

PERBANDINGAN METODE ALGORITMA C4.5 DAN NAIVE BAYES UNTUK MEMPREDIKSI PENJUALAN KOSMETIK PADA TOKO JOLITA

M Khabib Al Fatach ¹, Arief Wibowo ²

^{1,2}Universitas Budi Luhur, Magister Ilmu Komputer
2311600189@student.budiluhur.ac.id

ABSTRAK

Tujuan dari penelitian ini adalah untuk membantu memprediksi masalah yang terjadi di toko Jolita, yang menjual produk kecantikan. Peneliti melakukan penelitian dengan menggunakan dua algoritma—algoritma C4.5 dan Naive Bayes—untuk membantu menentukan produk mana yang paling diminati pelanggan karena toko sering kehabisan stok produk, membuat pelanggan tidak dapat mendapatkan produk yang mereka cari. Digunakan algoritma C4.5 dan Naive Bayes untuk mendapatkan data penjualan toko Jolita, evaluasi kinerja menunjukkan bahwa produk Wardah adalah yang paling diminati. Hasil pengujian confusion matrix C4.5 menunjukkan akurasi sebesar 100% untuk data produk Wardah dengan 51 prediksi paling laku menggunakan data kolom harga dan nama barang dan kategori. Dibandingkan dengan produk lain seperti Kahf, Makeover, dan OMG, pengujian Naive Bayes menunjukkan akurasi sebesar 98% untuk 51 data yang menunjukkan bahwa Wardah adalah produk Dari dua algoritma, C4.5 memiliki hasil akurasi yang lebih baik daripada algoritma Naive Bayes.

Keyword : *Toko Jolita, C4.5, Naive Bayes.*

1. PENDAHULUAN

Seiring perkembangan produk kecantikan, kosmetik, semakin memiliki banyak peminat di kalangan remaja dan ibu-ibu. Saat ini, kosmetik dan perawatan kulit telah menjadi kebutuhan sehari-hari untuk memperindah penampilan mereka. Pecinta produk kecantikan ini sering berubah sesuai dengan tren. Menurut penelitian yang dilakukan pada toko Jolita, pengunjung sering tidak memiliki cukup stok barang saat mereka ingin membeli produk tertentu, yang tentunya berdampak pada kepuasan pelanggan. karenanya penelitian Peneliti menggunakan algoritma C4.5 dan Naive Bayes untuk memprediksi penjualan di toko Jolita dan memastikan bahwa stok selalu tersedia dan kebutuhan pelanggan terpenuhi. Tujuannya adalah untuk membantu toko Jolita memprediksi stok produk di masa depan sehingga tidak terjadi lagi kekurangan atau kehabisan persediaan barang.

Sebagai salah satu toko kosmetik yang berkembang, Toko Jolita menghadapi tantangan dalam mengidentifikasi pola penjualan dan memperkirakan permintaan produk di masa depan. Perencanaan promosi yang tepat sasaran, pengelolaan stok yang lebih baik, dan layanan pelanggan yang lebih baik adalah beberapa keuntungan yang dapat dihasilkan dari kemampuan untuk memprediksi penjualan dengan akurat. Untuk mencapai tujuan ini, dibutuhkan teknik yang efisien untuk menganalisis data penjualan saat ini.[1]

Algoritma C 4.5, yang dikembangkan oleh Ross Quinlan, digunakan untuk memproses data mining dengan membuat pohon keputusan, juga disebut sebagai pohon keputusan. Algoritma ini dikenal kuat untuk menghitung nilai kecenderungan untuk sekumpulan data tertentu. Algoritma C4.5 dapat digunakan untuk mempelajari prediksi stok kosmetik yang ada dan nilai prediksi ketersediaan

kosmetik yang dijual. Selain itu, data penjualan kosmetik pada toko Jolita dapat diproses lebih lanjut menggunakan Naive Bayes. Ini secara khusus mengolah data grossales dan penjualan untuk mengetahui preferensi pembeli untuk produk yang paling banyak terjual dan diminati; ini terkait langsung dengan stok kosmetik di toko Jolita[2].

2. TINJAUAN PUSTAKA

Banyak peneliti dan akademisi dari berbagai disiplin ilmu telah melakukan penelitian sebelumnya tentang penulisan tugas akhir. Penelitian lain, "Aplikasi Naive Bayes dalam Memprediksi Penjualan dan Persediaan Kain Jumutan di Toko Batiq Colet Tuan Kentang Palembang", menggunakan data mining dan algoritma c4.5 dan Naive Bayes, menunjukkan hasil dengan menggunakan aplikasi Rapidminer versi 9.9 berdasarkan data penjualan dan peredian dari tahun 2018 hingga 2020. Dengan metode Naive Bayes, hasil tersebut diperoleh. Penelitian menunjukkan bahwa hasil prediksi searah dengan data lapangan rill [3].

Dalam penelitian lain yang berjudul "Perbandingan Algoritma Klasifikasi Data Mining Pada Prediksi Penyakit Diabetes," terdapat beberapa kelemahan yang perlu dicermati. Salah satunya adalah keterbatasan dalam data yang digunakan, baik dari segi jumlah maupun keragamannya, yang dapat mengakibatkan hasil prediksi kurang mewakili populasi yang lebih luas atau kasus yang lebih kompleks. Selain itu, pemilihan algoritma yang diuji mungkin belum mencakup semua metode yang relevan atau paling efektif untuk prediksi diabetes, sehingga perbandingan yang dihasilkan mungkin tidak sepenuhnya menyeluruh. Kualitas data dalam penelitian ini juga menjadi aspek penting, di mana data yang tidak lengkap atau

adanya bias dalam pengumpulan dapat berdampak pada akurasi hasil prediksi[4].

2.1. Google Colab

Google Colab, atau dikenal juga sebagai Google Colaboratory, adalah layanan gratis dari Google yang mendukung pembelajaran mesin. Platform ini memungkinkan pengguna untuk menulis dan menjalankan kode Python, serta mengunggah, menyimpan, dan berbagi kode serta data mereka. Google Colab juga memungkinkan kolaborasi secara real-time dengan orang lain, serta menyediakan akses ke mesin virtual yang dilengkapi dengan CPU, GPU, dan TPU (Tensor Processing Unit) untuk menjalankan kode Python[5].

2.2. Data Mining

Data mining" mengacu pada proses mengekstrak informasi dari kumpulan data yang sangat besar dengan menggunakan teknik statistik, matematika, dan kecerdasan buatan untuk mengekstrak aturan, trend, politik, dan model prediksi yang sangat akurat. Data mining adalah proses pengumpulan, penggunaan, dan penggunaan data historis untuk menemukan keteraturan pola dalam kumpulan data [6].

2.3. Algoritma C4.5

Dengan atribut yang memiliki nilai gain tertinggi sebagai titik utama dalam klasifikasi, algoritma C4.5 digunakan untuk membentuk pohon keputusan. Cabang-cabang dalam pohon keputusan merepresentasikan pertanyaan klasifikasi, sedangkan daunnya menunjukkan kelas atau kelompok yang ditentukan. Salah satu tujuan utama dari algoritma C4.5 adalah mengelompokkan data ke dalam kelas tertentu melalui proses klasifikasi[7]. Dalam eksplorasi data, pohon keputusan sangat berguna untuk mengungkap hubungan tersembunyi antara variabel target dan berbagai variabel input potensial. Salah satu keunggulan utama pohon keputusan adalah kemampuannya menyederhanakan proses pengambilan keputusan yang kompleks, sehingga membantu pengambil keputusan lebih mudah memahami algoritma tersebut. Algoritma ini dapat menyelesaikan masalah secara sistematis dengan membentuk pohon keputusan [8].

2.4. Algoritma Naive Bayes

Naive Bayes adalah klasifikasi probabilitas sederhana yang menggunakan penjumlahan frekuensi dan kombinasi nilai dan kumpulan data yang diberikan untuk menghitung rentang probabilitas. Menurut Naive Bayes, nilai atribut tertentu dalam kelas tidak bergantung pada nilai atribut lainnya. Ketika digunakan pada database yang cukup besar, algoritma naive bayes dianggap memiliki kecepatan dan akurasi yang tinggi.[9]

2.5. Data

Data perlu diproses meskipun pada awalnya mungkin tidak bermanfaat bagi penerimanya. Data bias mencakup keadaan, gambar, suara, huruf, angka, matematika, bahasa, atau simbol lainnya yang dapat berfungsi sebagai alat untuk mengamati lingkungan, objek, kejadian, atau ide. Data dapat didefinisikan sebagai fakta atau elemen yang belum terstruktur tetapi memiliki makna yang terkait dengan peristiwa nyata[10].

2.6. Basis Data

Basis data adalah sekumpulan informasi yang disimpan secara sistematis dalam komputer dan dapat diperiksa oleh program komputer. Basis data terdiri dari tabel yang terdiri dari field-field, dan sebuah basis data dapat terdiri dari lebih dari satu tabel. Bagian penting dari sistem informasi kontemporer adalah basis data, yang memungkinkan pengelolaan data yang efisien untuk memenuhi berbagai kebutuhan bisnis dan operasional[11].

2.7. Prediksi/Forecasting

Prediksi, juga disebut forecasting, adalah seni dan ilmu memprediksi apa yang akan terjadi di masa depan[12]. Peramalan melibatkan pengumpulan data historis, seperti penjualan tahun lalu, dan menggunakan model matematika untuk memproyeksikan kejadian yang akan datang. Peramalan, atau prediksi, adalah penggunaan data untuk menentukan sasaran yang dikehendaki dan mengestimasi kejadian yang akan datang dengan tingkat kemungkinan terjadi yang besar dan dapat diterima [13].

2.8. Knowledge Discovery in Database (KDD)

Knowledge Discovery In Databases (KDD) adalah proses non-trivial yang mencakup pencarian dan identifikasi pola dalam data; pola-pola ini harus valid, baru, berguna, dan mudah dipahami. KDD melibatkan teknik integrasi dan penemuan ilmiah, serta penafsiran dan visualisasi pola dari berbagai kumpulan data [12]. Berikut tahapan dari KDD:

a. Data Selection

Sebelum tahap penggalian informasi dalam KDD dimulai, data yang dipilih dari sekumpulan data operasional harus dipilih dan disimpan dalam suatu berkas yang berbeda dari basis data operasional.

b. Pre-processing/ Cleaning

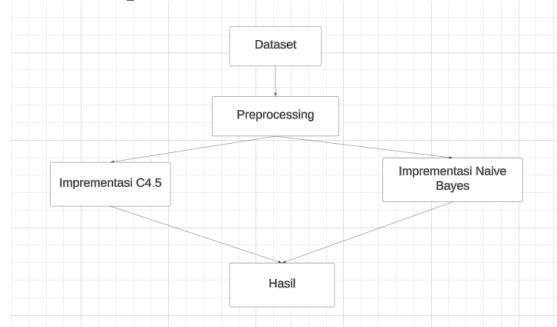
Sebelum proses data mining dapat dilakukan, data yang menjadi fokus dalam KDD harus

- melalui tahap pembersihan. Proses ini mencakup penghapusan duplikasi, pengecekan konsistensi data, serta perbaikan kesalahan. Selain itu, data yang ada juga dapat 'diperkaya' dengan menambahkan data atau informasi lain dari sumber eksternal.
- c. Transformation
Proses ini melibatkan transformasi data yang telah dipilih agar siap digunakan dalam data mining. Tahap ini sangat kreatif dan disesuaikan dengan jenis atau pola informasi yang ingin ditemukan dalam basis data.
 - d. Data mining
Data mining merupakan proses untuk menemukan pola atau informasi penting dari data tertentu dengan menggunakan berbagai teknik atau metode. Terdapat beragam teknik, metode, atau algoritma yang bisa diterapkan dalam data mining, dan pemilihan teknik atau algoritma yang sesuai sangat tergantung pada tujuan dan proses keseluruhan KDD.
 - e. Interpretation/ Evaluation
Tahap ini adalah bagian dari proses KDD, yang melibatkan penentuan apakah pola atau informasi yang ditemukan sesuai atau bertentangan dengan fakta atau hipotesis yang sudah ada sebelumnya. Pola informasi yang dihasilkan dari data mining perlu disajikan dalam format yang mudah dipahami oleh pihak yang berkepentingan.

3. METODE PENELITIAN

Karena penelitian ini merupakan jenis kuantitatif, penulis menggunakan data yang berisi angka sebagai bahan penelitian. Jenis data yang digunakan termasuk primer dan berkala, yang merupakan data yang dikumpulkan dari instansi dalam jangka waktu tertentu [4].

3.1. Tahapan Penelitian



Gambar 1. Diagram Metodologi

- a. Dataset
Dataset yang akan digunakan adalah data penjualan dari Oktober 2023 hingga Desember 2024 dari Jolita Store
- b. Preprocessing Data
Tahap pertama dalam proses preprocessing data adalah perubahan data mentah menjadi data yang siap untuk diproses melalui perbaikan dan transformasi data. Proses transformasi ini mengubah data menjadi bentuk yang akan dianalisis menggunakan perhitungan C4.5 dan Naive Bayes.
- c. Implementasi algoritma C4.5
C4.5 adalah algoritma yang digunakan untuk membuat pohon keputusan dari atribut dengan prioritas tertinggi yang dipilih berdasarkan nilai entropy yang menjadi poros atribut klasifikasi. Untuk membagi dataset dalam kelas yang berbeda dengan efektif, entropy sendiri digunakan untuk memilih fitur terbaik. J. Ross Quinlan mengembangkan algoritma induksi pohon keputusan ID3 ini. Algoritma ini memiliki kelebihan untuk menghilangkan perhitungan data yang tidak diperlukan dan membuat daerah yang sebelumnya luas menjadi lebih sederhana, spesifik, dan dapat disesuaikan.
- d. Algoritma Naive Bayes
Naive Bayes adalah algoritma pembelajaran mesin yang didasarkan pada teorema Bayes dan digunakan untuk memprediksi kemungkinan di masa yang akan datang berdasarkan pengalaman masa lalu.
- e. Hasil
Pada tahap ini, hasil dievaluasi dan dinilai keakuratannya. Kemudian, hasil dievaluasi dengan mengukur akurasi, ketepatan, dan recall

2.9. Klasifikasi

Dalam data mining, klasifikasi, juga dikenal sebagai klasifikasi, adalah model di mana klasifikasi digunakan untuk memprediksi kategori label tertentu. Misalnya, klasifikasi dapat digunakan untuk data aplikasi pinjaman uang dengan label "aman" atau "beresiko", data marketing dengan label "ya" atau "tidak", atau data medis dengan label "terapi A", "terapi B", atau "terapi C". Untuk memenuhi kebutuhannya, kategori dapat ditunjukkan dengan nilai yang tidak memiliki arti tertentu.[14]

Proses klasifikasi data terdiri dari dua tahap. Pendidikan adalah tahap di mana data dipelajari dengan menggunakan algoritma klasifikasi. Klasifikasi adalah tahap kedua, di mana data uji digunakan untuk mengevaluasi ketepatan aturan klasifikasi. Jika keakuratan yang dikondisikan dan diperkirakan dapat diterima, aturan tersebut dapat digunakan untuk mengklasifikasikan kumpulan data lainnya [15].

2.10. Teorema Bayes

Dalam data mining, metode statistika Bayes sangat penting. Metode ini didasarkan pada jumlah keuntungan yang terjadi antara berbagai pilihan klasifikasi yang berbeda yang didasarkan pada probabilitas [16]. Menurut Bayes (terutama Naive Bayes), independensi yang kuat pada fitur berarti bahwa keberadaan atau tidaknya fitur lain dalam data yang sama tidak berkaitan dengan fitur pada data tersebut. Sebagai contoh, dalam mengklasifikasikan hewan berdasarkan fitur seperti penutup kulit, cara reproduksi, berat, dan menyusui, ada ketergantungan pada fitur menyusui karena hewan yang menyusui biasanya melahirkan, sementara hewan yang bertelur biasanya tidak menyusui. Namun, dalam pendekatan Bayes, fitur-fitur ini dianggap independen, sehingga tidak ada hubungan yang terlihat di antara mereka [17].

4. HASIL DAN PEMBAHASAN

4.1. Dataset

Penelitian ini menggunakan data penjualan dari toko Jolita yang berlokasi di Komp.Ruko Taman Kb.Jeruk, Jl.Meruya Ilir Raya No.12 RT.4/RW.7, Meruya Selatan, Kec. Kembangan, Kota Jakarta Barat, Daerah Khusus Ibu Kota Jakarta 11640, dari Oktober 2023 hingga Desember 2023. Dataset ini diproses menggunakan Google Colab[18].



Gambar 2. Pemanggilan Dataset



Gambar 3. Menampilkan Dataset

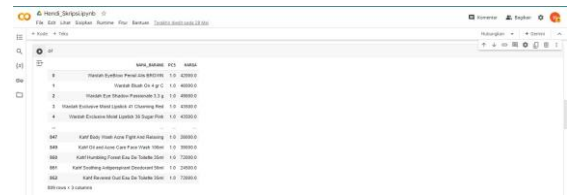
Pemanggilan dataset akan dilakukan dengan mengimport Pandas dan numpy ke dalam dataset, kemudian melakukan pengecekan dataset.

4.2. Pre-Processing



Gambar 4. Preprocessing

Preprocessing, langkah penting dalam pengolahan data, dilakukan untuk meningkatkan kualitas data untuk menghasilkan hasil data mining yang lebih akurat



Gambar 5. Hasil Preprocessing

Pada tahap preprocessing ini, dataset akan dihapus dari "unnammde", tanggal, dan kode dari kolomnya. Selanjutnya, dataset akan diperiksa untuk memastikan apakah memiliki nilai null atau kosong. Selain itu, kolom nama barang diubah menjadi nama_barang, jumlah diubah menjadi harga, dan harga diubah menjadi type float. dan menciptakan kategori baru untuk kolom dengan nama awalan seperti wardah, kahf, dan lain-lain.

4.3. Menyiapkan Data Untuk Di Proses Algoritma



Gambar 6. Pelabelan Otomatis

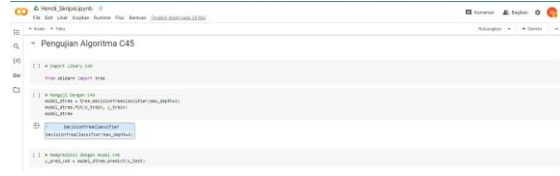
Memanfaatkan label encoder untuk melabelkan dataset: kita memilih cat (kategori) dan nama_barang sebagai kolom yang akan dilabeli menjadi nubur untuk mengubah data dari karakter menjadi angka



Gambar 7. Menampilkan hasil pelabelan

Grupkan data untuk melihat data, yang menunjukkan bahwa dataset mengandung enam kategori produk, seperti kristal, kaft, dll. Setelah itu, buat data x yang memiliki data nama_barnag, pcs, dan harga. Menggabungkan data y ke dalam data cat(kategori). Kemudian, dengan menggunakan metode import train_test_split dari sklear, data dibagi menjadi 80% data pelatihan dan 20% data pengujian.

4.4. Pengujian Algoritma Dengan C4.5



Gambar 8. Pengujian algoritma

Mengimport algoritma C4.5 dari sklear dan menguji dataset dengan C4.5. Dengan data uji yang belum diuji, memprediksi model C4.5.

4.5. Pengujian Algoritma Naive Bayes



Gambar 9. Pengujian algoritma Naive Bayes

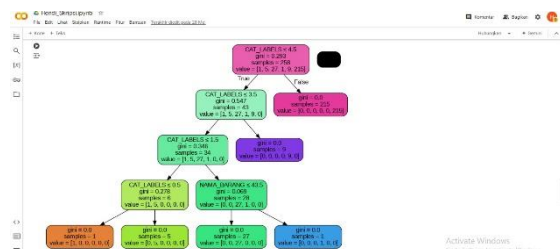
Mengimport algoritma Naive Bayes dari sklear, menguji dataset dengan Naive Bayes, dan kemudian menggunakan data pengujian yang belum diuji untuk memprediksi model Naive Bayes.

4.6. Hasil Pengujian C4.5 dan Naive Bayes



Gambar 10. Ekspor grafik pohon keputusan

Beberapa temuan menarik dihasilkan dari pengujian algoritma C4.5 dan Naive Bayes untuk memprediksi penjualan kosmetik di Toko Jelita. Untuk membuat pohon keputusan, algoritma C4.5 digunakan. Ini menghasilkan model yang dapat digambarkan sebagai grafik pohon Keputusan.



Gambar 11. Pohon keputusan

Poin keputusan ini menjelaskan bagaimana berbagai variabel, seperti kategori produk, harga, dan periode promosi, mempengaruhi penjualan kosmetik. Grafik pohon keputusan ini menunjukkan

jalur pengambilan keputusan yang diambil oleh algoritma berdasarkan data yang tersedia. Hasil pengujian menunjukkan bahwa C4.5 sangat akurat dan mudah dipahami oleh manajemen Toko Jelita, yang memudahkan pengambilan keputusan strategis.

5. KESIMPULAN DAN SARAN

Hasil penelitian ini menunjukkan bahwa penggunaan algoritma C4.5 dalam data mining terbukti efektif untuk memprediksi penjualan kosmetik di Toko Jolita, sementara algoritma Naive Bayes mampu memberikan prediksi yang cukup akurat terkait ketersediaan stok kosmetik. Perbandingan antara kedua algoritma ini memberikan pemahaman tentang kinerja masing-masing dalam konteks prediksi yang berbeda, di mana C4.5 lebih unggul dalam memprediksi penjualan, sedangkan Naive Bayes menunjukkan performa yang baik dalam memprediksi stok. Disarankan agar Toko Jolita memilih algoritma yang paling sesuai dengan kebutuhan prediksi mereka, dan jika memungkinkan, mengkombinasikan kedua metode ini untuk meningkatkan akurasi serta efisiensi dalam manajemen penjualan dan stok kosmetik.

DAFTAR PUSTAKA

- [1] Rian Pratama, B. Huda, E. Novalia, and H. Kabir, "Perbandingan Algoritma C4.5 dan Naive Bayes dalam Menentukan Persediaan Stok," METIK JURNAL, vol. 6, no. 2, pp. 115–122, Dec. 2022, doi: 10.47002/metik.v6i2.379.
- [2] Nurani and Afif, "Perbandingan Kinerja Algoritma Naive Bayes dan C4.5 Untuk Klasifikasi Harga Pangan," 2020.
- [3] Niken Putri Setyadi, "Penerapan Data Mining Untuk Prediksi Hasil Produksi Karet Menggunakan Algoritma Decision Tree C4.5," Teknologipintar.org, vol. 2, no. 7, 2022.
- [4] Y. F. Wijaya and A. Triayudi, "Perbandingan Algoritma Klasifikasi Data Mining Pada Prediksi Penyakit Diabetes," Journal of Computer System and Informatics (JoSYC), vol. 5, no. 1, pp. 165–174, Nov. 2023, doi: 10.47065/josyc.v5i1.4614.
- [5] R. Gelar Guntara, "Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7," Jurnal Teknologi Dan Sistem Informasi Bisnis, vol. 5, no. 1, pp. 55–60, Feb. 2023, doi: 10.47233/jteksis.v5i1.750.
- [6] J. Mantik et al., "PENERAPAN NAIVE BAYES DALAM MEMPREDIKSI PENJUALAN DAN PERSEDIAAN KAIN JUMPUTAN PADA TOKO BATIQ COLET TUAN KENTANG PALEMBANG," 2022.
- [7] M. Zulham, D. Saripurna, and Mhd. Z. Siambaton, "Aplikasi Diagnosa Penyakit

- Hepatitis dengan Menggunakan Metode Teorema Bayes dan Certainty Factor,” *Blend Sains Jurnal Teknik*, vol. 2, no. 1, pp. 1–15, Apr. 2023, doi: 10.56211/blendsains.v2i1.241.
- [8] S. P. Dewi, N. Nurwati, and E. Rahayu, “Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor,” *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 4, pp. 639–648, Mar. 2022, doi: 10.47065/bits.v3i4.1408.
- [9] W. A. Firmansyah, U. Hayati, and Y. A. Wijaya, “ANALISA TERJADINYA OVERFITTING DAN UNDERFITTING PADA ALGORITMA NAIVE BAYES DAN DECISION TREE DENGAN TEKNIK CROSS VALIDATION,” 2023.
- [10] S. Wahono and H. Ali, “PERANAN DATA WAREHOUSE, SOFTWARE DAN BRAINWARE TERHADAP PENGAMBILAN KEPUTUSAN (LITERATURE REVIEW EXECUTIVE SUPPORT SISTEM FOR BUSINESS),” *JEMSI JURNAL EKONOMI MANAJEMEN SISTEM INFORMASI*, vol. 3, no. 2, 2021, doi: 10.31933/jemsi.v3i2.
- [11] I. R. Mukhlis and R. Santoso, “Perancangan Basis Data Perpustakaan Universitas Menggunakan MySQL dengan Physical Data Model dan Entity Relationship Diagram,” *Journal of Technology and Informatics (JoTI)*, vol. 4, no. 2, pp. 81–87, Apr. 2023, doi: 10.37802/joti.v4i2.330.
- [12] H. I. Mariami. Dicky Nofriansyah, “Implementasi Data Mining Untuk Pengelempokan Buku Di Perpustakaan Yayasan Nurul Islam Indonesia Baru Dengan Metode K-Means Clustering,” *Jurnal CyberTech*, vol. 1, no. 1, 2021, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jct/index>
- [13] S. Chris Limong and D. Susilawati, “Analisis Perancangan Sistem untuk Kepuasan Pelanggan pada UD. Shinta Elektronik dengan Menggunakan Metode Algoritma C4.5,” 2020. [Online]. Available: <https://jurnal.buddhidharma.ac.id/index.php/algorithm/index>
- [14] A. Setiawan, “Perbandingan Penggunaan Jarak Manhattan, Jarak Euclid, dan Jarak Minkowski dalam Klasifikasi Menggunakan Metode KNN pada Data Iris,” *Jurnal Sains dan Edukasi Sains*, vol. 5, no. 1, pp. 28–37, May 2022, doi: 10.24246/juses.v5i1p28-37.
- [15] I. Romli and A. T. Zy, “Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5,” 2020.
- [16] S. Janu et al., “ANALISIS PERBANDINGAN ALGORITMA NAIVE BAYES DAN C.45 DALAM KLASIFIKASI DATA MINING UNTUK MEMPREDIKSI KELULUSAN,” 2021.
- [17] H. Derajad Wijaya and S. Dwiasnati, “Implementasi Data Mining dengan Algoritma Naive Bayes pada Penjualan Obat,” *JURNAL INFORMATIKA*, vol. 7, no. 1, 2020, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [18] J. Faran and R. T. Aldisa, “Analisis Data Mining dalam Komparasi Average Linkage AHC dan K-Means Clustering untuk Dataset Facebook Live Sellers,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 4, p. 2041, Oct. 2023, doi: 10.30865/mib.v7i4.6892.