blok read excel, kemudian read excel digabungkan dengan blok yang divalidasi. Berikut uji validasi menggunakan rapidminer.



Gambar 4. Proses Awal Rapid Miner



Gambar 4 Proses Validasi Rapid miner

#### 4. Kesimpulan

Berdasarkan dari analisa data dan pembahasan maka peneliti dapat mengambil kesimpulan yaitu sebagai berikut :

- a. Sistem prediksi kelulusan dapat mengoptimalkan terjadinya penumpukan data sehingga dapat membantu sekolah untuk mengidentifikasi siswa yang berisiko tidak lulus dan yang berpotensi lulus.
- b. Sistem prediksi kelulusan dengan menggunakan metode decision tree yang dihasilkan dapat digunakan untuk melakukan prediksi kelulusan, bahwa akurasi Rapid miner dalam teknik bisnis sepeda motor dan multimedia mencapai 98.49%, sementara dalam akomodasi perhotelan dan akuntansi, akurasi Rapid miner mencapai 99.05%.
- c. Implementasi data mining untuk memprediksi kelulusan siswa di SMK Iptek Tangsel bukan hanya sekadar penggunaan teknologi, tetapi juga merupakan langkah strategis untuk mengatasi berbagai kesulitan dalam pengelolaan data dan pengambilan keputusan. Dengan memahami tantangan yang ada dan memilih metode yang tepat, sekolah dapat lebih siap untuk menerapkan solusi yang efektif dan berkelanjutan dalam meningkatkan kualitas pendidikan. Rekomendasi untuk pengembangan lebih lanjut akan membantu memastikan bahwa sistem ini tetap relevan dan efektif dalam mendukung keberhasilan siswa di masa depan

## Daftar Rujukan

- [1] Adinugroho, S., & Yuita, A. S. (2018). Implementasi Data mining Menggunakan WEKA (1st ed.). UB Press.
- [2] Ali Ma, F., Pratama, A., Sholihin, I., & Rizki Rinaldi, A. (2021). Penerapan Model Prediksi Menggunakan Algoritma C.45 Untuk Prediksi Kelulusan Siswa SMK Wahidin. *Jurnal Data Science & Informatika*, 1(1), 16–20.
- [3] Arhami, M., & Nasir, M. (2020). Data mining Algoritma dan Implementasi. ANDI.
- [4] Arthasia, M. (2022). IMPLEMENTASI DATA MINING UNTUK PREDIKSI PENJUALAN SOUVENIR MENGGUNAKAN ALGORITMA C4.5. Skripsi.
- [5] Asroni, A., Fitri, H., & Prasetyo, E. (2018). Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik). *Semesta Teknik*, 21(1), 60–64. <https://doi.org/10.18196/st.211211>
- [6] Bulo, E. (2020). Data mining Untuk Perguruan Tinggi (Pertama). CV BUDI UTAMA.
- [7] Fibo, M., Ikhbal, D., & Kurniadi, D. (2021). Menentukan Penjurusan Siswa Dengan Menggunakan Metode Decision tree Algoritma C4.5 (Studi Kasus : SMA Negeri 2 Padang). *JAVIT (Jurnal Vokasi Informatika)*, 1(3). <http://javit.ppi.unp.ac.id>
- [8] Irawan, M. D., & Simargolang, S. A. (2018). Implementasi E-Arsip Pada Program Studi Teknik Informatika. *Jurnal Teknologi Informasi*, 2(1), 67. <https://doi.org/10.36294/jurti.v2i1.411>
- [9] Joko, S. (2019). Data mining : Algoritma dan Implementasi dengan Program an PHP (1st ed.). PT Elex Media Komputindo.
- [10] Kuriawan, A. (2019). Implementasi Data mining Algoritma C4.5 Untuk Memprediksi Kelulusan Uji Kompetensi Smk Teknik Komputer Dan Jaringan (Tkj) (Study Kasus: Smk Pembangunan Daerah Lubuk Pakam). *Jurnal Majalah Ilmiah Informasi Dan Teknologi Ilmiah (INTI)*, 7(1), 5–13. <http://ejournal.stmik-udidarma.ac.id/index.php/inti/article/view/1801>
- [11] Mahmud, R., & Hartanto, A. (2020). Penerapan Data mining Rekomendasi Laptop Menggunakan Algoritma Apriori. *Juisi*, 06(02), 21–30.
- [12] Merawati, D., & Rino. (2019). Penerapan data mining penentu minat Dan bakat siswa Smk dengan metode C4 . 5. *JURNAL ALGOR*, 1(1).
- [13] Muslim, M. A., Prasetyo, B., M, E. L. H., H, A. J., irqotussa`adah, R, S. H., & Nurzahputra, A. (2019). DATA MINING ALGORITMA C4.5 (E. Listiana & N. Cahyani (eds.); 1st ed.).
- [14] Mustika, Ardilla, Y., Manuhutu, A., Ahmad, N., Hasbi, I., Gunto, Manuhutu, A. M., Ridwan, M., & Hozairi. (2021). Data mining dan Aplikasinya (N. Risnawati (ed.); 1st ed.). Widina Bhakti Persada Bandung.
- [15] Normah, Rifai, B., Vambudi, S., & Maulana, R. (2022). Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE. *Jurnal Teknik Komputer AMIK BSI*, 8(2), 174–180. <https://doi.org/10.31294/jtk.v4i2>
- [16] Putri, R. E. (2021). Implementasi data mining untuk prediksi efektivitas pada mesin injection menggunakan algoritma c4. 5 studi kasus: pt. tridaya artaguna santara. Skripsi, 5. <https://repository.uinjkt.ac.id/dspace/handle/123456789/57846>  
[https://repository.uinjkt.ac.id/dspace/bitstream/123456789/57846/1/RIZKY\\_EVITA\\_PUTRI-FST.pdf](https://repository.uinjkt.ac.id/dspace/bitstream/123456789/57846/1/RIZKY_EVITA_PUTRI-FST.pdf)
- [18] Rizmayanti, A. I., Hidayati, N., Nugraha, F. S., & Gata, W. (2021). Penerapan Data mining Untuk Memprediksi Kompetensi Siswa Menggunakan Metode Decission Tree ( Studi Kasus Smk Multicomp Depok ). *Swabumi*, 9(1), 9–18. <https://doi.org/10.31294/swabumi.v9i1.8363>