

# Klasifikasi Jenis Produk Sepatu Terlaris di TokoPedia Menggunakan Prediksi K-Nearest Neighbor

Azzah Luthfiatussundus<sup>\*1</sup>, Bayu Priyatna<sup>2</sup>, Baenil Huda<sup>3</sup>

<sup>1,2,3</sup>Sistem Informasi, Teknik Ilmu Komputer, Universitas Buana Perjuangan Karawang

E-mail: <sup>\*1</sup>[si19.azzahluthfiatussundus@mhs.ubpkarawang.ac.id](mailto:si19.azzahluthfiatussundus@mhs.ubpkarawang.ac.id),

<sup>2</sup>[bayu.priyatna@ubpkarawang.ac.id](mailto:bayu.priyatna@ubpkarawang.ac.id), <sup>3</sup>[baenil88@ubpkarawang.ac.id](mailto:baenil88@ubpkarawang.ac.id)

## Abstrak

*Sepatu yaitu jenis alas kaki yang melindungi seluruh bagian kaki, namun jenis sepatu ada banyak sekali, tergantung dari tujuan pemakainya. Dengan banyak model sepatu, salah satu toko penjual sepatu menghadapi beberapa masalah. Diantaranya yaitu pada era digitalisasi ini, banyak toko-toko lain yang sudah melakukan penjualan online di situs web atau sering disebut dengan e-commerce salah satunya TokoPedia, jika penjual masih melakukan penjualan secara offline di era digitalisasi ini maka semakin banyak pesaing toko sepatu yang sudah melakukan pembukaan toko di internet. Sebelum toko sepatu ingin melakukan penjualan secara online perlu dilakukan prediksi jenis sepatu yang diminati oleh pelanggan agar toko dapat bersaing dengan toko lain yang berada di TokoPedia. Algoritma prediksi penelitian ini adalah algoritma K-NN. Proses perhitungan dan pengukuran akurasi role menggunakan aplikasi Orange Data Mining dan Microsoft Excel. Penelitian ini menggunakan model jenis sepatu yang diminati dengan algoritma prediksi KNN dengan nilai presisi 0,531. Jenis sepatu yang diminati yaitu sepatu Sneakers Pria (1333), ini merupakan hasil dari total data set sebanyak 5165 data*

**Kata Kunci**—K-NN, Sepatu, TokoPedia, Orange Data Mining, E-commerce

## 1. PENDAHULUAN

Sepatu adalah alas kaki yang bersifat universal. Sepatu bisa digunakan hanya untuk pemanis penampilan kita saat olahraga, acara resmi, atau saat berpergian[1]. Sepatu juga merupakan sepatu yang melindungi seluruh bagian kaki, mulai dari alas kaki hingga mata kaki. Jenis sepatu ada bermacam-macam, tergantung dari tujuan si pengguna sepatu itu sendiri. Seiring perkembangan teknologi yang sangat kompleks dan berdampak banyak pada kehidupan manusia. ini termasuk dunia bisnis internet, yang disebut e-commerce, dan segala metode belanja atau perdagangan online menggunakan internet dengan situs web[2]. Dengan adanya perkembangan teknologi ini yang berdampak pada pebisnis yang mulai beralih ke toko online atau disebut dengan e-commerce untuk mengenalkan produk pada masyarakat luas[3]. Toko yang ingin memulai berniaga di e-commerce TokoPedia dapat menggunakan prediksi penjualan untuk memprediksi produk yang diminati oleh pembeli dan membuat keputusan produksi yang tepat[4]. Bahkan di era digitalisasi, masih banyak penjual sepatu yang menjual produknya di toko fisik dan mengandalkan hafalan pelanggan, voting, dan data penjualan sebelumnya. Semakin banyak pesaing toko bermunculan maka mengharuskan merencanakan strategi penjualan[5], e-commerce dapat digunakan oleh penjual untuk memperluas pemasaran mereka, tetapi mereka juga membutuhkan prediksi penjualan untuk melihat berapa banyak pelanggan potensial yang ada dan produk apa yang disukai pelanggan. Proses pengumpulan data penting dalam jumlah besar, atau “data besar”, adalah dikenal sebagai “data mining”[6]. Proses ini sering menggunakan sejumlah metode, seperti penggunaan matematika, statistik, dan teknologi kecerdasan buatan. Selain itu ada juga metode k-nearest neighbor (KNN). Algoritma klasifikasi K-NN langsung membagi objek menjadi label. Karena mengklasifikasikan berdasarkan seberapa dekat dengan objek lain

(tetangga), metode ini lebih mudah dipahami daripada metode lainnya. Algoritma ini hanya menyimpan dan mengklasifikasikan sampel data. Algoritma k-nearest neighbor memiliki keunggulan dalam menghasilkan data yang kuat atau terdefinisi dengan baik, dan bekerja dengan baik bila diterapkan pada data yang cukup besar. Dengan menggunakan prediksi dengan metode k-nearest neighbor dapat membantu memperbaiki performa toko untuk menjadi lebih baik dan dapat bersaing dengan toko lain.

## 2. METODE PENELITIAN

### 2.1. Data mining

Data mining, yang bahkan dapat disebut sebagai siklus untuk mengikuti, adalah metode yang paling umum untuk memisahkan data menjadi informasi yang berguna yang dapat digunakan untuk memangkas biaya dan meningkatkan keuntungan[7]. Proses pemecahan data menjadi informasi yang berguna berdasarkan berbagai evaluasi atau perspektif dikenal sebagai data mining. Koneksi atau contoh ratusan atau ribuan bidang dari kumpulan data[8]. Pengumpulan data dan pemanfaatan data historis untuk mengidentifikasi keteraturan dan hubungan dalam kumpulan data besar disebut knowledge discovery in a database atau KDD[9].

### 2.2. K-Nearest Neighbor

Sebuah metode yang dikenal sebagai K-Nearest Neighbor (K-NN) menggunakan algoritma yang diawasi yang sebagian besar kategori pada K-NN digunakan untuk mengklasifikasikan. Algoritma K-Nearest Neighbor (K-NN) memiliki keunggulan menghasilkan dalam memberikan hasil yang akurat sesuai perhitungan aplikasi[10]. Metode K-Nearest Neighbor dapat dihitung dengan cara berikut :

1. Tentukan parameter k
2. Menentukan jarak antara semua data pelatihan dan data yang akan dievaluasi. Gunakan rumus Euclidean Distance untuk menentukan jarak.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

3. Urutkan jarak yang dihasilkan dan pilih jarak yang terpendek ke-k
4. Pilih kelas yang tepat
5. Mencari tetangga terdekat kelas sebagai kelas data untuk menentukan nomor dan mengaturnya.

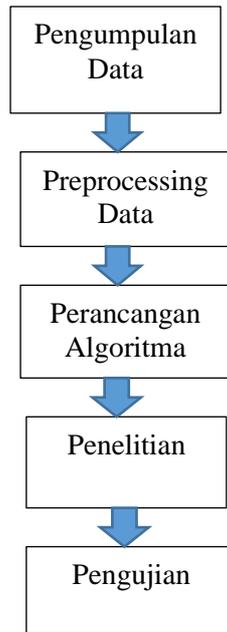
### 2.3. Orange Data Mining

Orange data mining adalah perangkat lunak sumber terbuka untuk memproses, menganalisis, dan data mining[11]. Tidak seperti program data mining informasi lainnya, orange menekankan representasi, dan bahkan dapat disebut sebagai program visual[12]. Orange menyediakan banyak widget yang bisa diletakan di kanvas atau papan gambar dan dihubungkan dengan widget lainnya. Media kanvas ini memudahkan pengguna untuk memanipulasi data dan menangani analisis data secara intuitif. Setiap widget yang ada memiliki fungsinya masing masing dan dapat menerima masukan dan keluaran.

2.4. Tahapan penelitian

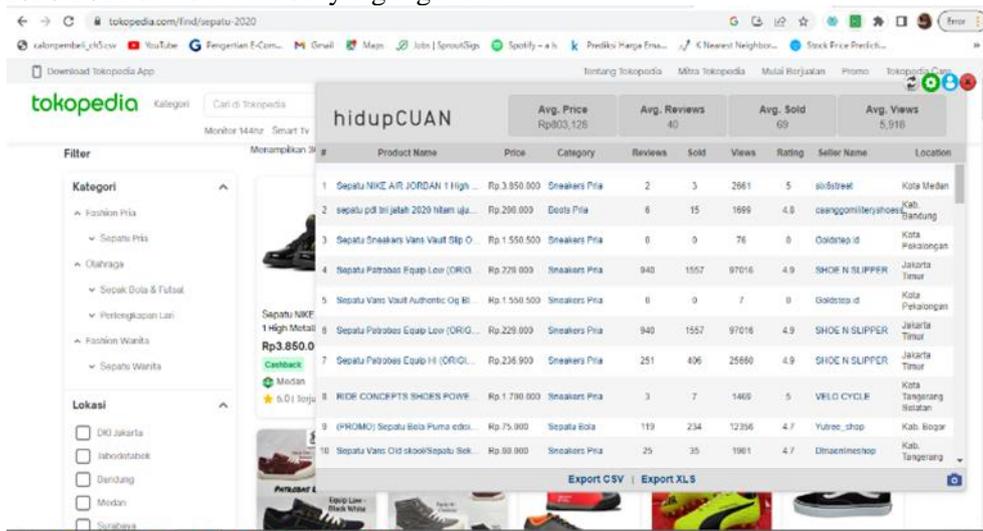
Penelitian ini dilakukan untuk memulai pengumpulan data. Data preprocessing dilakukan setelah data terkumpul. Orange Data Mining kemudian mengolah data tersebut menggunakan algoritma K-Nearest Neighbor untuk menghasilkan nilai akurasi klasifikasi[13]. Berikut tahapan yang dilakukan dalam penelitian ini:

Tabel 1. Tahapan



2.4.1. Pengumpulan data

Produk sepatu TokoPedia digunakan untuk pengumpulan data dalam penelitian ini [13]. data set yang digunakan dalam penelitian ini diperoleh dari plugin web scraping situs belanja online TokoPedia. Data yang digunakan berjumlah sebesar 5165 data training produk sepatu dari tahun 2020-2022 sumber data set yang digunakan.



Gambar 1. plugin web scrapping situs belanja online

#### 2.4.2. *Preprocessing data*

Sebelum menggunakan data dalam metode atau teknik data mining, beberapa langkah harus diselesaikan, termasuk pengolahan data. Pencarian data dengan missing value merupakan bagian dari tahap preprocessing. Artinya data yang diterima tidak normal. Fase ini membuat label/kelas dan mengubah atribut string menjadi angka. Pada fase pra-pemrosesan data, atribut diberi peran target dari atribut dengan memilih beberapa atribut dari kumpulan data. Maka setelah ditentukan processing, metode K-NN digunakan untuk menentukan prosedur klasifikasi, dan digunakan nilai parameter  $K=5$ ,  $K=7$ ,  $K=9$ ,  $K=10$ ,  $K=3$ , dan  $K=4$ . Hasilnya dikumpulkan untuk memastikan seberapa baik klasifikasi dan pengukuran dilakukan, khususnya dengan presisi, recall, dan akurasi.

#### 2.4.3. *Perancangan algoritma*

Tahapan desain arsitektur sistem pemodelan adalah dimana struktur data, arsitektur perangkat lunak, tampilan antarmuka, dan algoritma dirancang. Tujuannya yaitu untuk mendapatkan gambaran yang lebih baik tentang apa yang diperlukan. Untuk menentukan KNN, algoritma K-NN menggunakan jarak terpendek antara query instance dengan data training. Metode Euclidean distance merupakan salah satu metode untuk menentukan jarak tetangga. Jarak sering dihitung dengan menggunakan jarak euclidean. Ukuran yang dapat digunakan sebagai interpretasi kedekatan jarak antara dua objek diuji dengan menggunakan jarak Euclidean.

Semakin besar hasil rumus diatas, maka semakin besar kemiripan antara dua objek, dan sebaliknya, semakin kecil hasil ini, semakin dekat kemiripan antara objek. Objek yang dimaksud adalah data training dan data testing. Nilai k terbaik untuk algoritma ini ditentukan oleh jumlah data. Namun, nilai k yang terbesar tidak selalu merupakan nilai k yang terbaik.

Validasi dengan cross validation dengan jumlah (k) tes yaitu 2,5,7,10,20 . Proses pengujian cross-validation dijalankan di orange data mining[14], pemodelan dilakukan menggunakan metode K-NN, selanjutnya memasukan apply model dan performa. Evaluasi akurasi prediksi model dikenal sebagai validasi. Teknik validasi secara acak membagi data menjadi beberapa bagian yang disebut validasi silang. Validasi silang melakukan sejumlah konvolusi atau pembagian data saja.confusion matrix dan model akurasi dihasilkan melalui prosedur validasi silang.

#### 2.4.4. *Penelitian*

Sebelum menerapkan proses implementasi data pada orange data mining, pembagian dataset yang dibuat menjadi data pengujian dan data pelatihan. Data training yaitu informasi histori penjualan dengan hasil prediksi. Data testing adalah data dalam bentuk informasi yang belum diperoleh hasil prediksi dari 5165 dataset, dibagi menjadi dua bagian, yaitu 5150 data untuk data training dan 15 data untuk data testing. Alasan mengapa hanya 15 data yang digunakan untuk pengujian data karena proses pengujian menggunakan k-fold dengan nilai k 20, kemudian data uji target hasil prediksi tidak dimasukkan , sehingga belum ada hasil prediksi karena dibangun di aplikasi orange data miningdengan data traininguntuk menguji sistem prediksi.berikut adalah hasil pengujian data pada orange data mining.

#### 2.4.5. *Gambar dan Tabel Pengujian*

Algoritma K-Nearest Neighbor diuji dalam penelitian ini. Untuk menjamin kinerja algoritma K-NN, digunakan kumpulan data untuk menguji akurasi, presisi, dan recall algoritma. Algoritma k-nearest neighbor digunakan untuk menghitung akurasi nilai-nilai dalam tahapan ini. Performansi yang diperoleh

dari metode K-NN membuktikan keakuratan model algoritma. Pada pengujian data training sebanyak 5150 data.

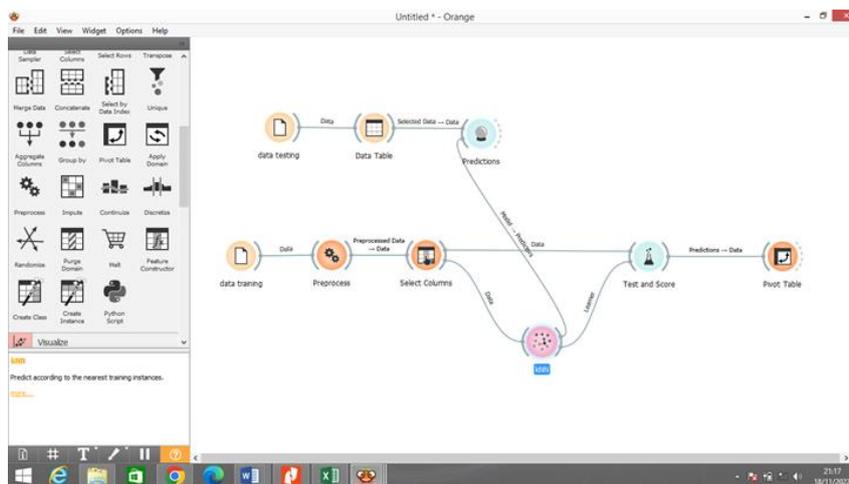
### 3. HASIL DAN PEMBAHASAN

#### 3.1. Penerapan Orange Data Mining

Dalam ulasan ini, 5150 data persiapan dan 15 data pengujian untuk menyelesaikan pengujian menggunakan persetujuan silang digunakan untuk mengelompokkan jenis sepatu di TokoPedia, penulis melakukan preprocessing sebelum melakukan validasi silang untuk menjamin kualitas data sebelum digunakan untuk analisis data[15].

Dalam penelitian ini dilakukan untuk mengetahui jenis sepatu yang diminati oleh pelanggan dengan menggunakan metode k-nearest Neighbor yang dilakukan dengan menggunakan orange data mining dengan melakukan beberapa widget yang digunakan file yang berisi data testing dan data training, preprocessing untuk menghilangkan missing value, select column, KNN, prediction, test and score, pivot table.

#### 3.2. Pengujian



Gambar 2. model k-nearest neighbor

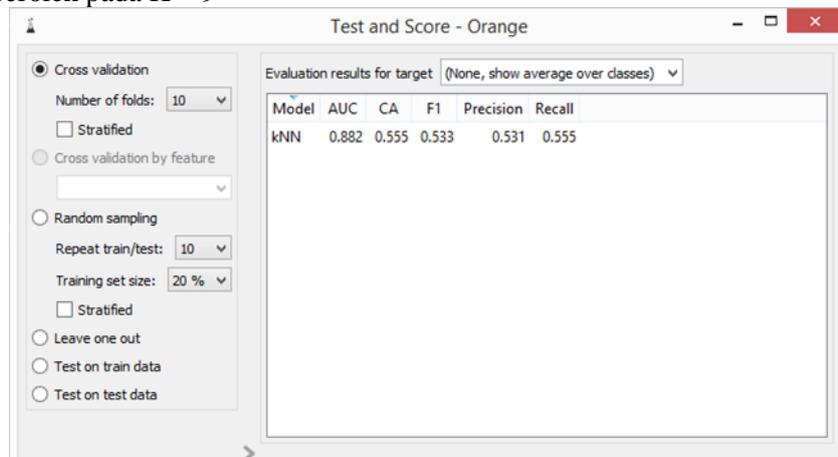
Berdasarkan hasil dataset 5150 algoritma K-Nearest Neighbor. 15 data uji disajikan sebagai kumpulan data latih, dengan klasifikasi jenis sepatu TokoPedia sebagai dasar. Klasifikasi data dengan nilai parameter K= 5, K= 4, K= 8, K= 9 serta validasi silang k-flod 10 dengan tabel dibawah ini digunakan untuk menentukan akurasi metode K-Nearest Neighbor,

Tabel 2. Perbandingan akurasi

No	K	Accuracy	Precision	Recall
1	5	0.547	0.529	0.547
2	4	0.539	0.522	0.539
4	8	0.550	0.527	0.550
5	9	0.555	0.531	0.555

Hasil pengujian setiap parameter K1-K10 menunjukkan bahwa nilai akurasi tertinggi diperoleh pada K = 9 dengan nilai akurasi 0,555, nilai presisi 0,531, dan nilai recall 0,555. Nilai terendah pada K = 4 dengan nilai akurasi 0,539, nilai presisi 0,522, dan nilai recall 0,539. Seperti

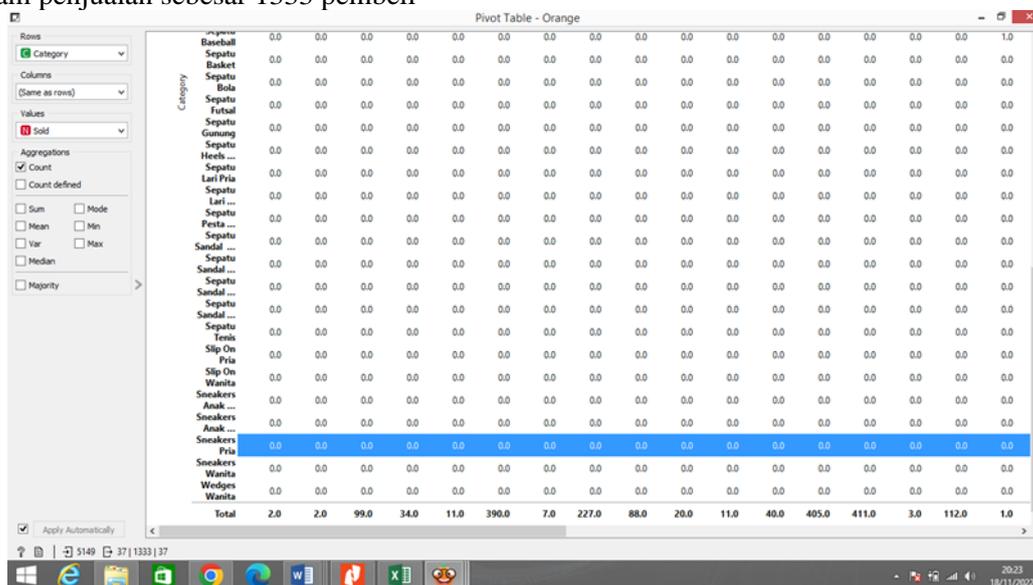
yang terlihat, hasil pengujian untuk setiap parameter K1-K10 menunjukkan bahwa nilai akurasi tertinggi diperoleh pada K = 9



Gambar 3. hasil test and score

Dari gambar dapat dilihat bahwa hasil akurasi dari k-nearest neighbor yaitu 0,555. Setelah diketahui hasil nilai k-nearest neighbor dengan nilai presisi, maka dapat dilihat jenis sepatu yang paling diminati berdasarkan algoritma k-nn sebagai berikut jenis sepatu yang paling diminati pelanggan

Dari gambar diatas menunjukkan sepatu sneakers pria lah yang memiliki angka tertinggi dalam penjualan sebesar 1333 pembeli



Gambar 4. hasil penjualan

#### 4. KESIMPULAN

Berdasarkan penelitian penulis terhadap sistem prediksi penjualan sepatu TokoPedia, maka penulis dapat menarik kesimpulan sebagai berikut :

1. Telah berhasil dibuat dataset dari data penjualan yang diperoleh dari tokoPedia dengan total data 5165 data untuk data training menggunakan 5150 data dan untuk data testing menggunakan 15 data secara acak
2. Dari hasil algoritma k-nn dengan nilai presisi 0,531
3. Prediksi penjualan sepatu di tokoPedia dengan aplikasi orange data mining dilakukan menggunakan 2 variabel yaitu kategori sepatu dan produk terjual sehingga diperoleh jenis sepatu yang diminati.
4. Hasil akurasi knn dengan 20 folds cross validation mencapai 0,555
5. Dari training data 5150 diperoleh jenis sepatu yang paling diminati adalah sepatu sneaker pria dengan jumlah yang terjual mencapai 1333

#### DAFTAR PUSTAKA

- [1] A. Madya, "Laporan tugas akhir foto sepatu beazt dalam fotografi still life," 2022.
- [2] N. K. Marchyta, "Antecedents of Online Experience on E-Commerce Platform in Indonesia," *J. Manaj. Pemasar.*, vol. 16, no. 1, pp. 32–36, 2022, doi: 10.9744/pemasaran.16.1.32-36.
- [3] L. E. Saputri and A. Utomo, "Pengaruh Brand Image, Kualitas Produk, Dan Harga Terhadap Keputusan Pembelian Sepatu Converse Di Surakarta," *Excellent*, vol. 8, no. 1, pp. 92–103, 2021, doi: 10.36587/exc.v8i1.911.
- [4] B. Hardiyanto and F. Rozi, "PREDIKSI PENJUALAN SEPATU MENGGUNAKAN METODE K-NEAREST NEIGHBOR."
- [5] S. Diwandari and A. T. Hidayat, "Predicting Analysis of User's Interest from Web Log Data in e-Commerce using Classification Algorithms," *J. Ilmu Komput. dan Inf. (Journal Comput. Sci. Inf.*, vol. 1, no. 33, pp. 33–38, 2022, [Online]. Available: <http://dx.doi.org/10.21609/jiki.v15i1.1024>
- [6] S. S. Hilabi *et al.*, "TechnoXplore Jurnal Ilmu Komputer & Teknologi Informasi ISSN : 2503-054X Vol 4 No: 1, April 2019," *J. Ilmu Komput. Teknol. Inf.*, vol. 4, no. 1, pp. 28–37, 2019.
- [7] L. Husna, S. Syahputra, and B. S. Ginting, "Penerapan data mining menggunakan metode K-Means cluster untuk pengelompokkan data perizinan Madrasah Diniyah Taklimiyah Awwaliyah (MDTA) studi kasus Kementerian Agama Stabat," *J. Inform. Kaputama*, vol. 6, no. 3, 2022.
- [8] R. Fitra and I. Rusdi, "Penerapan Metode Algoritma K-Nearest Neighbor Menggunakan Rapidminer Studio Pada Klasifikasi Status Sosial Ekonomi Studi Kasus: Kelurahan Kapuk Muara Rt ...," *Smart Comp Jurnalnya Orang Pint. ...*, pp. 653–660, 2022, [Online]. Available: <http://ejournal.poltektegal.ac.id/index.php/smartcomp/article/view/4250>
- [9] H. Gunawan and V. Purwayoga, "Data Mining Menggunakan Algoritma K-Means Clustering Untuk Mengetahui Potensi Penyebaran Virus Corona Di Kota Cirebon," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 1, pp. 1–8, 2022, doi: 10.32736/sisfokom.v11i1.1316.
- [10] H. Putri, A. I. Purnamasari, A. R. Dikananda, O. Nurdiawan, and S. Anwar, "Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera Menggunakan Metode NAÏVE BAYES dan KNN," *Build. Informatics, Technol. Sci.*, vol. 3, no. 3, pp. 331–337, 2021, doi: 10.47065/bits.v3i3.1093.
- [11] P. Val, I. Framework, U. Mengevaluasi, and K. Sistem, "Jurnal saintesa," vol. 2, pp. 1–7,

- 2022.
- [12] H. Hozairi, A. Anwari, and S. Alim, "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes," *Netw. Eng. Res. Oper.*, vol. 6, no. 2, p. 133, 2021, doi: 10.21107/nero.v6i2.237.
  - [13] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
  - [14] Aji Priyambodo and Prihati Prihati, "Evaluasi Ekstraksi Fitur Klasifikasi Teks Untuk Peningkatan Akurasi Klasifikasi Menggunakan Naive Bayes," *Elkom J. Elektron. dan Komput.*, vol. 13, no. 1, pp. 159–175, 2020, doi: 10.51903/elkom.v13i1.277.
  - [15] S. P. Dewi, N. Nurwati, and E. Rahayu, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor," *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.