



ANALISIS DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK CLUSTERING PENJUALAN STUDI KASUS DAPUR BU IPUNG

Asep Muhidin¹, Stefanus Kevin Alfandara², Nurhadi Surojudin³, Dedi Afandi⁴

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa

¹asep.muhidin@pelitabangsa.ac.id, ²alfandara.kevin@gmail.com

³Nurhadi.surojudin@pelitabangsa.ac.id, ⁴afandi.ihbs@gmail.com

Abstrak

Dapur Bu Ipung merupakan industri rumahan yang bergerak dalam pembuatan makanan donat yang berlokasi di perumahan Perum Cikarang Permai. Tidak hanya Dapur Bu Ipung saja, masih cukup banyak pembisnis lainnya yang bergerak di bidang serupa. Penelitian ini diharapkan dapat bermanfaat untuk membantu konsumen melihat produk yang kurang laris, laris, dan sangat laris pada penjualan Dapur Bu Ipung. Data yang terkumpul dari tahun 2016 sampai dengan tahun 2021 adalah 570 data transaksi penjualan. Tujuan dari penelitian ini untuk mengelompokkan data penjualan kedalam sebuah cluster dengan metode Data Mining Algoritma *K-Means Clustering*. Data penjualan di kelompokkan berdasarkan kemiripan data tersebut sehingga data dengan karakteristik yang sama akan dalam satu *cluster*. Sehingga dapat di kelompokkan menggunakan algoritma *K-Means* menjadi beberapa kriteria sangat laku, laku dan kurang laku. *Cluster* yang terbentuk setelah di lakukan proses *K-Means Clustering* terbagi menjadi 3 *Cluster*. Cluster-0 jumlah anggota 9 dengan presentase 64,28% dikategorikan Kurang Laris, Cluster-1 jumlah anggota 1 dengan presentase 7.14% dikategorikan Sangat Laris, dan Cluster-2 jumlah anggota 4 dengan presentase 27,57% dikategorikan Laris. Dari proses clustering diatas dapat di peroleh validasi DBI (*Davies Bouldin Index*) dengan nilai 0,451.

Kata kunci: Data Mining, Clustering, K-Means, Davies Bouldin Index

Abstract

Ibu Ipung kitchen is a home industry that is engaged in making donuts located in Perum Cikarang Permai housing. Not only Bu Ipung's Kitchen, there are still quite a few other businessmen who are engaged in similar fields. This research is expected to be useful to help consumers see the products that are not selling well, selling well, and selling very well in the sales of Dapur Bu Ipung. The data collected from 2016-2021 is 570 sales transaction data. The purpose of this study is to group sales data into a cluster using the K-Means Clustering Algorithm Data Mining method. Sales data are grouped based on the similarity of the data so that data with the same characteristics will be in one cluster. So that it can be grouped using the K-Means algorithm into several criteria that are very good, sellable and less well-sold. The cluster that is formed after the K-Means Clustering process is carried out is divided into 3

Clusters Cluster 0 with 9 members with a percentage of 64.28% categorized as Less Selling, Cluster 1 number of members 1 with a percentage of 7.14% is categorized as very popular, and Cluster 2 with a percentage of 4 members with a percentage of 27.57% is categorized as popular, from the clustering process above, DBI validation (Davies Bouldin Index) can be obtained with a value of 0.451.

Keywords: : Data mining, Clustering, K-Means, Davies Bouldin Index

1. Pendahuluan

Dalam dunia bisnis yang selalu dinamis dan penuh persaingan para pelakunya harus selalu memikirkan cara-cara untuk survive dan jika mungkin pengembangan skala bisnis mereka. Untuk mencapai hal tersebut, ada 3 kebutuhan bisnis yang dapat di lakukan, yaitu penambahan jenis maupun peningkatan

kapasitas produk, pengurangan biaya operasional perusahaan, dan peningkatan efektifitas pemasaran serta keuntungan, agar bisa memenuhi kebutuhan-kebutuhan bisnis di atas, banyak cara yang dapat di tempuh salah satunya adalah dengan melakukan analisis dana usaha. Dapur Bu Ipung merupakan home industri yang bergerak di bidang makanan. Tidak hanya Dapur Bu Ipung, masih cukup banyak perusahaan-perusahaan lain yang bergerak di bidang yang sama. Hal tersebut tentu saja menimbulkan persaingan bisnis antar pembisnis. Maka dari itu Dapur Bu Ipung melakukan serangkaian promosi-promosi aktif yang seragam, agar tidak kalah bersaing dengan perusahaan-perusahaan lainnya

Data penjualan yang sudah ada akan dikelola atau dianalisa untuk mengetahui tingkat kecenderungan konsumen di setiap tujuan pemasaran produk pada faktor ketertarikannya. Dari pengolahan data tersebut akan di peroleh satu pola konsumsi masyarakat terhadap produk dari perusahaan tersebut. Ketersediaan data yang cukup banyak, kebutuhan akan informasi (atau pengetahuan) sebagai pendukung pengambilan keputusan untuk membuat bussines solution serta dukungan infrastruktur di bidang teknologi informasi, maka teknologi informasi big data merupakan metode yang tepat di gunakan pada saat ini dalam pengambilan keputusan di dunia bisnis.

Banyaknya ragam jenis kue yang dijual dan banyaknya permintaan membuat Dapur Bu Ipung kesulitan dalam memenuhi permintaan dengan masih manualnya pencatatan transaksi dan manualnya proses manajemen penyetokan 2 barang melalui pengecekan manual data transaksi yang terjadi membuat kurang efisien dan akurat dalam penyetokan barang sehingga dalam pengalaman yang sudah terjadi, terjadi kekosongan barang karena stok yang tidak tersedia, tidak hanya permasalahan dalam stok barang yang kurang atau mengalami kekosongan barang, dalam pengalaman yang sudah berjalan penentuan penyediaan yang masih dilakukan secara manual dari menganalisis secara manual transaksi penjualan yang masih manual, membuat stok barang terlalu banyak sehingga membuat keuangan tidak berjalan dengan lancar karena barang yang ter-stok terlalu banyak, dengan teknologi informasi yang semakin maju ini, banyak teknik-teknik pengolahan data yang terlahir untuk menyelesaikan masalah-masalah yang dihadapi. Salah satu teknik pengolahan data yang sering dipakai yaitu teknik data mining.

Melihat dari permasalahan-permasalahan yang sudah dipaparkan sebelumnya, penulis mengusulkan teknik data mining untuk menyelesaikan masalah yang ada dengan mengolah data historis transaksi penjualan yang sudah ada. Dalam permasalahan yang ada penulis menggunakan metode K-Means, dengan mengklaster atau mengelompokkan barang-barang yang terjual menjadi 3 bagian yaitu sangat laku, laku dan kurang laku, dengan pembagian seperti ini diharapkan memudahkan Dapur Bu Ipung dalam menyusun strategi dalam manajemen penyetokan bahan dengan pengelompokan barang yang terjual dengan kriteria sangat laku, maka menyetok barang dalam jumlah 3

banyak, dengan kriteria laku, maka menyetok barang dengan jumlah sedang dan kriteria kurang laku, maka menyetok barang dengan jumlah sedikit. Maka dari itu penelitian yang akan di ambil yaitu “ANALISIS DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK CLUSTERING PENJUALAN DAPUR BU IPUNG”.

2. Landasan Pemikiran

2.1. Pengertian Analisis

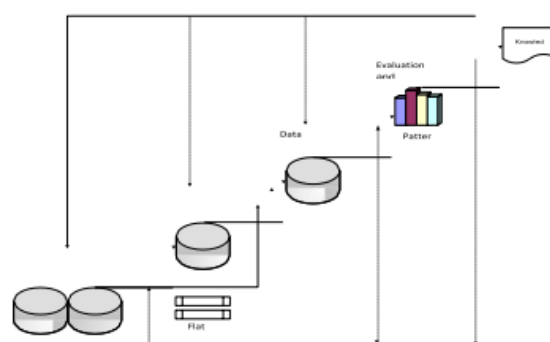
Menurut Jogiyanto Analisis dapat didefinisikan sebagai penguraian dari suatu sistem informasi yang utuh ke dalam bagian-bagian komponennya dengan maksud untuk mengidentifikasi dan mengevaluasi permasalahan, kesempatan, hambatan yang terjadi dan kebutuhan yang diharapkan sehingga dapat diusulkan perbaikannya. [1]

2.2. Data mining

Definisi data mining secara formal adalah proses mengekstrak informasi yang valid, bermanfaat, tak dikenal, dan dapat dipahami dari data dan menggunakannya untuk membuat keputusan bisnis. [2]

2.3. Knowledge Discovery and Database

Penelitian ini menggunakan metode Knowledge Discovery and Database merupakan suatu proses untuk mendapatkan suatu informasi pada sebuah data yang tersembunyi menjadi sebuah pola-pola karakteristik dari data tersebut.



Gambar 2.1 Tahap penemuan Knowledge pada Data mining (KDD) [3]:

- Data Cleaning
- Data Integration
- Data Selection
- Data Transformation
- Data Mining
- Interpretation / Evaluasi
- Knowledge Presentation

2.4. Clustering

Pada *Clustering* dapat dibedakan menjadi dua tujuan, yaitu clustering untuk pemahaman dan clustering untuk penggunaan. Jika tujuan untuk pemahaman maka cluster yang terbentuk harus menangkap struktur alami data. Biasanya proses *clustering* dalam tujuan ini hanya sebagai proses awal

untuk kemudian dilanjutkan dengan pekerjaan ini seperti *summarization* (rata-rata, *standar deviasi*), pelabelan kelas pada setiap kelompok untuk kemudian digunakan sebagai data latih klasifikasi, dan sebagainya. Sementara jika tujuannya untuk penggunaan, biasanya tujuan utama untuk mencari *prototype cluster* yang paling representative terhadap data dan memberikan abstraksi dan setiap objek data dalam cluster di mana sebuah data terletak didalamnya.[4]

2.5. Clustering K-Means

Algoritma K-Means merupakan algoritma pengelompokan iterative yang melakukan partisi set data ke dalam sejumlah K *cluster* yang sudah ditetapkan di awal. Algoritma K-Means sederhana untuk diimplementasikan dan dijalankan, relative cepat, mudah beradaptasi, umum penggunaannya dalam praktek. Secara historis, K-Means menjadi salah satu algoritma yang paling penting dalam bidang data mining.

K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok. Metode ini mempartisi ke dalam cluster atau kelompok sehingga data yang memiliki karakteristik yang sama (*High intra class similarity*) dikelompokkan ke dalam satu cluster yang sama dan yang memiliki karakteristik yang berbeda (*Low inter class similarity*) dikelompokkan pada kelompok yang lain. Proses *clustering* dimulai dengan mengidentifikasi data yang akan dicluster, X_{ij} ($i=1, \dots, n; j=1, \dots, m$) dengan n adalah jumlah data yang akan dikluster dan m adalah jumlah variabel. Pada awal iterasi, pusat setiap cluster ditetapkan secara bebas (sembarang), C_{kj} ($k=1, \dots, k; j=1, \dots, m$). Kemudian dihitung jarak antara setiap data dengan setiap pusat cluster. Untuk melakukan penghitungan jarak data ke- i (x_i) pada pusat cluster ke- k (c_k), diberi nama (d_{ik}), dapat digunakan formula *Euclidean*, seperti pada persamaan (1), yaitu:

$$\text{Min} \sum_{k=1}^k d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{ij})^2}$$

Suatu data akan menjadi anggota dari *cluster* ke- k apabila jarak data tersebut ke pusat *cluster* ke- k bernilai paling kecil jika dibandingkan dengan jarak ke pusat *cluster* lainnya.

Hal ini dapat dihitung dengan menggunakan persamaan (2) Selanjutnya, kelompokkan data-data yang menjadi anggota pada setiap *cluster*.

$$c_{ij} = \frac{\sum_{i=1}^n x_{ij}}{p}$$

Nilai pusat *cluster* yang baru dapat dihitung dengan cara mencari nilai rata-rata dari data-data yang menjadi anggota pada *cluster* tersebut, dengan menggunakan rumuspada persamaan (3):

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{ij})^2}$$

Dimana $x_{ij} \in$ cluster ke - k

$p =$ banyaknya anggota cluster ke k

2.6. Rapid Miner

Rapidminer merupakan sebuah aplikasi yang digunakan untuk mengolah data dengan berbagai teknik dan metode dalam data mining, sehingga data dapat menjadi informasi yang berguna untuk proses pengambilan keputusan. Rapidminer menyediakan prosedur data mining dan machine learning didalam fiturnya, termasuk extraction, transformation, loading, data pre-processing, visualisasi, modeling serta evaluasi.

Rapidminer memudahkan proses perhitungan data dalam skala besar menggunakan operator yang tersedia pada fiturnya. Operator-operator tersebut digunakan untuk mengekstrak data agar menjadi sebuah informasi. Pengujian data menggunakan Rapidminer juga dapat ditampilkan secara visual melalui grafik. [5].

2.7. Validasi Davies Bouldin Index (DBI)

Davies Bouldin Index (DBI) merupakan cara validasi cluster yang dibuat oleh D.L. Davies. DBI adalah fungsi rasio dari jumlah distribusi di dalam cluster untuk pemisahan antar *cluster*. Pengukuran menggunakan DBI bertujuan untuk memaksimalkan jarak *inter-cluster*. Dalam penelitian ini, DBI digunakan untuk melakukan validasi data pada setiap cluster. DBI dapat dihitung menggunakan persamaan [6]:

$$R_{i,j} = \frac{\text{var}(C_i) + \text{var}(C_j)}{\|c_i - c_j\|}$$

$$R_i = \max_{j=1, \dots, k, i \neq j} R_{ij}$$

$$DB = \frac{1}{k} \sum_{i=1}^k R_i$$

$$\text{var}(x) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

Keterangan:

R : jarak antar cluster

Var : variance dari data

X : data ke- i

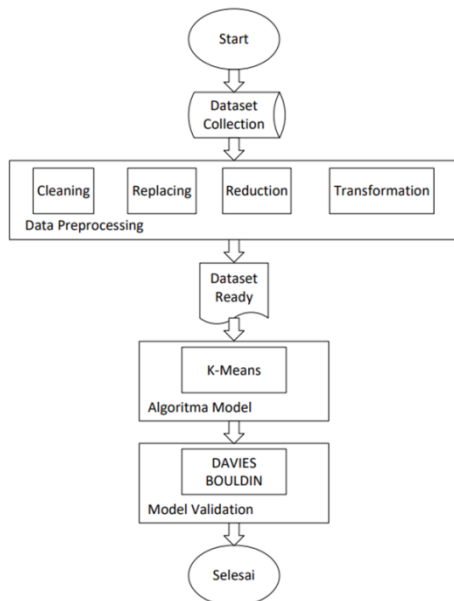
\bar{x} : rata-rata dari tiap cluster

DB : validasi Davies Bouldin

Dengan menggunakan Davies Bouldin Index suatu *cluster* akan dianggap memiliki skema *clustering* yang optimal jika memiliki *Index Davies Bouldin* minimal.

3. Metode Penelitian

Analisa klustering penjualan pada penelitian ini menggunakan tahapan yang sistematis dan terarah, sebagaimana yang digambarkan berikut :



Gambar 3.1 Metode Penelitian

3.1 Pengumpulan Data (Dataset Collection)

Dataset yang digunakan pada penelitian ini adalah data penjualan tahun 2016 sampai 2021. Pengumpulan data berdasarkan data yang di butuhkan pada penelitian. Dengan data ini diharapkan dapat memperoleh data-data yang relevan, akurat dan terpercaya.

Tabel 3.1 Dataset Penjualan

Total				
Nama Kue	Jumlah Transaksi	Minimal Order	Jumlah Terjual	Rata-Rata Penjualan
Bolu Kukus	99	15	1485	15
Bolu Keju	30	25	750	25
Bolu Kukus	43	25	1075	25
Cenil	18	20	360	20
Fruit Pie	75	20	1500	20
Kelpon	17	25	425	25
Kue Ku Jeruk	97	25	2425	25
Kue Lumpur	82	25	2050	25
Pie Brownies	2	20	40	20
Putu Ayu	11	25	275	25
Putu Belanda	35	25	875	25
Soes Buah	180	20	3600	20
Soes Via	38	20	760	20
Talam Pandan	48	25	1200	25

3.2 Preprocessing / Cleaning

Pada penelitian ini data dikumpulkan dan diidentifikasi untuk selanjutnya dilakukan *preprocessing* yakni pembersihan data (*cleaning*) dan melakukan *replacing* (*missing values*). Setelah melalui tahap preprocessing data maka data tersebut telah siap untuk digunakan untuk kemudia diproses ke tahap berikutnya.

3.3 Penerapan Algoritma

Pada penelitian ini algoritma yang digunakan ialah algoritma K-Means. Pada tahapan ini penggunaan algoritma K-Means digunakan untuk mendapatkan hasil *cluster* data yang akan dibagi menjadi 3 cluster.

3.4 Evaluasi

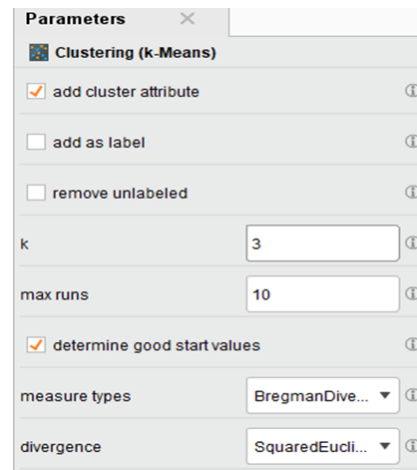
Validasi bertujuan untuk melihat perbandingan hasil dari model atau metode yang digunakan dengan

hasil yang telah ada sebelumnya. Proses evaluasi dari akurasi algoritma pada penelitian diukur menggunakan model *Davies Bouldin Index*.

4. Pembahasan

4.1 Proses Klustering pada RapidMiner

Paraameter yang digunakan untuk proses klustering adalah sebagai berikut :

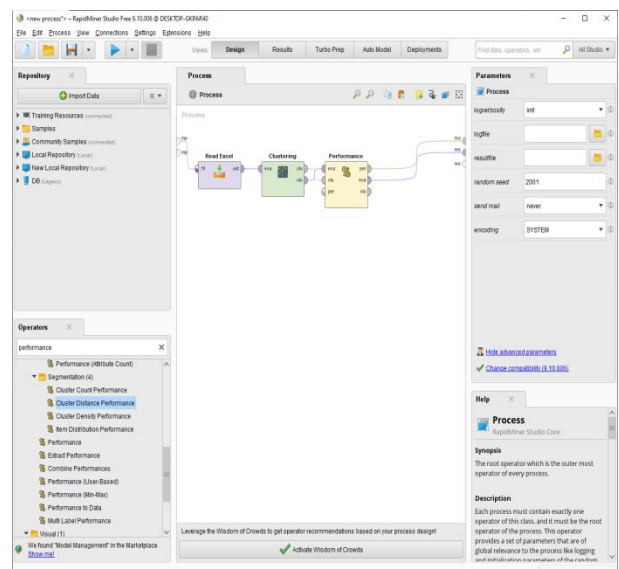


Gambar 4.1 Parameter K-Mean yang digunakan

Pada gambar 4.1, nilai k merupakan nilai yang digunakan untuk menentukan jumlah *cluster* yang akan dibentuk. Disini jumlah cluster yang akan dibentuk adalah sebanyak 3 cluster sesuai dengan tingkatan penjualan yaitu kurang laris, laris dan paling laris.

Desain model algoritma K-mean yang dibuat pada aplikasi Rapidminer sebagai berikut :

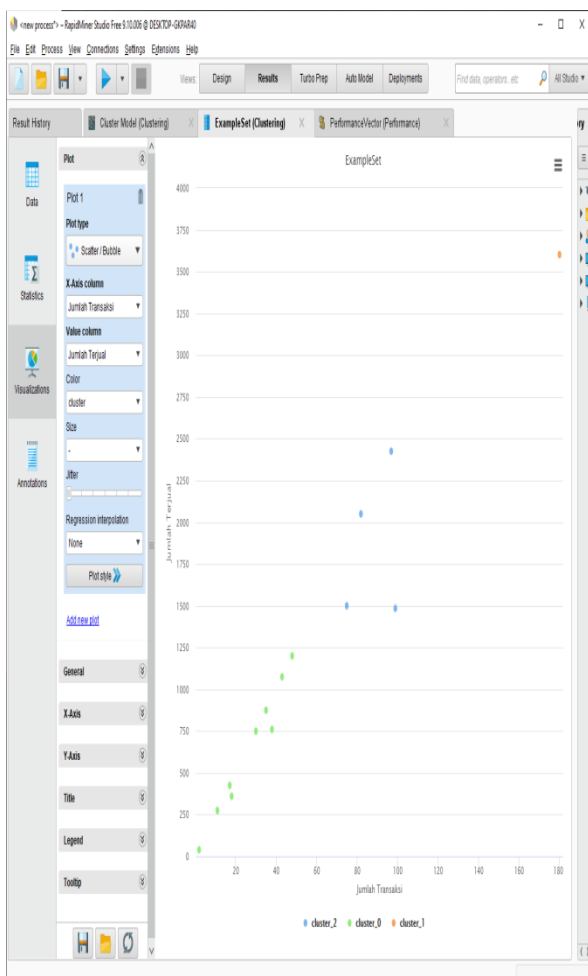
Gambar 4.2 Desain model K-Mean



Ada 3 proses yang akan dilakukan yaitu:

- 1) *Retrieve data test*
Tahapan ini dilakukan operasi penginputan dataset berupa file berekstensi .xlsx dataset Penjualan Dapur Bu Ipung.
- 2) *Clustering*
Tahapan ini dilakukan operasi *clustering* sebagai algoritma yang dilakukan pada penelitian ini.
- 3) *Performance*
Tahapan ini dilakukan operasi pencarian nilai *davies bouldin index*.

Setelah dijalankan, dan melalui proses 3 tahapan di atas, maka didapatkan beberapa output dan hasil pengujian oleh rapidminer yaitu sebagai berikut :



Gambar 4.3 Grafik Kluster dengan Scartter/Bubble

Visualisasi data menggambarkan hasil pengelompokan data dengan bentuk grafik titik dalam tiga warna untuk tiap cluster. Warna hijau mengartikan cluster-0 dengan jumlah 9 item, warna orange mengartikan cluster-1 dengan 1 item, dan pada warna biru mengartikan cluster-2 dengan 4 item.

Attribute	cluster_0	cluster_1	cluster_2
Jumlah Transaksi	26.889	180	88.250
Jumlah Terjual	640	3000	1895
Rata-Rata Pospalan	23.333	20	21.250

Adapun titik pusat tiap cluster digambarkan sebagai berikut :

Gambar 4.4 Centroid table

Nilai *centroid table* akan menjadi acuan perhitungan tiap-tiap dataset dengan mengukur nilai dengan masing-masing titik pusat cluster.

5. Penutup

Dari hasil penelitian yang telah dilakukan oleh peneliti, maka dapat diperoleh kesimpulan sebagai berikut:

- Metode Clustering dengan Algoritma K-Means dapat digunakan untuk pengelompokan data penjualan Dapur Bu Ipung berdasarkan jumlah penjualan yang di kelompokkan menjadi 14 item yaitu cluster 0 kurang laris, cluster 1 paling laris dan cluster 2 laris. Sehingga pihak pemilik Dapur Bu Ipung dapat mengantisipasi penjualan yang kurang laris ini bagaimana supaya menjadi laris.
- Metode Clustering ini dapat digunakan untuk membantu pihak pemilik Dapur Bu Ipung dalam melakukan pengelompokan penjualannya. Berdasarkan hasil pengujian dari cara manual dengan cara software menggunakan rapidminer v.9.10.006 mendapatkan hasil cluster 0 jumlah item 9 (kurang laris), cluster 1 jumlah item 1 (sangat laris) dan cluster 2 jumlah item 4 (laris). Hasil perhitungan menggunakan algoritma k-means DBI (Davies Bouldin Index) pada penelitian kali ini menunjukkan nilai - 0,451

Referensi

- [1] Yuhefizar, S. Hanik Mujiati, "Analisis Dan Perancangan Sistem Informasi Stok Obat Pada Apotek Arjowinangun," *Indones. J. Comput. Sci. - Speed FTI UNSA*, vol. 9330, no. 2, pp. 1–6, 2013.
- [2] H. Yulianton, "Data Mining untuk Dunia Bisnis Keputusan Informasi," *J. Teknol. Inf. Din.*, vol. XIII, no. 1, pp. 9–15, 2008.
- [3] D. Anggarwati, O. Nurdiawan, I. Ali, and D. A. Kurnia, "Penerapan Algoritma K-Means Dalam Prediksi Penjualan," *J. Data Sci. Inform. (Jdsi)*, vol. 1, no. 2, pp. 58–62, 2021.
- [4] A. Rijali, "Analisis Data Kualitatif," *Alhadharah J. Ilmu Dakwah*, vol. 17, no. 33, p. 81, 2019, doi: 10.18592/alhadharah.v17i33.2374.
- [5] 7) Marakas dan O'Brien (2017, "Bab Ii Landasan Teori," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 8–24, 2018.
- [6] Patel, "濟無No Title No Title No Title," pp. 9–25, 2019.