

Analisis Klasifikasi Layanan Postpaid menggunakan Algoritma KNN, Decision Tree, dan Naïve Bayes

Miftahul Hasanah*¹, Nurmalitasari², Nurchim³

^{1,2}Sistem Informasi, Ilmu Komputer, Universitas Duta Bangsa Surakarta

³Teknik Informatika, Ilmu Komputer, Universitas Duta Bangsa Surakarta

Email: *202020865@mhs.udb.ac.id, nurmalitasari@udb.ac.id, nurchim@udb.ac.id

(Naskah masuk: 11 Juni 2024, diterima untuk diterbitkan: 20 April 2025)

Abstrak: Layanan postpaid adalah salah satu layanan yang pembayarannya dilakukan setelah pelanggan memakai produk yang disediakan oleh perusahaan, tagihan ini dibayarkan setiap akhir bulan. Klasifikasi layanan postpaid memegang peranan penting dalam strategi bisnis penyedia layanan telekomunikasi karena memungkinkan perusahaan untuk mengidentifikasi dan memahami pola penggunaan pelanggan, sehingga dapat mengembangkan paket layanan yang lebih tersegmentasi dan tepat sasaran, serta meningkatkan efisiensi operasional dan loyalitas pelanggan. Penelitian ini bertujuan untuk membandingkan kinerja setiap algoritma dalam mengidentifikasi pola penggunaan dan segmen pelanggan, sehingga penyedia layanan telekomunikasi dapat menentukan metode yang paling akurat dan efisien diantara tiga model yang sudah diuji. Data yang diteliti adalah data penjualan layanan postpaid pada bulan Juli dan Agustus yang terdiri dari 1.768 data dengan total 5 variabel. Penelitian ini menunjukkan hasil dari pengujian yang dilakukan memperoleh nilai akurasi tertinggi untuk model KNN dan Naïve Bayes sebesar 100% dan untuk model Decision Tree dengan sebesar 98%. Ini menunjukkan bahwa model KNN dan Naïve Bayes adalah model yang paling cocok untuk menganalisis data penjualan layanan postpaid. Pada penelitian selanjutnya dapat memasukkan algoritma klasifikasi lain seperti Support Vector Machine (SVM), Neural Networks, atau Logistic Regression untuk perbandingan performa yang lebih komprehensif.

Kata Kunci – KNN; decision tree; naïve bayes; klasifikasi; penjualan

Classification Analysis of Postpaid Services using KNN, Decision Tree, and Naïve Bayes Algorithms

Abstract: Postpaid service is one of the services for which payment is made after the customer uses the product provided by the company, this bill is paid at the end of each month. Postpaid service classification plays an important role in the business strategy of telecommunication service providers because it allows companies to identify and understand customer usage patterns, so that they can develop more segmented and targeted service packages, as well as improve operational efficiency and customer loyalty. This study aims to compare the performance of each algorithm in identifying usage patterns and customer segments, so that telecommunications service providers can determine the most accurate and efficient method among the three models that have been tested. The data studied is postpaid service sales data in July and August consisting of 1,768 data with a total of 5 variables. This research shows the results of the tests carried out obtained the highest accuracy value for the KNN and Naïve Bayes models of 100% and for the Decision Tree model with 98%. This shows that the KNN and Naïve Bayes models are the most suitable models for analyzing postpaid service sales data. Future research can include other classification algorithms such as Support Vector Machine (SVM), Neural Networks, or Logistic Regression for a more comprehensive performance comparison.

Keywords - KNN; decision tree; naïve bayes; classification; sales

1. PENDAHULUAN

Dalam industri telekomunikasi yang kompetitif, layanan postpaid atau lebih dikenal pascabayar menjadi salah satu pilar penting bagi penyedia layanan untuk mempertahankan dan meningkatkan pangsa pasar. Layanan postpaid menawarkan berbagai keunggulan seperti

pembayaran yang fleksibel dan akses tanpa batas yang menarik bagi pelanggan dengan kebutuhan komunikasi yang intensif [1]. Namun, tantangan utama yang dihadapi oleh penyedia layanan adalah memahami pola penggunaan pelanggan yang beragam dan mengembangkan strategi pemasaran yang tepat sasaran.

Analisis klasifikasi postpaid memainkan peranan krusial dalam mengidentifikasi segmen pelanggan berdasarkan perilaku dan karakteristik mereka. Dengan demikian, penyedia layanan dapat menyesuaikan penawaran produk dan layanan secara lebih efektif. Untuk mencapai tujuan ini, penggunaan metode klasifikasi yang akurat dan efisien sangat diperlukan.

Banyak penelitian yang membahas terkait analisis klasifikasi, yang pertama dilakukan oleh Wibowo menghasilkan kesimpulan bahwa penggunaan model Naïve Bayes memperoleh nilai akurasi paling rendah yakni 79% dan model Neural Network dengan nilai paling tinggi yaitu 94,1% [2]. Selanjutnya, pada penelitian yang dilakukan oleh Permana menghasilkan nilai cross validation dan T-test sebesar 79,29% untuk model Decision Tree, model KNN sebesar 66,69%, dan model Naïve Bayes sebesar 64,21% [3]. Kemudian mengutip dari penelitian yang dilakukan oleh Iqbal menunjukkan besaran nilai akurasi yang diperoleh dari model Naïve Bayes adalah 85,56%, model Decision Tree sebanyak 7,78%, dan model KNN 71,67% [4]. Lalu pada penelitian yang dilakukan oleh Nurchim dengan memanfaatkan analisis klasifikasi untuk mengelompokkan berdasarkan topik yang dicari menggunakan model NER dengan menghitung presisi, perolehan, dan skor f-1 menghasilkan nilai akurasi sebesar 86,96% [5]. Dan pada penelitian yang dilakukan oleh Hasanah menggunakan analisis klustering yang setara dengan analisis klasifikasi dengan metode K-Means menghasilkan nilai perhitungan yang telah dilakukan dengan tiga cara yakni, secara manual, aplikasi mobile, dan simulasi RapidMiner dengan perolehan hasil yang sama, yaitu kluster 1 meliputi provinsi Jabar, Jateng, dan Jatim kemudian sisanya terdapat 220 data yang masuk ke dalam kluster 0 [6].

Penelitian ini memfokuskan pada penggunaan tiga metode klasifikasi yang berbeda, yakni algoritma K-Nearest Neighbor (KNN) dikenal sebagai metode non-parametrik yang sederhana namun efektif dalam mengklasifikasikan data berdasarkan kedekatan antar-pelanggan [7], kemudian algoritma Decision Tree menyediakan model yang mudah dipahami dan diinterpretasikan, memungkinkan identifikasi faktor-faktor penting yang mempengaruhi keputusan pelanggan [3], dan algoritma Naïve Bayes menawarkan pendekatan probabilistik yang cepat dan seringkali cukup akurat dalam memprediksi kelas berdasarkan atribut yang ada [8].

Dengan membandingkan kinerja dari ketiga algoritma klasifikasi, penelitian ini bertujuan untuk menemukan metode yang paling sesuai untuk klasifikasi pelanggan postpaid dalam konteks penyedia layanan telekomunikasi. Analisis ini diharapkan dapat membantu dalam pembuatan strategi pemasaran yang lebih baik, meningkatkan retensi pelanggan, dan mengoptimalkan pendapatan perusahaan.

2. METODE PENELITIAN

Penelitian ini ditujukan untuk menilai dan membandingkan antara performa ketiga model klasifikasi saat menganalisis data pelanggan layanan postpaid.

2.1. Pengumpulan Data

Data penelitian berasal dari penyedia layanan telekomunikasi yang nama perusahaannya tidak dapat kami sebutkan kaitannya dengan privasi. Data ini adalah rekapan penjualan pada bulan Juli dan Agustus 2023 yang terdiri dari 1.768 data memuat 4 variabel independen dan 1 variabel dependen, yakni target penjualan bulan Agustus, laporan penjualan bulan Juli dan Agustus, selisih penjualan pada bulan Juli dan Agustus, dan kategori capai target atau tidak target.

2.2. Pra-Pemrosesan Data

Data akan diolah melalui pra-pemrosesan untuk memastikan kualitas dan konsistensinya. Tahapan pra-pemrosesan meliputi:

- 1) Pengecekan Missing Value
Data yang hilang akan diimputasi menggunakan metode yang sesuai, seperti imputasi rata-rata atau median.
- 2) Normalisasi Data
Data numerik akan dinormalisasi untuk memastikan setiap fitur memiliki nilai yang setara, sehingga tidak ada data yang dengan tingkatan yang berbeda.
- 3) Pembagian Data Testing dan Data Training
Data dibagi menjadi dua dengan rasio 70:30.

2.3. Implementasi Metode Klasifikasi

- 1) K-Nearest Neighbors (KNN)
Metode KNN akan diterapkan dengan berbagai nilai-k untuk menentukan nilai k yang optimal berdasarkan kinerja model pada data validasi.
- 2) Decision Tree
Metode decision tree akan diterapkan dengan menggunakan algoritma CART (Classification and Regression Trees). Parameter seperti kedalaman maksimum pohon dan jumlah minimum sampel per daun akan dioptimalkan menggunakan rasio gini.
- 3) Naive Bayes
Metode naive bayes akan diterapkan dengan asumsi distribusi Gaussian untuk fitur numerik. Model akan dievaluasi berdasarkan kinerja klasifikasinya pada data pengujian.

2.4. Evaluasi Model

Kinerja setiap model akan dievaluasi menggunakan metrik-metrik berikut:

- 1) Nilai akurasi adalah persentase tingkat kebenaran dari prediksi yang diuji.
- 2) Nilai precision, recall, dan f1-score adalah metrik untuk mengukur keefektifitasan model dalam mendeteksi kelas minoritas.

2.5. Analisis Perbandingan

Hasil evaluasi dari ketiga metode akan dibandingkan untuk menentukan metode yang paling efektif dalam klasifikasi pelanggan layanan postpaid. Analisis ini akan mencakup diskusi mengenai kelebihan dan kekurangan setiap metode serta rekomendasi untuk penerapan praktis di industri telekomunikasi.

2.6. Validasi dan Verifikasi

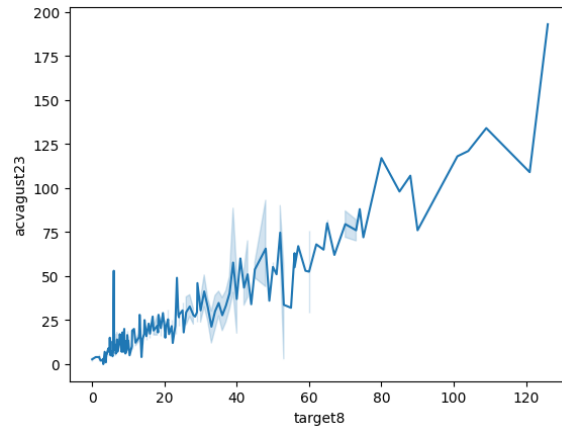
Untuk memastikan keandalan hasil penelitian, validasi silang (cross-validation) akan digunakan selama proses pelatihan model. Selain itu, teknik bootstrapping akan digunakan untuk mengestimasi ketepatan model pada data yang tidak terlihat.

Dengan mengikuti metode penelitian yang sistematis ini, diharapkan dapat diperoleh wawasan yang komprehensif mengenai kinerja dan keefektifan algoritma KNN, decision tree, dan naive bayes dalam klasifikasi pelanggan layanan postpaid.

3. HASIL DAN PEMBAHASAN

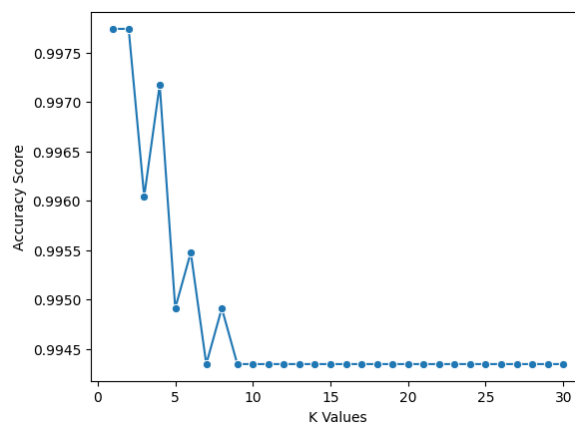
3.1. K-Nearest Neighbor

Berikut ini adalah grafik dari KNN yang menunjukkan perbandingan antara target penjualan pada bulan Agustus dan pencapaian penjualan pada bulan Agustus. Untuk target penjualan dibawah angka 20 mengalami kenaikan pencapaian penjualan dibawah angka 25, sedangkan untuk target dibawah angka 60 menunjukan penurunan karena pencapaian penjualan berada dibawah angka 50. Sedangkan untuk target penjualan dibawah 150 mengalami kenaikan karna pencapaian penjualan berada dibawah angka 200.



Gambar 1. Grafik penjualan

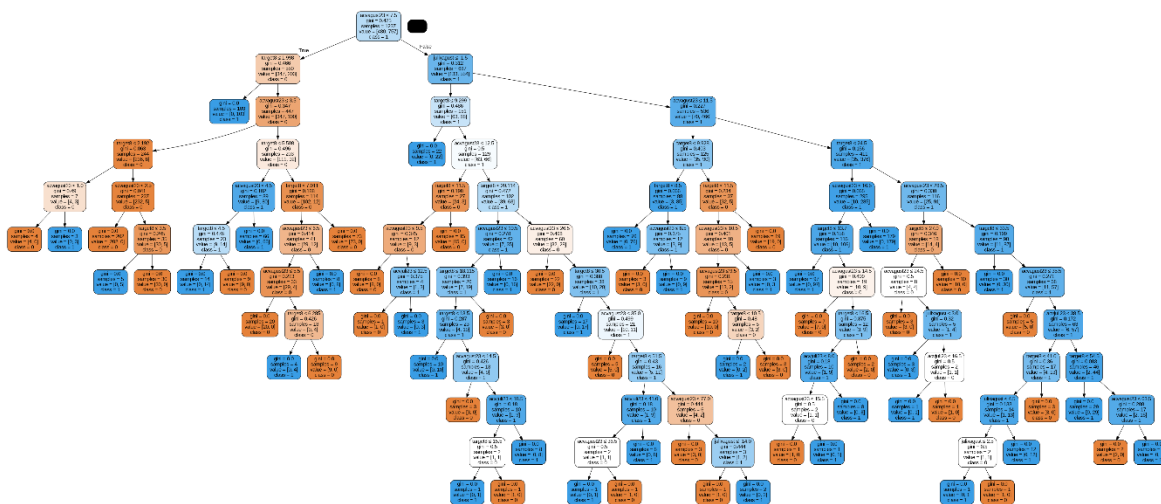
Dari grafik dibawah ini kita dapat melihat bahwa nilai-k 1 dan nilai k-2 berada di atas 0.9975, sedangkan nilai-k 7, 9, 10 sampai 30 berada di bawah 0.9945 dikarenakan keduanya sama untuk mendapatkan skor terbaik, kita akan menggunakan nilai k yang lebih kecil untuk meminimalisir lebih banyak titik data yang jauh dari aslinya jika kita memakai nilai k yang lebih tinggi.



Gambar 2. Grafik nilai-k

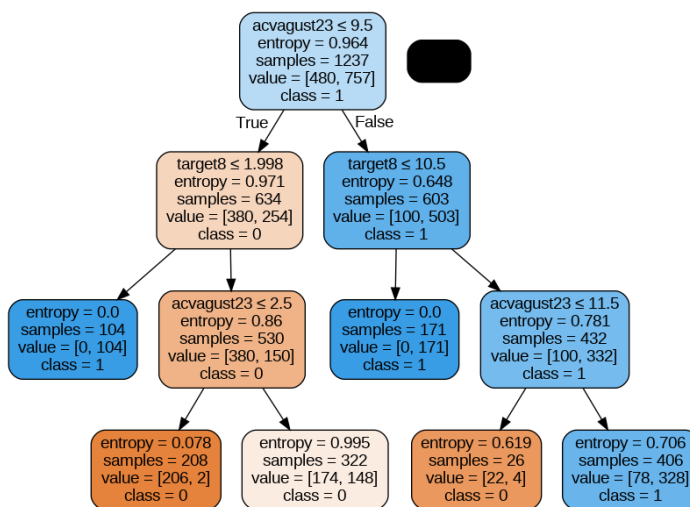
3.2. Decision Tree

Dalam bagan Decision Tree dibawah ini, setiap node internal memiliki aturan keputusan yang membagi data dan disebut sebagai rasio Gini yang mengatur node. Gambaran ini dapat dikatakan sebagai node murni ketika semua catatannya termasuk dalam kelas yang sama, node tersebut dikenal sebagai node daun[9]. Pada gambar ini, pohon yang dihasilkan tidak dipangkas. Pohon yang tidak dipangkas ini tidak dapat dijelaskan dan tidak mudah untuk dipahami. Selanjutnya kita akan memangkas pohon ini.



Gambar 3. Grafik pohon

Setelah melakukan pemangkasan, bagan Decision Tree menjadi tidak terlalu rumit, lebih mudah dijelaskan, dan lebih mudah dipahami dibandingkan plot model pohon keputusan sebelumnya. Pada bagan dibawah ini dilihat dari node pertama, nilai entropy dari variabel *acv Agust 23* (penjualan di bulan Agustus) memiliki nilai 0.964, nilai ini menunjukkan tingkat ketidakpastian data karna tercampur aduk dengan data yang lain. Kemudian, nilai sample sebesar 1237 yang menunjukkan 1237 instance dipertimbangkan dalam node ini, lalu value terbagi menjadi dua nilai, yaitu 480 dan 757 artinya sebanyak 480 data masuk kedalam kelas 0 dan sebanyak 757 data masuk kedalam kelas 1. Lalu kelas menunjukkan angka 1, yang artinya dominan kelas dalam node ini adalah kelas 1, kemudian node ini terbagi menjadi dua kondisi, yakni true (kondisi data terpenuhi) dan false (kondisi data tidak terpenuhi).



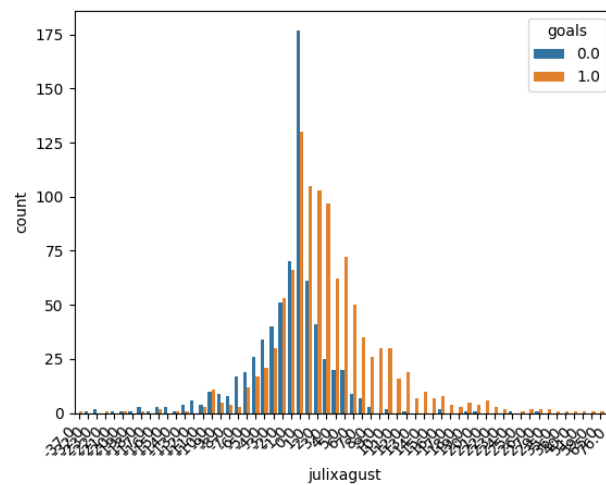
Gambar 4. Grafik pohon setelah dipangkas

Node ini terus dipecah lebih lanjut berdasarkan nilai dari fitur tertentu (feature) dengan ambang batas tertentu (threshold). Ketika sebuah instance melewati decision tree, keputusan pada setiap node didasarkan pada fitur dan ambang batas tersebut, mengarahkan instance ke cabang yang sesuai sampai mencapai daun (leaf) yang memberikan prediksi kelas akhir berdasarkan majority class dalam node tersebut.

3.3. Naïve Bayes

Berikut adalah hasil visualisasi dari dataset penjualan layanan postpaid menggunakan model Naïve Bayes. Perbandingan ini menggunakan 2 variabel yaitu *julix Agust* sebagai selisih dari

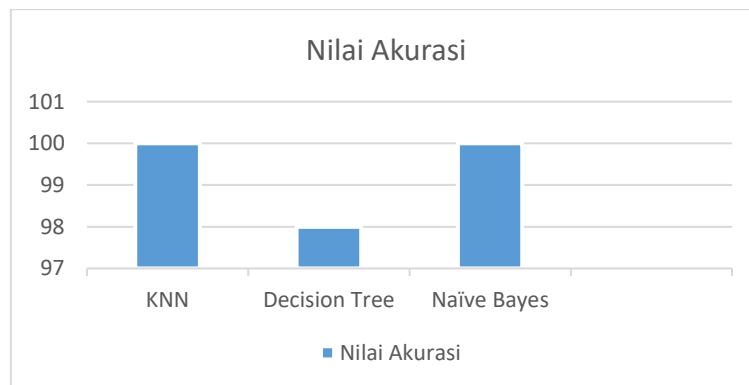
penjualan bulan Juli dan Agustus, sedangkan goals adalah kategori yang menunjukkan apakah penjualan telah mencapai target (angka 1) dan belum mencapai target (angka 0).



Gambar 5. Grafik Naïve Bayes

3.4. Perbandingan Akurasi

Pengujian menggunakan tiga model klasifikasi menghasilkan masing-masing nilai akurasi sebagai berikut.



Gambar 6. Grafik nilai akurasi

Metode KNN menunjukkan akurasi yang sempurna dalam mengklasifikasikan data postpaid. Ini berarti bahwa model KNN berhasil memprediksi semua instance dalam dataset dengan benar. KNN bekerja dengan baik dalam kasus ini mungkin karena data memiliki distribusi yang memungkinkan tetangga terdekat (neighbors) memberikan informasi yang cukup untuk klasifikasi yang akurat.

Decision Tree menunjukkan akurasi yang sedikit lebih rendah dibandingkan dengan KNN dan Naive Bayes, yaitu 98%. Meskipun sedikit lebih rendah, ini tetap merupakan tingkat akurasi yang sangat tinggi. Decision Tree cenderung memberikan hasil yang sangat baik dalam mengidentifikasi pola-pola dalam data, namun bisa sedikit rentan terhadap overfitting, terutama jika tidak diatur dengan baik.

Naive Bayes juga menunjukkan akurasi yang sempurna dalam klasifikasi data postpaid. Keunggulan Naive Bayes dalam kasus ini mungkin karena asumsi independensi antar fitur yang mendasarinya sesuai dengan karakteristik dataset yang digunakan. Naive Bayes sering kali sangat efektif dalam klasifikasi teks atau data yang memiliki probabilitas distribusi yang jelas.

Baik KNN maupun Naive Bayes memberikan akurasi 100%, menunjukkan bahwa kedua metode ini mampu mengklasifikasikan dataset tanpa kesalahan. Pernyataan ini didukung penelitian yang dilakukan oleh Putry dalam hasil uji untuk menganalisis data pada penyakit Diabetes Mellitus

menggunakan algoritma KNN dan Naïve Bayes [10]. Hal ini juga menunjukkan bahwa dataset yang digunakan sangat sesuai dengan karakteristik kedua metode tersebut. KNN mungkin lebih unggul dalam situasi di mana data memiliki batasan yang jelas dan terpisah, sedangkan Naive Bayes mungkin lebih unggul dalam situasi di mana fitur-fitur yang ada memberikan informasi probabilistik yang kuat.

4. KESIMPULAN

Pengujian dalam penelitian ini menggunakan tiga algoritma klasifikasi untuk menganalisis data penjualan pada layanan postpaid. Data yang digunakan adalah data penjualan layanan postpaid pada bulan Juli dan Agustus yang memuat sebanyak 1.768 data. Penelitian ini menghasilkan nilai akurasi untuk model KNN dan Naïve Bayes sebesar 100% dan Decision Tree sebesar 98%. Dari hasil penelitian yang diperoleh, model KNN dan Naïve Bayes sangat cocok untuk dengan dataset yang diteliti. Selain tiga metode yang digunakan, penelitian lebih lanjut dapat memasukkan algoritma klasifikasi lain seperti Support Vector Machine (SVM), Neural Networks, atau Logistic Regression untuk perbandingan performa yang lebih komprehensif.

DAFTAR PUSTAKA

- [1] R. M. P. Ramos and M. A. A. Trinidad, "Animo Repository The Effect of Conditional Cash Transfers on the Prepaid and Postpaid Expenditures of Internet and Cellular Services : The Case of Filipino Households," 2023.
- [2] A. Wibowo Putra, W. Widiyono, A. Saifudin, A. Darmawan Soma, and E. Budihartono, "Naïve Bayes, Neural Network dan K-Nearest Neighbor untuk Klasifikasi Topik Tugas Akhir," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 11, no. 4, pp. 661–667, 2022, doi: 10.30591/smartcomp.v11i4.4251.
- [3] A. P. Permana, K. Ainiyah, and K. F. H. Holle, "Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 6, no. 3, pp. 178–188, 2021, doi: 10.14421/jiska.2021.6.3.178-188.
- [4] M. Iqbal, A. Davy Wiranata, R. Suwito, and R. Faiz Ananda, "KLIK: Kajian Ilmiah Informatika dan Komputer Perbandingan Algoritma Naïve Bayes, KNN, dan Decision Tree terhadap Ulasan Aplikasi Threads dan Twitter," *Media Online*, vol. 4, no. 3, pp. 1799–1807, 2023, doi: 10.30865/klik.v4i3.1402.
- [5] N. Nurchim, N. Nurmalitasari, and Z. A. Long, "Indonesian news classification application with named entity recognition approach," *J. Infotel*, vol. 15, no. 2, pp. 130–134, 2023, doi: 10.20895/infotel.v15i2.909.
- [6] H. Hasanah, Nurmalitasari, and N. A. Sudiby, "Implementasi Data Mining Clustering untuk Mengetahui Potensi Produktifitas Kacang Tanah di Indonesia," *NJCA (National J. Comput. Its Appl.)*, vol. 5, no. 1, pp. 1–7, 2020, [Online]. Available: <http://journal.csnu.or.id/index.php/njca/article/view/148>
- [7] E. Utami, A. Y. Alwanda, A. S. wahyudi Hidayat, H. Sanjaya, and D. A. Dzulhijjah, "Perbandingan Metode SVM dan KNN pada Analisis Sentimen Twitter," pp. 767–772, 2023.
- [8] I. R. Pratama, M. Maimunah, and E. R. Arumi, "Sistem Klasifikasi Penjualan Produk Alat Listrik Terlaris Untuk Optimasi Pengadaan Stok Menggunakan Naïve Bayes," *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 2135, 2022, doi: 10.30865/mib.v6i4.4418.
- [9] D. Ade *et al.*, "Analisis Penilaian Kinerja Pekerja Building Maintenance Menggunakan Metode Klasifikasi Dengan Algoritma Decision Tree," vol. 4, no. 1, pp. 129–137, 2024.
- [10] N. M. Putry, "Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus," *EVOLUSI J. Sains dan Manaj.*, vol. 10, no. 1, 2022, doi: 10.31294/evolusi.v10i1.12514.