



## *The Comparison of K-Means, Naïve Bayes and Decision Tree Algorithm in Predicting Fuel Oil Sales*

### **Perbandingan Algoritma K-Means, Naïve Bayes dan Decision Tree dalam Memprediksi Penjualan Bahan Bakar Minyak**

Usman Arfan<sup>1</sup>, Novita Paraga<sup>2</sup>

<sup>1,2</sup>Program Studi Informatika, STMIK Pesat Nabire, Indonesia

E-Mail: <sup>1</sup>usmanarfanpesat@gmail.com, <sup>2</sup>nnovita620@gmail.com

Received Jun 18th 2024; Revised Jul 26th 2024; Accepted Jul 29th 2024  
Corresponding Author: Usman Arfan

#### **Abstract**

*This research discusses the application of the K-Means algorithm to cluster data on fuel oil (BBM) sales at the Wadio Nabire gas station, as well as evaluating the accuracy of the Decision Tree and Naive Bayes models using data mining in the Orange application. The main objective of this research is to optimize fuel distribution by understanding consumption patterns and customer segmentation. This research identifies the main problem in sub-optimal fuel distribution, which causes excess or shortage of stock at the Wadio Nabire gas station, as well as consumer dissatisfaction due to long queues and increased operational costs. By using data mining techniques such as clustering and sales predictions, this research aims to provide solutions for more efficient distribution planning and more effective marketing strategies to increase fuel sales and customer satisfaction.*

*Keywords: Data Mining, Decision Tree, K-Means, Naïve Bayes, Orange App*

#### **Abstrak**

Penelitian ini membahas penerapan algoritma K-Means untuk melakukan clustering terhadap data penjualan Bahan Bakar Minyak (BBM) di Stasiun Pengisian Bahan Bakar Umum (SPBU) Wadio Nabire, serta evaluasi akurasi model Decision Tree dan Naive Bayes menggunakan data mining pada aplikasi Orange. Tujuan utama penelitian ini adalah untuk mengoptimalkan distribusi BBM dengan memahami pola konsumsi dan segmentasi pelanggan. Penelitian ini mengidentifikasi masalah utama dalam distribusi BBM yang tidak optimal, yang menyebabkan kelebihan atau kekurangan stok di SPBU Wadio Nabire, serta ketidakpuasan konsumen akibat antrean panjang dan kenaikan biaya operasional. Dengan menggunakan teknik data mining seperti clustering dan prediksi penjualan, penelitian ini bertujuan untuk memberikan solusi dalam perencanaan distribusi yang lebih efisien dan strategi pemasaran yang lebih efektif untuk meningkatkan penjualan BBM dan kepuasan pelanggan.

Kata kunci: Aplikasi Orange, Data Mining, Decision Tree, K-Means, Naïve Bayes

#### **1. PENDAHULUAN**

Pengelolaan Bahan Bakar Minyak (BBM) di Stasiun Pengisian Bahan Bakar Umum (SPBU) merupakan aspek yang sangat penting bagi industri BBM di Indonesia. Konsumsi BBM yang meningkat dari tahun ke tahun seiring dengan pertumbuhan ekonomi dan perkembangan jumlah kendaraan, menyebabkan kebutuhan pengelolaan BBM yang lebih efektif dan efisien [1]. *Clustering* adalah teknik analisis data yang dapat digunakan dengan tepat untuk segmentasi pelanggan dan memahami pola konsumsi BBM. Dengan mengelompokkan pelanggan ke dalam cluster tertentu, pelanggan yang lebih disukai dan yang kurang disukai dapat dibedakan sehingga dapat dirumuskan strategi yang tepat dalam meningkatkan penjualan [2]. Salah satu permasalahan utama dalam pengelolaan distribusi BBM di SPBU adalah ketidakefektifan distribusi BBM, yang mengakibatkan kelebihan stok di beberapa SPBU dan kekurangan stok di SPBU lainnya. Hal ini dapat disebabkan oleh pola konsumsi BBM yang tidak merata di setiap SPBU, serta kurangnya pemahaman terhadap preferensi dan kebutuhan konsumen BBM di masing-masing SPBU [3]. *Clustering* adalah salah satu solusi yang dapat diterapkan untuk mengoptimalkan distribusi BBM di SPBU. Dengan melakukan clustering, SPBU dapat mengelompokkan pelanggan berdasarkan pola konsumsi BBM mereka, dan kemudian merumuskan

strategi pemasaran yang sesuai untuk masing-masing segmen [4]. Selain itu, teknik prediksi permintaan juga dapat digunakan untuk memperkirakan volume penjualan BBM di setiap SPBU, sehingga distribusi BBM dapat diatur dengan lebih baik [5].

Ketidakpuasan konsumen akibat kekurangan pasokan BBM di SPBU tertentu, meningkatnya biaya logistik dan transportasi dalam upaya menyeimbangkan distribusi BBM antar SPBU [6]. Masalah distribusi bahan bakar minyak (BBM) yang tidak optimal dapat membawa dampak negatif yang signifikan bagi konsumen. Salah satu isu utama yang sering terjadi adalah antrean panjang SPBU yang menyebabkan kekecewaan konsumen. Situasi ini dapat timbul akibat ketidakseimbangan antara ketersediaan stok dan permintaan, di mana jumlah konsumen yang membutuhkan bahan bakar dalam waktu bersamaan melebihi kapasitas layanan yang ada [7]. Permasalahan distribusi bahan bakar yang tidak optimal juga dapat mengakibatkan kelangkaan pasokan bahan bakar di berbagai wilayah. Kenaikan harga bahan bakar juga menjadi dampak ikutan dari kondisi ini, di mana petani dan masyarakat rentan terkena imbas kenaikan biaya operasional. Selain itu, ada kekhawatiran akan jaminan ketersediaan bahan bakar fosil di masa depan. Oleh karena itu, diperlukan upaya untuk mengoptimalkan distribusi bahan bakar agar dapat memenuhi permintaan konsumen secara memadai dan mencegah terjadinya antrean panjang serta kekecewaan. Salah satu solusi yang dapat dipertimbangkan adalah pemanfaatan energi alternatif terbarukan, seperti biogas, sebagai substitusi bahan bakar fosil [8]. Selain dampak terhadap konsumen, distribusi bahan bakar yang tidak optimal juga dapat memberikan dampak negatif terhadap reputasi dan citra SPBU serta perusahaan penyedia bahan bakar. Antrian panjang dan kekurangan stok yang terjadi di SPBU dapat menimbulkan kekecewaan dan ketidakpuasan konsumen, yang pada akhirnya dapat menurunkan kepercayaan masyarakat terhadap layanan dan kredibilitas SPBU tersebut. Situasi ini juga dapat berimbas pada reputasi dan citra perusahaan penyedia bahan bakar secara keseluruhan, khususnya jika permasalahan terjadi di berbagai lokasi SPBU di bawah naungannya. Oleh karena itu, optimalisasi distribusi bahan bakar menjadi penting tidak hanya untuk memenuhi kebutuhan konsumen, tetapi juga untuk menjaga reputasi dan memperkuat brand image perusahaan. Untuk mengatasi permasalahan ini, perusahaan penyedia bahan bakar perlu merancang sistem penjadwalan dan rute pengiriman bahan bakar yang efisien ke setiap SPBU agar dapat meminimalisir terjadinya kekurangan stok dan antrean panjang [9].

Berbagai faktor, baik internal maupun eksternal, dapat mempengaruhi tingkat permintaan bahan bakar di SPBU. Faktor internal SPBU, seperti lokasi, jam buka, jenis layanan, dan promosi, dapat memengaruhi daya tarik konsumen. Di sisi lain, faktor eksternal seperti jumlah kendaraan, kondisi lalu lintas, dan musim juga turut mempengaruhi permintaan bahan bakar [10]. Faktor kebijakan pemerintah, seperti penetapan harga, subsidi, dan regulasi terkait energi, juga berperan penting dalam mempengaruhi dinamika permintaan.

Untuk mengatasi dampak negatif distribusi BBM yang tidak optimal, diperlukan beberapa solusi yang komprehensif. Salah satu solusi yang penting adalah mengembangkan sistem yang dapat memprediksi permintaan BBM secara akurat berdasarkan berbagai faktor yang mempengaruhinya, baik internal maupun eksternal. Selain itu, perlu adanya perencanaan distribusi BBM yang optimal untuk memastikan ketersediaannya di berbagai wilayah dan mencegah terjadinya kekurangan pasokan, sehingga dapat memenuhi kebutuhan konsumen secara merata. Perencanaan distribusi BBM yang optimal dapat dilakukan dengan mengelompokkan SPBU berdasarkan pola penjualan, sehingga alokasi dapat dilakukan dengan tepat sesuai kebutuhan. Untuk membangun solusi yang komprehensif, teknik data mining seperti clustering dan prediksi penjualan dapat dimanfaatkan untuk membangun sistem yang dapat memprediksi permintaan BBM secara akurat [11]. Dengan pengelompokan SPBU berdasarkan pola penjualan, perencanaan distribusi BBM yang optimal dapat dilakukan agar alokasi dapat dilakukan sesuai dengan kebutuhan masing-masing wilayah.

Teknik data mining menawarkan beberapa keunggulan dalam membangun solusi untuk mengatasi permasalahan distribusi BBM yang tidak optimal. Melalui teknik clustering, SPBU dapat dikelompokkan berdasarkan tingkat penjualan dan karakteristik konsumen, sehingga alokasi BBM dapat dilakukan secara lebih tepat sesuai kebutuhan masing-masing wilayah. Selain itu, teknik prediksi penjualan dapat digunakan untuk memperkirakan dan mengantisipasi permintaan BBM di masa mendatang, sehingga perusahaan dapat melakukan perencanaan distribusi yang lebih efektif. Sementara itu, teknik aturan asosiasi dapat membantu mempelajari pola pembelian konsumen, yang dapat digunakan untuk menyusun strategi pemasaran dan promosi yang lebih efektif [12]. Prediksi Penjualan untuk mengantisipasi permintaan hasil prediksi dapat membantu perusahaan dalam mengalokasikan sumber daya distribusi secara lebih efisien. Aturan Asosiasi untuk mempelajari pola pembelian hasil aturan asosiasi dapat membantu meningkatkan kepuasan konsumen melalui strategi pemasaran yang lebih efektif. Teknik data mining menawarkan beberapa keunggulan dalam membangun solusi untuk mengatasi permasalahan distribusi BBM yang tidak optimal. Melalui teknik prediksi penjualan, SPBU dapat memprediksi kebutuhan BBM secara akurat, sehingga dapat mempersiapkan stok yang sesuai dan mencegah kekurangan pasokan [13]. Sehingga dapat menyiapkan pasokan yang memadai dan mencegah kekurangan. Selain itu, analisis pola pembelian konsumen juga dapat digunakan untuk mengembangkan program pemasaran yang dapat meningkatkan loyalitas pelanggan. Penerapan teknik data mining seperti clustering, prediksi penjualan, dan analisis aturan asosiasi memiliki potensi besar dalam membangun solusi untuk mengatasi permasalahan distribusi BBM yang tidak optimal.

Bagi penelitian penelitian ini sangat penting karena memiliki urgensi yang signifikan dalam memberikan kontribusi penting bagi optimalisasi operasional SPBU di wilayah Papua Tengah. Dalam konteks ini, memahami pola dan tren penjualan BBM melalui algoritma prediktif yang efektif dapat membantu manajemen SPBU dalam merencanakan stok, menghindari kekurangan atau kelebihan pasokan, serta meningkatkan efisiensi operasional dan kepuasan pelanggan. Selain itu, mengingat kondisi geografis dan logistik yang unik di Papua Tengah, penelitian ini dapat memberikan panduan praktis yang lebih tepat sasaran untuk pengelolaan distribusi BBM di daerah tersebut. Perbandingan antara K-Means, Naïve Bayes, dan Decision Tree akan mengidentifikasi algoritma yang paling akurat dan andal, yang pada gilirannya dapat diadopsi oleh SPBU lain di wilayah serupa, sehingga berkontribusi pada stabilitas ekonomi dan kesejahteraan masyarakat setempat.

Penelitian ini memiliki kelebihan dibandingkan dengan penelitian sebelumnya karena dalam menyajikan perbandingan langsung antara tiga algoritma yang berbeda seperti model K-Means, Naïve Bayes, dan Decision Tree, dalam konteks memprediksi penjualan BBM di SPBU Wadio Nabire. Pendekatan ini dapat memberikan wawasan mendalam tentang keefektifan masing-masing algoritma dalam aplikasi praktis yang spesifik. Namun, kelemahannya mungkin terletak pada kompleksitas analisis yang diperlukan untuk membandingkan kinerja ketiga algoritma tersebut secara menyeluruh, serta tantangan dalam memastikan bahwa data yang digunakan mencerminkan kondisi yang representatif di lapangan, khususnya dalam konteks varian lokal seperti Papua Tengah.

## 2. METODOLOGI PENELITIAN

### 2.1. K-Means

K-Means adalah salah satu algoritma clustering yang paling sederhana dan populer. Algoritma ini digunakan untuk membagi data ke dalam beberapa kelompok (*cluster*) berdasarkan kemiripan fitur atau atributnya. Tujuannya adalah untuk meminimalkan variabilitas dalam cluster dan memaksimalkan perbedaan antar *cluster*. Tidak ada rumus matematis yang kompleks yang perlu diimplementasikan secara langsung dalam aplikasi. Algoritma K-Means menghitung *centroid* untuk setiap *cluster* yang ditemukan berdasarkan jarak Euclidean antara titik data dan *centroid*. Berikut rumus dasar jarak Euclidean digunakan untuk menentukan kedekatan titik data ke centroid, ditunjukkan pada persamaan 1 [21].

$$d(x, c) = \sqrt{\sum_{i=1}^n (x_i - c_i)^2} \tag{1}$$

adapun fungsi objective (*Within-Cluster Sum of Squares - WCSS*), di mana  $K$  adalah jumlah cluster,  $n_j$  adalah jumlah titik dalam cluster  $j$   $x_i^{(j)}$  adalah titik data ke- $i$  dalam cluster  $j$ , dan  $c_j$  adalah centroid dari cluster  $j$ .

$$WCSS = \sum_{j=1}^K \sum_{i=1}^{n_j} \|x_i^{(j)} - c_j\|^2 \tag{2}$$

Algoritma K-Means sangat efektif dan umum digunakan dalam berbagai aplikasi seperti segmentasi pasar, pengelompokan dokumen, dan pengelompokan gambar.

### 2.2. Naive Bayes

Naive Bayes adalah algoritma klasifikasi yang didasarkan pada Teorema Bayes dengan asumsi independensi (*naive*) antara fitur-fitur yang ada. Meskipun asumsi ini jarang benar dalam praktik, Naive Bayes sering kali memberikan hasil yang baik, terutama dalam kasus klasifikasi teks dan analisis sentimen. Teorema Bayes memberikan cara untuk menghitung probabilitas suatu hipotesis berdasarkan informasi yang tersedia. Persamaan naïve bayes ditunjukkan pada persamaan 3 [22].

$$P(C_k | x) = \frac{P(C_k) \cdot P(x|C_k)}{P(x)} \tag{3}$$

Keterangan:

- $P(C_k | x)$  : probabilitas kelas  $C_k$  yang diberikan fitur  $x$  (probabilitas posterior).
- $P(C_k)$  : probabilitas awal (prior) dari kelas  $C_k$
- $P(x | C_k)$  : probabilitas fitur  $x$  yang diberikan kelas  $C_k$  (*likelihood*)
- $P(x)$  : probabilitas total fitur  $x$

$$P(x_1, x_2, \dots, x_n | c_k) = P(x_1 | c_k) \cdot P(x_2 | c_k) \cdot \dots \cdot P(x_n | c_k) \quad (4)$$

### 2.3. Decision Tree

Decision Tree adalah algoritma *machine learning* yang menggunakan struktur pohon untuk membuat keputusan berdasarkan serangkaian aturan yang diperoleh dari fitur data. Decision Tree dapat digunakan untuk tugas klasifikasi maupun regresi. Struktur pohon terdiri dari node, cabang (branch), dan daun (leaf), di mana setiap node internal mewakili pengujian pada suatu fitur, setiap cabang mewakili hasil pengujian, dan setiap daun mewakili label kelas atau nilai regresi. Rumus dan kriteria pemilihan atribut, informasi gain (Entropy): digunakan dalam algoritma ID3 dan C4.5. Entropi mengukur ketidakhomogenan atau ketidakpastian suatu dataset

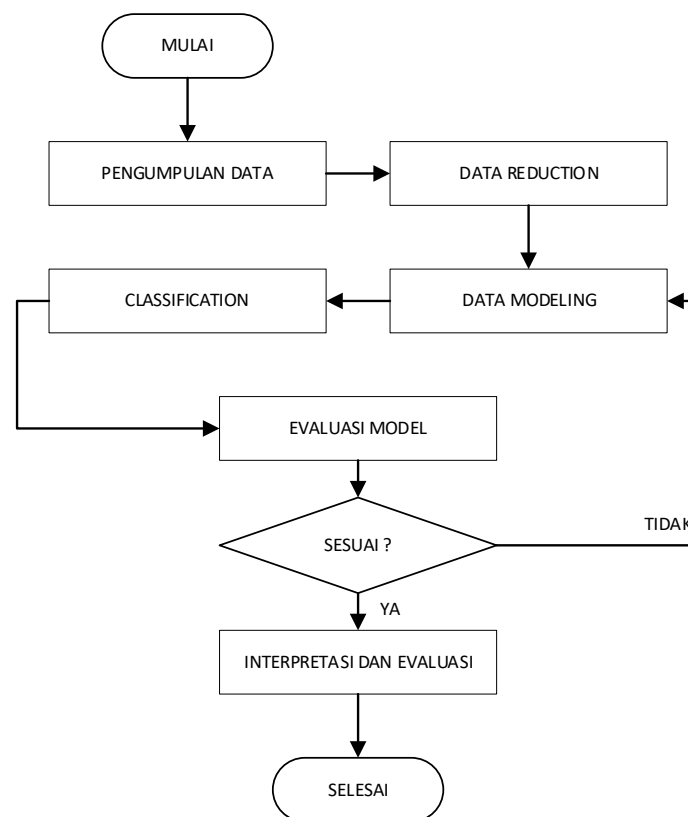
$$Entropy(S) = - \sum_{i=1}^c p_i \log_2(p_i) \quad (5)$$

di mana  $S$  adalah dataset,  $c$  adalah jumlah kelas, dan  $p_i$  adalah proporsi contoh dalam kelas  $i$ . Informasi gain adalah pengurangan dalam entropi [23]:

$$Information\ Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (6)$$

### 2.4. Alur Penelitian

Gambar 1 menunjukkan flowchart alur penelitian, dimulai dari mengumpulkan data dari tempat penelitian dalam hal ini SPBU Wadio, kemudian mengurangi volume data dengan memilih fitur penting dan menghilangkan yang tidak relevan, tanpa kehilangan informasi penting (*Data Reduction*). selanjutnya menerapkan algoritma dan teknik data mining seperti clustering untuk membangun model (*Data Modeling*). langkah berikutnya memisahkan data ke dalam kategori atau kelas yang sudah ditentukan sebelumnya, algoritma yang digunakan dalam hal ini untuk perbandingan ialah Decision Tree dan Naive Bayes (*Classification*).



**Gambar 1.** Metodologi Penelitian

Langkah berikutnya menilai kinerja model menggunakan metrik seperti akurasi, presisi, recall dan F1-score (Evaluasi Model). Jika tidak sesuai maka akan melakukan langkah ulang dari *Data Modeling* tetapi jika sudah

sesuai maka akan dilakukan penafsiran hasil model untuk membuat keputusan atau tindakan berdasarkan temuan (Interpretasi dan Evaluasi).

### 3. HASIL DAN PEMBAHASAN

Perusahaan sudah mempertimbangkan strategi persediaan dan pemasaran untuk meningkatkan penjualan bahan bakar minyak secara tidak langsung. Namun, saat bahan bakar minyak dipasarkan, ada beberapa masalah yang muncul. Tidak jelas berapa banyak yang sudah terjual, apakah itu tertinggi, sedang, atau terendah. Untuk memberikan gambaran, penelitian menggali informasi dari data yang telah dikumpulkan dengan menggunakan algoritma yang dapat mengklasifikasikan penjualan ke dalam kategori tertentu. data yang digunakan ialah persediaan dan penjualan tahun 2023. Untuk memudahkan, data penjualan bahan bakar minyak diberikan dalam satuan per liter.

**Tabel 1.** Data Persediaan dan Penjualan BBM Tahun 2023

Tahun 2023	Januari		Februari		Maret		April	
	Stok	Jual	Stok	Jual	Stok	Jual	Stok	Jual
Pertalite	166178	165390	120517	119678	120517	119903	120517	115275
Pertamax	37427	26885	34353	23125	35060	25236	41728	33878
Dexlite	80226	60949	83093	73171	90118	81578	88717	77853
Solar	224517	216929	168196	167844	160517	160502	165517	155390

Tahun 2023	Mei		Juni		Juli		Agustus	
	Stok	Jual	Stok	Jual	Stok	Jual	Stok	Jual
Pertalite	124895	124137	120517	119822	120517	119889	120517	114138
Pertamax	39652	31276	23621	19842	19744	18399	41332	30426
Dexlite	98666	81976	112288	99648	84380	75869	96319	81999
Solar	169978	168480	160517	152956	167532	166388	160517	159751

Tahun 2023	September		Oktober		November		Desember	
	Stok	Jual	Stok	Jual	Stok	Jual	Stok	Jual
Pertalite	126155	116914	129369	123357	125360	116995	137954	136964
Pertamax	39509	31442	39727	31343	40211	39462	48517	40788
Dexlite	85875	80677	132921	126256	173086	168404	115125	107784
Solar	160517	159636	160517	159371	160517	159443	150517	143513

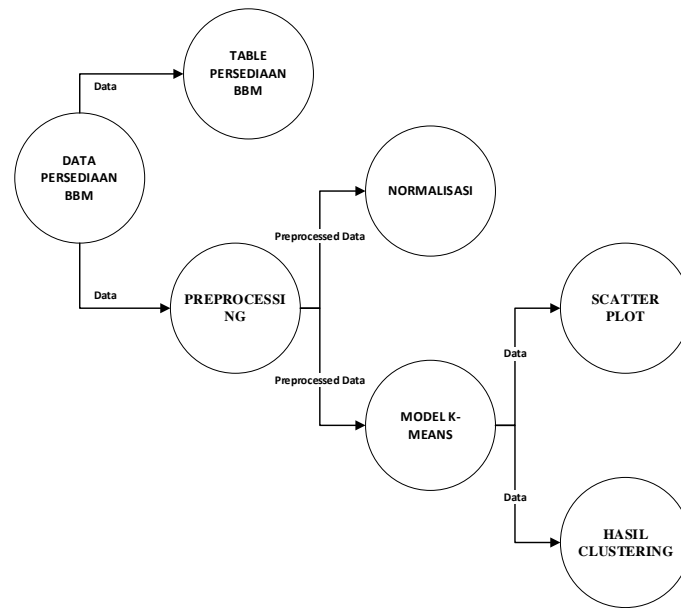
#### 3.1. Data Modeling K-Means

Algoritma ini mengelompokkan data ke dalam sejumlah cluster yang telah ditentukan sebelumnya, berdasarkan kedekatan atau kemiripan antar data. Berikut adalah tahapan-tahapan yang dilakukan dalam algoritma K-Means.

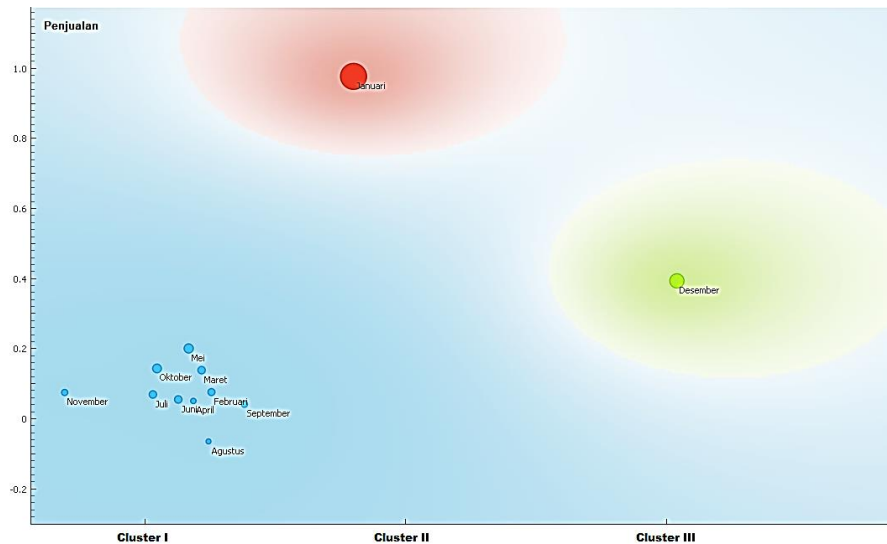
Alur kerja (*workflow*) dalam perangkat lunak Orange untuk analisis data menggunakan algoritma K-Means. Berikut adalah penjelasan setiap komponen dalam alur kerja ini. Data persediaan BBM, Ini adalah sumber data yang digunakan untuk analisis. Data ini kemungkinan besar berisi informasi terkait penjualan BBM. Tabel persediaan BBM, Komponen ini digunakan untuk melihat dan mengeksplorasi data yang telah dimuat. Preprocess, Tahap ini melibatkan pembersihan dan persiapan data sebelum analisis lebih lanjut. Proses preprocessing mungkin mencakup penanganan nilai yang hilang, pengkodean variabel, dan lain-lain.

Normalisasi, Langkah ini menormalkan data, yang berarti menyesuaikan skala dari berbagai fitur data sehingga berada dalam rentang yang sama. Ini penting untuk algoritma seperti K-Means yang sensitif terhadap skala data. Model K-Means, Algoritma clustering K-Means digunakan untuk mengelompokkan data ke dalam beberapa cluster berdasarkan kesamaan fitur. *Scatter Plot*, Visualisasi ini digunakan untuk melihat hasil clustering dengan memplot data dalam dua dimensi, memungkinkan identifikasi pola dan distribusi cluster. Hasil Clustering, Komponen ini menampilkan hasil clustering dalam bentuk tabel atau visualisasi lain yang memudahkan interpretasi hasil. Save Data, Langkah terakhir adalah menyimpan hasil analisis atau clustering ke dalam file untuk penggunaan lebih lanjut. Secara keseluruhan, alur kerja ini menunjukkan proses analisis data dari tahap awal (memuat data) hingga tahap akhir (menyimpan hasil clustering) menggunakan Orange. Modeling K-Means Clustering pada Aplikasi Orange ditunjukkan pada gambar 2.

Menentukan jumlah cluster K yang diinginkan. Ini adalah parameter yang harus dipilih sebelum menjalankan algoritma. dalam hal ini (Tinggi, Sedang, Rendah). Setelah itu Inisialisasi *Centroid*, memilih secara acak K titik sebagai centroid awal dari *cluster*. Centroid adalah pusat dari *cluster* yang nantinya akan dioptimalkan. Menetapkan setiap data ke cluster terdekat untuk setiap titik data dalam dataset: Hitung jarak antara titik data dan setiap centroid. Jarak yang biasa digunakan adalah jarak Euclidean dan Tetapkan titik data ke cluster dengan centroid terdekat (yakni yang memiliki jarak terkecil). Visualisasi scatter plot hasil *clustering* pada aplikasi orange ditunjukkan pada gambar 3.



**Gambar 2.** Modeling K-Means Clustering Pada Aplikasi Orange



**Gambar 3.** Scatter Plot Hasil Clustering pada Aplikasi Orange

Grafik scatter plot di atas menunjukkan hasil *clustering* penjualan Bahan Bakar Minyak (BBM) menggunakan algoritma K-Means. Berikut penjelasan mengenai gambar tersebut:

1. Sumbu X (Cluster): Menunjukkan kluster hasil dari algoritma K-Means.
2. Sumbu Y (Penjualan): Menunjukkan nilai penjualan BBM.
3. Titik-titik pada Grafik: Mewakili bulan-bulan dalam setahun. Titik biru (C1): Merupakan kluster 1, yang berisi bulan-bulan dengan penjualan rendah hingga sedang. Bulan-bulan dalam kluster ini adalah April, Mei, Juni, Juli, Agustus, September, Oktober, dan November. Titik merah (C2): Merupakan kluster 2, yang hanya berisi bulan Januari dengan penjualan yang sangat tinggi. Titik hijau (C3): Merupakan kluster 3, yang hanya berisi bulan Desember dengan penjualan yang juga cukup tinggi, namun tidak setinggi bulan Januari.

Dari grafik 3, dapat disimpulkan bahwa: Bulan Januari memiliki pola penjualan yang unik dan jauh lebih tinggi dibandingkan bulan-bulan lainnya, sehingga masuk ke dalam kluster tersendiri (C2). Bulan Desember juga memiliki penjualan yang signifikan dan berada di kluster tersendiri (C3), namun tidak setinggi bulan Januari. Bulan-bulan lainnya cenderung memiliki penjualan yang lebih rendah dan konsisten, sehingga dikelompokkan bersama dalam kluster 1 (C1). Analisis ini menunjukkan adanya perbedaan signifikan dalam penjualan BBM pada bulan-bulan tertentu yang mungkin disebabkan oleh faktor-faktor musiman atau promosi khusus. Data Hasil *Clustering* Menggunakan Algoritma K-Means ditunjukkan pada tabel 2.

**Tabel 2.** Data hasil *clustering* menggunakan algoritma K-Means

Jenis BBM	Bulan	Hasil Tingkat Penjualan (K-Means)	Kategori Clustering
Pertalite	Februari	Rendah	C1
Pertalite	Maret	Rendah	C1
Pertalite	April	Rendah	C1
Pertalite	Mei	Rendah	C1
Pertalite	Juni	Rendah	C1
Pertalite	Juli	Rendah	C1
Pertalite	Agustus	Rendah	C1
Pertalite	September	Rendah	C1
Pertalite	Oktober	Rendah	C1
Pertalite	November	Rendah	C1
Pertamax	Juni	Rendah	C1
Pertamax	Juli	Rendah	C1
Dexlite	Januari	Rendah	C1
Dexlite	Februari	Rendah	C1
Dexlite	Maret	Rendah	C1
Dexlite	April	Rendah	C1
Dexlite	Mei	Rendah	C1
Dexlite	Juli	Rendah	C1
Dexlite	Agustus	Rendah	C1
Dexlite	September	Rendah	C1
Solar	Maret	Rendah	C1
Solar	April	Rendah	C1
Solar	Juni	Rendah	C1
Solar	Agustus	Rendah	C1
Solar	September	Rendah	C1
Solar	Oktober	Rendah	C1
Solar	November	Rendah	C1
Solar	Desember	Rendah	C1
Pertalite	Januari	Tinggi	C3
Pertamax	November	Tinggi	C3
Pertamax	Desember	Tinggi	C3
Dexlite	November	Tinggi	C3
Solar	Januari	Tinggi	C3

Berdasarkan hasil clustering dengan model K-Means pada penjualan BBM Tahun 2023 dapat disimpulkan sebagai berikut: Penjualan tertinggi jenis pertalite ada di bulan Januari 2023 (C3) sedangkan penjualan terendah ada di bulan Februari sampai dengan November 2023 (C1). Penjualan tertinggi jenis pertamax ada di bulan November dan Desember 2023 (C3) sedangkan penjualan terendah ada di bulan Juni dan Juli 2023 (C1). Penjualan tertinggi jenis dexlite ada di bulan November 2023 (C3) sedangkan penjualan terendah ada di bulan Januari sampai dengan Juni, Agustus dan September 2023 (C1). Penjualan tertinggi jenis solar ada di bulan Januari 2023 (C3) sedangkan penjualan terendah ada di bulan Maret, April, Juni, Agustus sampai dengan Desember 2023 (C1).

### 3.2. Decision Tree dan Naive Bayes (Classification)

Pada bagian ini penelitian akan melakukan prediksi dengan membandingkan 2 algoritma untuk melihat tingkat akurasi dari masing-masing algoritma. setelah penelitian melakukan clustering pada data penjualan untuk mencari tingkatan penjualan, sekarang penelitian akan melakukan prediksi penjualan ditahun berikutnya. Data asli yang digunakan untuk proses klasifikasi ditunjukkan pada tabel 3.

**Tabel 3.** Data asli yang akan diolah pada aplikasi orange

No	Nama	Tipe	Peran	Nilai
1	Jenis BBM	Kategori	Fitur	Dexlite, Pertalite, Pertamax, Solar
2	Bulan	Kategori	Fitur	Januari s/d Desember
3	Status Penjualan Perbulan	Kategori	Fitur	Naik, Turun
4	Kategori Tingkat Penjualan	Kategori	Target	Rendah, Sedang, Tinggi
5	Subsidi Pemerintah	Kategori	Fitur	Tidak, Ya

Dikarenakan yang akan menjadi output ialah tingkat akurasi perhitungan dari kedua algoritma antara Decision Tree dan Naive Bayes, maka penelitian akan menjadikan tingkat penjualan yang didapat dari algoritma K-Means sebagai target untuk melakukan prediksi. Data training ditunjukkan pada tabel 4.

**Tabel 4.** Data *training* yang akan diolah pada aplikasi orange

No	Nama	Tipe	Peran	Nilai
1	Jenis BBM	Kategori	Fitur	Dexlite, Peralite, Pertamina, Solar
2	Bulan	Kategori	Fitur	Januari s/d Desember
3	Status Penjualan Perbulan	Kategori	Fitur	Naik, Turun
4	Subsidi Pemerintah	Kategori	Fitur	Tidak, Ya

Berdasarkan hasil *test and score* antara model naïve bayes dan tree, maka ditemukan hasil evaluasi model seperti tabel 5.

**Tabel 5.** Test and Score Decision Tree dan Naive Bayes

Model	AUC	CA	F1	Prec	Recall
Naïve Bayes	0.695	0.500	0.526	0.598	0.500
Tree	0.639	0.625	0.612	0.611	0.625

Berdasarkan tabel diatas *Area Under Curve* (AUC): Mengukur kemampuan model untuk membedakan antara kelas. Nilai yang lebih tinggi lebih baik (Naive Bayes 0.695 dan Decision Tree 0.639). CA (*Classification Accuracy*): Persentase prediksi yang benar dari total prediksi, (Naive Bayes 0.500 (50%) dan Decision Tree 0.625 (62,5%)). *F1 Score*: Harmonic mean dari precision dan recall, memberikan gambaran keseimbangan antara keduanya (Naive Bayes 0.526 dan Decision Tree 0.612). Precision: Proporsi prediksi positif yang benar-benar positif (Naive Bayes 0.598 dan Decision Tree 0.611). Recall: Proporsi data aktual positif yang berhasil diprediksi dengan benar (Naive Bayes 0.500 dan Decision Tree 0.625).

Decision Tree memiliki nilai *Classification Accuracy* (CA) dan F1 Score yang lebih tinggi daripada Naive Bayes, menunjukkan bahwa model ini lebih akurat dan seimbang dalam prediksi sedangkan Naive Bayes memiliki nilai AUC yang lebih tinggi, menunjukkan bahwa dalam hal membedakan antara kelas, Naive Bayes sedikit lebih baik.

Jika berdasarkan F1 Score, Ketika membandingkan dengan Naive Bayes, Decision Tree memiliki nilai F1 Score yang lebih tinggi (0.612), artinya Decision Tree lebih baik dari Naive Bayes dalam hal F1 Score serta probabilitas bahwa Decision Tree memiliki performa lebih baik dalam F1 Score dibandingkan Naive Bayes adalah 0.990, yang menunjukkan keyakinan tinggi bahwa Decision Tree lebih baik dalam hal ini. Evaluasi confusion matrix ditunjukkan pada gambar 4.

		Predicted			
		Rendah	Sedang	Tinggi	Σ
Actual	Rendah	14	7	7	28
	Sedang	4	7	4	15
	Tinggi	1	1	3	5
Σ		19	15	14	48

**Confusion Matrix Naive Bayes**

		Predicted			
		Rendah	Sedang	Tinggi	Σ
Actual	Rendah	23	3	2	28
	Sedang	7	6	2	15
	Tinggi	2	2	1	5
Σ		32	11	5	48

**Confusion Matrix Decision Tree**

**Gambar 10.** Confusion Matrix Naïve Bayes dan Decision Tree

Pada gambar 10 menjelaskan dua confusion matrix yang membandingkan performa model Naive Bayes dan Decision Tree dalam mengklasifikasikan data ke dalam tiga kategori: Rendah, Sedang, dan Tinggi. Berikut adalah penjelasan mengenai masing-masing confusion matrix dan bagaimana mereka dapat digunakan untuk mengevaluasi kinerja model. Yang pertama Confusion Matrix Naive Bayes, diagonal utama (14,7,3): jumlah prediksi yang benar untuk masing-masing kelas, off diagonal: jumlah kesalahan prediksi misalnya, 7 prediksi yang sebenarnya rendah diklasifikasikan sebagai sedang dan evaluasi kinerja model: accuracy  $(14+7+3)/48 = 24/48 \approx 0.5$  atau 50%. Yang kedua Confusion Decision Tree, diagonal utama (23,6,1): Jumlah prediksi yang benar untuk masing-masing kelas, off diagonal: jumlah kesalahan prediksi Misalnya, 3 prediksi yang sebenarnya rendah diklasifikasikan sebagai sedang dan evaluasi kinerja model: accuracy  $(23+6+1)/48 = 30/48 \approx 0.625$  atau 62.5%.

Analisis perbandingan yang bisa dijelaskan ialah *Accuracy*, Decision Tree memiliki accuracy yang lebih tinggi (62.5%) dibandingkan dengan Naive Bayes (50%) sedangkan kesalahan prediksi, Decision Tree



cenderung membuat lebih sedikit kesalahan dalam memprediksi kelas rendah, tetapi memiliki kesulitan dalam memprediksi kelas tinggi.

#### 4. DISKUSI

Menurut jurnal yang ditulis oleh Jenny Veronika dan Andri dari Universitas Bina Darma Palembang, penerapan algoritma algoritma neural network untuk memprediksi penjualan bahan bakar minyak, kebutuhan akan bahan bakar minyak bervariasi. Jurnal tersebut menjelaskan bahwa algoritma neural network adalah sistem pemrosesan informasi yang dibuat dengan meniru cara otak manusia menyelesaikan masalah, yang dapat digunakan untuk memprediksi penjualan bahan bakar minyak [14].

Jurnal yang ditulis oleh Popy Meilina dari Universitas Muhammadiyah Jakarta membahas penerapan data mining dengan algoritma klasifikasi menggunakan Decision Tree dan Regresi, yang menjelaskan bahwa proses data mining melibatkan proses decision tree, regresi, dan k means untuk clustering, yang ditampilkan dalam bentuk scatter plot. Kelurahan Bajing Kulon, Kedawung, Pekuncen, dan Pesanggarahan adalah kelurahan yang paling membutuhkan bantuan berdasarkan pemilihan pohon. Dalam regresi, atribut per kelurahan digunakan, dengan variabel x (pendidikan atau pekerjaan) dan variabel y (penghasilan) [15].

Dalam jurnalnya yang berjudul "Penerapan data mining menggunakan algoritma naive bayes untuk klasifikasi data penentuan hasil penjualan dalam strategis pemasaran", Muhammad Fakhrul Rozi, Rosyidah Siregar, dan Nenna Irsa Syahputri dari Universitas Harapan Medan menjelaskan bahwa algoritma ini memungkinkan peneliti untuk memprediksi hasil penjualan barang melalui pembagian data berdasarkan kategori yang digunakan. Peneliti membagi data penjualan selama 2,5 tahun terakhir menjadi 900 data training dan 209 data pengujian. Hasil pengujian pengklasifikasi adalah 0.598 pada akurasi, 0.493 pada akurasi pengklasifikasi, 0.472 pada akurasi F1, 0.557 pada ketepatan, dan 0.493 pada recall. Untuk pengujianya, peneliti menggunakan aplikasi orange dan memvisualisasikan hasil dari prediksi penjualan selama setengah tahun berikutnya [16].

Hasil penelitian sebelumnya yang telah dijelaskan oleh penelitian di atas menunjukkan bahwa ada kesamaan antara proses mengekstraksi informasi bermanfaat dan pola tersembunyi dari data yang sangat besar yang menggunakan algoritma dan algoritma tertentu. Tujuan utama penggunaan data mining oleh penelitian adalah untuk menemukan hubungan, pola, dan tren dalam data yang dapat digunakan untuk membuat keputusan yang lebih baik tentang penjualan BBM di SPBU Wadio Nabire di masa mendatang. Untuk menentukan tingkat penjualan tinggi, sedang, atau rendah, penelitian menggunakan teknik data mining clustering K-Means. Kemudian, menggunakan perbandingan algoritma Decision Tree dan Naive Bayes, memisahkan data ke dalam kategori atau kelas yang sudah ditetapkan sebelumnya. Ini digunakan untuk memprediksi tingkat penjualan di tahun berikutnya dengan membuat tiga kelompok: tinggi, sedang, dan rendah.

Adapun pengembangan yang ingin penelitian sampaikan berdasarkan hasil penelitian ini, beberapa hal yang bisa dikembangkan untuk penelitian dan aplikasi lebih lanjut adalah penggunaan data yang lebih besar dan variatif: penelitian ini dapat diperluas dengan menggunakan dataset yang lebih besar dan lebih bervariasi untuk mendapatkan hasil yang lebih *generalizable*. Data yang lebih kaya dapat membantu dalam mendapatkan *insight* yang lebih mendalam. Eksplorasi algoritma clustering lainnya: selain K-Means, algoritma clustering lain seperti dbscan atau *agglomerative clustering* dapat dieksplorasi untuk melihat apakah mereka memberikan hasil yang lebih baik dalam konteks data penjualan BBM [17]. Penyesuaian *hyperparameter*: melakukan *tuning hyperparameter* pada model decision tree dan naive bayes dapat membantu meningkatkan akurasi dan performa model [18]. Teknik seperti *cross-validation* dapat digunakan untuk menemukan kombinasi *hyperparameter* yang optimal. Integrasi dengan analisis bisnis: hasil clustering dan klasifikasi dapat digunakan lebih lanjut dalam analisis bisnis untuk strategi pemasaran yang lebih efektif, misalnya memahami segmen pelanggan dapat membantu dalam merancang kampanye promosi yang lebih terarah [19]. Implementasi model ensemble: menggabungkan beberapa model (*ensemble methods*) seperti *random forest* atau *gradient boosting* dapat dieksplorasi untuk melihat apakah pendekatan ini dapat meningkatkan performa klasifikasi dibandingkan dengan model individual. Penggunaan teknik pra-pemrosesan data yang lebih lanjut: menerapkan teknik pra-pemrosesan data yang lebih canggih seperti normalisasi, penghapusan *outlier*, dan *feature engineering* dapat membantu meningkatkan kualitas *clustering* dan klasifikasi [20].

#### 5. KESIMPULAN

Penelitian ini mengkaji penerapan algoritma K-Means untuk clustering penjualan BBM serta mengevaluasi akurasi model Decision Tree dan Naive Bayes pada data mining menggunakan aplikasi Orange. Berdasarkan hasil eksperimen dan analisis yang dilakukan, diperoleh beberapa temuan penting sebagai berikut, clustering dengan K-Means: algoritma K-Means berhasil mengelompokkan data penjualan bbm ke dalam beberapa cluster yang representatif. Hasil clustering ini membantu dalam mengidentifikasi pola dan segmentasi pelanggan yang berbeda berdasarkan data penjualan. Evaluasi model klasifikasi: decision tree dan naive bayes digunakan untuk mengevaluasi akurasi klasifikasi data penjualan bbm. Penggunaan aplikasi orange: aplikasi orange terbukti efektif dan *user-friendly* dalam penerapan algoritma K-Means serta evaluasi model decision

tree dan naive bayes. Alat visualisasi dan antarmuka yang intuitif membantu dalam proses eksplorasi dan analisis data.

Penelitian ini memiliki kelebihan dalam menyajikan perbandingan langsung antara tiga algoritma yang berbeda seperti model K-Means, Naïve Bayes, dan Decision Tree, dalam konteks memprediksi penjualan BBM di SPBU Wadio Nabire. Pendekatan ini dapat memberikan wawasan mendalam tentang keefektifan masing-masing algoritma dalam aplikasi praktis yang spesifik. Namun, kelemahannya mungkin terletak pada kompleksitas analisis yang diperlukan untuk membandingkan kinerja ketiga algoritma tersebut secara menyeluruh, serta tantangan dalam memastikan bahwa data yang digunakan mencerminkan kondisi yang representatif di lapangan, khususnya dalam konteks varian lokal seperti Papua Tengah.

## REFERENSI

- [1] A. Falakhi, "Pengolahan Data Pelanggan Dengan Teknik Clustering K-Means Di Aplikasi Weka," *Journal Computer Science and Information Systems : J-Cosys*, vol. 3, no. 2, pp. 54–60, Jul. 2023, doi: <https://doi.org/10.53514/jco.v3i2.394>.
- [2] T. M. Dista and F. F. Abdulloh, "Clustering Pengunjung Mall Menggunakan Algoritma K-Means dan Particle Swarm Optimization," *Jurnal Media Informatika Budidarma*, vol. 6, no. 3, p. 1339, Jul. 2022, doi: <https://doi.org/10.30865/mib.v6i3.4172>.
- [3] F. Aulia Safrin, "Analisis Strategi Pemasaran Menggunakan Business Model Canvas dalam Meningkatkan Penjualan (Studi pada UMKM Bobabox Medan)," 2022. doi: <https://doi.org/10.57251/reg.v1i3.266>.
- [4] A. H. Stmik and A. Malang, "Segmentasi Pelanggan Internet Service Provider (ISP) Berbasis Pillar K-Means," *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 13, no. 2, 2019, doi: <https://doi.org/10.32815/jitika.v13i2.413>.
- [5] Riski Putri Anjayani and Intan Rike Febriyanti, "Analisis Strategi Bauran Pemasaran untuk Meningkatkan Volume Penjualan (Studi Pada UMKM YJS Sukabumi)," *Jurnal Multidisiplin Madani*, vol. 2, no. 9, pp. 3610–3617, Sep. 2022, doi: <https://doi.org/10.55927/mudima.v2i9.1175>.
- [6] D. Anjainah and S. Monalisa, "Analisis Rekomendasi Produk Berdasarkan Segmentasi Pelanggan Menggunakan Algoritma DBSCAN dan FP-Growth Product Recommendation Analysis Based on Customer Segmentation Using DBSCAN and FP-Growth Algorithm," 2022. Accessed: Jul. 25, 2024. [Online]. Available: <https://core.ac.uk/download/pdf/548563124.pdf>
- [7] I. Hoerunisa and D. Sukanta, "Penerapan Model Antrian Multi Channel-Single Phase pada SPBU Sempu Jurong Cikarang Utara," 2021. doi: <https://doi.org/10.33592/unistek.v8i1.1202>.
- [8] Suroto Munahar, Bagiyo Condro Purnomo, and Andi Widiyanto, "Implementasi Pemanfaatan Biogas Sebagai Energi Pendingin Susu Di Kecamatan Cepogo Kabupaten Boyolali," *Jati Emas (Jurnal Aplikasi Teknik dan Pengabdian Masyarakat)*, vol. 3, 2019, Accessed: Jul. 25, 2024. [Online]. Available: <https://www.academia.edu/download/74653390/154.pdf>
- [9] I. Surjandari, A. Rachman, F. Dianawati, and R. P. Wibowo, "Oil fuel delivery optimization for multi product and multi depot: the case of petrol station replenishment problem (PSRP)," in *International Conference on Graphic and Image Processing (ICGIP 2011)*, Y. Xie and Y. Zheng, Eds., SPIE, 2011, p. 82853Q. doi: <https://doi.org/10.1117/12.914444>.
- [10] M. A. Noor and P. M. A. Saputra, "Emisi Karbon dan Produk Domestik Bruto: Investigasi Hipotesis Environmental Kuznets Curve (EKC) pada Negara Berpendapatan Menengah di Kawasan ASEAN," *Jurnal Wilayah dan Lingkungan*, vol. 8, no. 3, pp. 230–246, Dec. 2020, doi: <http://dx.doi.org/10.14710/jwl.8.3.230-246>.
- [11] K. Pola *et al.*, "Clustering Sales Patterns of Best Selling and Less Selling Products at El Jhon Bengkulu Stores Using the K-Medoid Method," *JURNAL KOMITEK*, vol. 2, no. 2, pp. 637–642, 2022, doi: <https://doi.org/10.53697/jkomitek.v2i2.1048>.
- [12] E. Dwika Putra, M. Husni Rifqo, and D. Hardianto, "Apriori Algorithm Implementation on Market Basket Analysis (MBA) of Mobile Phone Accessories Implementasi Algoritma Apriori Pada Market Basket Analysis (MBA) Aksesoris Telepon Seluler," *JURNAL KOMITEK*, vol. 2, no. 2, pp. 373–382, 2022, doi: [10.53697/jkomitek.v2i2](https://doi.org/10.53697/jkomitek.v2i2).
- [13] N. Kadek, C. Bandinithya, D. #1, A. Wirdiani, D. Made, and S. Arsa, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Klasifikasi Kecanduan Smartphone pada Pelajar Sekolah Menengah Atas menggunakan Algoritma Machine Learning Berbasis Feature Weighting," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 8, no. 1, 2022, Accessed: Jul. 25, 2024. [Online]. Available: <https://pdfs.semanticscholar.org/0f5c/62833c7dba58ae233a9a18c9587d7a352478.pdf>
- [14] J. Veronika, "Penerapan Algoritma Algoritma Neural Network Untuk Memprediksi Penjualan Bahan Bakar Minyak," 2022. doi: <https://doi.org/10.51519/journal022>.
- [15] P. Meilina, "Penerapan Data Mining Dengan Algoritma Kalsifikasi Menggunakan Decision Tree Dan Regresi," 2015. doi: <https://doi.org/10.24853/jurtek.7.1.11-20>.

- [16] F. Rozi, M. Fakhrol Rozi, R. Siregar, N. I. Syahputri, F. Teknik, and D. Komputer, "Penerapan Data Mining Menggunakan Algoritma Naive Bayes Untuk Klasifikasi Data Penentuan Hasil Penjualan Dalam Strategi Pemasaran," Online, 2023. doi: <https://doi.org/10.62712/juktisi.v2i2.137>.
- [17] Y. Heryadi and T. Wahyono, "Machine learning konsep dan implementasi," *Yogyakarta: Gava Media*, 2020.
- [18] W. Budiharto, "Machine learning & computational intelligence," *ANDI, Yogyakarta*, 2016.
- [19] Wahyono and Teguh, "Fundamental of Python for Machine Learning: Dasar-Dasar Pemrograman Python untuk Machine Learning dan Kecerdasan Buatan," *Gava Media*, 2018, Accessed: Jul. 25, 2024. [Online]. Available: <https://www.researchgate.net/>
- [20] W. Budiharto, "Pemrograman Python Untuk Ilmu Komputer Dan Teknik," *ANDI, Yogyakarta*, 2018.
- [21] A. R. Rizalde, H. A. Mubarak, G. Ramadhan, and Mohd. A. Fatan, "Comparison of K-Means, BIRCH and Hierarchical Clustering Algorithms in Clustering OCD Symptom Data," *Public Research Journal of Engineering, Data Technology and Computer Science*, vol. 1, no. 2, pp. 102–108, Feb. 2024, doi: 10.57152/predatecs.v1i2.1106.
- [22] A. F. Lubis et al., "Classification of Diabetes Mellitus Sufferers Eating Patterns Using K-Nearest Neighbors, Naive Bayes and Decision Tree," *Public Research Journal of Engineering, Data Technology and Computer Science*, vol. 2, no. 1, pp. 44–51, Apr. 2024, doi: 10.57152/predatecs.v2i1.1103.
- [23] A. I. Putri et al., "Implementation of K-Nearest Neighbors, Naive Bayes Classifier, Support Vector Machine and Decision Tree Algorithms for Obesity Risk Prediction," *Public Research Journal of Engineering, Data Technology and Computer Science*, vol. 2, no. 1, pp. 26–33, Apr. 2024, doi: 10.57152/predatecs.v2i1.1110.