

## PENERAPAN K-MEANS CLUSTERING DAN CROSS-INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM) UNTUK MENGELOMPOKAN PENJUALAN KUE

Muhammad Rafi Muttaqin<sup>1,\*</sup>, Teguh Iman Hermanto<sup>2</sup>, Muhamad Agus Sunandar<sup>3</sup>

<sup>1,2,3</sup> Program Studi Teknik Informatika, Sekolah Tinggi Teknologi Wastukencana, Purwakarta, Indonesia

\*Corresponding Author: [rafi@stt-wastukencana.ac.id](mailto:rafi@stt-wastukencana.ac.id)

### Abstrak

Permasalahan yang terjadi pada Toko Kue adalah terdapatnya sis kue yang tidak terjual selama waktu tertentu sehingga kue menjadi tidak bisa dikonsumsi karena sudah kadaluarsa. Produsen kue harus dapat memperkirakan kue yang akan habis dan tidak pada waktu tertentu sehingga dapat memproduksi kue secara efektif dan efisien. Tujuan penelitian ini adalah untuk membantu produsen agar dapat memprediksi jenis kue yang laris atau habis pada waktu tertentu menggunakan metode data mining. Metode yang digunakan adalah algoritma K-Means. Data yang digunakan berasal dari histori penjualan kue selama satu tahun. Proses data mining yang digunakan dalam penelitian ini adalah Cross-Industry Standard Process for Data Mining (CRISP-DM). Hasil yang diperoleh adalah terbentuknya kelompok penjualan kue yang banyak dibeli oleh konsumen dalam setiap tanggal. Pengelompokan ini dibedakan menjadi 3, yaitu penjualan rendah, menengah, dan tinggi. Hal ini akan membantu produsen untuk mempersiapkan produknya dengan lebih efektif dan efisien sehingga mengurangi hasil produksi yang terbuang percuma. Jika kue masuk kelompok penjualan rendah maka jumlah produk kue tersebut sedikit. Begitu sebaliknya, jika ada kue yang masuk ke kelompok penjualan tinggi, maka produsen akan memproduksi kue tersebut dengan jumlah banyak.

**Kata kunci:** CRISP-DM; K-Means; Penjualan

### Abstract

The problem that occurs at the Cake Shop is that there are cakes that are not sold for a certain time so that the cake cannot be consumed because it has expired. Cake manufacturers must be able to predict which cakes will run out and not at a certain time so that they can produce cakes effectively and efficiently. The purpose of this study is to help producers predict the types of cakes that are selling or running out at a certain time using data mining methods. The method used is the K-Means algorithm. The data used comes from the history of cake sales for one year. The data mining process used in this research is the Cross-Industry Standard Process for Data Mining (CRISP-DM). The results obtained are the formation of cake sales groups that are purchased by consumers on each date. This grouping is divided into 3, namely low, medium, and high sales. This will help producers to prepare their products more effectively and efficiently thereby reducing wasted production. If the cake is included in the low sales group, the number of cake products is small. On the other hand, if a cake is included in the high sales group, the producer will produce the cake in large quantities.

**Keywords:** CRISP-DM; K-Means; Sales

## 1. Pendahuluan

Seiring dengan pertumbuhan bisnis di era globalisasi dan kemajuan di bidang teknologi informasi yang cepat memberikan pengaruh yang cukup besar baik dalam bidang industri maupun jasa. Hal ini juga membawa suatu perubahan besar dalam tingkat persaingan perusahaan, sehingga pelaku-pelaku usaha tersebut harus selalu menciptakan berbagai strategi untuk bertahan. Dalam rangka menghadapi persaingan bisnis dan meningkatkan pendapatan perusahaan, pimpinan maupun manajemen dalam suatu perusahaan tersebut dituntut untuk dapat mengambil keputusan yang tepat dalam menentukan strategi penjualan. Untuk dapat melakukan hal tersebut, perusahaan membutuhkan sumber informasi yang cukup banyak untuk

dapat dianalisis lebih lanjut. Pihak eksekutif perusahaan mengharapkan adanya teknologi yang mampu menghasilkan suatu informasi yang siap digunakan untuk membantu mereka dalam mengambil keputusan strategis perusahaan. Mereka ingin mengetahui produk apa yang harus ditingkatkan, seberapa besar pencapaian hasil yang diperoleh oleh perusahaan. Untuk memenuhi kebutuhan tersebut terdapat banyak cara, diantaranya menggunakan teknik *data mining* [1].

*Data mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan terkait dari berbagai database besar [2]. Pendekatan dasar dalam *data mining* adalah merangkum data dan mengambil informasi yang bermakna dan yang sebelumnya tidak diketahui [3]. *Clustering* adalah salah satu permasalahan di dalam pengenalan pola, pemrosesan citra [4], *machine learning*, dan statistika [5][6].

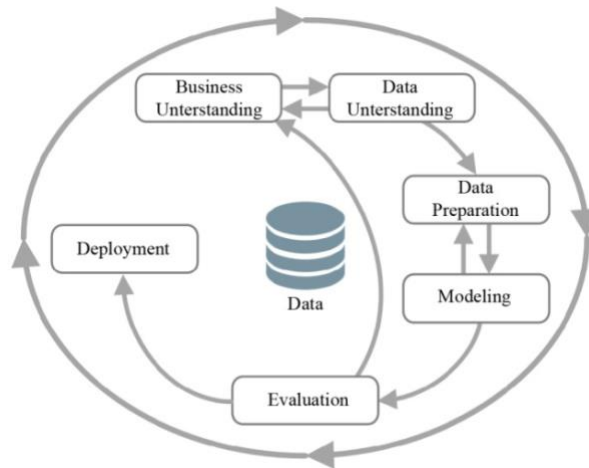
K-means diperkenalkan pertama kali oleh MacQueen pada awal tahun 1967 sebagai algoritma *clustering* pertama kali. Makna “K” mewakili jumlah kumpulan data yang akan diklusterisasi, “means” adalah nilai rata-rata, yang mewakili nilai rata-rata dari kelas data yang menjadi pusat kluster. K-Means adalah sebuah metode *clustering* data non-hirarki yang dikelompokkan ke dalam satu atau lebih *cluster*. K-Means bekerja cepat dan efisien, sehingga banyak digunakan dalam klusterisasi data [7]–[15].

Pada penelitian ini akan diambil sebuah Toko Kue yang berada di Wilayah Kabupaten Purwakarta sebagai studi kasus. Pada penelitian yang dilakukan oleh Yulia Darmi dan Agus Setiawan, metode K-Means Clustering digunakan untuk mengelompokkan penjualan produk yang laku terjual dan yang tidak laku terjual [16]. Metode K-means juga digunakan untuk menentukan penjualan sebuah produk yang paling banyak dicari oleh konsumen, seperti yang dilakukan oleh Nur Halimah Lintang Suryani [17]. Pada penelitian ini pun akan menggunakan metode *K-Means Clustering* untuk mengelompokkan penjualan kue berdasarkan tanggal nya untuk setiap bulan. Tujuan dari penelitian ini adalah untuk mengetahui potensi penjualan kue pada setiap tanggal dari setiap bulan. Dengan demikian produsen dapat mempertimbangkan berapa banyak jumlah kue yang harus diproduksi pada tanggal tersebut.

## 2. Metode Penelitian

Metode yang akan digunakan dalam penelitian ini adalah metode *Cross Industry Standard Process for Data Mining* (CRISP-DM). Metode CRISP-DM (Gambar 1) terdiri dari enam (6) fase [18] yaitu :

1. Fase Pemahaman Bisnis (*Business Understanding Phase*)
2. Fase Pemahaman Data (*Data Understanding Phase*)
3. Fase Persiapan Data (*Data Preparation Phase*)
4. Fase Pemodelan (*Modeling Phase*)
5. Fase Evaluasi (*Evaluation Phase*)
6. Fase Penyebaran (*Deployment Phase*)



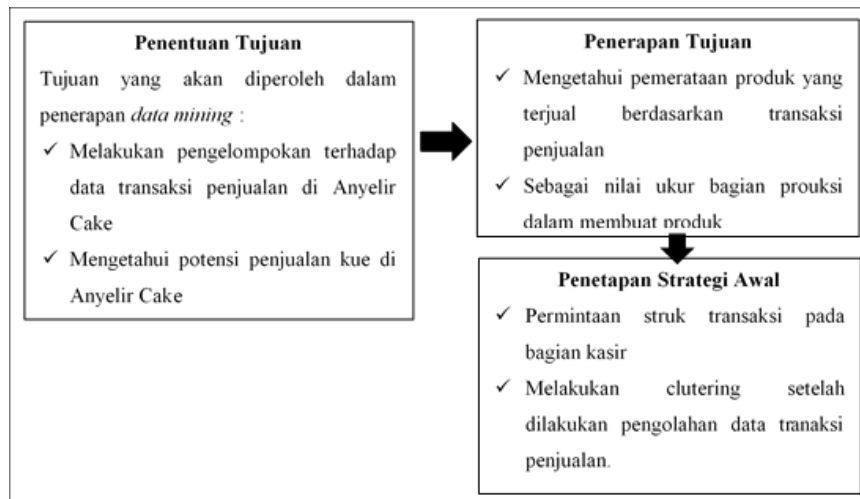
Gambar 1. Proses CRISP-DM

### 3. Hasil dan Pembahasan

#### 3.1. Fase Pemahaman Bisnis

Berdasarkan hasil observasi, diketahui bahwa Anyelir Cake merupakan perusahaan yang bergerak di bidang *bakery* dengan produk utamanya yaitu *birthday cake* yang terbagi menjadi *blackforest*, *rainbow* dan *tart* (lapis Surabaya), aneka bolu yang terbagi menjadi bolu panggang dan bolu kukus, serta aneka roti yang terbagi menjadi roti manis dan roti asin. Banyaknya produk yang dihasilkan tersebut terkadang tidak merata terjual sehingga ada saja produk yang terbuang.

*Business understanding phase* atau fase pemahaman bisnis merupakan tahap pertama dalam metode CRISP-DM, dimana pada fase ini penulis melakukan penentuan tujuan proyek dan kebutuhan, menerjemahkan tujuan dan batasan, serta menyiapkan strategi awal untuk mencapai tujuan. Langkah-langkah pemahaman bisnis tersebut dapat dilihat pada Gambar 2.



Gambar 2. Langkah-langkah pemahaman bisnis

#### Penentuan Tujuan Bisnis

Tahap menentukan tujuan bisnis yaitu tahap menemukan faktor-faktor penting yang terlibat dalam penelitian yang direncanakan. Tujuan bisnis bukan merupakan tujuan langsung penelitian, namun lebih sebagai tujuan jangka panjang. Hasil *data mining* mendeskripsikan sebuah

pengetahuan yang akan menjadi tolak ukur Anyelir Cake, khususnya pada bagian produksi dalam menentukan produk yang perlu diproduksi dalam jumlah banyak maupun sedikit.

Tujuan bisnis berdasarkan rataan transaksi penjualan di Anyelir Cake pada tahun 2017 antara lain untuk mencari, mengelompokkan jenis kue apa saja yang paling banyak terjual sehingga dapat menemukan strategi untuk meminimalisir bagian produksi agar tidak memproduksi berlebih produk yang lama terjual dan sebaliknya tidak kurang dalam memproduksi produk yang sering atau cepat terjual. Pengolahan data transaksi penjualan juga dapat dijadikan sebagai salah satu acuan untuk melakukan pengolahan data kembali pada tahun-tahun berikutnya yang bertujuan sama yaitu menemukan strategi untuk meminimalisir produksi yang berlebih.

**Penerapan Tujuan**

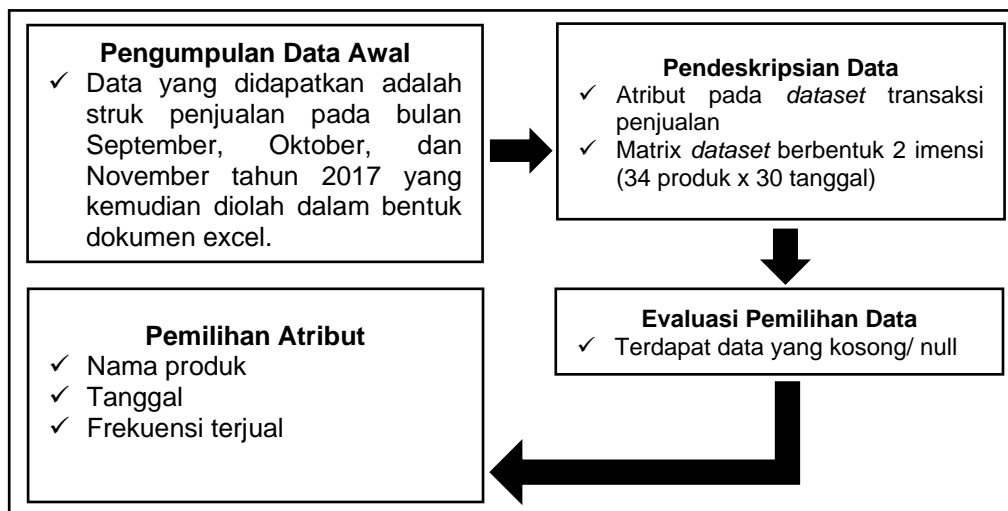
Pada tahap ini dibutuhkan pemahaman terhadap tujuan bisnis dan menerjemahkan ke dalam tujuan *data mining*. Berdasarkan tujuan bisnis yang telah ditemukan dan dipaparkan pada poin sebelumnya, maka dapat penulis simpulkan bahwa penerapan tujuan dari pemahaman bisnis yang ada adalah mengetahui pemerataan produk yang terjual berdasarkan transaksi penjualan dan menjadikannya nilai ukur atau rekomendasi pada bagian produksi dalam menentukan sasaran produk berdasarkan pengelompokan yang telah dilakukan sebelumnya.

**Penetapan Strategi Awal**

Strategi awal dalam menerapkan tujuan dilakukannya *data mining* adalah melakukan permintaan struk data transaksi penjualan pada bagian kasir Anyelir Cake Purwakarta. Pengolahan berdasarkan data transaksi penjualan tersebut akan membentuk beberapa kelompok yang kemudian digunakan dalam analisis *clustering*. Hasil *clustering* tersebut dapat digunakan sebagai informasi yang akurat dalam menyusun strategi produksi.

**3.2. Fase Pemahaman Data**

Berdasarkan data transaksi penjualan yang telah didapat, tahap selanjutnya dalam metode CRISP-DM yaitu pemahaman terhadap kebutuhan data terkait dengan pencapaian tujuan dalam menentukan strategi produksi yang efektif dan efisien. Data transaksi penjualan tersebut merupakan *dataset* yang didapat dari bagian kasir berupa struk penjualan yang kemudian diubah ke dalam bentuk dokumen dengan format excel dan memiliki jumlah 1020 *record*. Langkah-langkah dalam fase pemahaman data (*data understanding phase*) dapat dilihat pada Gambar 3.



Gambar 3. Langkah-langkah pemahaman data.

**Pengumpulan Data Awal**

Pengumpulan data awal yaitu melakukan pengumpulan data yang dibutuhkan untuk mendukung dalam melakukan pemahaman data. Adapun sumber data utama yang digunakan

dalam penelitian ini adalah *dataset* transaksi penjualan Anyelir Cake pada tahun 2017 bulan September, Oktober, dan November dalam bentuk struk transaksi. *Dataset* tersebut kemudian dibuat dalam bentuk excel untuk mempermudah dalam pengolahan selanjutnya guna menentukan strategi produksi.

### Pendeskripsian Data

Tahap mendeskripsikan data dilakukan dengan memahami *dataset* transaksi penjualan dimana *dataset* tersebut dideskripsikan menjadi pendataan atribut yang terdiri dari nama produk dan tanggal dengan matriks 2 dimensi (34 produk x 30 tanggal).

### Evaluasi Pemilihan Data

Hasil evaluasi terhadap kualitas data yaitu menemukan beberapa nilai kosong/*null* yang disebut dengan missing value pada atribut dalam *dataset* transaksi penjualan. Atribut yang baik adalah atribut yang berisi nilai relevan atau tidak *missing value*, dimana syarat tersebut merupakan syarat awal yang harus dilakukan dalam *data mining*, sehingga akan diperoleh *dataset* transaksi penjualan yang bersih untuk digunakan dalam tahap pengolahan data. Relevan atau tidaknya sebuah data, dapat ditentukan oleh keluaran (*output*) yang dihasilkan. Sehingga pada tahap selanjutnya akan dilakukan persiapan data dengan cara menghilangkan nilai kosong / *missing value* yang terdapat dalam *dataset* transaksi penjualan agar dapat diolah dan mendapatkan hasil yang relevan sesuai dengan tujuan awal yang ditentukan.

### Pemilihan Atribut

Atribut pada *dataset* transaksi penjualan tidak seluruhnya digunakan dalam proses pengolahan data, karena telah disesuaikan dengan fokus penelitian. Atribut yang digunakan adalah produk, tanggal penjualan, dan frekuensi penjualan. Contoh data berdasarkan atribut yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Contoh data berdasarkan atribut yang digunakan

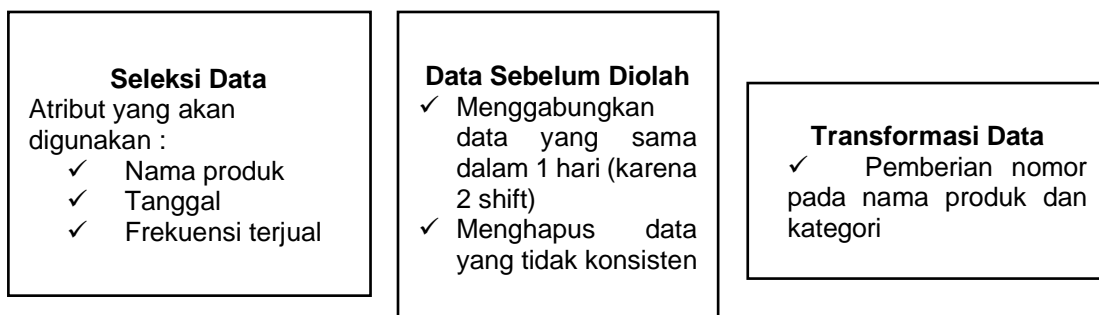
Produk	Tanggal	Frekuensi
Brownies Kukus Cokelat	1	13
Brownies Kukus Blueberry Besar	1	1
Brownies Kukus Keju	1	10
Brownies Oven	1	7
Lapis Pandan B	1	8
Lapis Pandan K	1	12
Lapis Pelangi	1	6
Banana	1	4
Brownies Kukus Tiramisu	1	1
Bolu Tape	1	3
Bolu Ketan Item	1	9
Choco Cake	1	1
Cake Kentang	1	3
Eggless Cokelat	1	3
Lapis Keju	1	3
Lapis Cokelat	1	3
Donat	1	7
Bolu Apel	1	1
Bolu Caramel	1	1
Lapis Ubi Besar	1	1
Lapi Ubi Kecil	1	1
Bika Ambon	1	1
Mamon Cake	1	1
Lapis Legit	1	2
Proll Tape	1	1

Produk	Tanggal	Frekuensi
Wingko	1	5
Roti Cokelat	1	1
Roti Keju	1	1
Roti Tabur	1	1
Roti Jagung	1	1
Roti Susu	1	2
Roti Sosis	1	1
Roti Abon	1	1
Roti Pisang Cokelat	1	1

### 3.3. Fase Persiapan Data

Persiapan data mencakup semua kegiatan untuk membangun dataset penjualan yang akan diterapkan ke dalam alat pemodelan, dari data mentah berupa dataset transaksi penjualan dan selanjutnya akan dilakukan proses *data mining*. Fungsi utamanya yaitu khusus sebagai alat pemodelan *clustering*. Persiapan data merupakan tahap yang padat dengan aktivitas pengolahan data. Tahap membuat dataset transaksi penjualan baru yaitu tahap persiapan data akhir untuk dilakukan pemodelan *clustering data mining*.

Pada proses ini akan dilakukan persiapan data transaksi penjualan yang telah dilakukan pada taha evalasi dan dilakukan perbaikan data berdasarkan hasil evaluasi tersebut. Pada tahap ini telah dilakukan penetapan tujuan bisnis dan tujuan *data mining* dalam menentukan potensi penjualan produk berdasarkan atribut, sehingga pada tahap persiapan data merupakan implementasi awal dari penerapan tujuan tersebut. Fase persiapan data dapat dilihat pada Gambar 4.



Gambar 4. Langkah-langkah persiapan data.

#### Seleksi Data

Rancangan dataset transaksi penjualan harus menyesuaikan dengan apa yang telah dirumuskan pada tahap pemahaman bisnis terutama pada perumusan tujuan *data mining* yaitu mendeskripsikan pola pada pengelompokan karakteristik transaksi penjualan berdasarkan data dari Toko Kue Anyelir. Tahap pemilihan data yang digunakan dalam melakukan analisis berdasarkan pada beberapa atribut, termasuk relevansinya dengan tujuan *data mining*. Atribut yang digunakan adalah nama produk, kategori, tanggal, dan frekuensi penjualan. Atribut tersebut telah diseleksi berdasarkan dataset struk transaksi penjualan asli dan telah disesuaikan dengan kebutuhan penelitian.

#### Pengolahan Data Mentah

Pada tahap ini penulis memastikan data transaksi yang dipilih telah layak dilakukan untuk proses pengolahan. Kegiatan antara lain membersihkan dan memperbaiki data yang rusak, menghapus data yang tidak digunakan, menyeragamkan data yang dianggap sama namun memiliki nilai yang berbeda atau membuatnya konsisten. Hasil evaluasi pada tahap pemahaman data menunjukkan adanya beberapa data dengan kualitas buruk. Data tersebut perlu dibersihkan dengan cara diperbaiki, dirubah, dan dihapus.

### Transformasi Data

Agar dataset transaksi penjualan tersebut dapat diolah menggunakan *k-means clustering*, maka data yang berjenis data nominal seperti nama produk harus diinisialisasi terlebih dahulu dalam bentuk angka. Inisialisasi nama produk dapat dilihat pada Tabel 2.

Tabel 2. Inisialisasi nama produk.

Produk	Inisialisasi
Brownies Kukus Cokelat B	1
Brownies Kukus Blueberry B	2
Brownies Kukus Keju B	3
Brownies Oven	4
Lapis Pandan B	5
Lapis Pandan K	6
Lapis Pelangi	7
Bolu Pisang Keju	8
Brownies Tiramisu	9
Bolu Tape	10
Bolu Ketan Item	11
Choco Caket	12
Cake Kentang	13
Eggless Cokelat	14
Lapis Keju	15
Lapis Cokelat	16
Donat Variasi	17
Bolu Apel	18
Bolu Caramel	19
Lapis Ubi Besar	20
Lapis Ubi K	21
Bika Ambon	22
Mamon Cake	23
Lapis Legit	24
Froll Tape	25
Wingko Keju	26
Roti Cokelat	27
Roti Keju Iris	28
Roti Keju Tabur	29
Roti Jagung	30
Roti Susu	31
Roti Sosis	32
Roti Daging Manis	33
Roti Pisang Cokelat	34

### 3.4. Fase Pemodelan

Dalam tahap ini data yang telah dikumpulkan dan ditransformasikan mulai dikelompokkan menggunakan algoritma *K-means Clustering*. Untuk dapat melakukan pengelompokan data tersebut menjadi beberapa *cluster* perlu dilakukan beberapa langkah yaitu :

1) Tentukan jumlah cluster yang diinginkan. Dalam penelitian ini data yang ada akan dikelompokkan menjadi tiga cluster.

2) Tentukan titik pusat dari setiap *cluster*. Dalam pembahasan ini hanya digunakan sampel data sebagai contoh dalam proses perhitungan manual. Sampel data yang digunakan dapat dilihat pada Tabel 3.

Tabel 3. Sampel data.

Nama Produk	Frekuensi
Brownies Kukus Cokelat B	29
Brownies Kukus Blueberry B	5
Brownies Kukus Keju B	12
Brownies Oven	9
Lapis Pandan B	20
Lapis Pandan K	28
Lapis Pelangi	16
Bolu Pisang Keju	6
Brownies Tiramisu	3
Bolu Tape	7
Bolu Ketan Item	11
Choco Caket	5
Cake Kentang	7
Eggless Cokelat	7
Lapis Keju	5
Lapis Cokelat	7
Donat Variasi	9
Bolu Apel	3
Bolu Caramel	3
Lapis Ubi Besar	3
Lapis Ubi K	11
Bika Ambon	3
Mamon Cake	5
Lapis Legit	4
Froll Tape	5
Wingko Keju	7
Roti Cokelat	5
Roti Keju Iris	3
Roti Keju Tabur	15
Roti Jagung	7
Roti Susu	4
Roti Sosis	7
Roti Daging Manis	9
Roti Pisang Cokelat	5

Kemudian titik pusat awal ditentukan secara random dan didapatkan titik pusat dari setiap *cluster* dapat dilihat pada Tabel 4.

Tabel 4. Titik awal cluster

Titik Pusat Awal	Frekuensi Terjual
Cluster 1	5
Cluster 2	12
Cluster 3	28



3) Setelah diketahui nilai K dan pusat kluster awal, selanjutnya mengukur jarak antara pusat kluster menggunakan Euclidean Distance, kemudian didapatkan matriks jarak yaitu C1, C2, dan C3 sebagai berikut :

Rumus Euclidean Distance :

$$D_e = \sqrt{(X_i - S_i)^2}$$

Perhitungan jarak data pertama dengan pusat cluster pertama adalah :

$$D_e = \sqrt{(29 - 5)^2} = 24$$

Perhitungan jarak data pertama dengan pusat cluster kedua adalah :

$$D_e = \sqrt{(29 - 12)^2} = 27$$

Perhitungan jarak data pertama dengan pusat cluster ketiga adalah :

$$D_e = \sqrt{(29 - 28)^2} = 1$$

Tabel hasil perhitungan jarak kluster untuk iterasi 0 dapat dilihat pada Tabel 5.

Tabel 5. Hasil perhitungan jarak pusat kluster (iterasi 0)

Produk	Frekuensi Terjual	C1	C2	C3	Jarak Terpendek	Cluster yang diikuti
1	29	24	17	1	1	C3
2	5	0	7	23	0	C1
3	12	7	0	16	0	C2
4	9	4	3	19	3	C2
5	20	15	8	8	8	C3
6	28	23	16	0	0	C3
7	16	11	4	12	4	C2
8	6	1	6	22	1	C1
9	3	2	9	25	2	C1
10	7	2	5	21	2	C1
11	11	6	1	17	1	C2
12	5	0	7	23	0	C1
13	7	2	5	21	2	C1
14	7	2	5	21	2	C1
15	5	0	7	23	0	C1
16	7	2	5	21	2	C1
17	9	4	3	23	3	C2
18	3	2	9	25	2	C1
19	3	2	9	13	2	C1
20	3	2	9	21	2	C1
21	11	6	1	24	1	C2
22	3	2	9	21	2	C1
23	5	0	7	19	0	C1
24	4	1	8	23	1	C1
25	5	0	7	23	0	C1
26	7	2	5	21	2	C1
27	5	0	7	23	0	C1
28	3	2	9	25	2	C1
29	15	10	3	13	3	C2
30	7	2	5	21	2	C1
31	4	1	8	24	1	C1
32	7	2	5	21	2	C1
33	9	4	3	19	3	C2
34	5	0	7	23	0	C1

4) Setelah diketahui anggota setiap kluster, kemudian pusat kluster baru dihitung berdasarkan data anggota setiap kluster sesuai dengan rumus pusat anggota kluster.

Hasil perhitungan titik pusat klaster (iterasi 1) dapat dilihat pada Tabel 6. Setelah mendapatkan pusat klaster yang baru, maka hasil ulang perhitungan jaraknya dapat dilihat pada Tabel 7.

Tabel 6. Titik pusat awal klaster (iterasi 1)

Titik Pusat Awal	Frekuensi Terjual
Cluster 1	5
Cluster 2	11,8571429
Cluster 3	25,6666667

Tabel 7. Hasil perhitungan jarak pusat klaster (iterasi 1)

Produk	Frekuensi Terjual	C1	C2	C3	Jarak Terpendek	Cluster yang diikuti
1	29	24	17.14286	3.333333	3.333333	C3
2	5	0	6.857143	20.66667	0	C1
3	12	7	0.142857	13.66667	0.142857	C2
4	9	4	2.857143	16.66667	2.857143	C2
5	20	15	8.142857	5.666667	5.666667	C3
6	28	23	16.14286	2.333333	2.333333	C3
7	16	11	4.142857	9.666667	4.142857	C2
8	6	1	5.857143	19.66667	1	C1
9	3	2	8.857143	22.66667	2	C1
10	7	2	4.857143	18.66667	2	C1
11	11	6	0.857143	14.66667	0.857143	C2
12	5	0	6.857143	20.66667	0	C1
13	7	2	4.857143	18.66667	2	C1
14	7	2	4.857143	18.66667	2	C1
15	5	0	6.857143	20.66667	0	C1
16	7	2	4.857143	18.66667	2	C1
17	9	4	2.857143	20.66667	2.857143	C2
18	3	2	8.857143	22.66667	2	C1
19	3	2	8.857143	10.66667	2	C1
20	3	2	8.857143	18.66667	2	C1
21	11	6	0.857143	21.66667	0.857143	C2
22	3	2	8.857143	18.66667	2	C1
23	5	0	6.857143	16.66667	0	C1
24	4	1	7.857143	20.66667	1	C1
25	5	0	6.857143	20.66667	0	C1
26	7	2	4.857143	18.66667	2	C1
27	5	0	6.857143	20.66667	0	C1
28	3	2	8.857143	22.66667	2	C1
29	15	10	3.142857	10.66667	3.142857	C2
30	7	2	4.857143	18.66667	2	C1
31	4	1	7.857143	21.66667	1	C1
32	7	2	4.857143	18.66667	2	C1
33	9	4	2.857143	16.66667	2.857143	C2

Setelah diketahui anggota setiap klaster iterasi 1, kemudian pusat klaster baru dihitung kembali berdasarkan data anggota setiap klaster iterasi 1. Hasil perhitungan titik pusat awal untuk iterasi 2 dapat dilihat pada Tabel 8. Hasil perhitungan ulang untuk jarak setiap anggota dengan pusat klaster untuk iterasi 2 dapat dilihat pada Tabel 9. Pada perhitungan ini iterasi berhenti pada iterasi ke 2, karena kelompok data hasil klastering 1 dan 2 sama dan telah mencapai status stabil atau konvergen.

Tabel 8. Titik pusat awal klaster (iterasi 2)

Titik Pusat Awal	Frekuensi Terjual
Cluster 1	5
Cluster 2	11,8571429
Cluster 3	25,6666667

Tabel 9. Hasil perhitungan jarak pusat klaster (iterasi 2)

Produk	Frekuensi Terjual	C1	C2	C3	Jarak Terpendek	Cluster yang diikuti
1	29	24	17.14286	3.333333	3.333333	C3
2	5	0	6.857143	20.66667	0	C1
3	12	7	0.142857	13.66667	0.142857	C2
4	9	4	2.857143	16.66667	2.857143	C2
5	20	15	8.142857	5.666667	5.666667	C3
6	28	23	16.14286	2.333333	2.333333	C3
7	16	11	4.142857	9.666667	4.142857	C2
8	6	1	5.857143	19.66667	1	C1
9	3	2	8.857143	22.66667	2	C1
10	7	2	4.857143	18.66667	2	C1
11	11	6	0.857143	14.66667	0.857143	C2
12	5	0	6.857143	20.66667	0	C1
13	7	2	4.857143	18.66667	2	C1
14	7	2	4.857143	18.66667	2	C1
15	5	0	6.857143	20.66667	0	C1
16	7	2	4.857143	18.66667	2	C1
17	9	4	2.857143	20.66667	2.857143	C2
18	3	2	8.857143	22.66667	2	C1
19	3	2	8.857143	10.66667	2	C1
20	3	2	8.857143	18.66667	2	C1
21	11	6	0.857143	21.66667	0.857143	C2
22	3	2	8.857143	18.66667	2	C1
23	5	0	6.857143	16.66667	0	C1
24	4	1	7.857143	20.66667	1	C1
25	5	0	6.857143	20.66667	0	C1
26	7	2	4.857143	18.66667	2	C1
27	5	0	6.857143	20.66667	0	C1
28	3	2	8.857143	22.66667	2	C1
29	15	10	3.142857	10.66667	3.142857	C2
30	7	2	4.857143	18.66667	2	C1
31	4	1	7.857143	21.66667	1	C1
32	7	2	4.857143	18.66667	2	C1
33	9	4	2.857143	16.66667	2.857143	C2
34	5	0	6.857143	20.66667	0	C1

**3.5. Analisis Hasil**

Berdasarkan implementasi algoritme *K-Means Clustering* menggunakan aplikasi Orange, berikut adalah hasil analisisnya :

1) Hasil analisis menggunakan aplikasi orange berikut ini dibedakan berdasarkan tanggal. Hasil klastering tanggal 10 dapat dilihat pada Tabel 10, hasil klastering tanggal 20 dapat dilihat pada Tabel 11 dan hasil klastering tanggal 30 dapat dilihat pada Tabel 12.

Tabel 10. Hasil klastering untuk tanggal sepuluh (10)

Cluster	Nama Produk
C1	Bika Ambon, Bolu Apel, Bolu Caramel, Bolu Ketan Item, Bolu Pisang Keju, Bolu Tape, Brownies Kukus Keju B, Brownies Oven, Brownies Tiramisu, Choco Caket, Donat Variasi, Eggless Cokelat, Froll Tape, Lapis Cokelat, Lapis Keju, Lapis Legit, Mamon Cake, Roti Cokelat, Roti Daging Manis, Roti Jagung, Roti Keju Iris, Roti Keju Tabur, Roti Susu
C2	Brownies Kukus Blueberry B, Cake Kentang, Lapis Pelangi, Lapis Ubi Besar, Lapis Ubi K, Roti Pisang Cokelat, Roti Sosis, Wingko Keju
C3	Brownies Kukus Cokelat B, Lapis Pandan B, Lapis Pandan K

Pada Tabel 10, menunjukkan bahwa klaster 1 terdiri atas 23 produk, klaster 2 terdiri atas 8 produk, dan klaster 3 terdiri atas 3 produk. Klaster 1 merupakan klaster dengan tingkat penjualan rendah, klaster 2 dengan tingkat penjualan menengah, dan klaster 3 dengan tingkat penjualan tinggi.

Tabel 11. Hasil klastering untuk tanggal dua puluh (20)

Cluster	Nama Produk
C1	Bika Ambon, Bolu Apel, Bolu Caramel, Bolu Ketan Item, Brownies Kukus Blueberry B, Brownies Tiramisu, Choco Caket, Donat Variasi, Eggless Cokelat, Lapis Cokelat, Lapis Keju, Lapis Legit, Lapis Ubi K, Roti Cokelat, Roti Daging Manis, Roti Jagung, Roti Keju Iris, Roti Sosis, Wingko Keju
C2	Bolu Pisang Keju, Bolu Tape, Cake Kentang, Froll Tape, Lapis Pelangi, Lapis Ubi Besar, Mamon Cake, Roti Pisang Cokelat, Roti Susu
C3	Brownies Kukus Cokelat B, Brownies Kukus Keju B, Brownies Oven, Lapis Pandan B, Lapis Pandan K, Roti Keju Tabur

Pada Tabel 11 menunjukkan bahwa klaster 1 terdiri atas 19 produk, klaster 2 terdiri atas 9 produk, dan klaster 3 terdiri dari 6 produk. Klaster 1 merupakan klaster dengan tingkat penjualan rendah, klaster 2 dengan tingkat penjualan menengah, dan klaster 3 dengan tingkat penjualan tinggi.

Tabel 12. Hasil klastering untuk tanggal tiga puluh (30)

Cluster	Nama Produk
C1	Bika Ambon, Bolu Apel, Bolu Caramel, Bolu Ketan Item, Bolu Pisang Keju, Bolu Tape, Brownies Kukus Blueberry B, Brownies Oven, Brownies Tiramisu, Choco Caket, Donat Variasi, Eggless Cokelat, Froll Tape, Lapis Cokelat, Lapis Legit, Lapis Ubi Besar, Mamon Cake, Roti Cokelat, Roti Daging Manis, Roti Jagung, Roti Sosis, Wingko Keju
C2	Brownies Kukus Cokelat B, Brownies Kukus Keju B, Cake Kentang, Lapis Keju, Lapis Pelangi, Lapis Ubi K, Roti Keju Iris, Roti Pisang Cokelat, Roti Susu

C3 Lapis Pandan B, Lapis Pandan K, Roti Keju Tabur

Tabel 12 menunjukkan bahwa klaster 1 terdiri atas 19 produk, klaster 2 terdiri atas 9 produk dan klaster 3 terdiri atas 3 produk. Klaster 1 merupakan klaster dengan tingkat penjualan rendah, klaster 2 dengan tingkat penjualan menengah, dan klaster 3 dengan tingkat penjualan tinggi.

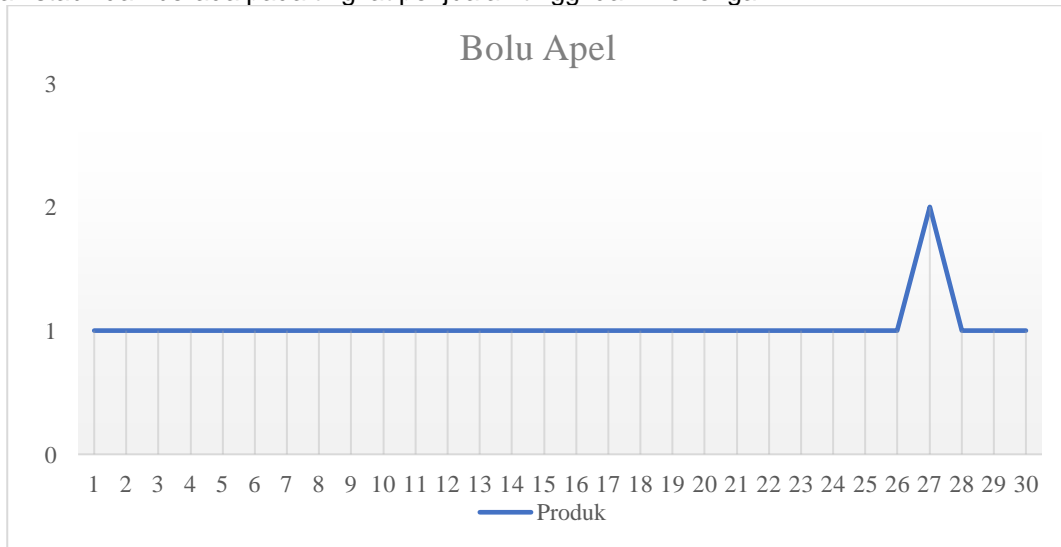
2) Hasil analisis dengan menggabungkan hasil analisis per tanggal dapat dilihat pada Tabel 13. Tabel 13 menunjukkan bahwa tidak semua produk berada pada suatu klaster di setiap tanggal atau dengan kata lain di beberapa tanggal setiap produk memiliki kemungkinan berada pada klaster 1, klaster 2, ataupun klaster 3.

Tabel 13. Hasil klastering gabungan (tanggal 10, 20, dan 30)

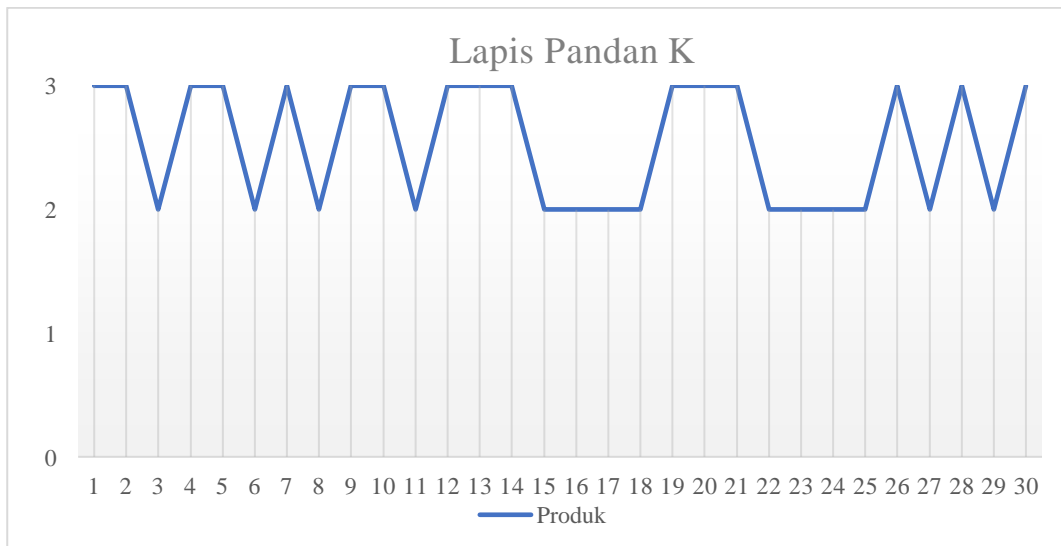
Tanggal	C1 (Rendah)	C2 (Sedang)	C3 (Tinggi)		
10	Bika Ambon, Bolu Apel, Bolu Caramel, Bolu Ketan Item, Bolu Pisang Keju, Bolu Tape, Brownies Kukus Keju B, Brownies Oven, Brownies Tiramisu, Choco Caket, Donat Variasi, Eggless Cokelat, Froll Tape, Lapis Cokelat, Lapis Keju, Lapis Legit, Mamon Cake, Roti Cokelat, Roti Daging Manis, Roti Jagung, Roti Keju Iris, Roti Keju Tabur, Roti Susu	Brownies Kukus Blueberry B, Cake Kentang, Lapis Pelangi, Lapis Ubi Besar, Lapis Ubi K, Roti Pisang Cokelat, Roti Sosis, Wingko Keju	Brownies Kukus Cokelat B, Lapis Pandan B, Lapis Pandan K		
	20	Bika Ambon, Bolu Apel, Bolu Caramel, Bolu Ketan Item, Brownies Kukus Blueberry B, Brownies Tiramisu, Choco Caket, Donat Variasi, Eggless Cokelat, Lapis Cokelat, Lapis Keju, Lapis Legit, Lapis Ubi K, Roti Cokelat, Roti Daging Manis, Roti Jagung, Roti Keju Iris, Roti Sosis, Wingko Keju	Bolu Pisang Keju, Bolu Tape, Cake Kentang, Froll Tape, Lapis Pelangi, Lapis Ubi Besar, Mamon Cake, Roti Pisang Cokelat, Roti Susu	Brownies Kukus Cokelat B, Brownies Kukus Keju B, Brownies Oven, Lapis Pandan B, Lapis Pandan K, Roti Keju Tabur	
		30	Bika Ambon, Bolu Apel, Bolu Caramel, Bolu Ketan Item, Bolu Pisang Keju, Bolu Tape, Brownies Kukus Blueberry B, Brownies Oven, Brownies Tiramisu, Choco Caket, Donat Variasi, Eggless Cokelat, Froll Tape, Lapis Cokelat, Lapis Legit, Lapis Ubi Besar, Mamon	Brownies Kukus Cokelat B, Brownies Kukus Keju B, Cake Kentang, Lapis Keju, Lapis Pelangi, Lapis Ubi K, Roti Keju Iris,	Lapis Pandan B, Lapis Pandan K, Roti Keju Tabur

Tanggal	C1 (Rendah)	C2 (Sedang)	C3 (Tinggi)
	Cake, Roti Cokelat, Roti Daging	Roti Pisang	
	Manis, Roti Jagung, Roti Sosis,	Cokelat, Roti	
	Wingko Keju	Susu	

3) Hasil analisis berupa grafik berfungsi untuk melihat perubahan kluster pada suatu produk selama satu bulan. Grafik hasil analisis untuk salah satu produk kue dapat dilihat pada Gambar 5 dan Gambar 6. Berdasarkan Gambar 5, produk Bolu Apel cenderung stabil pada tingkat penjualan rendah. Gambar 6 menunjukkan bahwa grafik produk Kue Lapis Pandan K cenderung tidak stabil dan berada pada tingkat penjualan tinggi dan menengah.



Gambar 5. Grafik kluster untuk produk kue Bolu Apel



Gambar 6. Grafik kluster untuk produk kue Lapis Pandan K.

#### 4. Kesimpulan

Penelitian ini menghasilkan tiga kluster, dengan kluster 1 sebagai kluster tingkat penjualan yang rendah, kemudian kluster 2 adalah kluster untuk tingkat penjualan menengah, dan kluster 3 merupakan kluster untuk tingkat penjualan yang tinggi. Strategi produksi yang dapat

dilakukan adalah bagi produk dengan potensi stabil, maka produk tersebut dapat disesuaikan jumlah produksinya. Jika suatu produk secara stabil berada pada kluster rendah, maka bagian produksi dapat membuat produk dalam jumlah yang sedikit. Begitu juga sebaliknya jika suatu produk secara stabil berada pada kluster tinggi, maka bagian produksi dapat menaikkan jumlah produk yang diproduksi lebih banyak dari yang lain.

## Referensi

- [1] S. Butsianto and N. T. Mayangwulan, "Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 3, no. 3, pp. 187–201, 2020, doi: 10.32672/jnkti.v3i3.2428.
- [2] E. Turban, J. E. Aronson, and T.-P. Liang, *Decision Support System and Intelligent Systems Edisi Bahasa Indonesia Jilid 1*. Yogyakarta: Andi, 2005.
- [3] P. Harliana and R. Rahim, "Comparative Analysis of Membership Function on Mamdani Fuzzy Inference System for Decision Making," *J. Phys. Conf. Ser.*, vol. 930, no. 1, 2017, doi: 10.1088/1742-6596/930/1/012029.
- [4] R. Rahim, T. Afriliansyah, H. Winata, D. Nofriansyah, Ratnadewi, and S. Aryza, "Research of Face Recognition with Fisher Linear Discriminant," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 300, no. 1, 2018, doi: 10.1088/1757-899X/300/1/012037.
- [5] D. Siregar, D. Arisandi, A. Usman, D. Irwan, and R. Rahim, "Research of Simple Multi-Attribute Rating Technique for Decision Support," *J. Phys. Conf. Ser.*, vol. 930, no. 1, 2017, doi: 10.1088/1742-6596/930/1/012015.
- [6] A. S. Ahmar *et al.*, "Modeling Data Containing Outliers using ARIMA Additive Outlier (ARIMA-AO)," *J. Phys. Conf. Ser.*, vol. 954, no. 1, 2018, doi: 10.1088/1742-6596/954/1/012010.
- [7] L. J. Tao, L. Y. Hong, and H. Yan, "The improvement and application of a K-means clustering algorithm," *Proc. 2016 IEEE Int. Conf. Cloud Comput. Big Data Anal. ICCCBDA 2016*, no. 3, pp. 93–96, 2016, doi: 10.1109/ICCCBDA.2016.7529540.
- [8] A. P. Windarto, "Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method," *Int. J. Artif. Intell. Res.*, vol. 1, no. 2, p. 26, 2017, doi: 10.29099/ijair.v1i2.17.
- [9] F. Saeed, N. Salim, A. Abdo, and H. Hentabli, "Graph-based consensus clustering for combining multiple clusterings of chemical structures," *Mol. Inform.*, vol. 32, no. 2, pp. 165–178, 2013, doi: 10.1002/minf.201200110.
- [10] J. Ding, J. Shi, and F. X. Wu, "Quality assessment of tandem mass spectra by using a weighted K-means," *Clin. Proteomics*, vol. 5, no. 1, pp. 15–22, 2009, doi: 10.1007/s12014-009-9025-4.
- [11] E. Borrayo, R. Machida-Hirano, M. Takeya, M. Kawase, and K. Watanabe, "Principal components analysis - K-means transposon element based foxtail millet core collection selection method," *BMC Genet.*, vol. 17, no. 1, pp. 1–10, 2016, doi: 10.1186/s12863-016-0343-z.
- [12] E. D. Übeyli and E. Doğdu, "Erratum: Automatic detection of erythematous-squamous diseases using k-means clustering (Journal of Medical Systems 34:2 (179-184) DOI: 10.1007/s10916-008-9229-6)," *J. Med. Syst.*, vol. 36, no. 2, p. 1007, 2012, doi: 10.1007/s10916-010-9534-8.
- [13] P. K. Sharma and G. Holness, "L2 -norm transformation for improving k-means clustering: Finding a suitable model by range transformation for novel data analysis," *Int. J. Data Sci. Anal.*, vol. 3, no. 4, pp. 247–266, 2017, doi: 10.1007/s41060-017-0054-1.
- [14] P. J. G. Lisboa, T. A. Etchells, I. H. Jarman, and S. J. Chambers, "Finding reproducible cluster partitions for the k-means algorithm," *BMC Bioinformatics*, vol. 14, no. SUPPL.1, pp. 1–19, 2013, doi: 10.1186/1471-2105-14-S1-S8.
- [15] O. Shamir and N. Tishby, "Stability and model selection in k-means clustering," *Mach. Learn.*, vol. 80, no. 2–3, pp. 213–243, 2010, doi: 10.1007/s10994-010-5177-8.
- [16] Y. D. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2017, doi: 10.37676/jmi.v12i2.418.

- [17] N. H. L. Suryani, "Penerapan Data Mining Dalam Produk Penjualan Pada Intan Permai Menggunakan Metode K-Means Clustering," pp. 0–1, 2017.
- [18] S. Huber, H. Wiemer, D. Schneider, and S. Ihlenfeldt, "DMME: Data mining methodology for engineering applications - A holistic extension to the CRISP-DM model," *Procedia CIRP*, vol. 79, pp. 403–408, 2019, doi: 10.1016/j.procir.2019.02.106.