

Klasifikasi Kepuasan Pengguna Aplikasi Kopi Kenangan Menggunakan Algoritma Decision Tree

¹**Enggar Rahma Agustin, ²Intan Sulistyowati, ³Erna Daniati**

¹⁻³ Sistem Informasi, Universitas Nusantara PGRI Kediri

E-mail: ¹rahmaenggar5@gmail.com ²intansulistyowati27@gmail.com

³ernadaniati@unpkediri.ac.id

Penulis Korespondens : Erna Daniati

Abstrak— Di tengah persaingan industri minuman yang semakin digital, aplikasi Kopi Kenangan hadir sebagai solusi praktis untuk memesan minuman favorit para pelanggan. Namun, kenyamanan dan kemudahan yang ditawarkan aplikasi belum tentu menjamin kepuasan pengguna. Untuk itu, penting dilakukan analisis lebih lanjut guna memahami sejauh mana aplikasi ini mampu memenuhi harapan penggunanya. Penelitian ini menggunakan algoritma Decision Tree untuk mengklasifikasikan tingkat kepuasan pengguna berdasarkan berbagai faktor, seperti kemudahan penggunaan, kecepatan layanan, hingga fitur yang tersedia. Melalui pendekatan ini, diperoleh gambaran mengenai faktor-faktor utama yang memengaruhi kepuasan, serta masukan berbasis data untuk pengembangan layanan yang lebih baik ke depannya.

Kata Kunci—aplikasi Kopi Kenangan, Decision Tree, kepuasan pengguna, klasifikasi, layanan digital

Abstract—*In the midst of competition in the increasingly digital beverage industry, the Kopi Kenangan application is here as a practical solution to order customers' favorite drinks. However, the convenience and convenience offered by the application does not necessarily guarantee user satisfaction. For this reason, it is important to conduct further analysis to understand the extent to which this application is able to meet the expectations of its users. This study uses the Decision Tree algorithm to classify user satisfaction levels based on various factors, such as ease of use, speed of service, and available features. Through this approach, an overview of the main factors that affect satisfaction is obtained, as well as data-driven inputs for better service development in the future.*

Keywords—*Kopi Kenangan application, Decision Tree, user satisfaction, classification, digital services*

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

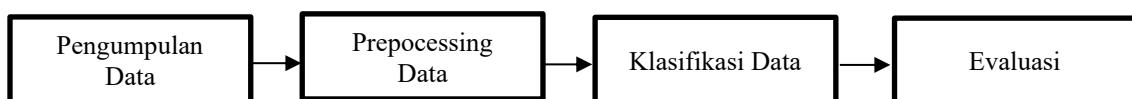
Perkembangan teknologi digital telah secara signifikan mengubah masyarakat dalam mengonsumsi makanan dan minuman, terutama melalui aplikasi mobile. Salah satu aplikasi yang sangat populer di Indonesia adalah "Kopi Kenangan" [1]. Dikelola oleh PT Bumi Berkah Boga, kopi kenangan merupakan kedai kopi lokal independen yang tengah mengalami pertumbuhan pesat di tanah air. Mereka menawarkan beragam pilihan minuman kopi

berkualitas, mulai dari espresso hingga non-espresso, serta berbagai pastry seperti roti dan kue kering. PT Bumi Berkah Boga mengintegrasikan teknologi aplikasi dalam operasionalnya dengan mengadopsi konsep "New Retail" [2]. Di mana perbedaan antara interaksi online dan offline semakin samar. Dengan aplikasi Kopi Kenangan, konsumen dapat memesan kopi dan berbagai minuman lain tanpa perlu mengantre di outlet. Aplikasi ini juga dilengkapi dengan beragam fitur menarik, termasuk promosi, cashback, Kenangan VIP, Daily CheckIn, layanan pesan-antar, chat langsung, serta berbagai metode pembayaran yang lengkap [3]. Penelitian ini akan mencakup pengumpulan dan pemrosesan data dari ulasan di Google Play Store, menggunakan Algoritma Decision Tree dengan bantuan software Google Colab [4].

Ulasan pengguna memiliki peranan yang sangat penting dalam mengidentifikasi masalah serta mengevaluasi kinerja aplikasi. Melalui penerapan metode decision tree guna mengelompokan respon subjektif dari komentar konsumen, PT Bumi Berkah mencapai pemahaman menyeluruh mengenai tanggapan dan kepuasan konsumen terhadap aplikasi tersebut [5]. Perolehan dari pengkajian ini, akan berkontribusi pada pengembangan dan perbaikan aplikasi Kopi Kenangan agar bisa menangani keperluan dan keinginan pelanggan di zaman modern. PT Bumi Berkah berkomitmen terus mengembangkan beragam pembaruan kreatif serta meningkatkan infrastruktur layanan yang kreatif serta kolaboratif.

II. METODE

Peneliti menggunakan kumpulan data yang diambil dari aplikasi Google Playstore dan memanfaatkan Google Colab sebagai alat untuk melakukan ekstraksi data. Setelah mengumpulkan data yang diperlukan, langkah berikutnya adalah melakukan proses pembersihan data (Preprocessing). Selanjutnya, klasifikasi akan dilakukan dengan menerapkan metode Decision Tree, yang diakhiri dengan tahap evaluasi, mengikuti alur penelitian yang ditunjukkan dalam gambar 1.



gambar 1. Proses penelitian

2.1 Pengambilan Data

Peneliti memanfaatkan Google Colab sebagai sarana untuk mengumpulkan data pada tahap awal penelitian. Data yang diperoleh berasal dari ulasan pengguna aplikasi Kopi Kenangan yang terdapat di Playstore.

2.2 Preprocessing Data

Preprocessing data merupakan tahap pembersihan data dengan mengecek data yang duplikat serta memastikan konsistensi [6]. Preprocessing data penting dalam analisis data mining untuk membersihkan, mengubah format, dan mempersiapkan data agar lebih mudah dan akurat [7].

2.2.1 Data Transformation

Tahap transformasi bertujuan untuk memproses data dengan skala, melakukan normalisasi, dan mereduksi. Salah satu tahapan dalam proses transformasi data yaitu mengubah tipe data

dari non-numerik menjadi numerik, selain itu, pada tahap ini juga dilakukan penanganan terhadap nilai yang hilang (missing value) dan data penyimpangan (outlier)[8].

2.3 Klasifikasi Data

Klasifikasi adalah teknik penambangan data yang mengelompokkan data ke dalam kategori atau kelas yang telah ditentukan [9][10]. Proses klasifikasi data dapat dilakukan dengan berbagai metode, salah satunya adalah menggunakan pohon keputusan (Decision Tree) [11].

2.4 Decision Tree

Decision Tree adalah Struktur flowchart yang berbentuk seperti pohon, di mana setiap bagian disebut sebagai cabang. Setiap simpul internal, terdapat tes atribut, sedangkan setiap cabang menunjukkan hasil tes, dan simpul daun menunjukkan distribusi kelas [12][13].

2.5 Data Mining

Data mining adalah suatu proses yang memanfaatkan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstrak dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang relavan dari berbagai basis data yang besar [14].

2.6 Web Scraping

Proses pengambilan dataset dilakukan dengan cara web scraping. Proses web scraping menggunakan bahasa pemrograman Python dengan Google Colab sebagai coding environment-nya. Data yang berhasil dikumpulkan terdiri dari 829 ulasan pengguna mengenai aplikasi kopi kenangan yang terdapat di Google Play store. Contoh ulasan yang diperoleh dari proses web scraping dapat dilihat pada Gambar 2.

2.7 Evaluasi

Tahap Evaluasi merupakan tahap paling akhir yang dilakukan untuk mengukur kinerja model klasifikasi dengan menghitung nilai accuracy, precision, dan recall.

III. HASIL DAN PEMBAHASAN

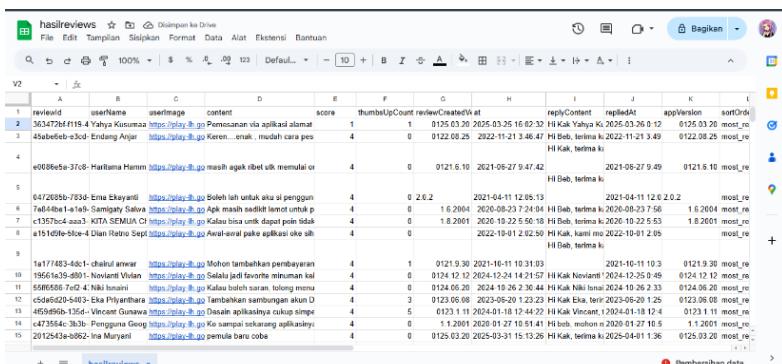
3.1 Pengumpulan Data

Dataset diperoleh dengan cara *scraping data* menggunakan bahasa *python* kemudian disimpan dalam bentuk csv/excel. Alur visualisasi pengumpulan data ditunjukkan pada gambar 2.



Gambar 2. Proses Pengumpulan Data

Scraping merupakan metode yang digunakan untuk mengumpulkan informasi dari situs web tertentu, baik secara manual maupun otomatis. Teknik ini berfokus pada pengambilan data dan ekstrasi data [15]. Setelah proses scraping selesai, kami menemukan sebanyak 829 data yang memberikan beragam ulasan dan bintang. Hasil dataset tersebut dapat dilihat pada Gambar 3.



Gambar 3. Tampilan Atas Dataset

3.2 Preprocessing Data

Preprocessing data adalah tahap pembersihan data dengan memeriksa duplikat data dan tingkat konsistensi [16]. Dalam proses preprocessing data memiliki tahapan yang perlu dilakukan seperti melihat missing value atau outliernya dan pembersihan data [17].

3.2.1 Proses Trasformation Data

Proses pembersihan data teks dilakukan pada kolom content dengan fungsi *clean_text*, yang menghapus angka, tanda baca, spasi berlebih, serta stopwords, dan mengubah semua teks menjadi huruf kecil. Hasilnya disimpan dalam kolom *cleaned_content* untuk analisis lebih lanjut. Selain itu, pencarian missing value dilakukan menggunakan *isnull().sum()*, dan nilai kosong pada kolom numerik seperti reviewCreatedVersion, appVersion, replyContent, dan repliedAt diisi dengan nilai rata-rata masing-masing kolom guna menjaga konsistensi data saat pemodelan.

Setelah melakukan pembersihan data teks dan penambahan mean (rata-rata), langkah selanjutnya yaitu mentransformation data dari kolom yang bertipe objek (teks) kedalam numerik (int). Kode untuk perubahan tipe numerik (int) dapat dilihat pada Gambar 4.

```
# Konversi kolom teks menjadi panjang karakter
text_columns = ['userName', 'content', 'replyContent', 'sortOrder', 'appId']
for col in text_columns:
    df_review[col] = df_review[col].apply(lambda x: len(str(x)) if pd.notnull(x) else 0)
```

Gambar 4. Kode Perubahan Tipe Data

```
df_review.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 829 entries, 0 to 828
Data columns (total 13 columns):
 #   Column      Non-Null Count  Dtype  
 ---  --          --          --      
 0   reviewId    829 non-null    object  
 1   userName    829 non-null    object  
 2   userImage   829 non-null    object  
 3   content     829 non-null    object  
 4   score       829 non-null    int64  
 5   thumbsUpCount 829 non-null    int64  
 6   reviewCreatedVersion 793 non-null    object  
 7   at          829 non-null    object  
 8   replyContent 725 non-null    object  
 9   repliedAt   725 non-null    object  
 10  appVersion   783 non-null    object  
 11  sortOrder    829 non-null    object  
 12  appId       829 non-null    object  
dtypes: int64(2), object(11)
memory usage: 84.3+ KB
```

Gambar 5. Sebelum Perubahan

```
df_review.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 829 entries, 0 to 828
Data columns (total 11 columns):
 #   Column      Non-Null Count  Dtype  
 ---  --          --          --      
 0   userName    829 non-null    int64  
 1   content     829 non-null    int64  
 2   thumbsUpCount 829 non-null    int64  
 3   at          829 non-null    int64  
 4   reviewCreatedVersion 829 non-null    int64  
 5   replyContent 829 non-null    int64  
 6   repliedAt   829 non-null    int64  
 7   appVersion   829 non-null    int64  
 8   sortOrder    829 non-null    int64  
 9   score       829 non-null    int64  
 10  appId       829 non-null    int64  
dtypes: int64(11)
memory usage: 71.4 KB
```

Gambar 6. Setelah Perubahan

Dari Gambar 5 (Sebelum Perubahan) menunjukkan bahwa mayoritas kolom dalam DataFrame masih bertipe data object, termasuk beberapa kolom yang seharusnya bersifat

numerik, seperti thumbsUpCount, replyCount, dan reviewCreatedVersion. Sementara itu pada gambar 6 (Setelah Perubahan) memperlihatkan tampilan setelah perubahan tipe data dilakukan. Kolom-kolom yang sebelumnya berformat object kini telah berhasil dikonversi menjadi tipe data yang lebih sesuai, seperti int64 untuk data numerik dan datetime64[ns] untuk informasi tanggal/waktu (at).

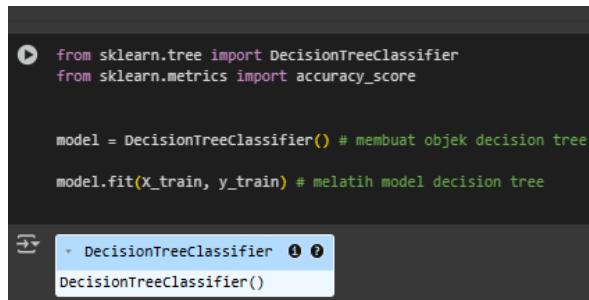
3.3 Klasifikasi Data

Sebelum memasuki tahap klasifikasi data, tahap awal yang perlu dilakukan yaitu pengembangan model klasifikasi melalui pembagian data (data splitting). Dataset dibagi menjadi dua bagian utama, data latih (training set) dan data uji (testing set) dengan proporsi 50:50. Proses ini dilakukan menggunakan fungsi `train_test_split` dari `sklearn.model_selection`, untuk memastikan bahwa data yang digunakan untuk pelatihan dan pengujian model terpisah dengan baik. Variabel input (fitur) disimpan di dalam `X`, sedangkan label target kepuasan pengguna disimpan dalam `y`. Dengan cara ini, model akan dilatih menggunakan sebagian besar data dan diuji pada data yang sebelumnya tidak pernah diakses, sehingga hasil evaluasi tetap dapat mencerminkan performa model secara lebih adil dan realistik. Kode untuk pengujian data dapat dilihat pada Gambar 7.

```
[ ] # import library train_test_split
from sklearn.model_selection import train_test_split

# data splitting
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.5)
```

Gambar 7. Tahapan Split Data



```
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score

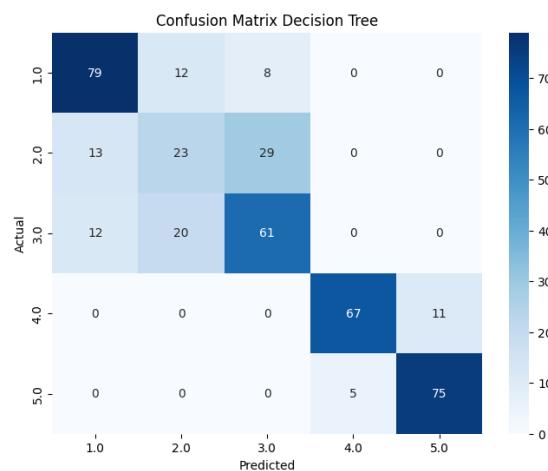
model = DecisionTreeClassifier() # membuat objek decision tree
model.fit(X_train, y_train) # melatih model decision tree
```

Gambar 8. Pelatihan Model Dengan Decision Tree

Pada Gambar 8, langkah krusial berikutnya adalah proses pelatihan model (training) yang dilakukan dengan memanggil fungsi `model.fit(X_train, y_train)`. Fungsi ini bertugas untuk memproses data pelatihan dan mengevaluasi nilai atribut pada `X_train`, serta mempelajari pola-pola yang ada dalam data tersebut berdasarkan label yang sesuai pada `y_train`. Hasil dari proses ini adalah terbentuknya struktur pohon keputusan, di mana setiap node merepresentasikan pertanyaan atau kondisi logis terkait atribut input, hingga akhirnya mengarah pada suatu keputusan klasifikasi. Setelah itu, kita melanjutkan ke proses prediksi untuk data uji.

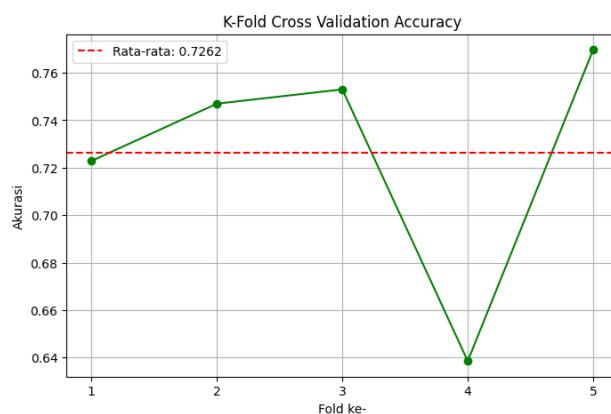
3.4 Evaluasi

Tahap evaluasi menjadi tahap akhir untuk mengetahui hasil dari klasifikasi yang telah dilakukan pada proses diatas.



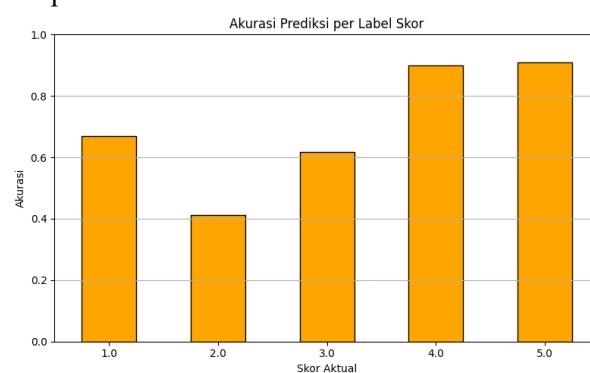
Gambar 9. Confusion Matrix Hasil Accuracy

Gambar 9 menunjukkan confusion matrix dari model Decision Tree dalam memprediksi skor kepuasan pengguna. Menurut jumlah prediksi benar yang dominan, model sangat akurat dalam memprediksi skor tinggi (4 dan 5), tetapi akurasi menurun pada skor 1 hingga 3, di mana banyak prediksi salah ke skor yang sama.



Gambar 10. K-Fold Cross Validation

Hasil evaluasi model menggunakan metode K-Fold Cross Validation sebanyak lima putaran ditunjukkan pada grafik gambar 10. Terlihat bahwa akurasi model bervariasi di setiap putaran, dengan nilai tertinggi sekitar 0,77 dan nilai terendah sekitar 0,64. Akurasi rata-rata dari kelima putaran adalah 0,7262, yang ditunjukkan oleh garis merah putus-putus. Hasil ini menunjukkan bahwa model itu cukup stabil untuk memprediksi tingkat kepuasan pengguna secara umum, meskipun ada variasi.



Gambar 11. Grafik akurasi perlabel

Akurasi prediksi model untuk setiap label skor kepuasan pengguna ditunjukkan pada grafik gambar 11. Terlihat bahwa model sangat akurat dalam memprediksi skor 4 dan 5 (di atas 85%), menunjukkan bahwa model dapat mengidentifikasi ulasan pengguna yang puas dengan baik. Sebaliknya, akuratnya menurun pada skor 2 dan 3, khususnya pada skor 2 yang hanya sekitar 41%, menunjukkan bahwa model masih kesulitan membedakan ulasan dengan nada netral atau kurang puas. Secara umum, model lebih dapat diandalkan untuk mengidentifikasi pengguna yang sangat puas dengan aplikasi.

IV. KESIMPULAN

Dengan menggunakan algoritma Decision Tree untuk mengklasifikasikan tingkat kepuasan pengguna aplikasi Kopi Kenangan, penelitian ini menunjukkan bahwa metode ini cukup akurat untuk mengidentifikasi kecenderungan kepuasan pengguna. Terbaik menurut model adalah prediksi kepuasan tinggi (4 dan 5), yang menunjukkan bahwa sebagian besar pengguna puas dengan layanan aplikasi. Bidang teknik industri, khususnya yang terkait dengan pengembangan layanan digital berbasis data, mendapatkan metode klasifikasi alternatif untuk memahami persepsi pengguna secara otomatis. Hasil ini juga mendukung upaya untuk meningkatkan layanan dan proses pengambilan keputusan yang didasarkan pada evaluasi pengguna yang objektif.

UCAPAN TERIMA KASIH

Dengan penuh rasa syukur, kami mengucapkan terima kasih kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya yang memungkinkan kami menyelesaikan penelitian ini dengan baik. Ucapan terima kasih yang sama kami sampaikan kepada dosen pembimbing kami, yang dengan sabar memberikan arahan dan bimbingan. Kami menghargai setiap pihak yang telah berkontribusi, baik secara langsung maupun tidak langsung dalam kelancaran penelitian ini. Kami juga mengucapkan terima kasih kepada para pengguna aplikasi Kopi Kenangan, karena ulasan mereka menjadi sumber data utama dalam penelitian ini. Semoga hasil dari penelitian ini dapat memberikan manfaat, khususnya dalam pengembangan layanan digital di masa yang akan datang.

DAFTAR PUSTAKA

- [1] M. A. Setyawan, “Strategi kopi kenangan dalam membangun loyalitas pelanggan,” p. 129, 2021, [Online]. Available: <https://repository.uinjkt.ac.id/dspace/handle/123456789/56471>
- [2] P. T. Bumi and B. Boga, “PENGARUH BUDAYA ORGANISASI TERHADAP RETENSI KARYAWAN PADA THE INFLUENCE OF ORGANIZATIONAL CULTURE ON EMPLOYEE RETENTION AT PT . BUMI BERKAH BOGA,” vol. 1, no. 2, pp. 40–44, 2025.
- [3] F. I. Komputer and U. D. Nuswantoro, “Menggunakan Metode Decision Tree Pada,”

- vol. 7, no. 6, pp. 3830–3833, 2010.
- [4] D. S. Al Azkiah, E. Erizal, and F. N. Hikmah, “Perbandingan Algoritma SVM dan Decision Tree Dalam Klasifikasi Kepuasan Pengguna Aplikasi Migo E-Bike di Playstore,” *Build. Informatics, Technol. Sci.*, vol. 6, no. 1, pp. 158–167, 2024, doi: 10.47065/bits.v6i1.5344.
- [5] D. Fitriono, S. A. Wardani, M. N. B. Al, A. Ristyawan, and E. Daniati, “Perbandingan Metode Algoritma Decission Tree dan K-Nearest Neighbors untuk Memprediksi Kualitas Air yang dapat dikonsumsi,” vol. 8, pp. 475–484, 2024.
- [6] E. Herdika *et al.*, “Klasifikasi Risiko Kambuhnya Kanker Tiroid Menggunakan Algoritma Random Forest,” vol. 8, pp. 63–74, 2024.
- [7] Ihsan Zulfahmi, “Analisis Sentimen Aplikasi PLN Mobile Menggunakan Metode Decission Tree,” *J. Penelit. Rumpun Ilmu Tek.*, vol. 3, no. 1, pp. 11–21, 2023, doi: 10.55606/juprit.v3i1.3096.
- [8] M. R. Muzaki, I. M. Nurfajriana, P. A. Ilahi, A. Ristyawan, and E. Daniati, “Implementasi Data Mining dengan Algoritma K- Nearest Neighbors untuk Memprediksi Risiko Diabetes Menggunakan Chatbot Telegram,” vol. 8, pp. 310–318, 2024.
- [9] S. Bahri and A. Lubis, “Metode Klasifikasi Decision Tree Untuk Memprediksi Juara English Premier League,” *J. Sintaksis*, vol. 2, no. 1, pp. 63–70, 2020.
- [10] F. K. Wardhani, N. Kamilatutsaniya, A. Alamsyah, E. Daniati, and A. Ristyawan, “Perbandingan Algoritma Naive Bayes Dan Desicion Tree Dalam Pengujian Data Anemia Menggunakan,” *Inotek*, vol. 8, 2024, [Online]. Available: <https://proceeding.unpkediri.ac.id/index.php/inotek/>
- [11] A. Tangkelayuk, “The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 1109–1119, 2022, doi: 10.35957/jatisi.v9i2.2048.
- [12] K. Amarya, A. C. A. G, R. Achmad, E. Daniati, and A. Ristyawan, “Analisa Perbandingan Algoritma Classification Berdasarkan Komposisi Label,” vol. 8, pp. 32–40, 2024.
- [13] D. Putriani, A. P. A. Prayogi, A. I. Shofyana, A. Ristyawan, and E. Daniati, “Prediksi Customer Churn Menggunakan Algoritma Decision Tree,” *Inotek*, vol. 8, pp. 85–94, 2024.
- [14] I. Romli, “Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Klasifikasi Penyakit Ispa,” *Indones. J. Bus. Intell.*, vol. 4, no. 1, p. 10, 2021, doi: 10.21927/ijubi.v4i1.1727.
- [15] F. Djiwadikusumah, G. H. Irawan, and R. Haekal Al-Fadilah, “Web scraping situs e-commerce menggunakan teknik parsing dom,” *J. Siliwangi*, vol. 7, no. 2, pp. 52–57, 2021, [Online]. Available: <https://jurnal.unsil.ac.id/index.php/jssainstek/article/view/4223/1958>
- [16] K. Sentimen *et al.*, “Klasifikasi Sentimen Pengguna Aplikasi Livin ByMandiri Pada Playstore Menggunakan Algoritma Naive Bayes,” *Agustus*, vol. 8, pp. 2549–7952, 2024.
- [17] H. T. Santoso, F. A. Felmidi, A. Nur, A. Ristyawan, and E. Daniati, “Analisis Kinerja Algoritma Data Mining pada Klasifikasi Tingkat Obesitas dengan K-Fold Cross Validation dan AUC,” vol. 8, pp. 113–122, 2024.