



Application of the K-Nearest Neighbor Algorithm for Predicting Digital Product Sales

Penerapan Algoritma K-Nearest Neighbor untuk Prediksi Penjualan Produk Digital

Bayu Susilo^{1*}, Nur Ariesanto Ramdhan², Otong Saeful Bachri³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Teknik,
Universitas Muhadi Setiabudi Brebes, Indonesia

E-Mail: ¹bayususilo2311@gmail.com, ²ariesantoramdhan@gmail.com,
³otongsaifulbahriumus@gmail.com

Received Jul 5th 2024; Revised Aug 24th 2024; Accepted Sept 10th 2024
Corresponding Author: Bayu Susilo

Abstract

This research aims to predict the best-selling digital products at PT. Global Indo Multimedia uses the K-Nearest Neighbor (K-NN) method with the Rapidminer application. And also digital product transaction data for 6 months, from August 2023 to January 2024, is used for analysis. The Knowledge Discovery in Databases (KDD) process is applied in data mining of digital product transaction data for 6 months from August 2023 to January 2024 where based on the KDD stages it is then applied to data mining using the K-Nearest Neighbor method with a value of $K = 3$ indicating that the algorithm K-Nearest Neighbor can predict the number of digital product transactions for the next 2 months. Based on the trials carried out, it shows that the K-NN algorithm can be used in classifying digital product transaction data at PT. Global Indo Multimedia and Based on testing, the accuracy results were 95.24%, 3 True Selling Products, 17 True Loss Products, and 1 True Medium Product. This means that the data set can later be used at the next stage as valid data for use. With this high level of accuracy, the K-NN model can be a solution for predicting digital product transactions based on pre-existing sales data.

Keywords: Digital Product Prediction, Data Mining, K-Nearest Neighbor (K-NN), Rapidminer

Abstrak

Penelitian ini bertujuan untuk memprediksi produk digital terlaris di PT. Global Indo Multimedia menggunakan metode K-Nearest Neighbor (K-NN) dengan aplikasi Rapidminer. Dan juga data transaksi produk digital selama 6 bulan, dari Agustus 2023 hingga Januari 2024, digunakan untuk analisis. Proses Knowledge Discovery in Databases (KDD) diterapkan dalam data mining data transaksi produk digital selama 6 bulan dari bulan Agustus 2023 sampai bulan Januari 2024 dimana berdasarkan tahapan KDD lalu diterapkan pada data mining menggunakan metode K-Nearest Neighbor dengan nilai $K = 3$ menunjukkan bahwa algoritma K-Nearest Neighbor dapat memprediksi jumlah transaksi produk digital untuk 2 bulan yang akan datang. Berdasarkan uji coba yang dilakukan menunjukkan bahwa algoritma K-NN dapat digunakan dalam klasifikasi data transaksi produk digital pada PT. Global Indo Multimedia dan Berdasarkan pengujian didapatkan hasil akurasi sebesar 95,24%, 3 Produk True Laris, 17 Produk True Rugi, dan 1 Produk True Sedang. Yang artinya data set nantinya dapat digunakan pada tahapan selanjutnya sebagai data yang valid untuk digunakan. Dengan tingkat akurasi yang cukup tinggi ini, model K-NN dapat menjadi salah satu solusi untuk memprediksi transaksi produk digital berdasarkan data penjualan yang telah ada sebelumnya.

Kata Kunci: Data Mining, K-Nearest Neighbor (K-NN), Prediksi Produk Digital, Rapidminer

1. PENDAHULUAN

Peramalan (*forecasting*) merupakan salah satu metode yang dapat membantu memperkirakan jumlah permintaan barang di masa periode yang akan datang. Secara definisi peramalan merupakan suatu proses memprediksi secara sistematis mengenai kemungkinan apa yang terjadi dimasa yang akan datang dengan berdasarkan informasi data yang telah didapatkan dimasa lalu ataupun sekarang bertujuan agar menurunkan kesalahan yang akan dialami [1]. Dalam dunia bisnis yang selalu dinamis dan penuh persaingan, para pelaku

UMKM, Konter, Toko, dan lainnya harus senantiasa memikirkan cara-cara untuk terus survive dan jika mungkin mengembangkan skala bisnis mereka [2]. Karena pada era digitalisasi yang pesat ini banyak masyarakat maupun khalayak umum sudah semakin cerdas dalam medayagunakan kemudahan ketika berinteraksi maupun bertransaksi [3] di sisi lain forecasting juga membantu perusahaan dalam melakukan perencanaan penyediaan stok. misalnya prediksi stok yang merupakan data di masa lalu yang digunakan untuk keperluan estimasi data yang akan datang [4] karena prediksi ini memberikan output terbaik bagi perusahaan yang mana sehingga dapat meminimalisir kesalahan perencanaan serta dapat ditekankan seminimal mungkin. Peramalan memiliki bagian dan peranan penting dalam menentukan sebuah perencanaan dan mengambil keputusan dengan baik dalam penjualan. Dalam dunia bisnis khususnya dalam hal usaha yang berhubungan dengan produksi sangat penting untuk memperkirakan hal-hal yang akan terjadi di masa mendatang yang dilakukan dengan menghimpun data-data masa lampau, kemudian diolah menggunakan metode peramalan sebagai acuan dalam pengambilan keputusan yang tepat [5] PT.

Global Indo Multimedia merupakan usaha dagang yang bergerak dibidang pembayaran elektronik. Perusahaan ini beralamat di Jalan Randu Gede, Kecamatan Kersana, Kabupaten Brebes. Produk yang dijual oleh PT. Global Indo Multimedia yaitu: Penjualan Pulsa, *E-Wallet*, *Top Up Game* dan Token Listrik yang mana semuanya menggunakan *host to host system*. Apalagi sekarang bertambahnya jumlah pelanggan jaringan seluler dari waktu ke waktu, tidak dapat dipungkiri bahwa hal ini juga berdampak pada peningkatan bisnis jual beli pulsa seluler. Layanan Pulsa merupakan salah satu jenis layanan yang memungkinkan pembelian pulsa telepon genggam dan paket data internet untuk pengguna telepon pintar atau smartphone[6]. Penelitian ini di latar belakang oleh kemajuan perkembangan teknologi komunikasi dan informasi yang sangat pesat dan semakin murah sehingga membuat masyarakat untuk pulsa telepon selular menjadi kebutuhan yang wajib pada masa kini. Dari bermacam produk pulsa yang ada pada PT. Global Indo Multimedia seperti Telkomsel, Pulsa XL, Pulsa Indosat, Pulsa 3, dan lainnya [7].

Strategi peramalan merupakan salah satu strategi yang cukup signifikan untuk meningkatkan keuntungan bisnis. Peramalan transaksi penjualan dapat membantu para pengambil keputusan dalam suatu bisnis untuk menentukan strategi penjualan yang akan diterapkan[8]. Dalam hal ini PT. Global Indo Multimedia memerlukan strategi peramalan supaya bisa menaikkan jumlah penjualan produk digital. Permasalahan utama yang dihadapi dalam penelitian ini adalah adanya kesulitan yang dialami oleh PT. Global Indo Multimedia dalam menentukan produk digital yang paling diminati oleh konsumen. Hal ini terkadang juga disebabkan oleh tingginya volume transaksi yang harus diperiksa secara manual, dan tidak adanya sistem otomatis yang dapat memberikan informasi secara akurat tentang produk terlaris. Seiring dengan peningkatan jumlah permintaan dan keragaman produk yang dijual (pulsa, e-wallet, top up game, token listrik). perusahaan belum memiliki metode yang efektif untuk mengidentifikasi tren produk terlaris berdasarkan data historis penjualan. Kondisi ini menghambat perusahaan dalam melakukan perencanaan stok yang optimal dan mengurangi kesalahan perencanaan penjualan.

Terdapat beberapa penelitian terdahulu diantaranya yaitu Analisis Penjualan Produk Paket Kuota Internet Dengan Metode K- Nearest Neighbor. Penelitian ini dilakukan oleh Dedi Handoko, Heru Satria Tambunan, dan Jaya Tata hardinata pada tahun 2020. Penelitian ini pemodelan menggunakan algoritma k-nearest neighbor dengan menggunakan data yang diolah berdasarkan tahapan knowledge discovery in database (KDD). Berdasarkan hasil perhitungan data mining menggunakan teknik klasifikasi dan algoritma k-nearest neighbor, didapatkan hasil prediksi penjualan kartu paket kuota internet dengan 4 jenis paket internet dan data yang digunakan pada tahun 2017 – 2019 yang dimana atribut terdiri dari Nama Produk, Kuant, dan Bulan kemudian berdasarkan nilai akurasi terhadap prediksi penjualan produk untuk tahun 2020 yaitu sebesar 71,43% dan telah diuji menggunakan tools Rapid Miner 5.0 [9]. Perbedaan penelitian yang akan dilakukan adalah subjek penelitian ini dilakukan di PT. Akses Lintas Nusantara sedangkan penelitian ynung dilakukan penulis di PT. Global Indo Multimedia

Penerapan Data Mining Pada Jumlah Penduduk Miskin di Indonesia. Penelitian yang dilakukan oleh Sudibyo pada tahun 2020 membahas tentang Aplikasi yang menggali informasi jumlah penduduk miskin di Indonesia dengan menggunakan data tahun 2006 hingga Maret 2020. Clustering dilakukan dengan metode k-means kemudian dihasilkan peramalan jumlah penduduk. miskin. Hasil analisis menunjukkan metode k-means menghasilkan dua cluster. Berdasarkan proyeksi dengan metode double exponential smoothing dengan nilai Alpha sebesar 1,094 dan nilai Gamma sebesar 0,539, diperkirakan terdapat sekitar 27,4405 juta penduduk miskin di Indonesia pada periode tersebut [10]. Perbedaan dengan penelitian ini adalah menggunakan metode K-means sedangkan penelitian dibuat oleh penulis menggunakan metode K-Nearest Neighbor.

Penelitian berjudul "Implementasi Data Mining untuk Prediksi Peramalan Penjualan Produk Hj Karpel Menggunakan Metode Linear Regression" yang dilakukan oleh Muhamad Reza Athallah dan Anief Fauzan Rozi pada tahun 2022 menghasilkan peramalan penjualan menggunakan regresi linear. Hasil perhitungan tersebut digunakan untuk memberikan rekomendasi jumlah stok barang yang perlu disediakan pada bulan Januari 2023, yaitu 52 stok sajadah Iranshar. Angka 52 ini diperoleh dari pembulatan nilai y, sementara angka 55 diperoleh dengan menambahkan perkiraan kesalahan sebesar 5,667205% dari 52, yang menghasilkan nilai 2,9522, kemudian dibulatkan menjadi 3 dan ditambahkan pada nilai rekomendasi dari perhitungan regresi

linear menjadi 55 [11]. Perbedaannya terletak pada metode yang digunakan; penelitian ini menggunakan metode K-means, sedangkan penelitian yang dilakukan oleh penulis menggunakan metode K-Nearest Neighbor.

Penelitian berjudul Penerapan Algoritma K-Nearest Neighbor untuk Klasifikasi Jurusan Siswa di SMA Negeri 15 Pekanbaru oleh Qurotul A'yuniyah dan Muhammad Reza pada tahun 2023 menjelaskan bahwa dataset yang digunakan dalam penelitian ini terdiri dari 216 data siswa yang diperoleh dari SMA Negeri 15 Pekanbaru, yang kemudian dibagi menjadi 192 data training dan 24 data testing. Klasifikasi dilakukan menggunakan algoritma K-Nearest Neighbor (K-NN) dengan bantuan tools RapidMiner untuk memprediksi jurusan siswa berdasarkan data testing. Dari hasil pengujian dengan berbagai nilai k (k=3, k=7, k=8, k=9, dan k=10), didapatkan prediksi sebanyak 12 siswa masuk dalam kategori jurusan IPA dan 12 siswa dalam kategori IPS, berdasarkan jarak terdekat. Validasi terhadap prediksi menunjukkan bahwa algoritma K-NN memiliki tingkat akurasi yang tinggi, mencapai 93,52% [12]. Perbedaannya adalah penelitian ini dilakukan di SMA Negeri 15 Pekanbaru, sementara penelitian yang dilakukan oleh penulis dilakukan di PT. Global Indo Multimedia.

Penelitian berjudul Prediksi Penjualan Obat dan Alat Kesehatan Terlaris Menggunakan Algoritma K-Nearest Neighbor oleh Abdul Azis, Ahmad Turmudi Zy, dan Aswan S. Sunge menjelaskan bahwa prediksi penjualan obat dan alat kesehatan terlaris di Rumah Sakit As-Shofwan mencakup beberapa produk seperti Klip Obat Besar 15 x 10 cm, Klip Obat Uk. 13x87, Alkohol Swab, Kertas Puyer, Omeprazole 20 mg Nulab, Curcuma Sanbe Tablet, Ranitidine Inj, Antasida Tab, Cefixime 200 mg Tab Nulab, Betahistine 6 mg Tab, Ondan 4 mg Inj Interbat, Handscoon Latex Uk. M, dan Handscoon Latex Uk. L. Penelitian ini menggunakan Jupyter Notebook untuk melakukan prediksi penjualan dengan algoritma K-Nearest Neighbor. Hasil pengolahan dataset menunjukkan bahwa klasifikasi penjualan obat dan alat kesehatan terlaris dengan algoritma ini mencapai nilai akurasi, recall, dan precision masing-masing sebesar 1.0, menunjukkan performa yang sangat tinggi [13].

Proses transaksi di PT. Global Indo Multimedia saat ini menggunakan sistem host to host, namun belum ada sistem yang mampu mengidentifikasi produk yang paling diminati oleh konsumen, serta belum tersedia metode yang efektif untuk menentukan produk terlaris. Hal ini menjadi tantangan terutama karena adanya peningkatan penjualan produk akibat permintaan konsumen yang sangat beragam. Dengan adanya ratusan transaksi produk yang terjual, PT. Global Indo Multimedia mengalami kesulitan dalam memantau setiap transaksi secara manual. Oleh karena itu, diperlukan metode yang tepat untuk menentukan produk terlaris berdasarkan data transaksi dari Agustus 2023 hingga Januari 2024. Masalah utama yang dihadapi PT. Global Indo Multimedia adalah kesulitan dalam menentukan produk yang paling diminati konsumen. Sebagai solusi, diperlukan penerapan metode K-Nearest Neighbor untuk memprediksi penjualan produk digital di PT. Global Indo Multimedia guna mengetahui produk apa saja yang paling diminati oleh pelanggan.

2. TEORI PENDUKUNG DAN METODOLOGI

2.1. Digital Payment

Digital payment merupakan salah satu bentuk inovasi baru *financial technology* yang memberikan layanan baru bagi masyarakat mengenai transaksi pembayaran non-tunai yang praktis dan efisien serta dapat dilakukan hanya dengan melalui ponsel tanpa batasan waktu dan tempat. *Digital payment* lebih dikenal sebagai dompet *digital* atau uang ponsel dan juga merupakan pembayaran yang dilakukan menggunakan informasi *digital* melalui alat pembayaran elektronik [14] Digital Payments juga merupakan layanan elektronik untuk menyimpan data instrumen pembayaran antara lain: alat pembayaran dengan menggunakan kartu dan/atau uang elektronik, yang dapat juga menampung dana, untuk melakukan pembayaran. Saat ini, ada lima penerbit uang elektronik (electronic money) yang populer di Indonesia sebagai alat pembayaran nontunai yang dapat dimanfaatkan di berbagai sektor usaha, antara lain Gopay, OVO, Dana, LinkAja, ShopeePay, dan lainnya [15].

2.2. Host to Host System

Host to Host atau sering dikenal H2H adalah sistem antar server yang saling terhubung satu sama lain lebih sederhananya ialah komunikasi atau hubungan di dalam sebuah jaringan komputer yang terjadi antar host, yaitu komputer dengan perangkat lain yang terhubung satu sama lain. Teknologi internet semakin maju dalam perkembangan sistem distributed computing dengan munculnya *web services* [16]. Dengan arti lain suatu komunikasi antara komputer satu dengan lainnya yang saling terhubung yang saling terhubung satu sama lain dalam sebuah sistem pembayaran dimana menunjang konektivitas secara langsung dan secara *real time* serta otomatis tanpa memerlukan intervensi *manual*. Koneksi host-to-host sering juga disebut dengan komunikasi end-to-end, yaitu suatu bentuk komunikasi dalam jaringan komputer yang terjadi antar host, yaitu komputer dan perangkat lain yang saling terhubung [17] Web services juga merupakan salah satu paradigma baru dalam mengimplementasikan sistem terdistribusi melalui web yang menggunakan basis teknologi XML dan JSON, dengan standar protokol HTTP dan SOAP.

2.3. Definisi Data

Data merupakan komponen dasar dari informasi yang akan diproses lebih lanjut untuk menghasilkan informasi. Data adalah suatu istilah majemuk yang berarti fakta atau bagian dari fakta yang mengandung arti yang digabungkan dengan kenyataan, simbol-simbol, gambar-gambar, angka-angka, huruf-huruf, atau simbol-simbol yang menunjukkan suatu ide, objek, kondisi, atau situasi dan lain-lain. Informasi tersembunyi dalam sebuah basis data dan merupakan bagian dari *Knowledge Discovery in Databases (KDD)* [18] Untuk menemukan informasi dan pola yang berguna dalam data. Data merupakan suatu kejadian yang menggambarkan kenyataan yang terjadi yang dimasukkan dalam elemen input dan kemudian akan diproses menjadi output sehingga menghasilkan protokol komunikasi tertentu yang sering kita kenal dengan istilah Internet Protocol (IP) serta Transmission Control Protocol (TCP) [19].

2.4. Prediksi/ Peramalan

Prediksi (*Forecasting*) merupakan alat bantu yang penting dalam perancangan yang efektif dan efisien. Peramalan adalah prediksi, proyeksi atau estimasi tingkat kejadian yang tidak pasti dimasa yang akan datang. Prediksi juga dapat digunakan dalam pengklasifikasian, tidak hanya untuk memprediksi *time series*, karena sifatnya yang bisa menghasilkan *class* berdasarkan atribut yang ada [20]. Peramalan (*forecasting*) juga bisa dikatakan sebuah langkah awal terpenting pada perencanaan untuk sebuah organisasi bisnis dalam mengambil keputusan manajemen yang signifikan. Metode peramalan Single Eksponensial Smoothing yaitu suatu model ramalan pada data berkala (*time series*) yang dirancang sebagai data untuk membuat sebuah unsur trend [21].

2.5. Klasifikasi

Klasifikasi adalah sebuah proses untuk menemukan model yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang kelasnya tidak diketahui [22]. Di dalam klasifikasi diberikan sejumlah *record* yang dinamakan *training set*, yang terdiri dari beberapa atribut, atribut dapat berupa *continue* dan juga kategoris, salah satu atribut menunjukkan kelas untuk *record*. Klasifikasi adalah pengelompokan yang teratur dari pada sejumlah obyek, gagasan tersebut yang sama [23].

2.6. K-Nearest Neighbour (K-NN)

K-Nearest Neighbor (K-NN) adalah metode klasifikasi terhadap objek baru berdasarkan data training yang memiliki jarak tetangga terdekat (*nearest neighbor*) dengan objek baru tersebut. Algoritma K-NN juga merupakan metode yang digunakan untuk melakukan klasifikasi data berdasarkan jarak terdekat terhadap suatu objek data [24]. Dekat atau jauhnya *neighbor* biasanya dihitung berdasarkan jarak Euclidean [25] Berikut ini adalah langkah-langkah algoritma K-NN:

1. Tentukan nilai k (nilai tetangga terdekat)
2. Hitung jarak Euclidean kuadrat (contoh kueri) untuk setiap fitur dalam data pelatihan yang dihasilkan Pengurutan objek-objek data ke dalam kelompok yang memiliki jumlah jarak euclidean paling kecil.

$$RMSE = \sqrt{\sum \frac{(Y' - Y)^2}{n}} \quad (1)$$

Keterangan :

Y' : Nilai Prediksi

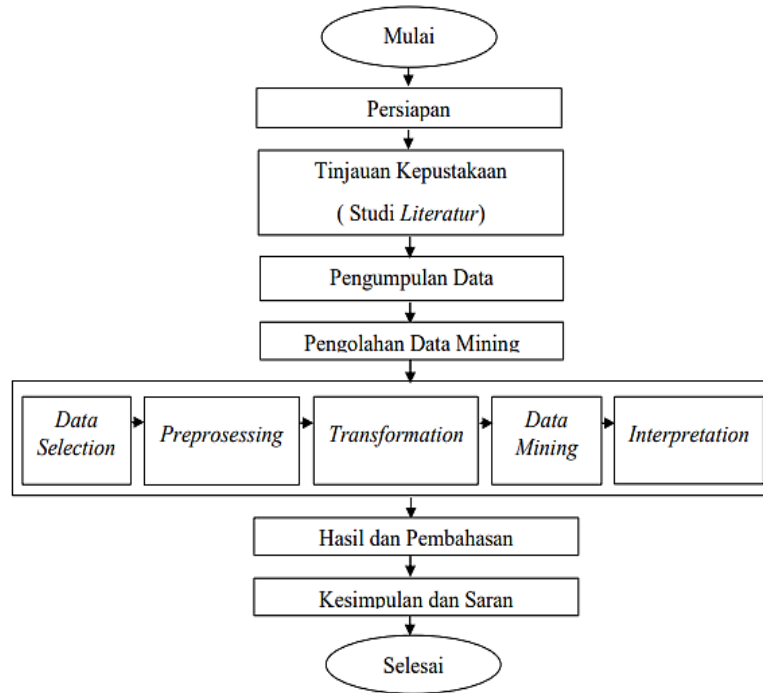
Y : Nilai Target

n : Jumlah Data

3. Mengambil label kelas y (klasifikasi K-NN)
4. Dengan menggunakan operator K-NN yang paling sesuai untuk digunakan, dimungkinkan untuk memprediksi nilai versi kueri yang dihitung

2.7. Metodologi Penelitian

Metode pada penelitian ini menggunakan pendekatan deskriptif analitis yang mana penelitian deskriptif analitis adalah suatu metode yang berfungsi untuk mendeskripsikan atau memberi gambaran terhadap objek yang diteliti melalui data ataupun sampel yang telah terkumpul sebagaimana adanya tanpa melakukan analisis dan membuat kesimpulan yang berlaku untuk umum. Dengan kata lain penelitian deskriptif analitis mengambil masalah atau memusatkan perhatian kepada masalah-masalah sebagaimana adanya saat penelitian ini dilakukan, hasil penelitian kemudian diolah dan dianalisis untuk diambil kesimpulannya. Lokasi penelitian adalah di PT. Global Indo Multimedia tepatnya di Jalan Randu Gede Desa Kemukten, Kecamatan. Kersana, Kabupaten Brebes. Tahapan penelitian merupakan tahap yang akan dijelaskan dalam melakukan penelitian. Metodologi penelitian dapat dilihat pada gambar 1.



Gambar 1. Tahapan Penelitian

Pengumpulan data yang dilakukan dalam pembuatan *data mining* prediksi penjualan pada penelitian ini dilakukan beberapa langkah-langkah yaitu: Studi literatur digunakan sebagai metode pengumpulan data dan informasi melalui *internet*, buku, jurnal, dan situs *website*. Selanjutnya Wawancara dilakukan untuk mengetahui lebih detail mengenai data melalui narasumber. Wawancara dilakukan langsung kepada karyawan yang bersangkutan guna untuk memprediksi penjualan produk di masa mendatang. Kemudian Observasi dilakukan langsung dengan mendatangi lokasi guna mengetahui data transaksi yang terjual dari bulan Agustus sampai Januari 2024 untuk dijadikan bahan penelitian, lamanya observasi dilakukan selama 2 bulan.

Pada tahapan ini proses pengumpulan data kebutuhan yang digunakan dalam implementasi *data mining* prediksi penjualan produk terlaris kode mitra LS31 di PT. Global Indo Multimedia, perangkat lunak apa saja yang digunakan yaitu Laptop: *Lenovo N3350*, *Processor AMD A4*, *RAM 4 GB*, *System Type 64-bit Operating Sistem*, *Hardisk 500 GB*, *Windows 2010 64-bit* sebagai sistem operasi, Aplikasi *Rapidminer* digunakan sebagai aplikasi *data mining*, *Microsoft excel* sebagai *input data manual*.

Adapun pengolahan data mining yang dilakukan pada penelitian ini yaitu, mengikuti tahapan dalam *Knowledge Discovery in Database (KDD)*, untuk menghasilkan informasi yang sesuai dengan urutan yang sudah ditentukan, berikut ini adalah tahapan-tahapannya adalah (1) Pemilihan (seleksi) data dari sekumpulan data operasional yang perlu dilakukan sebelum tahap penggalian informasi dalam *KDD* dimulai. Dan data hasil seleksi akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional. Data yang digunakan dalam penelitian ini berasal dari PT. Global Indo Multi Media, yaitu data transaksi penjualan produk dari bulan agustus 2023 sampai dengan bulan januari 2024. (2) Seluruh atribut dalam data penjualan, termasuk *field* seperti kode produk, nama produk, kuantitas, jumlah, dan bulan, akan diseleksi. Dari semua data tersebut, hanya tiga field yang akan dipilih dan digunakan dalam proses *Knowledge Discovery in Databases (KDD)*.

3. HASIL DAN PEMBAHASAN

3.1 *Data Selection*

Penggunaan data harian penjualan di rekap menjadi data bulanan, karena data yang akan dipakai dalam proses data mining adalah data di setiap bulannya dari bulan Agustus 2023 sampai Januari 2024. Data penjualan pun dikelompokkan menjadi dua kategori yaitu Laris dan Tidak Laris, dengan ketentuan penjualan yang lebih dari sama dengan 56 itu masuk kedalam kategori laris. Seleksi data dapat dilihat pada tabel 1.

Tabel 2. *Data Selection*

No.	Nama Produk	Kode Produk	Bulan	Kuantitas
1.	DANA 10 K	DANA10	Agustus	0
2.	DANA 100 K	DANA100	Agustus	3
3.	DANA 125 K	DANA125	Agustus	0
4.	DANA 20 K	DANA20	Agustus	1

No.	Nama Produk	Kode Produk	Bulan	Kuantitas
5.	DANA 200	DANA200	Agustus	2
....
51.	OVO 500 K	OVO500	Januari	0

3.2 Pre-processing/ Data Cleaning

Dalam tahapan ini tidak ada penggabungan atau pembersihan data karena data tidak memiliki *missing value*. Tahap persiapan atau pembersihan data ini dilakukan sebelum data tersebut digunakan dalam algoritma *K-NN* untuk klasifikasi atau regresi. Karena *K-NN* sangat bergantung pada jarak antara data, penting untuk memastikan bahwa data dalam format yang sesuai dan tidak mengandung outlier atau noise yang berlebihan yang bisa mempengaruhi hasil akhir. Sebuah produk dikatakan laris apabila lebih dari atau sama dengan 56 transaksi, dikatakan sedang apabila lebih dari atau sama dengan 28 transaksi, dan dikatakan rugi apabila kurang dari 28 transaksi. Pra proses data dapat dilihat pada tabel 2.

Tabel 2. Pre-processing

No.	Nama Produk	Agustus	September	Oktober	November	Desember	Januari	Keterangan
1.	DANA10	0	3	0	0	0	0	RUGI
2.	DANA100	3	0	0	0	2	1	RUGI
3.	DANA125	0	0	0	0	1	14	RUGI
4.	DANA20	1	0	0	0	0	0	RUGI
5.	DANA200	2	0	0	0	1	1	RUGI
6.	DANA50	2	0	0	0	0	0	RUGI
7.	GOJEK10	3	0	0	0	0	0	RUGI
8.	GOJEK100	5	7	11	0	0	0	RUGI
9.	GOJEK15	3	1	2	0	0	0	RUGI
10.	GOJEK20	4	1	2	0	0	0	RUGI
11.	GOJEK250	1	0	2	0	0	0	RUGI
12.	GOJEK50	4	6	9	0	0	0	RUGI
13.	IH20	0	0	23	2	22	28	LARIS
14.	LA100	0	0	0	0	0	1	RUGI
15.	LA40	0	1	0	0	0	0	RUGI
16.	LA50	0	1	0	0	0	0	RUGI
17.	LA90	0	0	0	0	0	1	RUGI
18.	MAXC100	0	2	0	0	0	0	RUGI
19.	MAXIM20	1	0	0	0	2	0	RUGI
20.	OVO100	3	12	19	1	24	9	LARIS
21.	OVO200	6	7	3	0	5	1	RUGI
22.	OVO50	7	22	15	1	7	4	RUGI
23.	PLNFULL100	10	31	16	0	32	33	LARIS
24.	PLNFULL20	77	153	156	2	105	117	LARIS
25.	PLNFULL200	7	9	15	1	12	21	LARIS
26.	PLNFULL50	100	216	235	7	110	162	LARIS
27.	PLNFULL500	1	0	5	0	2	5	RUGI
28.	SP100	4	134	5	0	5	5	LARIS
29.	SP25	0	4	2	0	4	2	RUGI
30.	SP5	4	3	10	1	3	9	SEDANG
31.	SP50	0	0	0	0	1	1	RUGI
32.	TH5	0	0	0	1	11	9	RUGI
33.	IH10	0	48	0	0	0	0	SEDANG
34.	IH100	2	1	0	0	0	0	RUGI
35.	LA10	0	1	0	0	0	0	RUGI
36.	SS10	0	152	0	0	0	0	LARIS
37.	SP10	5	4	9	0	108	0	LARIS
38.	SHD85	0	1	0	0	0	0	RUGI
39.	SHD55	1	0	0	0	0	0	RUGI
40.	SHD50	0	2	0	0	0	0	RUGI
41.	SHD30	0	1	0	0	0	0	RUGI
42.	SHD15	0	0	1	0	0	0	RUGI
43.	SHD100	0	1	3	0	0	0	RUGI
44.	PX10	0	285	0	0	0	0	LARIS
45.	OVO20	1	0	4	0	6	0	RUGI
46.	OVO300	0	1	1	0	0	0	RUGI
47.	OVO25	0	4	0	0	1	0	RUGI
48.	IH50	2	0	3	0	0	0	RUGI
49.	MAXIM30	1	0	0	0	0	0	RUGI
50.	OVO40	1	0	0	0	0	0	RUGI

No.	Nama Produk	Agustus	September	Oktober	November	Desember	Januari	Keterangan
51.	OVO500	0	0	1	0	0	0	RUGI

Keterangan:

DANA10	: DANA 10 K	PLNFULL200	: TOKEN LISTRIK 200K
DANA100	: DANA 100 K	PLNFULL50	: TOKEN LISTRIK 50 K
DANA125	: DANA 125	PLNFULL500	: TOKEN LISTRIK 500 K
DANA20	: DANA 20 K	SP100	: PULSA TSEL 100 K
DANA200	: DANA 200 K	SP25	: PULSA TSEL 25 K
DANA50	: DANA 50 K	SP5	: PULSA TSEL 5 K
GOJEK10	: GOJEK 10 K	SP50	: PULSA TSEL 50 K
GOJEK100	: GOJEK 100 K	TH5	: PULSA THREE 5 K
GOJEK15	: GOJEK 15 K	IH10	: PULSA INDOSAT 10 K
GOJEK20	: GOJEK 20 K	IH100	: PULSA INDOSAT 100 K
GOJEK250	: GOJEK 250 K	LA10	: LINK AJA 10 K
GOJEK50	: GOJEK 50 K	SS10	: PULSA TSEL PROMO 10 K
IH20	: PULSA INDOSAT 20 K	SP10	: PULSA TSEL 10 K
LA100	: LINK AJA 100 K	SHD85	: SHOPEE PAY 85 K
LA40	: LINK AJA 40 K	SHD55	: SHOPEE PAY 55 K
LA50	: LINK AJA 50 K	SHD50	: SHOPEE PAY 50 K
LA90	: LINK AJA 90 K	SHD30	: SHOPEE PAY 30 K
MAXC100	: MAXIM CUSTOMER 100 K	SHD15	: SHOPEE PAY 15 K
MAXIM20	: MAXIM DRIVER 20 K	SHD100	: SHOPEE PAY 100 K
OVO100	: OVO 100 K	PX10	: PULSA XL AXIS 10 K
OVO200	: OVO 200 K	OVO20	: OVO 20 K
OVO50	: OVO 50 K	PLNFULL100	: TOKEN LISTRIK 100 K
PLNFULL20	: TOKEN LISTRIK 20 K		

3.3 Transformation

Tahapan *Transformation* dimana data yang dipilih diubah dengan melakukan ringkasan (agregasi). Sehingga data tersebut dapat digunakan untuk data mining. tahap ini merupakan proses mengubah atau memodifikasi data dari bentuk aslinya menjadi bentuk lain yang lebih sesuai untuk analisis atau pemodelan. Proses ini dilakukan agar data lebih mudah digunakan, lebih informatif, atau supaya algoritma tertentu dapat beroperasi dengan lebih baik. Transformasi bisa mencakup berbagai jenis perubahan, tergantung pada tujuan spesifik dari analisis atau model. Dalam hal ini penulis melakukan pengujian dengan 60 % data latih dan 40% data uji. Transformasi data training dan data testing dapat dilihat pada tabel 3 dan tabel 4.

Tabel 3. Data Training

Produk	Agustus	September	Oktober	November	Desember	Januari	Keterangan
DANA10	0	3	0	0	0	0	RUGI
DANA100	3	0	0	0	2	1	RUGI
DANA125	0	0	0	0	1	14	RUGI
DANA20	1	0	0	0	0	0	RUGI
DANA200	2	0	0	0	1	1	RUGI
DANA50	2	0	0	0	0	0	RUGI
GOJEK10	3	0	0	0	0	0	RUGI
GOJEK100	5	7	11	0	0	0	RUGI
GOJEK15	3	1	2	0	0	0	RUGI
GOJEK20	4	1	2	0	0	0	RUGI
GOJEK250	1	0	2	0	0	0	RUGI
GOJEK50	4	6	9	0	0	0	RUGI
IH20	0	0	23	2	22	28	LARIS
LA100	0	0	0	0	0	1	RUGI
LA40	0	1	0	0	0	0	RUGI
LA50	0	1	0	0	0	0	RUGI
LA90	0	0	0	0	0	1	RUGI
MAXC100	0	2	0	0	0	0	RUGI
MAXIM20	1	0	0	0	2	0	RUGI
OVO100	3	12	19	1	24	9	LARIS
OVO200	6	7	3	0	5	1	RUGI
OVO50	7	22	15	1	7	4	LARIS
PLNFULL100	10	31	16	0	32	33	LARIS
PLNFULL20	77	153	156	2	105	117	LARIS
PLNFULL200	7	9	15	1	12	21	LARIS
PLNFULL50	100	216	235	7	110	162	LARIS

Produk	Agustus	September	Oktober	November	Desember	Januari	Keterangan
PLNFULL500	1	0	5	0	2	5	RUGI
SP100	4	134	5	0	5	5	LARIS
SP25	0	4	2	0	4	2	RUGI
SP5	4	3	10	1	3	9	SEDANG

Tabel 4. Data Testing

Produk	Agustus	September	Oktober	November	Desember	Januari	Keterangan
SP50	0	0	0	0	1	1	RUGI
TH5	0	0	0	1	11	9	RUGI
IH10	0	48	0	0	0	0	SEDANG
IH100	2	1	0	0	0	0	RUGI
LA10	0	1	0	0	0	0	RUGI
SS10	0	152	0	0	0	0	LARIS
SP10	5	4	9	0	108	0	LARIS
SHD85	0	1	0	0	0	0	RUGI
SHD55	1	0	0	0	0	0	RUGI
SHD50	0	2	0	0	0	0	RUGI
SHD30	0	1	0	0	0	0	RUGI
SHD15	0	0	1	0	0	0	RUGI
SHD100	0	1	3	0	0	0	RUGI
PX10	0	285	0	0	0	0	LARIS
OVO20	1	0	4	0	6	0	RUGI
OVO300	0	1	1	0	0	0	RUGI
OVO25	0	4	0	0	1	0	RUGI
IH50	2	0	3	0	0	0	RUGI
MAXIM30	1	0	0	0	0	0	RUGI
OVO40	1	0	0	0	0	0	RUGI
OVO500	0	0	1	0	0	0	RUGI

3.4 Pengujian Metode K-Nearest Neighbor

Dalam Pengujian ini penulis menggunakan *Confusion Matrix* setelah dilakukan perhitungan *Accuracy*, *Precision*, *Recall*, dan *Specificity*. *Confusion Matrix* digunakan untuk mengevaluasi kinerja model klasifikasi dalam pembelajaran mesin. Serta menunjukkan perbandingan antara prediksi yang dibuat oleh model dengan hasil sebenarnya, serta mengklasifikasikan hasilnya ke dalam empat kategori. Pengujian ini sangat berguna untuk melihat seberapa baik model mampu membedakan kelas-kelas di dalam data Berikut ini adalah hasil yang didapatkan dapat dilihat pada Tabel 5.

Tabel 5. Pengujian *Confusion Matrix*

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>Specificity</i>
95%	93,75 %	100%	100%

3.5 Pengujian Algoritma K-NN dengan Rapidminer

Pada tahapan ini terdiri dari beberapa operator diantaranya seperti *Read Excel*, *K-NN*, *Apply Model*, dan juga *Perfomance*. Pengujian ini dilakukan dengan menentukan nilai K adalah 3. Hasil dari pengujian confusion matriks dapat dilihat pada gambar 2.

accuracy: 95.24%

	true RUGI	true SEDANG	true LARIS	class precision
pred. RUGI	17	1	0	94.44%
pred. SEDANG	0	0	0	0.00%
pred. LARIS	0	0	3	100.00%
class recall	100.00%	0.00%	100.00%	

Gambar 2. Hasil *Accuracy*

3.6 Evaluasi Algoritma Rapidminer

Evaluasi perhitungan data akan menggunakan algoritma yang sudah dihitung otomatis oleh aplikasi *Rapidminer* berupa *Root Mean Square Error* (RMSE) yaitu penjumlahan kuadrat error atau selisih antara nilai sebenarnya dan nilai prediksi, kemudian membagi jumlah tersebut dengan banyaknya jumlah data testing dan

menghitung akarnya. Hasil nilai RMSE yang mendekati angka nol, maka semakin akurat nilai yang diprediksi yang mana mendekati dengan nilai aslinya. Nilai *RMSE* dari hasil proses dapat dilihat pada tabel 6.

Tabel 6. Nilai *RMSE* untuk semua produk *digital*

Produk	Jarak Euclidean Distance
DANA10	3,464101615
DANA100	5,099019514
DANA125	18,38477631
DANA20	2,000000000
DANA200	2,828427125
DANA50	2,449489743
GOJEK10	3,464101615
GOJEK100	25,25866188
GOJEK15	5,656854249
GOJEK20	6,92820323
GOJEK250	2,828427125
GOJEK50	20,59126028
IH20	83,85702117
LA100	1,414213562
LA40	2,000000000
LA50	2,000000000
LA90	1,414213562
MAXC100	2,449489743
MAXIM20	2,000000000
OVO100	73,91887445
OVO200	22,58317958
OVO50	60,99180273
PLNFULL100	133,0488632
PLNFULL20	669,0396102
PLNFULL200	69,61321714
PLNFULL50	913,9562353
PLNFULL500	12,80624847
SP100	202,0643462
SP25	11,40175425
SP5	31,27299154

3.6 Implementasi Sistem

Berikut ini adalah hasil dari pengujian sistem menggunakan pemrograman *PHP* :

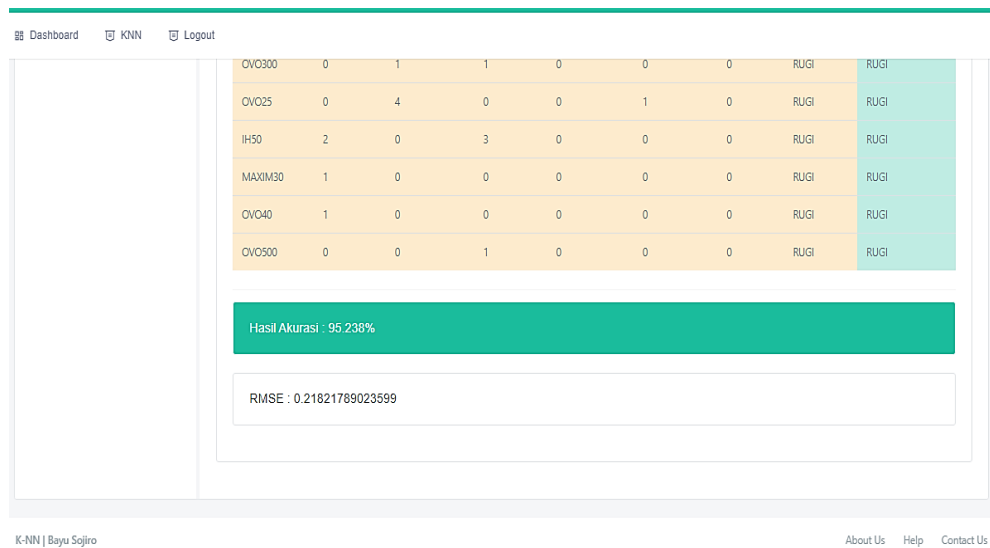
1. Hasil Prediksi

User mengklik prediksi lalu kemudian memasukkan kuantitas transaksi untuk memprediksi hasil jarak tetangga terdekat seperti pada Gambar 3.

Gambar 3. Tampilan Prediksi

2. Hasil *Perfomance*

User mengklik *Perfomance* lalu kemudian akan menampilkan halaman pengujian akurasi dan pilih akurasinya penulis memilih pengujian 60 % pada *option value* di halaman tersebut, tampilan hasil akurasi dapat dilihat pada gambar 4.



Gambar 4. Tampilan Hasil *Accuracy*

4. KESIMPULAN

Berdasarkan hasil dan pembahasan algoritma dalam penelitian ini, yang menggunakan aplikasi *RapidMiner* dan pemrograman PHP sederhana, dapat disimpulkan bahwa dengan memanfaatkan data transaksi produk digital selama 6 bulan, dari Agustus 2023 hingga Januari 2024, serta mengikuti tahapan KDD, dan menerapkan metode *K-Nearest Neighbor* dengan nilai $K = 3$, algoritma *K-Nearest Neighbor* dapat memprediksi jumlah transaksi produk digital untuk dua bulan ke depan. Berdasarkan uji coba yang dilakukan menunjukkan bahwa algoritma *K-NN* dapat digunakan dalam klasifikasi data transaksi produk digital pada PT. Global Indo Multimedia dan Berdasarkan pengujian didapatkan hasil akurasi sebesar 95,24% yang artinya data set dapat digunakan pada tahapan selanjutnya sebagai data yang valid untuk digunakan. Dengan tingkat akurasi yang cukup tinggi ini, model *K-NN* dapat menjadi salah satu solusi untuk memprediksi transaksi produk digital berdasarkan data penjualan yang telah ada sebelumnya. Dengan demikian, PT. Global Indo Multimedia dapat menggunakan model ini untuk membantu dalam pengambilan keputusan dan perencanaan strategi pemasaran yang lebih efektif. Meskipun algoritma *K-NN* telah berhasil menghasilkan akurasi yang cukup tinggi dalam klasifikasi data transaksi pada PT Global Indo Multimedia, namun agar memperoleh hasil yang lebih baik dari penelitian ini sebaiknya dilakukan penelitian lebih lanjut dengan menerapkan metode lainnya sehingga hasil yang didapatkan bisa lebih variatif dan dapat digunakan sebagai bahan perbandingan pada hasil yang didapatkan untuk hasil ini.

REFERENSI

- [1] S. A. Khusnul Wildan, "penentuan metode peramalan (forecasting) pada permintaan penjualan di cv. Lia tirta jaya prigen," vol. 2, no. 11, pp. 4077–4089, 2023.
- [2] A. Marlin, S. Ferneyanan, W. Latuwael, and I. S. Salampeppy, "analisis peramalan (forecasting) pendapatan warung putra," vol. 1, no. 3, pp. 56–60, 2024.
- [3] A. R. Hakim, "Eksistensi Transaksi Uang Digital Sebagai Alat Keuangan Modren Di Era Sekarang," no. 1, 2024.
- [4] F. M. Putri, "Jurnal Informatika Ekonomi Bisnis Tingkat Peramalan Penjualan Produk Bordir dan Sulaman Menggunakan Metode Trend Moment," vol. 4, pp. 34–38, 2022, doi: 10.37034/infv4i2.122.
- [5] F. K. Zega, T. Hartati, S. Hulu, S. Zebua, and E. Zebua, "Analisis Peramalan (Forecasting) Penjualan Tahu dengan Metode Single Moving Average untuk Mengoptimalkan Produksi pada Pabrik Tahu Nias," vol. 4, pp. 2931–2942, 2024.
- [6] N. Putu, L. Anastasya, and R. G. Artaninggrum, "analisis proses transaksi pulsa pada pt . lian media persada," vol. 6, pp. 523–532, 2023.
- [7] A. S. Wahyu Hadikristanto, "Penerapan Algoritma Naive Bayes Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan Pulsa," vol. 11, no. 2, pp. 115–122, 2020.
- [8] G. A. N. Pongdatu, E. Abinowi, and S. Wahyuddin, "Peramalan Transaksi Penjualan Dengan Metode

- Holt-Winter ' S Exponential Smoothing,” vol. 6, no. 3, pp. 1–6, 2020.
- [9] D. Handoko, H. Satria Tambunan, and J. T. Hardinata, “Analisis Penjualan Produk Paket Kuota Internet Dengan Metode K-Nearest Neighbor,” vol. 6, pp. 111–119, [Online]. Available: <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>
- [10] N. A. Sudibyoto *et al.*, “Penerapan Data Mining Pada Jumlah Penduduk,” vol. 1, no. 3, pp. 199–207, 2020.
- [11] J. Jurnal, S. Dan, T. Jsit, V. N. S. Hal, M. R. Athallah, and A. F. Rozi, “Implementasi Data Mining Untuk Prediksi Peramalan Penjualan Produk Hj Karpet Menggunakan Metode Linear Regression,” vol. 2, no. 3, pp. 180–187, 2022.
- [12] A. Qurotul and M. Reza, “Application Of The K-Nearest Neighbor Algorithm For Student Department Classification At 15 Pekanbaru State High School Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru,” vol. 3, no. 1, pp. 39–45, 2023.
- [13] V. No, J. Hal, A. K. Neighbor, A. Azis, A. Turmudi, and A. S. Sunge, “Prediksi Penjualan Obat Dan Alat Kesehatan Terlaris Menggunakan,” vol. 6, no. 1, pp. 117–124, 2024.
- [14] U. Tabanan, “Perkembangan Ekonomi Digital Terhadap Pertumbuhan Ekonomi Daerah Bali,” vol. 1, no. 4, 2023.
- [15] K. Samosir, T. Digital, and K. Samosir, “Pemanfaatan Pembayaran Digital Pada UMKM Di,” vol. 4, no. 1, pp. 768–776, 2023.
- [16] I. A. W. M. Arfa Andika Candra, “Implementasi Host To Host Bca Untuk Transaksi,” vol. 1, no. 3, pp. 866–870, 2023.
- [17] T. M. Sianturi and K. Ramli, “A Security Framework for Secure Host-to-Host Environments,” vol. 5, no. 158, pp. 380–386, 2022.
- [18] P. Mai, S. Tarigan, J. T. Hardinata, H. Qurniawan, M. Safii, and R. Winanjaya, “Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus : Toko Sinar Harahap),” vol. 12, no. 2, pp. 51–61, 2022.
- [19] M. A. A. C. (1) I. A. W. (2) Jurusan, “Sistem Informasi Berprestasi Berbasis Web Pada Smp Negeri 7 Kota Metro,” vol. 01, no. 01, 2021.
- [20] F. Teknik, U. Abdurrab, J. Riau, U. No, A. Hitam, and P. Sekaki, “Aplikasi Pendaftaran Mahasiswa Baru Menggunakan Metode Forecasting,” vol. 2, no. 1, 2022.
- [21] S. Forecasting, P. Produksi, D. Metode, and B. Indonesia, “NUSANTARA : Jurnal Ilmu Pengetahuan Sosial Single Eksponensial Smoothing Pada PT . FOOD,” vol. 9, no. 2, pp. 909–920, 2022.
- [22] H. Nurriqfi *et al.*, “Klasifikasi Penerima Bansos Menggunakan Algoritma Naive Bayes,” vol. 10, no. 1, pp. 683–695, 2023.
- [23] R. N. Ainum, N. Hidayat, and A. A. Soebroto, “Klasifikasi Buku Perpustakaan menggunakan Metode Naive Bayes,” vol. 6, no. 8, pp. 3726–3732, 2022.
- [24] K. N. N. Untuk, K. Seleksi, and P. Beasiswa, “Implementasi Algoritma Klasifikasi K-Nearest Neighbor,” vol. 6, no. 2, pp. 118–127, 2021.
- [25] dan M. T. C. Happy Andrian Dwi Fasnuari, Haris Yuana, “Penerapan Algoritma K-Nearest Neighbor (K-Nn) Untuk Klasifikasi Penyakit Diabetes Melitus Studi Kasus : Warga Desa Jatitengah,” vol. 16, no. 2, pp. 133–142, 2022.