



Prediksi Penjualan Barang Menggunakan Metode K-Means dan Regresi Linear

Henry Adam*, Tukino, Elfina Novalia, April Lia Hananto

Fakultas Ilmu Komputer, Sistem Informasi, Universitas Buana Perjuangan Karawang, Karawang, Indonesia

Email: ^{1*}si21.henryadam096@mhs.ubpkarawang.ac.id, ²tukino@ubpkarawangl.ac.id, ³elfinanovalia@ubpkarawang.ac.id,

⁴aprilialia@ubpkarawang.ac.id

Email Penulis Korespondensi: si21.henryadam096@mhs.ubpkarawang.ac.id

Abstrak—Analisis data penjualan memainkan peran penting dalam mendukung pengambilan keputusan bisnis, terutama untuk mengoptimalkan manajemen stok dan meningkatkan efisiensi operasional. permasalahan utama yang dihadapi oleh Vapestore XYZ di Karawang adalah kesulitan dalam memprediksi jumlah penjualan produk secara akurat, sehingga sering terjadi ketidakseimbangan antara persediaan dan permintaan pasar. Hal ini dapat menyebabkan kerugian akibat kelebihan stok maupun kekurangan barang. Saat ini, estimasi kebutuhan stok masih mengandalkan intuisi dan pengalaman pribadi, tanpa dukungan analisis data yang objektif. Penelitian ini bertujuan untuk membangun model prediksi penjualan dengan mengombinasikan metode *K-Means* untuk klusterisasi produk dan Regresi Linear untuk prediksi kuantitas penjualan. Data penjualan diambil langsung dari aplikasi POS toko, kemudian melalui tahapan pembersihan, pelabelan, dan klusterisasi menjadi tiga kelompok, yaitu “Kurang Laku”, “Laku”, dan “Sangat Laku”. Prediksi penjualan dilakukan menggunakan Regresi Linear dengan memanfaatkan hasil klusterisasi dan variabel waktu sebagai input. Evaluasi performa model dilakukan menggunakan metrik *error*, yaitu *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, dan *Root Mean Squared Error (RMSE)*. Berdasarkan hasil pengujian, model Regresi Linear yang dikembangkan memperoleh MAE sebesar 3,20, MSE sebesar 52,34, dan RMSE sebesar 7,23. Nilai-nilai *error* ini menunjukkan bahwa model mampu memberikan estimasi penjualan yang cukup mendekati data aktual, sehingga dapat diandalkan dalam perencanaan stok. Visualisasi hasil prediksi dalam bentuk tabel dan heatmap memudahkan identifikasi tren penjualan dan perbandingan performa antar produk. Temuan penelitian ini membuktikan bahwa kombinasi metode *K-Means* dan Regresi Linear efektif digunakan untuk mendukung pengambilan keputusan stok dan strategi pemasaran pada toko ritel vape. Pengembangan lebih lanjut disarankan dengan memperkaya dataset dan mengeksplorasi metode prediksi lain guna meningkatkan performa model.

Kata Kunci: Prediksi Penjualan; *K-Means*; Regresi Linear; *Data Mining*; Evaluasi Model

Abstract—Sales data analysis plays an important role in supporting business decision making, especially to optimise stock management and improve operational efficiency. the main problem faced by Vapestore XYZ in Karawang is the difficulty in accurately predicting the number of product sales, so there is often an imbalance between inventory and market demand. This can cause losses due to overstocks or shortages of goods. Currently, the estimation of stock requirements still relies on intuition and personal experience, without the support of objective data analysis. This research aims to build a sales prediction model by combining the *K-Means* method for product clustering and Linear Regression for sales quantity prediction. Sales data is taken directly from the store POS application, then goes through the stages of cleaning, labelling, and clustering into three groups, namely ‘Less Sold’, ‘Sold’, and ‘Very Sold’. Sales prediction is performed using Linear Regression by utilising the clustering results and time variables as inputs. Model performance evaluation is performed using error metrics, namely Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Based on the test results, the developed Linear Regression model obtained MAE of 3.20, MSE of 52.34, and RMSE of 7.23. These error values indicate that the model is able to provide sales estimates that are close enough to the actual data to be reliable in stock planning. Visualisation of the prediction results in the form of tables and heatmaps makes it easy to identify sales trends and compare performance between products. The findings of this study prove that the combination of *K-Means* and Linear Regression methods is effectively used to support stock decision making and marketing strategies in vape retail stores. Further development is recommended by enriching the dataset and exploring other prediction methods to improve model performance.

Keywords: Sales Prediction; *K-Means*; Linear Regression; *Data Mining*; Model Evaluation

1. PENDAHULUAN

Kemajuan teknologi informasi telah membawa perubahan besar di berbagai sektor, termasuk dunia bisnis dan perdagangan. Penggunaan teknologi dalam manajemen data dan informasi memungkinkan para pelaku usaha untuk membuat keputusan yang lebih akurat dan efisien. Salah satu metode yang kini banyak diterapkan adalah data mining, yakni proses mengekstraksi informasi berharga dari kumpulan data yang besar[1][2]. Dalam dunia bisnis, data mining dapat digunakan untuk menganalisis pola perilaku pelanggan, memantau tren penjualan, serta mendukung perencanaan strategi pemasaran dan pengelolaan persediaan barang..

Vapestore XYZ merupakan toko ritel yang bergerak di bidang penjualan produk vape dan aksesorisnya, Vapestore XYZ berlokasi di Karawang. Seiring meningkatnya minat masyarakat terhadap produk vape, toko ini mengalami peningkatan jumlah transaksi setiap bulannya. Namun, permasalahan yang dihadapi oleh pemilik usaha adalah ketidakmampuan dalam memprediksi jumlah penjualan barang secara akurat. Hal ini menyebabkan terjadinya ketidakseimbangan antara persediaan dan permintaan pasar, yang berpotensi mengakibatkan kerugian akibat kelebihan stok atau kekurangan barang.

Saat ini, Vapestore XYZ masih mengandalkan intuisi dan pengalaman pribadi dalam memperkirakan kebutuhan stok barang setiap bulannya. Pendekatan ini memiliki keterbatasan karena tidak berdasarkan pada analisis data yang objektif. Dikarenakan itu, diperlukan suatu sistem yang mampu menunjang dan menganalisis data penjualan historis untuk



menghasilkan prediksi penjualan produk yang lebih akurat. Metode yang dapat digunakan oleh peneliti untuk menghasilkan prediksi yang lebih akurat adalah kombinasi antara metode *K-Means* dan Regresi Linear.

Metode *K-Means* merupakan salah satu algoritma *unsupervised learning* yang digunakan untuk mengelompokkan data ke dalam sejumlah klaster berdasarkan tingkat kemiripan karakteristik antar data.[3][4][5] Dalam penelitian ini, metode *K-Means* digunakan untuk mengelompokkan produk berdasarkan pola kuantitas dari data penjualannya, sehingga dapat diketahui kelompok produk mana yang memiliki frekuensi tinggi, sedang, dan rendah. Setelah data dikelompokkan, langkah selanjutnya adalah melakukan prediksi terhadap penjualan di masa mendatang menggunakan metode Regresi Linear. Regresi Linear merupakan suatu metode statistik yang digunakan untuk memodelkan hubungan antara variabel independen dengan variabel dependen, di mana dalam konteks ini, jumlah penjualan berperan sebagai variabel yang diproyeksikan nilainya[6][7]. Proses pengolahan dan analisis data dilakukan menggunakan *Google Colab*, sebuah platform komputasi awan yang memungkinkan pemrosesan data secara efisien dan kolaboratif tanpa memerlukan perangkat keras khusus. Penggunaan *Google Colab* mendukung penerapan algoritma secara interaktif dan memudahkan dokumentasi serta reproduksi hasil penelitian.[8][9]

Beberapa penelitian yang sebelumnya telah peneliti baca, menunjukkan efektivitas penggunaan kombinasi metode *K-Means* dan Regresi Linear dalam berbagai bidang,[10][11] seperti manajemen persediaan, prediksi permintaan pasar, dan pengelolaan rantai pasok. Namun, belum banyak penelitian yang secara khusus menerapkan pendekatan ini dalam konteks toko retail vape di Indonesia, khususnya di wilayah Karawang. Oleh karena itu, penelitian ini diharapkan menjadi andil dalam pengembangan sistem pendukung keputusan berbasis *data mining* bagi pelaku usaha kecil dan menengah.

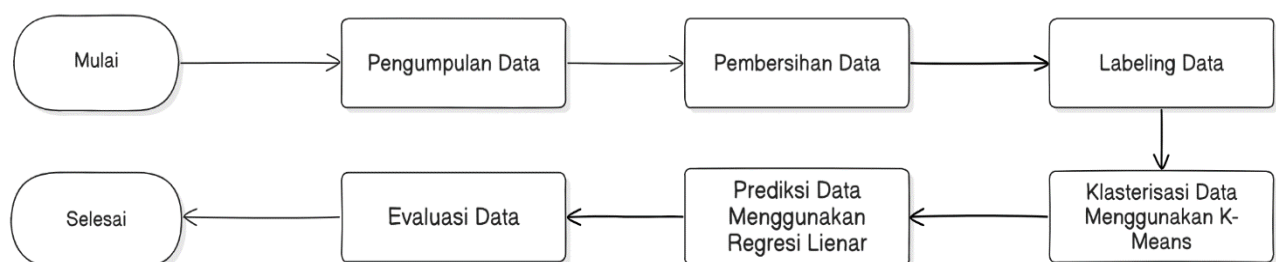
Tujuan dari penelitian ini adalah membangun model prediksi penjualan barang pada Vapestore XYZ di Karawang menggunakan pendekatan data mining dengan metode *K-Means* dan Regresi Linear. Kombinasi kedua metode ini dipilih karena proses klasterisasi dengan *K-Means* dapat mengelompokkan produk berdasarkan karakteristik penjualan yang serupa, sehingga pola-pola khusus pada setiap kelompok produk dapat diidentifikasi secara lebih spesifik dan dijadikan variabel tambahan dalam model Regresi Linear. Dengan demikian, model dapat menangkap variasi data yang lebih kompleks dan menghasilkan prediksi kuantitas penjualan yang lebih akurat [12]. Meskipun efektivitas kombinasi *K-Means* dan Regresi Linear telah dibuktikan dalam berbagai penelitian di bidang manajemen persediaan dan prediksi penjualan, sebagian besar studi sebelumnya hanya berfokus pada produk dengan karakteristik penjualan homogen atau belum diterapkan secara spesifik pada sektor retail vape di Indonesia, khususnya di Karawang [13]. Penelitian ini menawarkan pendekatan baru dengan mengintegrasikan hasil klasterisasi *K-Means* sebagai variabel input pada model Regresi Linear dan menerapkannya secara khusus pada toko vape, sehingga diharapkan dapat mengisi kekosongan literatur dan memberikan kontribusi nyata dalam pengembangan sistem prediksi penjualan berbasis data mining untuk sektor retail vape.

Struktur dari penelitian ini terdiri dari beberapa bagian, yaitu: bagian pertama merupakan pendahuluan yang menjelaskan latar belakang, rumusan masalah, dan tujuan penelitian. Bagian kedua membahas metodologi penelitian yang digunakan, termasuk proses pengumpulan dan pengolahan data. Bagian ketiga menyajikan hasil analisis dan pembahasan terhadap data yang diperoleh. Bagian keempat berisi kesimpulan dan saran yang dapat diberikan berdasarkan hasil penelitian. Dengan struktur ini, diharapkan pembaca dapat memahami secara menyeluruh proses dan hasil dari penelitian yang dilakukan.

2. METODOLOGI PENELITIAN

Metode yang digunakan oleh peneliti didalam penelitian ini adalah *K-Means* untuk mengklasterisasi penjualan produk dan Regresi Linear untuk memprediksi produk berdasarkan kuantitas. Berdasarkan studi terbaru oleh Muhammad Khaerudin et al. (2022), algoritma *K-Means* dan Regresi Linear efektif digunakan dalam mengelompokkan dan Memprediksi data untuk optimasi penjualan alat tulis kantor. Algoritma ini bekerja dengan meminimalkan jarak antar data terhadap pusat klaster untuk mengelompokkan data berdasarkan kemiripan tertentu.[13]

Sedangkan Regresi Linear digunakan untuk membuat model dari hubungan linier antara satu variabel bebas dan satu variabel terikat, guna menghasilkan estimasi prediktif terhadap data masa depan. Regresi Linear menjadi metode yang relevan dalam penelitian ini. Berikut tahapan penelitian yang ditampilkan pada Gambar 1 berikut.



Gambar 1. Tahapan Penelitian



Gambar 1 menunjukkan proses penelitian. Hal pertama yang dilakukan dalam penelitian ini adalah dengan melakukan pengumpulan data melalui *Olcera*, data diambil langsung dari aplikasi POS toko melalui koordinasi dengan pemilik toko untuk pengumpulan dataset. Tahapan berikutnya adalah *cleansing data* atau pembersihan data, data yang telah dikumpulkan kemudian dibersihkan melalui fase pembersihan data yang memiliki beberapa tahapan yaitu memeriksa data yang tidak memiliki nilai, cek duplikasi, dll. Selanjutnya dilakukan proses pelabelan data manual pada data latih digunakan sebagai *machine learning* untuk klasterisasi. Data yang telah diberi label kemudian dilakukan klasterisasi menjadi 3 klaster “Kurang Laku”, “Laku”, dan “Sangat Laku”. Data hasil klasterisasi akan di prediksi berdasarkan kuantitas untuk melihat produk akan terjual berapa dibulan-bulan selanjutnya. Setelah hasil prediksi data muncul, data akan di evaluasi menggunakan MAE, MSE, dan RMSE untuk dilihat seberapa akurat hasil prediksi tersebut.

2.1 Pengumpulan Data

Penelitian dilakukan pada salah satu toko *vapestore* di Karawang. Pada tahap ini peneliti melakukan koordinasi dengan pemilik toko, untuk mendapatkan ijin akses data penjualan melalui aplikasi POS *Olcera*. Peneliti mendapatkan data yang mencakup seluruh transaksi penjualan periode Januari 2024 hingga Oktober 2024. Total terdapat 12.707 transaksi yang melibatkan 670 produk unik. Data yang diekspor dalam format Excel berisi informasi tanggal transaksi, nama produk, dan jumlah item terjual. Untuk keperluan analisis, data transaksi tersebut diolah menjadi agregasi penjualan total produk terjual dalam tiap bulan. Data agregat ini digunakan sebagai input dalam proses klasterisasi menggunakan *K-Means* dan prediksi penjualan menggunakan Regresi Linear. Proses ini memastikan bahwa pola penjualan tiap produk dapat dianalisis secara spesifik dan prediksi penjualan di masa mendatang dapat dilakukan secara lebih akurat.

Data yang telah dikumpulkan selanjutnya akan melalui fase *cleansing data*. peneliti menggunakan *Google Colab* sebagai *platform* yang digunakan pada tahap-tahap berikutnya. *Google Colab (Collaboratory)* merupakan *platform* komputasi yang tersedia di *Google* dan memberikan akses kepada pengguna untuk menjalankan kode-kode Python secara interaktif tanpa perlu instalasi hardware atau software khusus. Dengan *Google Colab*, peneliti dapat mengakses sumber daya komputasi yang *powerful*, seperti *GPU* dan *TPU*, serta melakukan analisis data, pelatihan model machine learning, hingga visualisasi hasil secara kolaboratif dan efisien. Selain itu, *Google Colab* mendukung integrasi dengan *Google Drive* sehingga memudahkan penyimpanan dan berbagi data maupun hasil penelitian secara *real-time*. Penggunaan *Google Colab* dalam penelitian ini bertujuan untuk mempercepat proses pengolahan data, menguji berbagai model prediksi, serta memastikan reproducibility dan dokumentasi yang baik dalam setiap tahapan analisis.[11][12]

2.2 Pembersihan Data

Pada tahap ini, peneliti akan melakukan proses pembersihan data sebagai langkah awal sebelum melakukan pelabelan pada data penjualan. Proses pembersihan ini bertujuan untuk meningkatkan kualitas data agar analisis yang dilakukan nantinya menjadi lebih akurat dan valid. Peneliti akan mengidentifikasi dan menangani berbagai jenis *noise* yang sering ditemukan dalam dataset, seperti nilai yang hilang atau *missing value*, data duplikasi, kesalahan pengukuran, serta nilai ekstrem yang tidak sesuai dengan pola umum data. Dengan membersihkan data dari gangguan tersebut, peneliti memastikan bahwa dataset yang digunakan bebas dari kesalahan yang dapat mempengaruhi hasil analisis.

Selanjutnya, pembersihan data juga melibatkan proses validasi dan koreksi terhadap data yang bermasalah. Misalnya, *missing value* dapat diatasi dengan metode imputasi atau penghapusan data yang tidak lengkap, sedangkan data duplikasi akan dihapus untuk menghindari bias dalam pengolahan data. Kesalahan pengukuran dan nilai ekstrem juga dianalisis secara cermat untuk menentukan apakah perlu diperbaiki atau dihilangkan. Tahap pembersihan ini sangat penting karena kualitas data yang baik merupakan fondasi utama dalam membangun model prediksi atau melakukan analisis lebih lanjut, sehingga hasil penelitian dapat dipercaya dan memberikan kontribusi yang signifikan.[16]

2.3 Pelabelan Data

Pada tahap ini, peneliti melakukan proses pelabelan terhadap data penjualan dengan memberikan identifikasi yang jelas pada setiap entri data. Pelabelan mencakup informasi penting seperti nama produk, jenis produk, serta periode waktu saat produk tersebut terjual. Langkah ini sangat krusial untuk memastikan bahwa data yang digunakan dalam analisis memiliki struktur yang rapi dan mudah dipahami, sehingga memudahkan proses pengolahan dan interpretasi hasil penelitian.

Selain itu, pelabelan data juga berfungsi sebagai dasar untuk mengelompokkan dan mengkategorikan data secara sistematis. Dengan adanya informasi yang terperinci mengenai karakteristik produk dan waktu penjualan, peneliti dapat melakukan analisis tren, pola pembelian, serta evaluasi performa produk secara lebih akurat. Tahap ini menjadi fondasi penting dalam membangun model prediksi atau klasterisasi yang handal, karena kualitas data yang terlabel dengan baik akan sangat memengaruhi validitas dan reliabilitas hasil penelitian.[17]

2.4 Klasterisasi Data

Menurut artikel di *Jurnal Malcom* (2024), klasterisasi merupakan metode analisis data yang bertujuan untuk mengelompokkan objek-objek ke dalam beberapa kelompok atau klaster berdasarkan kemiripan karakteristik atau fitur yang dimiliki. Pendekatan ini digunakan untuk menemukan pola serta mengekstraksi informasi dari data berukuran besar dengan tingkat variasi yang tinggi, sehingga dapat mempermudah proses pengambilan keputusan yang berbasis data [18]. Pada tahap ini, peneliti melakukan klasterisasi produk menggunakan algoritma *K-Means*. Proses klasterisasi dilakukan terhadap seluruh data penjualan yang telah melalui tahap pembersihan, pelabelan, dan agregasi per bulan. Algoritma *K-*



Means diterapkan untuk mengelompokkan produk berdasarkan total kuantitas penjualan per produk setiap bulan selama periode penelitian. Jumlah kluster ditetapkan sebanyak tiga, yaitu “Kurang Laku”, “Laku”, dan “Sangat Laku”, dengan pertimbangan kebutuhan analisis bisnis dan kemudahan interpretasi hasil, sehingga segmentasi ini dinilai paling representatif untuk mendukung pengambilan keputusan stok dan strategi pemasaran. Proses inisialisasi pusat kluster dilakukan menggunakan metode *k-means++* dan metrik jarak yang digunakan adalah *Euclidean distance*, sesuai dengan pengaturan *default* pada *library scikit-learn* dan karakteristik data numerik yang dianalisis.

2.5 Prediksi Data

Prediksi data adalah proses memperkirakan nilai atau hasil di masa mendatang berdasarkan pola dan hubungan yang ditemukan dalam data historis. Pada penelitian ini, prediksi dilakukan terhadap jumlah penjualan produk di masa depan dengan memanfaatkan hasil klusterisasi sebagai salah satu variabel input. Metode yang diterapkan pada penelitian ini adalah Regresi Linear, yakni suatu metode statistik yang digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen dengan variabel dependen (misalnya hasil klusterisasi, periode waktu, dan variabel lain yang relevan) dengan variabel dependen berupa jumlah penjualan [10] [11]. Peneliti menggunakan data hasil klusterisasi pada tahap sebelumnya untuk melakukan prediksi data dengan menggunakan *Google Colab* dan memasukan algoritma regresi linear sebagai acuan model prediksi. Data diprediksi berdasarkan kuantitas penjualan produk di bulan-bulan sebelumnya.

Prediksi yang dihasilkan dari model Regresi Linear ini memberikan estimasi kuantitas penjualan produk pada periode mendatang, sehingga dapat membantu pemilik toko dalam melakukan perencanaan stok yang lebih akurat dan efisien. Dengan adanya prediksi berbasis data historis dan hasil klusterisasi, manajemen dapat mengidentifikasi produk-produk dengan potensi penjualan tinggi maupun rendah pada bulan-bulan berikutnya.[21], [22] Setelah proses klusterisasi selesai, data penjualan yang telah diagregasi per bulan dibagi menjadi dua bagian, yaitu data pelatihan (*training set*) dan data pengujian (*testing set*) dengan proporsi 80:20. Pembagian data dilakukan secara berbasis waktu, di mana 80% data awal (data dari Januari hingga Agustus 2024) digunakan sebagai data pelatihan untuk membangun model Regresi Linear, sedangkan 20% data akhir (data September dan Oktober 2024) digunakan sebagai data pengujian untuk mengevaluasi performa model.

2.6 Evaluasi Data

Evaluasi data merupakan tahap yang penting dalam proses penelitian data mining untuk mengukur tingkat akurasi dan performa model prediksi yang telah dibangun. Pada penelitian ini, evaluasi dilakukan terhadap hasil prediksi penjualan yang dihasilkan oleh model Regresi Linear setelah proses klusterisasi menggunakan algoritma K-Means. Evaluasi dilakukan dengan membandingkan nilai prediksi dengan nilai aktual pada data uji menggunakan beberapa metrik evaluasi yang umum digunakan, yaitu *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, dan *Root Mean Squared Error (RMSE)*. [23][24]

3. HASIL DAN PEMBAHASAN

3.1 Gambaran Umum Data

Penelitian ini memanfaatkan dataset penjualan yang telah dikumpulkan dari toko XYZ yang terdiri dari *Order Time*, *Order Source*, *Item Group*, *Item Name*, *QTY*, *Currency*, dan *Price*. Data penjualan ini diambil dari periode Januari sampai dengan September 2024. Berikut berbagai macam atribut yang menggambarkan karakteristik kehadiran mereka.

Tabel 1. Atribut Dataset

No	Atribut	Keterangan
1	<i>Order Time</i>	Waktu penjualan terjadi
2	<i>Order Source</i>	Tempat penjualan via aplikasi
3	<i>Item Group</i>	Jenis produk
4	<i>Item Name</i>	Nama produk
5	<i>QTY</i>	Kuantitas barang terjual
6	<i>Currency</i>	Nilai mata uang
7	<i>Price</i>	Harga barang

3.2 Preprocessing Data

Tabel 2. Dataset Sebelum *Preprocessing*

<i>Order No</i>	<i>Order Time</i>	<i>Order Source</i>	<i>Item Group</i>	<i>Item Name</i>	<i>QTY</i>	<i>Currency</i>	<i>Price</i>
3854230127000	2023-01-27	POINT OF	DEVICE				
13501	13:52:15	SALE	MOD	AEGIS T 200	1	IDR	820000
3854230102000	2023-01-02	POINT OF	LIQUID				
12744	11:43:22	SALE	SALT	AFLO ICE BLAST	1	IDR	50000



3854230101000	2023-01-01	POINT OF	LIQUID					
12717	18:09:24	SALE	SALT	BANANA MILK	1	IDR	50000	
3854230123000	2023-01-23	POINT OF	LIQUID					
13378	12:10:00	SALE	FREEBASE	BANANALICIOUS	1	IDR	140000	
3854230129000	2023-01-29	POINT OF	LIQUID	BANGSAWAN				
13571	01:17:42	SALE	FREEBASE	ISLAND OAK	1	IDR	100000	
3854230123000	2023-01-23	POINT OF	LIQUID	CANTALOUPE				
13401	22:11:37	SALE	SALT	FREEZE	1	IDR	50000	
3854230119000	2023-01-19	POINT OF						
13286	20:31:58	SALE	CARTRIDGE	CARTRIDGE AFLO	2	IDR	35000	

Sebelum dilakukannya klusterisasi data, dataset penjualan toko perlu melalui tahapan *preprocessing data* untuk memastikan kualitas dataset yang lebih baik dan meningkatkan akurasi hasil analisis. Tahapan *preprocessing* yang akan dilaksanakan dalam penelitian meliputi beberapa proses utama sebagai berikut:

- Pembersihan Data, pada tahap ini dataset yang telah dikumpulkan kemudian diperiksa untuk mengidentifikasi yang hilang (*missing values*) atau dataset yang tidak valid. Hasil dari proses pembersihan data yang dilakukan oleh peneliti menunjukkan bahwa tidak terdapat data yang hilang pada dataset yang digunakan, sehingga tidak diperlukan imputasi nilai.
- Cek Duplikasi dataset, pada tahap ini dataset diperiksa untuk mengidentifikasi dataset yang terduplikasi. Hasil pemeriksaan yang dilakukan oleh peneliti, dataset tidak menunjukkan adanya duplikasi.
- Penghapusan Data tidak relevan, peneliti melakukan penghapusan data kepada data yang tidak relevan untuk tahap klusterisasi.

Tabel 3. Dataset Setelah *Preprocessing*

Order Time	Item Group	Item Name	QTY	Price
2023-01-27				
13:52:15	DEVICE MOD	AEGIS T 200	1	820000
2023-01-02				
11:43:22	LIQUID SALT	AFLO ICE BLAST	1	50000
2023-01-01				
18:09:24	LIQUID SALT	BANANA MILK	1	50000
2023-01-23	LIQUID			
12:10:00	FREEBASE	BANANALICIOUS	1	140000
2023-01-29	LIQUID	BANGSAWAN		
01:17:42	FREEBASE	ISLAND OAK	1	100000
2023-01-23		CANTALOUPE		
22:11:37	LIQUID SALT	FREEZE	1	50000
2023-01-19				
20:31:58	CARTRIDGE	CARTRIDGE AFLO	2	35000

3.3 Implementasi Algoritma

Setelah melalui tahap *preprocessing*, dataset siap untuk dianalisis menerapkan metode klusterisasi dan prediksi. Penelitian ini menerapkan algoritma *K-Means* sebagai metode klusterisasi dataset dan Regresi Linear untuk prediksi hasil klusterisasi yang telah dilakukan berdasarkan kuantitas produk. Implementasi klusterisasi dan prediksi dijalankan menggunakan *Google Colabs* dengan bahasa pemrograman *Python*.

3.3.1 Klusterisasi Data

Proses klusterisasi dilakukan menggunakan algoritma *K-Means* dengan fitur utama berupa total kuantitas penjualan per produk setiap bulan. Produk dikelompokkan ke dalam tiga klaster, yaitu “Kurang Laku”, “Laku”, dan “Sangat Laku”, untuk memudahkan segmentasi berdasarkan performa penjualan. Hasil klusterisasi menunjukkan bahwa data penjualan toko terbagi dalam beberapa kelompok dengan distribusi jumlah anggota yang berbeda-beda. Tabel berikut menunjukkan jumlah anggota dalam setiap klaster untuk klusterisasi menggunakan *K-Means*.

Tabel 4. Distribusi Jumlah Anggota Tiap Klaster

Metode	Klaster 0	Klaster 1	Klaster 2
<i>K-Means</i>	278	63	2090

Tabel 4 menunjukkan hasil klusterisasi data penjualan dengan menerapkan metode *K-Means*, yang menghasilkan tiga klaster dengan jumlah anggota klaster yang berbeda. Setiap klaster merepresentasikan kelompok penjualan dengan karakteristik penjualan tertentu berdasarkan pola penjualan tertentu. Klaster 0 menunjukkan bahwa produk terjual dalam



periode perbulan itu diatas 5 *item* dan dibawah 15 *item*. Klaster 1 menunjukan bahwa produk terjual diatas 15 *item*. Klaster 2 menunjukan bahwa produk terjual dibawah 5 *item*.

Tabel 5. Dataset Hasil Klasterisasi

Periode	Item Name	QTY	Cluster_K-Means	Label
April 2024	Bequ Honeydew	13	0	Laku
April 2024	Cantaloupe Freeze	14	0	Laku
April 2024	Blondies Strawberry	5	2	Kurang Laku
April 2024	Bluenanarilla	1	2	Kurang Laku
April 2024	Cartridge Foom	69	1	Sangat Laku
April 2024	Cartridge Argus	32	1	Sangat Laku
April 2024	Cartridge Ursa Elite	9	0	Laku

Tabel 5 menampilkan hasil klasterisasi data penjualan produk pada periode bulan April 2024 menggunakan algoritma *K-Means*. Setiap baris pada tabel merepresentasikan satu jenis produk beserta jumlah penjualan (*QTY*), hasil klasterisasi (Cluster *K-Means*), dan label yang diberikan berdasarkan hasil klasterisasi tersebut. Produk-produk seperti *Bequ Honeydew*, *Cantaloupe Freeze*, dan *Cartridge Ursa Elite* termasuk dalam klaster 0 dengan label "Laku", sedangkan *Blondies Strawberry* dan *Bluenanarilla* masuk ke dalam klaster 2 dengan label "Kurang Laku". Sementara itu, *Cartridge Foom* dan *Cartridge Argus* tergolong dalam klaster 1 dengan label "Sangat Laku".

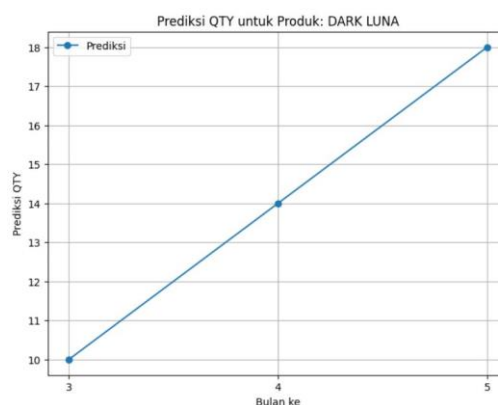
3.3.2 Prediksi Data

Model Regresi Linear dibangun dengan menggunakan label klaster hasil *K-Means* dan periode waktu sebagai variabel input, sehingga prediksi penjualan dapat mempertimbangkan segmentasi produk dan tren musiman. Hasil prediksi yang diperoleh menunjukkan bahwa model regresi linear memiliki kemampuan untuk memperkirakan jumlah penjualan produk pada periode mendatang dengan tingkat akurasi tertentu. Tabel berikut menyajikan nilai prediksi penjualan beserta nilai aktual untuk periode pengujian, sehingga dapat dilihat perbandingan dan performa model dalam memprediksi data penjualan.

Tabel 6. Hasil Prediksi Berdasarkan Kuantitas

No	Produk	Bulan ke	Prediksi_qty
1	CARTRIDGE OXVA SLIM PRO	11	101.266667
2	TICKET GUM SALT	12	92.0000
3	NAFAS SALT	12	58.806061
4	CARTRIDGE FOOM	11	57.084848
5	NAFAS SALT	11	56.266667

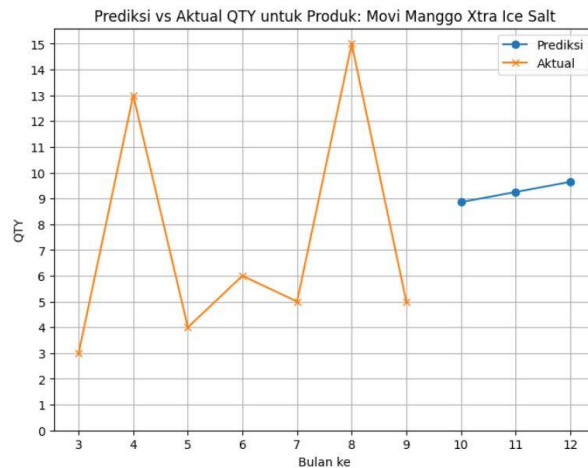
Tabel 6 menampilkan hasil prediksi penjualan berdasarkan kuantitas untuk lima produk berbeda pada periode waktu tertentu. Setiap baris pada tabel menunjukkan nama produk, bulan ke berapa prediksi dilakukan, serta jumlah penjualan yang diperkirakan (*Prediksi qty*). Seperti, produk "*CARTRIDGE OXVA SLIM PRO*" diprediksi akan terjual sebanyak 101,27 unit pada bulan ke-11, sedangkan "*TICKET GUM SALT*" diprediksi terjual sebanyak 92 unit pada bulan ke-12. Prediksi ini dihasilkan dari model peramalan yang digunakan dalam penelitian, sehingga memberikan gambaran tentang potensi penjualan di masa mendatang. Selain itu, tabel ini juga memperlihatkan adanya variasi prediksi penjualan antar produk dan periode. Produk "*NAFAS SALT*" muncul dua kali dengan prediksi penjualan yang berbeda pada bulan ke-11 dan ke-12, yaitu masing-masing sebesar 56,27 dan 58,81 unit. Hal ini menunjukkan bahwa model prediksi mampu memberikan estimasi yang spesifik untuk setiap produk pada waktu yang berbeda. Informasi dari tabel ini dapat dimanfaatkan oleh pelaku usaha untuk merencanakan stok, mengatur strategi pemasaran, serta mengambil keputusan bisnis berbasis data yang lebih akurat.



Gambar 2. Prediksi Berdasarkan Kuantitas untuk Barang "*DARK LUNA*"

Gambar 2 menggambarkan grafik prediksi jumlah penjualan (Prediksi QTY) untuk produk *DARK LUNA* selama tiga bulan berturut-turut, yaitu bulan ke-3, ke-4, dan ke-5. Sumbu horizontal (X) menunjukkan urutan bulan, sementara sumbu vertikal (Y) merepresentasikan jumlah penjualan yang diperkirakan. Setiap titik pada grafik menandakan nilai prediksi penjualan untuk bulan tertentu, dan garis yang menghubungkan titik-titik tersebut memperlihatkan pola atau tren perubahan penjualan dari waktu ke waktu.

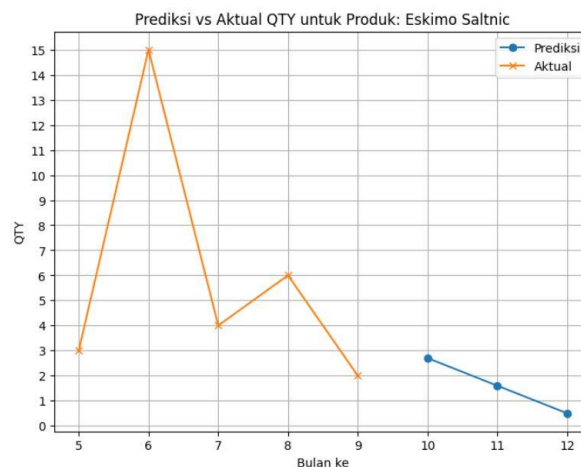
Dari grafik tersebut, terlihat adanya peningkatan jumlah penjualan produk *DARK LUNA* secara konsisten setiap bulannya. Prediksi penjualan pada bulan ke-3 sebesar 10 unit, kemudian meningkat menjadi 14 unit pada bulan ke-4, dan mencapai 18 unit pada bulan ke-5. Tren kenaikan yang berkelanjutan ini menunjukkan adanya permintaan yang meningkat terhadap produk tersebut dalam periode yang diamati. Informasi ini sangat penting bagi pelaku usaha untuk merencanakan pengelolaan stok dan strategi pemasaran secara lebih efektif, sehingga dapat mengantisipasi dan memenuhi kebutuhan pasar dengan lebih tepat.



Gambar 3. Prediksi dan Aktual kuantitas terjual dari produk “*MOVI MANGGO XTRA ICE SALT*”

Gambar 3 merupakan grafik perbandingan antara jumlah penjualan aktual dan hasil prediksi untuk produk “*Movi Manggo Xtra Ice Salt*” pada beberapa periode waktu. Sumbu horizontal (X) merepresentasikan urutan bulan, sedangkan sumbu vertikal (Y) menunjukkan jumlah penjualan (QTY). Data aktual digambarkan dengan garis berwarna oranye dan penanda berbentuk “x”, sementara hasil prediksi ditampilkan dengan garis berwarna biru dan penanda berbentuk lingkaran. Grafik ini memperlihatkan data penjualan aktual pada bulan ke-3 hingga bulan ke-9, serta prediksi penjualan untuk bulan ke-10 hingga bulan ke-12.

Dari grafik tersebut dapat diamati bahwa terdapat fluktuasi yang cukup signifikan pada data penjualan aktual, dengan puncak penjualan terjadi pada bulan ke-4 dan bulan ke-8. Sementara itu, hasil prediksi menunjukkan tren kenaikan penjualan yang lebih stabil pada periode bulan ke-10 hingga bulan ke-12. Perbedaan antara pola aktual dan prediksi ini menunjukkan bahwa model prediksi yang digunakan cenderung memberikan estimasi yang lebih moderat dibandingkan dengan variasi data aktual.



Gambar 4. Prediksi dan aktual kuantitas terjual dari produk “*Eskimo Saltnic*”

Perbedaan pola antara data aktual dan hasil prediksi ini juga mengindikasikan adanya kemungkinan faktor eksternal atau variabel lain yang mempengaruhi fluktuasi penjualan yang belum sepenuhnya tertangkap oleh model prediksi. Oleh karena itu, evaluasi lebih lanjut terhadap model dan penyesuaian parameter mungkin diperlukan agar prediksi dapat lebih akurat mencerminkan dinamika pasar yang sebenarnya. Selain itu, pemahaman mendalam terhadap

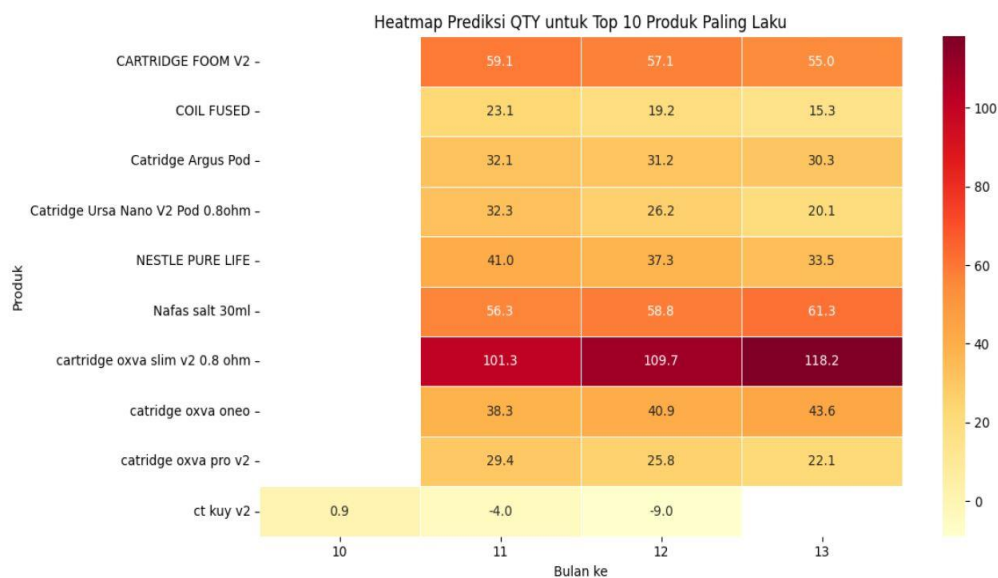
faktor-faktor yang memengaruhi penjualan dapat membantu dalam meningkatkan kualitas prediksi dan mendukung pengambilan keputusan bisnis yang lebih efektif.

Gambar 4 atas merupakan grafik perbandingan antara jumlah penjualan aktual dan hasil prediksi untuk produk "Eskimo Saltnic" pada beberapa periode waktu. Sumbu horizontal (X) menunjukkan urutan bulan, mulai dari bulan ke-5 hingga bulan ke-12, sedangkan sumbu vertikal (Y) merepresentasikan jumlah penjualan (QTY). Data penjualan aktual digambarkan dengan garis oranye dan penanda berbentuk "x", sementara hasil prediksi ditampilkan dengan garis biru dan penanda berbentuk lingkaran. Grafik ini memperlihatkan data aktual pada bulan ke-5 hingga bulan ke-9, serta prediksi penjualan pada bulan ke-10 hingga bulan ke-12.

Dari grafik tersebut, dapat diamati adanya fluktuasi yang cukup tajam pada data penjualan aktual, dengan puncak penjualan terjadi pada bulan ke-6, yaitu sebanyak 15 unit. Namun, setelah puncak tersebut, terjadi penurunan penjualan yang signifikan hingga bulan ke-9. Penurunan ini kemudian memengaruhi hasil prediksi pada bulan-bulan berikutnya, di mana model memproyeksikan tren penjualan yang terus menurun pada bulan ke-10, ke-11, dan ke-12. Temuan ini menunjukkan bahwa tren penurunan pada data historis sangat berpengaruh terhadap estimasi model untuk periode mendatang, sehingga penting bagi pelaku usaha untuk memperhatikan faktor-faktor penyebab penurunan agar dapat mengambil langkah strategis yang tepat dalam mengelola stok dan pemasaran produk.

Peneliti melakukan visualisasi prediksi berdasarkan kuantitas untuk 10 barang paling laku di toko, berikut adalah gambar *heatmap* yang dihasilkan. Visualisasi *heatmap* seperti ini sangat efektif untuk mengidentifikasi pola penjualan, tren permintaan, serta membandingkan performa antar produk secara cepat dan intuitif.

Visualisasi *heatmap* di bawah ini menampilkan hasil prediksi kuantitas penjualan untuk sepuluh produk terlaris di toko selama tiga bulan terakhir. Dengan memanfaatkan warna-warna gradasi dari kuning hingga merah, *heatmap* ini memudahkan identifikasi perbedaan volume penjualan antar produk dan periode waktu secara sekilas. Pola-pola yang muncul pada visualisasi ini dapat membantu peneliti dan manajer toko dalam memahami tren permintaan, mengamati fluktuasi penjualan, serta membandingkan performa setiap produk secara cepat dan intuitif. Dengan demikian, *heatmap* ini menjadi alat analisis yang sangat efektif untuk mendukung pengambilan keputusan strategis, seperti penentuan stok, promosi, dan pengembangan produk di masa mendatang.



Gambar 5. *Heatmap* prediksi 10 barang paling laku di toko XYZ

Gambar 5 merupakan *heatmap* prediksi kuantitas (*QTY*) penjualan untuk 10 produk paling laku pada periode bulan ke-10 hingga bulan ke-13. Setiap sel pada *heatmap* merepresentasikan hasil prediksi jumlah unit produk yang akan terjual pada bulan tertentu, dengan gradasi warna yang menunjukkan intensitas nilai prediksi-warna semakin gelap menandakan prediksi penjualan yang lebih tinggi. Sebagai contoh, produk "cartridge oxva slim v2 0.8 ohm" diprediksi mengalami penjualan tertinggi secara konsisten, dengan nilai prediksi meningkat dari 101,3 pada bulan ke-10 menjadi 118,2 pada bulan ke-13, sedangkan produk "ct kuy v2" justru menunjukkan prediksi penjualan yang sangat rendah hingga negatif, menandakan potensi penurunan atau bahkan tidak adanya penjualan pada periode tersebut.

Visualisasi *heatmap* seperti ini sangat membantu dalam mengidentifikasi pola penjualan, tren kenaikan atau penurunan permintaan, serta membandingkan performa antar produk secara cepat dan intuitif. Selain itu, *heatmap* juga dapat digunakan sebagai acuan dalam pengambilan keputusan terkait pengelolaan stok dan strategi pemasaran produk yang lebih efektif.



3.3.3 Evaluasi Hasil Regresi Linear

Untuk mengukur performa model prediksi yang dikembangkan, dilakukan evaluasi menggunakan *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, dan koefisien determinasi (R^2). Berdasarkan hasil pengujian pada data pengujian, diperoleh nilai MAE sebesar 3,20, MSE sebesar 52,34, RMSE sebesar 7,23, dan R^2 sebesar 0,61. Nilai MAE menunjukkan bahwa rata-rata kesalahan prediksi model adalah sekitar 3 unit penjualan per produk per bulan, sedangkan nilai R^2 sebesar 0,61 menunjukkan bahwa model mampu menjelaskan sekitar 61% variasi data penjualan yang ada.

3.4 Evaluasi Model

Berdasarkan hasil klusterisasi, produk-produk yang masuk dalam kluster “Sangat Laku” didominasi oleh liquid dengan rasa buah-buahan dan perangkat starter kit, yang memang menjadi favorit pelanggan di Vapestore XYZ. Hal ini menunjukkan bahwa model klusterisasi yang digunakan mampu mengidentifikasi pola penjualan yang relevan dengan tren pasar. Namun, terdapat beberapa produk yang berpindah kluster dari “Laku” ke “Kurang Laku” pada periode tertentu, yang dapat disebabkan oleh faktor musiman atau promosi yang tidak konsisten.

Hasil prediksi penjualan menggunakan Regresi Linear menunjukkan nilai MAE sebesar 3,20, yang berarti rata-rata kesalahan prediksi hanya sekitar tiga unit per produk per bulan. Nilai R^2 sebesar 0,61 juga menunjukkan bahwa model sudah cukup baik dalam menjelaskan variasi data penjualan, meskipun masih terdapat ruang untuk peningkatan akurasi, misalnya dengan menambah fitur lain seperti harga atau promosi.

Jika dibandingkan dengan penelitian serupa, hasil evaluasi model pada penelitian ini berada pada kisaran yang wajar, namun potensi perbaikan masih terbuka, terutama dalam mengatasi fluktuasi penjualan pada produk-produk dengan permintaan tidak stabil. Temuan ini membuktikan bahwa kombinasi K-Means dan Regresi Linear dapat diandalkan untuk mendukung pengambilan keputusan stok dan strategi pemasaran, namun pengembangan lebih lanjut diperlukan agar model lebih adaptif terhadap dinamika pasar.

4. KESIMPULAN

Evaluasi performa model Regresi Linear dilakukan menggunakan metrik *error*, yaitu *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, dan *Root Mean Squared Error (RMSE)*. Berdasarkan hasil pengujian, model regresi yang dikembangkan memperoleh MAE sebesar 3,20, MSE sebesar 52,34, dan RMSE sebesar 7,23. Nilai-nilai error tersebut menunjukkan rata-rata dan sebaran kesalahan prediksi model terhadap data aktual penjualan. Semakin kecil nilai MAE, MSE, dan RMSE, maka semakin baik performa model dalam memprediksi kuantitas penjualan. Berdasarkan seluruh rangkaian penelitian yang telah dilakukan, penerapan kombinasi metode *K-Means* untuk klusterisasi produk dan Regresi Linear untuk prediksi penjualan terbukti mampu memberikan gambaran segmentasi produk serta estimasi jumlah penjualan yang dapat diandalkan pada Vapestore XYZ. Proses analisis data historis, klusterisasi, dan prediksi yang dilakukan secara sistematis telah membantu mengidentifikasi produk-produk dengan performa penjualan berbeda dan memberikan dasar yang lebih kuat dalam pengambilan keputusan stok serta strategi pemasaran. Dengan demikian, model yang dikembangkan dalam penelitian ini dapat dijadikan referensi bagi pelaku usaha ritel vape dalam mengoptimalkan pengelolaan persediaan dan meningkatkan efisiensi operasional toko.

REFERENCES

- [1] R. Isnaen, P. Zaeni, and R. Renaldy, “IN-FEST 2024 Systematic Literature Review : Penerapan dan Teknik Data Mining pada Sistem Pengambilan Keputusan Pemasaran di Perusahaan IN-FEST 2024,” *Prosiding Seminar Nasional Informatika*, vol. 2, pp. 578–584, 2024.
- [2] I. T., “Penerapan Data Mining Dalam Dunia Bisnis Menggunakan Metode Clustering,” *J. Intsitution Sharia Financ.*, vol. 2, pp. 40–53, 2019.
- [3] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/tekno.v5i1.2019.17-24.
- [4] F. Dikarya and S. Muharni, “Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Universitas Terbaik Di Dunia,” *J. Inform.*, vol. 22, no. 2, pp. 124–131, 2022, doi: 10.30873/ji.v22i2.3324.
- [5] C. Science, “Implementasi Algoritma K-Means untuk Mengetahui Minat Siswa SMA Terhadap Mata Pelajaran Teknologi Informasi dan Komunikasi,” *Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 1, pp. 168–174, 2025, doi 10.57152/malcom.v5i1.1639
- [6] M. Djaka Permana, A. Lia Hananto, E. Novalia, B. Huda, and T. Paryono, “Klusterisasi Data Jamaah Umrah pada Tanurmutmainah Tour Menggunakan Algoritma K-Means,” *J. KomtekInfo*, vol. 10, no. 1, pp. 15–20, 2023, doi: 10.35134/komtekinfo.v10i1.332.
- [7] D. S. Annam, A. Hananto, F. Nurapriani, and T. Tukino, “Clustering User Sentiment Transportasi Online Gojek Dan Grab Dengan Metode K-Means,” *J. Tika*, vol. 8, no. 2, pp. 164–171, 2023, doi: 10.51179/tika.v8i2.2165.
- [8] A. H. Helyatin, N. Nadzirotul, F. Syarif, A. Khoiri, and M. Fauzen, “Perancangan Sistem Informasi Presensi Guru Berbasis Web,” *JUSTIFY : Jurnal Sistem Informasi*, vol. 3, no. 1, pp. 21–29, 2024, doi: 10.35316/justify.v3i1.5565.
- [9] M. D. Kurniawan, B. Priyatna, and F. Nurapriani, “Implementasi Algoritma K-Means Untuk Klusterisasi Data Obat Puskesmas Kotabaru,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 7, no. 2, pp. 882–890, 2023, DOI : <http://dx.doi.org/10.30645/j->



- sakti.v7i2.693.
- [10] F. P. Azizah, S. S. Hilabi, and A. Hananto, "Perbandingan Algoritma K-Means dan Hierarchical Untuk Klasterisasi Data Kehadiran Karyawan," *Jutisi*, vol. 14, no. 1, pp. 351–361, doi: <http://dx.doi.org/10.35889/jutisi.v14i1.2644>.
 - [11] T. P. Sari, A. L. Hananto, E. Novalia, T. Tukino, and S. S. Hilabi, "Implementasi Algoritma K-Means dalam Analisis Klasterisasi Penyebaran Penyakit Hiv/Aids," *Infotek J. Inform. dan Teknol.*, vol. 6, no. 1, pp. 104–114, 2023, doi: 10.29408/jit.v6i1.7423.
 - [12] W. W. Kristianto and C. Rudianto, "Penerapan Data Mining Pada Penjualan Produk Menggunakan Metode K-Means Clustering (Studi Kasus Toko Sepatu Kakikaki)," *J. Pendidik. Teknol. Inf.*, no. 5, pp. 90–98, 2020.
 - [13] M. Khaerudin, I. Zaenuddin, and Tukino, "Prediksi Barang Sering dan Jarang Terjual Dengan Menggunakan Algorithma K-Mean Clustering (Studi Kasus Toko Bina Mulia)," *J. Inform. Inf. Secur.*, vol. 3, no. 1, pp. 1–12, 2022, doi: 10.31599/jiforty.v3i1.1229.
 - [14] R. Nazar, "Implementasi Pemrograman Python Menggunakan Google Colab," *J. Inform. dan Komput.*, vol. 15, no. 1, pp. 50–56, 2024.
 - [15] R. A. A. Yanuar, "Sentimen Analisis Aplikasi Posaja Pada Google Playstore Untuk Peningkatan Pospay Superapp Menggunakan Support Vector Meachine" *Jurnal Teknik Informatika*, Vol. 16, No. 2, 1–7, 2024, [Online]. Available: <https://ejurnal.ulbi.ac.id/index.php/informatika/article/view/3533>
 - [16] N. P. A. Widiari, I. M. A. D. Suarjaya, and D. P. Githa, "Teknik Data Cleaning Menggunakan Snowflake untuk Studi Kasus Objek Pariwisata di Bali," *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*, vol. 8, no. 2, p. 137, 2020, doi: 10.24843/jim.2020.v08.i02.p07.
 - [17] H. Firda *et al.*, "Perbandingan Pelabelan Rating - based dan Inset Lexicon - based dalam Analisis Sentimen Menggunakan SVM (Studi Kasus : Ulasan Aplikasi GoBiz di Google Play Store)," *Jurnal Sistemasi* vol. 14, no. 2, pp. 516–528, 2025 DOI : <https://doi.org/10.32520/stmsi.v14i2.4795>.
 - [18] M. F. Fadhillah, A. Lovely, A. Suyoso, and I. Puspitasari, "Attributes Segmentasi Pelanggan dengan Algoritma Clustering Berdasarkan Atribut Recency , Frequency dan Monetary (RFM)," *Indonesian Journal of Machine Learning and Computer Science* vol. 5, no 1, pp. 48–56, 2025, DOI : <https://doi.org/10.57152/malcom.v5i1.1491>.
 - [19] A. Anggrawan, H. Hairani, and N. Azmi, "Prediksi Penjualan Produk Unilever Menggunakan Metode Regresi Linear," *J. Bumigora Inf. Technol.*, vol. 4, no. 2, pp. 123–132, 2022, doi: 10.30812/bite.v4i2.2416.
 - [20] K. Puteri and A. Silvanie, "Machine Learning untuk Model Prediksi Harga Sembako Dengan Metode Regresi Linear Berganda," *J. Nas. Inform.*, vol. 1, no. 2, pp. 82–94, 2020,.
 - [21] Y. Aqsho Ramadhan, A. Faqih, and G. Dwilestari, "Prediksi Penjualan Handphone di Toko X menggunakan Algoritma Regresi Linear," *J. Inform. Terpadu*, vol. 9, no. 1, pp. 40–44, 2023, doi: 10.54914/jit.v9i1.692.
 - [22] N. R. Setyoningrum, P. J. Rahimma, S. T. Teknologi, I. Tanjungpinang, and K. Tanjungpinang, "Implementasi Algoritma Regresi Linear Dalam Sistem Prediksi Pendaftar Mahasiswa Baru Sekolah Tinggi Teknologi Indonesia Tanjungpinang," *Pros. Semin. Nas. Ilmu Sos. dan Teknol.*, no. 4, pp. 13–18, 2022, [Online]. Available: <https://ejournal.upbatam.ac.id/index.php/prosiding/article/view/5200>
 - [23] T. Tukino and F. Fifi, "Penerapan Support Vector Machine Untuk Analisis Sentimen Pada Layanan Ojek Online," *J. Desain Dan Anal. Teknol.*, vol. 3, no. 2, pp. 104–113, 2024, doi: 10.58520/jddat.v3i2.59.
 - [24] T. Tukino and B. Huda, "Penerapan Algoritma K-Means Untuk Mendukung Keputusan Dalam Pemilihan Tema Tugas Akhir Pada Prodi Sistem Informasi Universitas Buana Perjuangan Karawang.," *Techno Xplore J. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 1, pp. 1–10, 2019, doi: 10.36805/technoxplore.v4i1.542.