

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Poltak Sianturi, Marsono, dan Rina Mahyuni melakukan penelitian ini dengan judul “Analisis Data Mining Menggunakan Algoritma K-Means Untuk Mengelompokkan Menu Makan Yang Potensial Pada Kafe Minum Kopi”. Kesulitan dalam menentukan cara mengelompokkan potensi penjualan menu makanan di Cafe Minum Kopi diselesaikan secara efisien dengan menggunakan algoritma *K-Means Clustering*. Metode ini membutuhkan informasi jumlah penjualan yang terjadi setiap bulannya (Sianturi & Mahyuni, 2020).

Penelitian yang berjudul “Analisis Penikmat Kopi Berdasarkan Cita Rasa Kopi Menggunakan Teknik Clustering Algoritma K-Means di Cirebon”. Setelah proses *clustering* berhasil diselesaikan, didapatkan hasil *cluster* 1 sampai 3. Dari hasil tersebut terlihat bahwa karakteristik penikmat kopi pada ketiga *cluster* tersebut didominasi oleh penikmat kopi laki-laki yang umumnya berusia antara 23 hingga 30 tahun dan bekerja sebagai karyawan swasta. Rasa favorit mereka adalah *vanila*, *floral* (seperti bunga melati), *nutty* (seperti almond), *fruity* (seperti apel), *roasted* (seperti tembakau), *spicy* (seperti cengkeh), dan *bitter* (seperti rasa lainnya)(Irma Purnamasari & Danar Dana, 2017).

Penelitian lain berjudul “Analisis Segmentasi Pasar Konsumen Kopi Hitam di Kota Padangsidempuan.” secara efektif Formula K-Means Penelitian “Analisis Segmenting, Targeting, dan Positioning dalam Penjualan Kopi di Cafe Kancakona Kopi Jember”, dengan menggunakan karakteristik demografis, geografis, psikografis, dan perilaku, BMT UGT Sidogiri Cabang Pembantu Mumbulsari secara efektif

menangani masalah teknik segmentasi pasar. BMT memberikan penekanan yang lebih besar pada variabel demografis saat mengembangkan strategi segmentasi pasar untuk tujuan memperkenalkan produk baru dan mengembangkan basis nasabah. Seperti yang ditunjukkan oleh peningkatan jumlah tabungan setiap tahunnya menjadi rata-rata 28,002%, jumlah nasabah pembiayaan menjadi 25,31%, dan DPK sebesar 16,30%, dampak dari penggunaan segmentasi pasar oleh BMT terkait dengan peningkatan konsumen (Harahap et al., 2017).

Masalah tersebut diselesaikan secara efektif dalam penelitian “Toko Tani Indonesia menggunakan Metode K-Means Clustering untuk menganalisa penjualan komoditas”. Menurut temuan penelitian, sistem operasi dan operasional di Toko Tani Indonesia di Kalimantan Barat masih bersifat manual. Selain itu, sulit bagi pemilik toko untuk memisahkan barang yang tidak laku dan yang laku. Karena ada begitu banyak barang yang masuk dan keluar setiap harinya, hal ini tentu saja menjadi tantangan tersendiri (Abid et al., 2022).

Sartika Sitohang, Darjat Saripurna, dan Widiarti Ristamaya melakukan penelitian “Data Mining untuk Pengelompokan Data Penjualan Kue.” Di Jofie Bakery, kami menggunakan Algoritma *K-means Clustering* secara efektif menyelesaikan masalah. Langkah terakhir dalam proses implementasi adalah implementasi sistem, di mana seluruh sistem akan dioperasikan. Untuk memastikan bahwa tidak ada masalah saat sistem digunakan, pengujian harus dilakukan sebelum sistem dapat digunakan dengan baik. Aplikasi resmi, instalasi, pengujian data, dan mulai menggunakan sistem baru adalah berbagai langkah dan prosedur yang membentuk proses implementasi yang dilakukan untuk menyelesaikan analisis (Saripurna & Ristamaya, 2021).

“Implementasi Metode K-Means untuk Mengelompokkan Produk yang Kurang Populer Menggunakan Data Penjualan” adalah judul penelitian yang dilakukan oleh Ardi Mardiana. secara efektif menyelesaikan masalah dengan pendekatan ini, yang melibatkan pengelompokan data menurut karakter, dengan karakter yang sama masuk ke dalam satu kelompok dan karakter data dibagi ke dalam berbagai kelompok. Keuntungan dari pengelompokan data ini adalah meminimalkan fungsi objektif yang dipilih selama proses pengelompokan, yang biasanya berusaha untuk memaksimalkan variasi data di antara kelompok-kelompok dan meminimalkan variasi data di dalam kelompok (*Metodologi K-Mean*, n.d.).

Penelitian lain yang berjudul “Data Mining untuk Mengelompokkan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan secara efektif menyelesaikan masalah tersebut. Data penjualan dari Dpom Coffee pada tahun 2021 menjadi dataset yang digunakan dalam penelitian ini. Terdapat dua variabel dan lima puluh lima data menu secara keseluruhan. Menemukan jumlah K, atau *cluster*, adalah tahap pertama dalam algoritma k-means. Selanjutnya, secara acak mengidentifikasi pusat *cluster*, atau *centroid*. Davies Blouldin Index (DBI) digunakan sebagai parameter untuk menilai akurasi atau kinerja algoritma *k-means*. Nilai DBI yang rendah dihasilkan oleh algoritma DBI ketika temuan klaster didasarkan pada jarak antar-klaster yang rendah (kemiripan antar-klaster yang tinggi) atau jarak antar-klaster yang tinggi (kemiripan antar-klaster yang rendah) (Triyandana et al., 2022).

Penelitian selanjutnya berjudul “Strategi Segmenting, Targeting, Positioning (STP).” berhasil menyelesaikan masalah tersebut. Dengan 54 responden, atau 49% dari keseluruhan responden, karakteristik responden mengindikasikan bahwa mereka mengetahui produk Kopi Poso paling banyak dari teman, keluarga, dan tetangga. Sembilan puluh tujuh persen (107 orang) dari responden survei berpendapat bahwa

harga produk Kopi Poso wajar. Pelanggan akan memilih barang yang mereka yakini dibuat dengan baik dan dengan harga yang wajar. Dengan 43 tanggapan, atau 39%, tingkat frekuensi maksimum konsumsi kopi dalam sehari adalah dua kali. Dari total 60 responden, atau 55%, alasan utama responden minum kopi adalah karena mereka menikmatinya. Bagi mereka yang menikmatinya, minum kopi telah menjadi gaya hidup, bukan hanya karena memberikan manfaat kesehatan (Wutabisu et al., 2021).

Penelitian lain berjudul “Penggalian Data Pemasaran Produk Menggunakan Metode Clustering” PT Cipta Niaga semesta mampu meramalkan jumlah stok produk yang dibutuhkan untuk bulan yang akan datang dengan mengelompokkan produk yang dipasarkan setelah masalah tersebut diatasi secara efektif dengan memanfaatkan program *Data Mining*. Bisnis akan dapat menentukan produk mana yang diminati oleh pelanggan dengan menggunakan aplikasi *Data Mining* untuk mendapatkan wawasan baru tentang penjualan produk. Sebagai contoh, implementasi program produk Torabika telah menunjukkan bahwa pelanggan lebih tertarik dengan rasa moka dan cappuccino daripada rasa lainnya, dan bahwa merek energen lebih populer dalam hal rasa coklat dan kacang hijau (Melladia et al., 2022).

Penelitian ini berjudul “Data Mining untuk Seleksi Calon Nasabah Menggunakan Algoritma K-Means” Calon klien sangat penting dalam pembuatan rencana bisnis karena mereka mendatangkan uang bagi perusahaan. Untuk itu, kami bertujuan untuk mengenal klien kami dengan baik. Bisnis dapat membuat keputusan tentang klien potensial ketika mereka memiliki pemahaman yang kuat tentang pembeli. Tantangan dalam menganalisis nilai klien adalah masalah yang umum terjadi. Menemukan pelanggan atau konsumen yang ideal dapat menjadi tantangan bagi banyak pemasar. Klien potensial dapat hilang sebagai akibatnya, yang akan sangat merugikan bisnis (Romadhona et al., 2022).

Penelitian berikut dengan judul, “Segmentasi Pelanggan Menggunakan Algoritma K-Means Berdasarkan Recency Frequency Monetary Model.” Pada tahun 2020, ketidakteraturan data akibat pandemi COVID19 berdampak pada penjualan perusahaan (Pramudiansyah, 2021).

Masalah tersebut berhasil diselesaikan dalam penelitian “Analisis Segmentasi Konsumen menggunakan Algoritma K-Means Clustering”. Data penjualan dengan jenis nominal - seperti jenis motor, jenis pembelian, dan pekerjaan - harus diinisialisasi terlebih dahulu sebelum k-means clustering dapat digunakan pada data tersebut. Data harus dikelompokkan sebelum memulai proses inisialisasi. (Eno Ketherin et al., 2018).

Penelitian ini dengan judul “Implementasi Algoritma K-Means untuk Pengelompokan Data Produktivitas Kelapa Sawit”, berhasil menyelesaikan permasalahan berdasarkan penelitian dimana hasil panen kelapa sawit dikelompokkan dengan menggunakan dua cluster; algoritma ini dapat digunakan untuk mengelompokkan hasil panen kelapa sawit. Selain itu, algoritma ini juga dapat digunakan sebagai alat rapidminer. Sebagai hasil dari penelitian ini, C0 dan C1 mampu mengumpulkan 13, dan hingga 35, informasi Blok Panen. Cara pengelompokan hasil panen kelapa sawit di Rapidminer cocok dengan hasil manual dalam hal ini (Ramadhani, 2023).

Penelitian berjudul “Algoritma K-Means Clustering untuk Segmentasi Pelanggan.” masalah pengumpulan data transaksi dalam jangka waktu yang ditentukan untuk diproses secara efektif diselesaikan. Bagian penjualan Fashion Viral Solo memberikan penulis data rekap transaksi online untuk bulan Februari 2023 (Ardi et al., 2023) yang telah dibersihkan untuk menghilangkan informasi dan atribut yang hilang. Pada bulan Februari 2023, penulis mengumpulkan total 851 ringkasan transaksi.

Erik Purnama dan Afifudin melakukan penelitian ini, “Analisis Segmentasi Pelanggan Mall.” Menerapkan Algoritma *Clustering K-Means* secara efektif menyelesaikan masalah Volume data meningkat secara eksponensial dengan data konsumen. Model pengelompokan semacam ini harus mampu menangani data yang sangat besar ini secara efisien. Prosedur *k-Means* diterapkan setelah bobot data ditentukan. Menemukan jumlah cluster adalah langkah pertama dalam perhitungan *k-Means*. Selanjutnya, jarak dihitung antara setiap titik objek yang dipilih dengan titik-titik lainnya(Erik, 2022).

Tabel 2.1 Penelitian Terdahulu

No	Nama	Judul	Metode	Hasil
1.	(Irma Purnamasari & Danar Dana, 2017)	Analisis Data Mining Menggunakan Algoritma K-Means Untuk Mengelompokkan Menu Makan Yang Potensial Pada Kafe Minum Kopi	Algoritma K-Means	Pendekatan K-Means Clustering yang membutuhkan jumlah penjualan bulanan dapat digunakan untuk menyelesaikan masalah dalam menentukan pengelompokan kemungkinan penjualan menu makanan di Cafe Minum Kopi.

2.	(Irma Purnamasari & Dinar Dana, 2017)	<p>Analisis Penikmat Kopi Berdasarkan Cita Rasa Kopi Menggunakan Teknik Clustering Algoritma K-Means di Cirebon</p>	Algoritma K-Means	<p>Setelah proses <i>clustering</i> berhasil diselesaikan, didapatkan hasil <i>cluster</i> 1 sampai 3. Dari hasil tersebut terlihat bahwa karakteristik penikmat kopi pada ketiga <i>cluster</i> tersebut didominasi oleh penikmat kopi laki-laki yang umumnya berusia antara 23 hingga 30 tahun dan bekerja sebagai karyawan swasta. Rasa favorit mereka adalah vanila, <i>floral</i> (seperti bunga melati), <i>nutty</i> (seperti almond), <i>fruity</i> (seperti apel), <i>roasted</i> (seperti tembakau), <i>spicy</i> (seperti cengkeh), dan <i>bitter</i> (seperti rasa lainnya)</p>
----	---------------------------------------	---	-------------------	---

3.	(Harahap et al., 2017)	Analisis Segmentasi Pasar Konsumen Kopi Hitam di Kota Padangsidempuan	Algoritma K-Means	dengan menggunakan karakteristik demografis, geografis, psikografis, dan perilaku, BMT UGT Sidogiri Cabang Pembantu Mumbulsari secara efektif menangani masalah teknik segmentasi pasar. BMT memberikan penekanan yang lebih besar pada variabel demografis saat mengembangkan strategi segmentasi pasar untuk tujuan
----	------------------------	---	-------------------	---

				<p>memperkenalkan produk baru dan mengembangkan basis nasabah. Seperti yang ditunjukkan oleh peningkatan jumlah tabungan setiap tahunnya menjadi rata-rata 28,002%, jumlah nasabah pembiayaan menjadi 25,31%, dan DPK sebesar 16,30%, dampak dari penggunaan segmentasi pasar oleh BMT terkait dengan peningkatan konsumen</p>
--	--	--	--	--

4.	(Abid et al., 2022)	Toko Tani Indonesia menggunakan Metode K-Means Clustering untuk menganalisa penjualan komoditas	Algoritma K-Means	sistem operasi dan operasional di Toko Tani Indonesia di Kalimantan Barat masih bersifat manual. Selain itu, sulit bagi pemilik toko untuk memisahkan barang yang tidak laku dan yang laku. Karena ada begitu banyak barang yang masuk dan keluar setiap harinya, hal ini tentu saja menjadi tantangan tersendiri
5.	(Saripurna & Ristamaya, 2021)	<i>Data Mining</i> Untuk Pengelompokan Data Penjualan Cake Dengan Menggunakan Algoritma K-means <i>Clustering</i> Pada Jofie Bakery	Algoritma K-Means	Langkah terakhir dalam proses implementasi adalah implementasi sistem, di mana seluruh sistem akan dioperasikan. Untuk memastikan bahwa tidak ada masalah saat sistem digunakan, pengujian harus dilakukan sebelum sistem dapat digunakan dengan baik.

				Aplikasi resmi, instalasi, pengujian data, dan mulai menggunakan sistem baru adalah berbagai langkah dan prosedur yang membentuk proses implementasi yang dilakukan untuk menyelesaikan analisis
6.	(<i>Metodologi K-Mean, n.d.</i>)	Implementasi Metode K-Means untuk Mengelompokkan Produk yang Kurang Populer Menggunakan Data Penjualan	Algoritma K-Means	secara efektif menyelesaikan masalah dengan pendekatan ini, yang melibatkan pengelompokan data menurut karakter, dengan karakter yang sama masuk ke dalam satu kelompok dan karakter data dibagi ke dalam berbagai kelompok. Keuntungan dari pengelompokan data ini adalah meminimalkan fungsi objektif yang dipilih selama proses pengelompokan, yang

				<p>biasanya berusaha untuk memaksimalkan variasi data di antara kelompok-kelompok dan meminimalkan variasi data di dalam kelompok</p>
7.	(Triyandana et al., 2022)	Data Mining untuk Mengelompokkan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan	Algoritma K-Means	<p>Menemukan jumlah K, atau <i>cluster</i>, adalah tahap pertama dalam algoritma k-means. Selanjutnya, secara acak mengidentifikasi pusat <i>cluster</i>, atau <i>centroid</i>.</p> <p>Davies Blouldin Index (DBI) digunakan sebagai parameter untuk menilai akurasi atau kinerja algoritma <i>k-means</i>. Nilai DBI yang rendah dihasilkan oleh algoritma DBI ketika temuan klaster didasarkan pada jarak antar-klaster yang rendah (kemiripan</p>

				antar-klaster yang tinggi) atau jarak antar-klaster yang tinggi (kemiripan antar-klaster yang rendah)
--	--	--	--	--

8.	(Wutabisu et al., 2021)	Strategi Segmenting, Targeting, Positioning (STP)	Algoritma K- Means	<p>Dengan 54 responden, atau 49% dari keseluruhan responden, karakteristik responden mengindikasikan bahwa mereka mengetahui produk Kopi Poso paling banyak dari teman, keluarga, dan tetangga. Sembilan puluh tujuh persen (107 orang) dari responden survei berpendapat bahwa harga produk Kopi Poso wajar. Pelanggan akan memilih barang yang mereka yakini dibuat dengan baik dan dengan harga yang wajar. Dengan 43 tanggapan, atau 39%, tingkat frekuensi maksimum konsumsi kopi dalam sehari adalah dua kali. Dari total 60 responden, atau 55%, alasan utama responden minum</p>
----	-------------------------	---	--------------------	--

				<p>kopi adalah karena mereka menikmatinya. Bagi mereka yang menikmatinya, minum kopi telah menjadi gaya hidup, bukan hanya karena memberikan manfaat kesehatan</p>
--	--	--	--	--

9.	(Melladia et al., 2022)	Penggalian Data Pemasaran Produk Menggunakan Metode Clustering	Algoritma K-Means	PT Cipta Niaga semesta mampu meramalkan jumlah stok produk yang dibutuhkan untuk bulan yang akan datang dengan mengelompokkan produk yang dipasarkan setelah masalah tersebut diatasi secara efektif dengan memanfaatkan program <i>Data Mining</i> . Bisnis akan dapat menentukan produk mana yang diminati oleh pelanggan dengan menggunakan aplikasi <i>Data Mining</i> untuk mendapatkan wawasan baru tentang penjualan produk. Sebagai contoh, implementasi program produk Torabika telah menunjukkan bahwa pelanggan lebih tertarik dengan rasa moka dan cappuccino daripada rasa
----	-------------------------	--	-------------------	---

				lainnya, dan bahwa merek energen lebih populer dalam hal rasa coklat dan kacang hijau
--	--	--	--	---

10.	(Romadhona et al., 2022)	Data Mining untuk Seleksi Calon Nasabah Menggunakan Algoritma K-Means	Algoritma k-means	Calon klien sangat penting dalam pembuatan rencana bisnis karena mereka mendatangkan uang bagi perusahaan. Untuk itu, kami bertujuan untuk mengenal klien kami dengan baik. Bisnis dapat membuat keputusan tentang klien potensial ketika mereka memiliki pemahaman yang kuat tentang pembeli. Tantangan dalam menganalisis nilai klien adalah masalah yang umum terjadi. Menemukan pelanggan atau konsumen yang ideal dapat menjadi tantangan bagi banyak pemasar. Klien potensial dapat hilang sebagai akibatnya, yang akan sangat merugikan bisnis
-----	--------------------------	--	-------------------	---

11.	(Pramudiansyah, 2021)	Segmentasi Pelanggan Menggunakan Algoritma K-Means Berdasarkan Recency Frequency Monetary Model	Algoritma K-Means	Pada tahun 2020, ketidakteraturan data akibat pandemi COVID19 berdampak pada penjualan perusahaan
12.	(Eno Ketherin et al., 2018)	Analisis Segmentasi Konsumen menggunakan Algoritma K-Means Clustering	Algoritma K-Means	Data penjualan dengan jenis nominal - seperti jenis motor, jenis pembelian, dan pekerjaan - harus diinisialisasi terlebih dahulu sebelum k-means clustering dapat digunakan pada data tersebut. Data harus dikelompokkan sebelum memulai proses inisialisasi

13.	(Ramadhani, 2023)	Implementasi Algoritma K-Means untuk Pengelompokan Data Produktivitas Kelapa Sawit	Algoritma K-Means	berhasil menyelesaikan permasalahan berdasarkan penelitian dimana hasil panen kelapa sawit dikelompokkan dengan menggunakan dua cluster; algoritma ini dapat digunakan untuk mengelompokkan hasil panen kelapa sawit. Selain itu, algoritma ini juga dapat digunakan sebagai alat rapidminer. Sebagai hasil dari penelitian ini, C0 dan C1 mampu mengumpulkan 13, dan hingga 35, informasi Blok Panen. Cara pengelompokan hasil panen kelapa sawit di Rapidminer cocok dengan hasil manual dalam hal ini
S1 4.	(Ardi et al., 2023)	Algoritma K-Means Clustering untuk	Algoritma K-Means	Pada bulan Februari 2023, penulis mengumpulkan

		Segmentasi Pelanggan		total 851 ringkasan transaksi.
15.	(Erik, 2022)	Analisis Segmentasi Pelanggan Mall.	Algoritma K-Means	<p>Menerapkan Algoritma Clustering K-Means secara efektif menyelesaikan masalah Volume data meningkat secara eksponensial dengan data konsumen. Model pengelompokan semacam ini harus mampu menangani data yang sangat besar ini secara efisien. Prosedur k-Means diterapkan setelah bobot data ditentukan.</p> <p>Menemukan jumlah cluster adalah langkah pertama dalam perhitungan k-Means. Selanjutnya, jarak dihitung antara setiap titik objek yang dipilih dengan titik-titik lainnya</p>

Mengutip penelitian “Analisis Data Mining Menggunakan Algoritma *K-Means* dalam Mengelompokkan Menu Makanan yang Potensial pada Cafe Minum Kopi” (Irma Purnamasari & Danar Dana, 2017). Penelitian Irma Purnamasari dan Danar hanya mengidentifikasi pengelompokan penjualan menu makanan, diperlukan analisis lebih lanjut untuk mengidentifikasi pola penjualan produk Kopi Flores Arabika Bajawa. *Orange* merupakan aplikasi pengolah algoritma *K-Means*. Algoritma *K-Means* akan digunakan sebagai pendekatan dengan bantuan aplikasi perangkat lunak.

2.2 Teori Penunjang

1. Data Mining

Data mining hanyalah sebuah langkah tambahan untuk mengekstraksi informasi penting yang tersirat dan tidak diketahui. Selain itu, *Data Mining* bersinggungan dengan berbagai topik, termasuk statistik, pembelajaran mesin, pengenalan pola, teknik komputasi, teknologi basis data, dan komputasi berkinerja tinggi (Mardi Hardjianto, 2021).

2. Orange Data Mining Tools

Orange adalah alat penggalian data yang digunakan dalam investigasi ini. Orange adalah alat penggalian data sumber terbuka. Orange mencakup widget yang berfungsi sebagai unit komputasi untuk membaca, memproses, memvisualisasikan, menganalisis, dan mengeksplorasi

data, di antaranya. Widget diatur sedemikian rupa sehingga membentuk alur kerja dan berkomunikasi satu sama lain dalam lingkungan Orange. Widget data memungkinkan Orange untuk mengubah data teks dan visual.

Analisis gambar adalah add-on yang diperlukan untuk menganalisis foto menggunakan Orange. Orange mengubah data gambar menjadi representasi vektor dengan melatih jaringan saraf dalam pada sejumlah besar gambar, sehingga menghasilkan data yang dapat dianalisis dan digunakan untuk pembelajaran mesin (Hartono et al., 2020).

3. Produksi

Pengertian produksi adalah menghasilkan pembuatan barang produk dan jasa dari bahan baku atau komponen dalam suatu periode waktu dan memiliki nilai tambah bagi perusahaan atau industri kecil dengan kata lain industri rumahan.

4. Kopi

Kopi adalah minuman yang penting bagi sebagian besar orang di seluruh dunia, tidak hanya untuk kenikmatan para peminum kopi, tetapi juga untuk nilai

ekonomi yang diberikan kepada negara-negara yang memproduksi dan mengekspor biji kopi.

5. Kopi Arabika

Salah satu dari dua spesies utama pohon kopi yang banyak ditanam adalah *Coffea arabica*, yang merupakan sumber kopi Arabika. Varietas kopi yang paling banyak dikonsumsi dan disukai di seluruh dunia adalah arabika. Dibandingkan dengan jenis kopi lainnya, termasuk Robusta, rasa biji kopi arabika lebih bernuansa dan kompleks. Kopi arabika biasanya memiliki tingkat keasaman yang lebih kuat dan aroma yang lebih beragam; beberapa jenis memiliki rasa mulai dari buah hingga bunga. Kopi Arabika sering dianggap sebagai kopi premium, dan merupakan dasar dari banyak kopi spesial.