

PENERAPAN ALGORITMA DECISION TREE DALAM KLASIFIKASI KETERLAMBATAN JAM MASUK KERJA KARYAWAN DENGAN RAPIDMINER

Niuita Sepha Kanaya*

Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan

*Email: niquitasephakanaya@gmail.com

Suvriadi Panggabean

Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan

Email : suvriadi@unimed.ac.id

Abstrak

Banyak elemen yang berkontribusi, termasuk jarak ke tempat kerja, waktu keberangkatan dan bangun tidur, cuaca, moda transportasi, dan tingkat lalu lintas, menjadi pertimbangan ketika menilai keterlambatan karyawan. Oleh karena itu, untuk mengukur dan meramalkan keterlambatan karyawan, diperlukan pendekatan kategorisasi. Data karyawan dari BPJS Kesehatan Cabang Medan diklasifikasikan dalam penelitian ini dengan menggunakan pendekatan Decision Tree. Data survei, yang meliputi informasi jam bangun, jam berangkat, waktu perjalanan, cuaca, transportasi, dan status keterlambatan, dikumpulkan dari 150 karyawan. Dengan menggunakan metodologi pembagian data pelatihan dan pengujian, program RapidMiner digunakan untuk pemrosesan dan analisis data. Dengan akurasi maksimum 94,00%, temuan akhir dari penelitian ini menunjukkan bahwa algoritma Decision Tree dapat mengkategorikan data secara efektif. Hal ini menunjukkan bahwa, dengan menggunakan data yang diamati, model yang dikembangkan dapat digunakan untuk membantu memprediksi kemungkinan keterlambatan karyawan.

Kata Kunci: Klasifikasi Keterlambatan Karyawan, Metode Decision Tree, RapidMiner

Abstract

There are several factors that affect the classification of employee tardiness, including time spent in the office, time spent on departure, time spent sleeping, weather conditions, transportation, and the level of congestion. Therefore, a classification method is needed to assess and evaluate employee performance. This research uses the Decision Tree method as a classification tool for data collected by BPJS Kesehatan Medan Branch. The dataset used is a survey of 150 respondents who were asked about waking hours, departure hours, weather, transportation, distance to the office, and tardiness status. Using the RapidMiner application, data were developed and analysed using techniques to separate test data from addition data. The final results of the study show that the Decision Tree algorithm can effectively classify data with an accuracy of up to 94.00%. This indicates that the constructed model can be used to help predict potential employee delays based on the observed variables.

Keywords: Classification of employee tardiness, Decision Tree Method, RapidMiner.

PENDAHULUAN

Karyawan merupakan elemen krusial yang berperan dalam pencapaian tujuan perusahaan. Hal ini disebabkan karena karyawan berperan sebagai motor penggerak utama dalam operasional perusahaan (Wahyudi, Rozi, & Putry, 2022). Dengan demikian, keberadaan karyawan memperkuat jalannya fungsi-fungsi organisasi di dalam perusahaan.

Perusahaan perlu menerapkan disiplin kerja agar mampu merancang dan menerapkan aturan beserta sanksi yang berlaku bagi karyawan (Suparman, Jajang, & Wahyudin, 2023). Artinya,

perusahaan harus menyosialisasikan peraturan yang telah ditetapkan hingga karyawan benar-benar memahami dan mampu menjalankannya. Upaya ini bertujuan untuk mencegah terjadinya pelanggaran dalam lingkungan kerja, sehingga masalah kedisiplinan dapat diminimalkan (Yulianti, Solehudin, Suyaman, & Hersona, 2023). Salah satu bentuk permasalahan yang umum terjadi adalah kebiasaan karyawan datang terlambat saat jam masuk kerja.

Keterlambatan kerja terjadi ketika karyawan datang melewati batas waktu yang telah ditetapkan oleh perusahaan. Kondisi ini dapat menghambat

jalannya proses bisnis, karena aktivitas kerja menjadi tertunda (Ika & Sitompul, 2022). Bagi karyawan, keterlambatan memberikan dampak negatif seperti penurunan indikator kinerja, terhambatnya pengembangan karier, hingga pemotongan gaji. Untuk mengatasi permasalahan ini, diperlukan suatu model yang mampu membantu perusahaan meningkatkan ketepatan waktu kerja karyawan dengan mengidentifikasi faktor-faktor yang paling berpengaruh terhadap keterlambatan (Dwi Anjani, Puspaningtyas Faeni, & Wahyu Setyawati, 2024). Dalam penelitian ini, teknik data mining digunakan untuk mengembangkan algoritma klasifikasi pohon keputusan untuk mengidentifikasi tren keterlambatan jam kerja.

Penambangan data adalah proses menganalisis dan mendapatkan informasi dari data secara otomatis menggunakan satu atau lebih algoritme pembelajaran mesin. Proses ini juga dikenal sebagai pembelajaran berbasis induksi, yaitu pembentukan konsep-konsep umum melalui pengamatan terhadap contoh-contoh khusus (Asyuti & Setyawan, 2023). RapidMiner, sebuah program open source yang dapat digunakan oleh semua orang, merupakan salah satu alat yang digunakan dalam pemrosesan data ini. RapidMiner menawarkan berbagai pendekatan deskriptif dan prediktif yang menjadikannya solusi untuk analisis pengolahan data (Setio, Saputro, & Bowo Winarno, 2020). Perangkat lunak ini dijalankan menggunakan bahasa pemrograman Java.

Penelitian ini mengacu pada penelitian sebelumnya yang menggunakan teknik data mining, seperti penelitian yang membandingkan model klasifikasi algoritma untuk keterlambatan sekolah. Dalam studi tersebut, variabel jam berangkat menjadi faktor paling berpengaruh terhadap keterlambatan, dan algoritma K-NN menunjukkan performa terbaik dengan tingkat akurasi sebesar 87,38% (Amirulloh & Taufiqurrochman, 2017).

Penelitian lainnya berjudul Optimasi metode Naïve Bayes dengan feature selection Information Gain untuk prediksi keterlambatan pembayaran sumbangan pembinaan pendidikan sekolah, menunjukkan bahwa, dengan tingkat akurasi 90%, algoritma Naïve Bayes yang dikombinasikan dengan pendekatan seleksi fitur Information Gain dapat memberikan hasil prediksi terbaik untuk

keterlambatan pembayaran uang sekolah (Muqorobin, Kusrini, & Luthfi, 2019).

KAJIAN TEORI (GUNAKAN STYLE SECTION)

Data Mining

Data mining adalah proses untuk menemukan informasi penting dari kumpulan data dalam jumlah besar dengan memanfaatkan teknik-teknik canggih seperti kecerdasan buatan, statistik, matematika, dan machine learning (Putri & Wijayanto, 2022). Tujuan utama dari proses ini adalah mengidentifikasi serta mengekstraksi pola-pola yang relevan dari database (Sudarsono, Leo, Santoso, & Hendrawan, 2021). Saat ini, data mining telah berkembang menjadi bidang ilmu multidisipliner yang diterapkan di berbagai sektor, termasuk biologi, keuangan, dan jaringan komputer (Triandini, Defit, & Nurcahyo, 2021).

Proses data mining terdiri dari beberapa tahapan utama sebagai berikut:

- Business Understanding, Tahap awal ini berfokus pada pemahaman terhadap kebutuhan bisnis dan merumuskan permasalahan secara tepat, agar keseluruhan proses data mining dapat diarahkan sesuai dengan tujuan yang ingin dicapai (Andry, Hartono, Honni, Chakir, & Rafael, 2022).
- Data Understanding, Meliputi pengumpulan data dari berbagai sumber serta eksplorasi awal untuk memahami struktur data, mengidentifikasi data yang hilang, dan menyoroti bagian-bagian penting yang relevan untuk analisis (Gheisari et al., 2022).
- Data Preparation, Tahap ini mencakup pemilihan atribut yang relevan, pembersihan data, dan transformasi agar data siap digunakan dalam tahap pemodelan. Kualitas data yang baik pada tahap ini sangat menentukan keberhasilan hasil akhir.
- Modelling, Pada tahap ini, algoritma seperti Decision Tree diterapkan pada data yang telah dipersiapkan untuk membangun model klasifikasi atau prediksi. Proses ini juga mencakup pengujian beberapa model guna memilih model terbaik (Jassim & Abdulwahid, 2021).
- Evaluation, Melibatkan penilaian performa model dengan menggunakan metrik seperti confusion matrix, precision, recall, dan AUC.

Teknik validasi seperti 10-fold cross validation diterapkan, dan visualisasi performa model dapat dilakukan menggunakan ROC curve (Dobesova, 2024).

Model Decission Tree

Decision Tree merupakan salah satu algoritma machine learning yang bekerja dengan menerapkan seperangkat aturan dalam struktur bercabang menyerupai pohon. Struktur ini menggambarkan berbagai kemungkinan hasil, alokasi sumber daya, tingkat utilitas, serta potensi konsekuensi atau risiko (Azam, Islam, & Huda, 2023). Algoritma ini beroperasi dengan menyusun pernyataan-pernyataan bersyarat yang membentuk cabang-cabang, di mana setiap cabang menggambarkan langkah-langkah pengambilan keputusan yang dapat menghasilkan suatu hasil tertentu. Proses klasifikasinya dilakukan dengan mengamati data pada setiap node untuk menentukan kelas akhir yang berada pada bagian leaves (Jijo & Abdulazeez, 2021). Metode ini menjadi sangat populer karena mudah dipahami oleh manusia, serta mampu menyederhanakan proses pengambilan keputusan yang kompleks menjadi lebih terstruktur dan sederhana (Mahbooba, Timilsina, Sahal, & Serrano, 2021).

Teknik pembelajaran mesin yang disebut Decision Tree mengatur keputusan ke dalam struktur percabangan yang menyerupai pohon. Teknik ini digunakan untuk meramalkan hasil, mengevaluasi risiko, dan memaksimalkan manfaat dengan menggambarkan proses pengambilan keputusan menggunakan aturan bersyarat. Node pohon masing-masing mewakili karakteristik, cabang-cabangnya mewakili hasil pengujian, dan daunnya mewakili kelas atau kategori hasil. (Mardiani, Rahmansyah, & Ningsih, 2023). Pohon keputusan, salah satu teknik kategorisasi yang paling populer dalam data mining, memiliki keuntungan karena mudah dipahami dan dianalisis. Menganalisis pengamatan pada node mengarah ke daun, yang merupakan tujuan klasifikasi akhir, di mana proses klasifikasi selesai. Pohon keputusan sering dibuat dengan menggunakan algoritme seperti ID3, C4.5, dan CART (Septhya et al., 2023). Pemilihan atribut terbaik untuk akar pohon didasarkan pada nilai information gain tertinggi, yang dihitung

menggunakan rumus berikut (Arnomo, Fajrin, Siyamto, & Sadikin, 2023):

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}_i$$

Keterangan:

S = Himpuan Kasus

A = Atribut

n = Jumlah partisi atribut A

|S_i| = Jumlah kasus pada partisi ke-i

|S| = Jumlah kasus dalam S

Untuk perhitungan entropi dapat dilihat pada persamaan 2.2 berikut ini:

$$\text{Entropy}(S) = \sum_{i=1}^n -p_i * \log_2 p_i$$

Keterangan:

S = Himpuan Kasus

A = Fitur

n = Jumlah partisi S

P_i = Proporsi dari S_i terhadap S

Recall dan Precision

Pengukuran recall dan akurasi adalah dua contoh bagaimana konsep pengukuran relevansi masih digunakan hingga saat ini. Dalam sistem temu kembali informasi, recall adalah persentase dokumen yang berhasil ditemukan selama pencarian (Prabowo, 2021). Rumus untuk menghitung recall adalah:

$$\text{Recall} = \frac{a}{a+c}$$

Keterangan:

a = Jumlah dokumen relevan yang berhasil ditemukan

b = Jumlah dokumen relevan yang tidak ditemukan

Ukuran recall dinyatakan dalam persentase, dari 1 hingga 100%.

Precision, di sisi lain, adalah persentase dokumen yang dianggap relevan dengan persyaratan pencarian informasi. Ini adalah proporsi jumlah total dokumen dalam kumpulan dokumen yang relevan dengan jumlah dokumen yang diambil (Yuranda, Sutabri, & Wahyuningsih, 2023).

Dengan rumus sebagai berikut:

$$\text{Precision} = \frac{a}{a+b}$$

Keterangan:

a = Jumlah dokumen relevan yang berhasil ditemukan

b = Jumlah dokumen relevan yang tidak ditemukan

Ukuran dalam precision dinyatakan dalam persentase, 1 sampai dengan 100%.

RapidMiner

RapidMiner adalah perangkat lunak yang digunakan untuk memproses dan menganalisis data. Aplikasi ini menawarkan solusi dalam melakukan analisis data mining, text mining, dan analisis prediksi (Dwi Lestari & Mulyawan, 2023). RapidMiner adalah sebuah platform perangkat lunak yang dirancang untuk mempermudah proses analisis data dan pengembangan model machine learning tanpa memerlukan kemampuan pemrograman yang kompleks. Dengan antarmuka grafis berbasis drag and drop, RapidMiner memungkinkan pengguna untuk membangun alur kerja analisis data secara intuitif, mulai dari tahap pembersihan dan transformasi data, pemilihan algoritma, pelatihan model, hingga evaluasi performa dan visualisasi hasil (Marzukhi, Awang, Alsagoff, & Mohamed, 2021).

Platform ini mendukung berbagai teknik analisis seperti klasifikasi, regresi, clustering, dan bahkan analisis teks, sehingga cocok digunakan baik oleh peneliti, akademisi, maupun praktisi di dunia industri. Keunggulan utama RapidMiner terletak pada kemudahan penggunaannya dan fleksibilitasnya dalam mengolah data, menjadikannya salah satu alat populer dalam dunia data science dan kecerdasan buatan (Rafi Nahjan, Nono, & Apriade, 2023).

METODE (GUNAKAN STYLE SECTION)

Penelitian ini memprediksi keterlambatan jam kerja karyawan dengan menggunakan strategi klasifikasi dan algoritma *Decision Tree*. Langkah *Knowledge Discovery in Database* (KDD) dalam pemrosesan data, yang melibatkan transformasi, pembersihan, dan pemilihan data, diselesaikan dengan RapidMiner hingga data siap untuk dianalisis.

Adapun tahapan penelitian ini meliputi:

1. Identifikasi dan perumusan masalah.
2. Studi pustaka terkait klasifikasi data dan algoritma *Decision Tree*.
3. Pengolahan data berdasarkan tahapan KDD.

4. Uji coba model *Decision Tree* dengan pembagian data training dan testing untuk memperoleh akurasi terbaik dengan RapidMiner.
5. Pembahasan hasil akurasi model klasifikasi.
6. Penarikan kesimpulan berdasarkan hasil penelitian.

Data yang digunakan berupa 150 data presensi karyawan dari Kantor Cabang BPJS Kesehatan, diperoleh melalui survei langsung. Variabel yang digunakan meliputi: jam bangun, jam berangkat, jarak ke kantor, jenis transportasi, tingkat kemacetan, cuaca, dan status keterlambatan sebagai label.

Teknik pengumpulan data dilakukan melalui:

1. Studi pustaka sebagai landasan teori.
2. Survei terhadap karyawan sebagai sumber data primer.
3. Seleksi dan pembersihan data sebelum tahap pengujian model.



Gambar 1. Flowchart Penelitian

Dimulai dengan proses identifikasi, perumusan masalah, dan studi literatur sebagai dasar pemahaman. Gambar 1 mengilustrasikan perkembangan tahapan penelitian. Proses *Knowledge Discovery in Database* (KDD), yang melibatkan pemrosesan, pembersihan, dan pemilihan data, kemudian diselesaikan. Setelah data tersedia, program RapidMiner digunakan untuk memprosesnya dengan menggunakan metode *Decision Tree* untuk kategorisasi. Setelah dilakukan uji akurasi terhadap hasil kategorisasi, komentar dan kesimpulan dibuat berdasarkan hasil penelitian.

HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini berasal dari data presensi karyawan yang bekerja di Kantor BPJS Kesehatan Cabang Medan. Proses pengambilan data dilakukan secara langsung melalui metode survei yang disebarluaskan kepada seluruh karyawan, sehingga diperoleh sebanyak 150 record data presensi yang mencerminkan perilaku

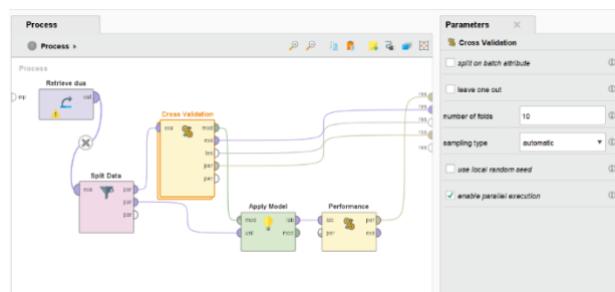
kehadiran karyawan, khususnya dalam hal keterlambatan jam masuk kerja.

Setelah data berhasil dikumpulkan, langkah selanjutnya adalah mempersiapkan data tersebut untuk tahap pengujian dan pemodelan. Tahap ini mencakup berbagai proses penting, seperti pemilihan tabel dan record yang relevan, serta identifikasi dan seleksi terhadap variabel-variabel data yang memiliki pengaruh terhadap keterlambatan. Selain itu, dilakukan pula proses pembersihan data (*data cleansing*) untuk menghilangkan data ganda, data kosong, atau variabel-variabel yang tidak signifikan dan tidak berkontribusi pada hasil akhir. Proses persiapan ini sangat krusial karena akan mempengaruhi kualitas model klasifikasi yang dibangun dalam tahap selanjutnya menggunakan algoritma *Decision Tree* di RapidMiner.

Nama	Jam Masuk	Jam Bangun	Jam Berangkat	Jarak	Transportasi	Kemacetan	Cuaca	Keterlambatan
A1	8:00	4:30	6:30	3.3	mobil	lancar	cerah	tidak terlambat
A2	8:00	6:00	7:15	4.8	motor	lancar	cerah	terlambat
A3	8:00	5:30	7:30	4.7	mobil	lancar	cerah	tidak terlambat
A4	8:00	7:00	7:30	5.3	mobil	lancar	cerah	terlambat
A5	8:00	7:00	7:30	4.1	mobil	lancar	cerah	terlambat
A6	8:00	6:45	6:15	4.6	motor	lancar	cerah	tidak terlambat
A7	8:00	7:00	7:30	6.9	motor	sedang	cerah	terlambat
A8	8:00	5:30	6:15	2.2	motor	lancar	cerah	tidak terlambat
A9	8:00	4:15	7:00	3	mobil	lancar	cerah	tidak terlambat

Gambar 2. Dataset Presensi Karyawan

1. Modelling



Gambar 4. Model Pengujian

Tahap berikutnya adalah proses pemodelan, yang terlihat pada gambar 3, yang menggambarkan pengujian model *decision tree*. Setelah melalui tahap data preparation, langkah selanjutnya adalah proses *Set Role* untuk menentukan label. Kemudian, dilakukan validasi dengan *10-fold cross-validation*

dalam tahap pelatihan (training). Untuk tahap

pengujian (testing), digunakan metode *apply model* dan *performance*. Proses pelatihan dan pengujian ini bertujuan untuk memperoleh *confusion matrix*, yang mencakup nilai tingkat akurasi, *class precision*, dan *class recall*.

2. Hasil Pengujian Algoritma Decission Tree

Penelitian ini menggunakan perangkat lunak RapidMiner untuk menerapkan pendekatan *Decision Tree* pada data keterlambatan jam kerja karyawan untuk melakukan percobaan klasifikasi. Sebanyak 150 sampel data yang telah melalui proses preprocessing untuk menjamin data tersebut bersih dan siap untuk digunakan dalam proses pemodelan digunakan untuk prosedur pengujian. Label ditentukan dengan menggunakan *Set Role*, model dilatih dengan menggunakan pendekatan *10-fold cross validation*, dan model serta *performance evaluator* digunakan untuk pengujian. *Matrix Confusion* yang memberikan ringkasan kinerja model klasifikasi, termasuk nilai akurasi, presisi, dan recall, adalah produk akhir dari prosedur ini. Gambar 4

precision: 94.00% +/- 9.66% (micro average: 93.75%) (positive class: terlambat)			
	true tidak terlambat	true terlambat	class precision
pred. tidak terlambat	66	6	91.67%
pred. terlambat	3	45	93.75%
class recall	95.65%	88.24%	

Gambar 3. Hasil Pengujian Algoritma Decission Tree

menunjukkan *confusion matrix* yang dihasilkan, yang berfungsi sebagai dasar untuk menilai seberapa baik model mengklasifikasikan keterlambatan karyawan. Tingkat akurasi data cukup baik, seperti yang ditunjukkan oleh hasil pengujian model pohon keputusan pada Gambar 4, yang menunjukkan hasil akurasi sebesar 94.00% dan nilai AUC sebesar 0.939%.

3. Pohon Keputusan

Berikut adalah visualisasi pohon keputusan hasil klasifikasi yang memprediksi keterlambatan karyawan berdasarkan atribut seperti jam berangkat, jarak ke kantor, cuaca, dan jenis transportasi yang digunakan.



Gambar 5. Pohon Keputusan

Output dari pohon keputusan Gambar 5 menunjukkan bahwa variabel "jarak" memiliki dampak terbesar pada keterlambatan jam kerja.

- a. Apabila jarak ≤ 3.5 km, maka diprediksi tidak terlambat, tanpa memperhitungkan variabel lain.
- b. Apabila jarak > 3.5 km, maka akan ditentukan oleh variabel jam berangkat:
 - Jika jam berangkat $\leq 07:00$, maka tidak terlambat.
 - Jika jam berangkat = 07:15, maka akan ditentukan oleh variabel jarak lebih lanjut:
 - Jarak > 4.5 km \rightarrow terlambat
 - Jarak ≤ 4.5 km \rightarrow akan ditentukan oleh variabel cuaca:
 - Cuaca = cerah \rightarrow kemungkinan terlambat atau tidak seimbang
 - Cuaca = hujan \rightarrow lebih cenderung terlambat
 - Jika jam berangkat = 07:30, maka akan ditentukan oleh variabel cuaca:
 - Cuaca = cerah \rightarrow ditentukan oleh transportasi:
 - Mobil \rightarrow kemungkinan terlambat dan tidak terlambat seimbang
 - Motor \rightarrow lebih cenderung terlambat
 - Cuaca = hujan \rightarrow terlambat
 - Jika jam berangkat = 07:45, maka pasti terlambat

Berikut ini merupakan hasil perhitungan manual untuk terhadap hasil akurasi tertinggi untuk *recall* dan *precision* yang dapat dilihat sebagai berikut ini:

a. Hasil *recall*

$$\text{Tidak terlambat: } \frac{66}{66+3} = \frac{66}{69} = 0,9565 \text{ (95,65\%)}$$

$$\text{Terlambat: } \frac{45}{46+6} = \frac{45}{51} = 0,8824 \text{ (88,24\%)}$$

b. Hasil *Precision*

$$\text{Tidak terlambat: } \frac{66}{66+6} = \frac{66}{72} = 0,9168 \text{ (91,68\%)}$$

$$\text{Terlambat: } \frac{45}{45+3} = \frac{45}{48} = 0,9375 \text{ (93,75\%)}$$

PENUTUP

SIMPULAN

Temuan penelitian akhir peneliti menunjukkan bahwa algoritma Decision Tree bekerja secara efektif dalam mengklasifikasikan data keterlambatan. Pendekatan Decision Tree memberikan hasil dengan tingkat akurasi sebesar 94,00% dan nilai AUC sebesar 0,939%. Berdasarkan model ini, faktor yang paling berpengaruh terhadap keterlambatan karyawan adalah cuaca, waktu keberangkatan, dan jarak. Dengan meneliti variabel-variabel yang mempengaruhi keterlambatan, temuan penelitian ini diyakini dapat membantu perusahaan untuk meningkatkan kedisiplinan karyawan.

DAFTAR PUSTAKA

- Amirulloh, I., & Taufiqurrochman. (2017). Komparasi Model Klasifikasi Algoritma Keterlambatan Siswa Masuk Sekolah. *Seminar Nasional Sains Dan Teknologi Fakultas Teknik Universitas Muhammadiyah Jakarta*, 2(November), 1-4.
- Andry, J. F., Hartono, H., Honni, Chakir, A., & Rafael. (2022). Data Set Analysis Using Rapid Miner to Predict Cost Insurance Forecast with Data Mining Methods. *Journal of Hunan University Natural Sciences*, 49(6), 167-175. <https://doi.org/10.55463/issn.1674-2974.49.6.17>
- Arnomo, S. A., Fajrin, A. A., Siyamto, Y., & Sadikin, S. F. N. (2023). Evaluasi Model Decision Tree Pada Keputusan Kelayakan Kredit. *Jurnal Desain Dan Analisis Teknologi*, 2(2), 200-206. <https://doi.org/10.58520/jddat.v2i2.39>
- Asyuti, S., & Setyawan, A. A. (2023). Data Mining Dalam Penggunaan Presensi Karyawan Denga Cluster Means. *Jurnal Ilmiah Sains Teknologi Dan Informasi*, 1(1), 01-10. Retrieved from <https://jurnal.alimspublishing.co.id/index.php/JITI/article/download/6/6>
- Azam, Z., Islam, M. M., & Huda, M. N. (2023). Comparative Analysis of Intrusion Detection Systems and Machine Learning-Based Model Analysis Through Decision Tree. *IEEE Access*, 11(July), 80348-80391.

<https://doi.org/10.1109/ACCESS.2023.329644>
4

Dobesova, Z. (2024). Evaluation of Orange data mining software and examples for lecturing machine learning tasks in geoinformatics. *Computer Applications in Engineering Education*, 32(4), 1-18. <https://doi.org/10.1002/cae.22735>

Dwi Anjani, N., Puspaningtyas Faeni, D., & Wahyu Setyawati, N. (2024). Beban Kerja, Burn Out, Jenjang Karir, Employee Engagement terhadap Loyalitas Karyawan Pada Kantor Kelurahan Harapan Jay Bekasi Utara. *PRIMAN; Pusat Publikasi Ilmu Manajemen*, 2(2), 50-60.

Dwi Lestari, P., & Mulyawan, M. (2023). Datamining Pada Penjualan Air Bersih Di Spam Akidah Menggunakan Algoritma K-Means Clustering Menggunakan Rapidminer. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 412-416. <https://doi.org/10.36040/jati.v7i1.6315>

Gheisari, M., Hamidpour, H., Liu, Y., Saedi, P., Raza, A., Jalili, A., ... Amin, R. (2022). Data Mining Techniques for Web Mining: A Survey. *Artificial Intelligence and Applications*, 1(1), 3-10. <https://doi.org/10.47852/bonviewwaia2202290>

Ika, N., & Sitompul, G. O. (2022). Pengaruh kompensasi dan disiplin kerja terhadap produktivitas kerja karyawan Bank KALBAR Pontianak. *Jurnal Ekonomi, Sosial Dan Humaniora*, 3(9), 22-29. Retrieved from <https://www.jurnaltelektiva.com/index.php/jurnal/article/view/738>

Jassim, M. A., & Abdulwahid, S. N. (2021). Data Mining preparation: Process, Techniques and Major Issues in Data Analysis. *IOP Conference Series: Materials Science and Engineering*, 1090(1), 012053. <https://doi.org/10.1088/1757-899x/1090/1/012053>

Jijo, B. T., & Abdulazeez, A. M. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(01), 20-28. <https://doi.org/10.38094/jastt20165>

Mahbooba, B., Timilsina, M., Sahal, R., & Serrano, M. (2021). Explainable Artificial Intelligence (XAI)

to Enhance Trust Management in Intrusion Detection Systems Using Decision Tree Model. *Complexity*, 2021. <https://doi.org/10.1155/2021/6634811>

Mardiani, E., Rahmansyah, N., & Ningsih, S. (2023). Komparasi Metode Knn, Naive Bayes, Decision Tree, Ensemble, Linear Regression Terhadap Analisis Performa Pelajar Sma. *Innovative: Journal Of...*, 3(2), 13880-13892. Retrieved from <http://j-innovative.org/index.php/Innovative/article/view/1949%0Ahttp://j-innovative.org/index.php/Innovative/article/download/1949/1468>

Marzukhi, S., Awang, N., Alsagoff, S. N., & Mohamed, H. (2021). RapidMiner and Machine Learning Techniques for Classifying Aircraft Data. *Journal of Physics: Conference Series*, 1997(1), 1-8. <https://doi.org/10.1088/1742-6596/1997/1/012012>

Muqorobin, M., Kusrini, K., & Luthfi, E. T. (2019). Optimasi Metode Naive Bayes Dengan Feature Selection Information Gain Untuk Prediksi Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan Sekolah. *Jurnal Ilmiah SINUS*, 17(1), 1. <https://doi.org/10.30646/sinus.v17i1.378>

Prabowo, T. T. (2021). Efektivitas Sistem Temu Kembali Informasi Perpustakaan Digital Institut Seni Indonesia (ISI) Yogyakarta dalam Tinjauan Recall dan Precision. *Media Pustakawan*, 28(1), 37-48. <https://doi.org/10.37014/medpus.v28i1.1087>

Putri, N. B., & Wijayanto, A. W. (2022). Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing. *Komputika : Jurnal Sistem Komputer*, 11(1), 59-66. <https://doi.org/10.34010/komputika.v11i1.4350>

Rafi Nahjan, M., Nono, H., & Apriade, V. (2023). Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 101-104. <https://doi.org/10.36040/jati.v7i1.6094>

Septhya, D., Rahayu, K., Rabbani, S., Fitria, V., Rahmaddeni, R., Irawan, Y., & Hayami, R. (2023). Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 15–19. <https://doi.org/10.57152/malcom.v3i1.591>

Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 64–71.

Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. *JBASE - Journal of Business and Audit Information Systems*, 4(1), 13–21. <https://doi.org/10.30813/jbase.v4i1.2729>

Suparman, D. R., Jajang, D., & Wahyudin. (2023). Pengaruh Motivasi terhadap Kinerja Karyawan Perusahaan PT Bekaert Indonesia Karawang. *Bisma: Jurnal Manajemen*, 9(1), 47–53. <https://doi.org/10.23887/bjm.v9i1.51314>

Triandini, M., Defit, S., & Nurcahyo, G. W. (2021). Data Mining dalam Mengukur Tingkat Keaktifan Siswa dalam Mengikuti Proses Belajar pada SMP IT Andalas Cendekia. *Jurnal Informasi Dan Teknologi*, 3(3), 167–173. <https://doi.org/10.37034/jidt.v3i3.120>

Wahyudi, W., Rozi, A., & Putry, M. (2022). Kompetensi dan kinerja karyawan: Peran moderasi iklim organisasional. *Jurnal Manajemen Maranatha*, 21(2), 165–176. <https://doi.org/10.28932/jmm.v21i2.4671>

Yulianti, T., Solehudin, Suyaman, D. J., & Hersona, S. (2023). Pengaruh Lingkungan Kerja Terhadap Disiplin Serta Dampaknya Terhadap Kinerja Karyawan. *Jurnal Swabumi*, 11(2), 173–185.

Yuranda, R., Sutabri, T., & Wahyuningsih, D. (2023). Pendekatan Macine Learning dalam Evaluasi Label Berita Berdasarkan Judul: Studi Kasus Media Online Rezky. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 12(3), 434–439. <https://doi.org/10.32736/sisfokom.v12i3.1808>