

Pengolahan dan Pengelompokan Data Makanan GoFood Menggunakan Algoritma K-Means Berdasarkan Harga Produk

Andhika Putra Pratama (119140224)
Teknik Informatika
Institut Teknologi Sumatera
Lampung Selatan, Indonesia
andhika.119140224@student.itera.ac.id

Andhika Wibawa (119140218)
Teknik Informatika
Institut Teknologi Sumatera
Lampung Selatan, Indonesia
andhika.119140218@student.itera.ac.id

Tri Aji Bagaskara (119140214)
Teknik Informatika
Institut Teknologi Sumatera
Lampung Selatan, Indonesia
tri.119140214@student.itera.ac.id

Salah satu sektor perekonomian yang paling terasa akan pemanfaatan aplikasi penjualan *online* adalah *food & beverage*. Gojek sebagai salah satu perusahaan yang menawarkan jasa di bidang *food & beverage* (GoFood) memiliki setidaknya 687 merchant dengan 23.900 produk yang tersebar di wilayah Medan, Surabaya, dan DKI Jakarta. Untuk dapat memperoleh informasi dari dataset GoFood tersebut, dilakukan proses visualisasi data serta *clustering* menggunakan K-means yang meliputi cakupan harga, jenis, serta sebaran produk/toko di tiap-tiap lokasi. Didapatkan hasil berupa sebaran harga, jumlah toko/produk, dan kategori produk serta 3 buah *cluster* harga (murah, mahal, sedang) dengan skor silhouette yaitu 0.62 yang ditampilkan dalam suatu peta interaktif.

Keywords — *clustering, data mining, gofood, k-means, makanan, minuman*

[1] I. PENDAHULUAN

Dalam beberapa tahun belakangan ini sektor perekonomian telah berkembang dengan adanya fenomena belanja menggunakan komunikasi digital seperti aplikasi penjualan secara *online*. Salah satu sektor perekonomian yang paling terasa akan pemanfaatan aplikasi penjualan *online* adalah *food & beverage* [1]. Salah satu aplikasi yang paling sering digunakan masyarakat adalah Gojek yang merupakan aplikasi digital yang pada hakikatnya menyediakan pelayanan antar jemput penumpang. Tetapi, seiring pertumbuhan permintaan konsumen akhirnya Gojek menyediakan fitur baru yaitu GoFood untuk layanan jasa transaksi *food & beverage* [2].

Pertumbuhan penggunaan layanan GoFood baik dari sisi pembeli dan penjual telah berkembang sangat pesat yang tak lepas dengan perkembangan dan pola gaya hidup masyarakat saat ini [3]. Hasil riset Nielsen menunjukkan rata-rata pembelian makanan secara online oleh masyarakat di Indonesia adalah 2,6 kali setiap minggunya [4]. Oleh sebab itulah pertumbuhan pengguna layanan GoFood dari sisi penjual juga mengalami peningkatan. Hal tersebut dapat dilihat dari data jumlah merchant yang terdaftar pada layanan GoFood dari 3 kota besar di Indonesia seperti Medan, Surabaya, dan DKI Jakarta yang mencapai 687 merchant dengan 23.900 produk yang dijual.

Berdasarkan latar belakang tersebut, untuk mendapatkan informasi *merchant* dan produk yang terdaftar pada layanan GoFood untuk area Medan, Surabaya, dan DKI Jakarta dilakukanlah penelitian “Pengolahan dan Pengelompokan Data Makanan GoFood Menggunakan Algoritma K-Means Berdasarkan Harga Produk”. Pengolahan data tersebut dilakukan untuk mendapatkan informasi berupa penyebaran produk yang dijual dan bagaimana penyebaran merchant

yang terdaftar. Pengelompokan produk juga dilakukan dengan metode K-Means untuk mengelompokkan data makanan berdasarkan harga produk.

[2] II. LANDASAN TEORI

A. Bisnis Online

Bisnis online atau yang biasa disebut sebagai e-commerce merupakan salah satu bentuk media usaha dengan memanfaatkan internet [5]. Berdasarkan data BPS dan APJII, pengguna internet di Indonesia telah mencapai 82 juta orang (pada tahun 2013) dan akan terus bertambah sehingga membantu para pebisnis menjadi lebih fleksibel [6]. Adanya bisnis *online* memungkinkan suatu toko untuk dapat berjalan sepenuhnya tanpa memerlukan waktu dan tempat yang spesifik. Pada tipe bisnis seperti ini, umumnya para penjual memanfaatkan media sosial dan suatu platform (pasar) untuk mempromosikan bisnis mereka [7]. Salah satu platform yang umum digunakan di Indonesia untuk bisnis online diantaranya adalah Gojek, Grab, Tokopedia, dan Shopee.

B. Layanan Gojek

Perusahaan Gojek pertama kali didirikan oleh Nadiem Makarim dan Michaelangelo Moran pada bulan Maret tahun 2014 [8]. Aplikasi Gojek awalnya dirancang sebagai pesaing dari perusahaan transportasi online lainnya seperti Grab dan Blue Bird. Namun, sampai saat ini Gojek telah berkembang dan melayani berbagai jasa lainnya seperti pemesanan makanan (GoFood), pengiriman (GoSend), tiket film (GoTix), dll secara online [9]. Per tahun 2019, aplikasi ini telah diunduh oleh setidaknya 125 juta orang dan telah melayani lebih dari 100 juta transaksi, yang diantaranya merupakan GoFood dengan mitra yang tersebar di 50 kota di Indonesia [9].

C. Data Mining

Data mining merupakan proses untuk menemukan hubungan dan tren yang bermakna dengan melakukan pemilihan data dalam jumlah yang besar yang disimpan dalam suatu *repository* menggunakan teknik statistika dan matematika. Dengan kata lain, *data mining* adalah proses untuk menemukan pola dan pengetahuan menarik dari data dalam jumlah yang besar [10].

Data mining memiliki beberapa tujuan, seperti *knowledge discovery* dan *pattern recognition*. *Knowledge discovery* berarti bahwa tujuan utama dari data mining adalah untuk mendapatkan pengetahuan yang masih tersembunyi di dalam bongkahan data, sedangkan pada *pattern recognition* berarti mencari pola yang tersembunyi di dalam bongkahan data

[10]. Data dari proses *mining* kemudian yang akan menghasilkan suatu informasi yang bersumber dari data yang lama [11]. Data mining memiliki dua kategori yaitu [12]:

- 1) *Descriptive mining*, merupakan suatu proses untuk mencari karakteristik dari data dalam satu basis data, contohnya *clustering association* dan *sequential mining*.
- 2) *Predictive mining*, merupakan proses untuk mencari pola dari data dengan menggunakan beberapa variabel lain, contohnya adalah klasifikasi.

D. Clustering

Terdapat 2 metode yang sering digunakan pada *machine learning*. *Supervised learning* digunakan untuk melakukan perbandingan antar kelas yang labelnya sudah ada dan tidak akan berubah, sedangkan *unsupervised learning* merupakan metode yang digunakan untuk melakukan pengelompokan data yang saling berhubungan dalam satu kluster dan data yang berbeda ke dalam *cluster* lainnya [13].

Clustering merupakan bagian dari *unsupervised learning* dan proses pengelompokan seperti *record* dan kelas/objek yang saling memiliki kemiripan. *Cluster* adalah suatu kelompok dari *record* yang memiliki kemiripan satu sama lain dan berbeda dengan *record* di *cluster* lain. *Clustering* mencoba untuk membagi seluruh kumpulan data menjadi kelompok yang relatif memiliki kemiripan, yang mana kemiripan *record* dalam satu kelompok bernilai maksimal, sedangkan kemiripan *record* dalam kelompok lain akan bernilai minimal [14]. Kesamaan pada objek ini biasanya diperoleh dari kedekatan nilai atribut yang menjelaskan objek data, sedangkan objek-objek data biasanya dipresentasikan sebagai sebuah titik dalam ruang multidimensi [14].

E. K-Means

K-means merupakan salah satu algoritma tertua dan yang paling banyak digunakan dalam *unsupervised learning* [15]. Algoritma *K-means* sangat bergantung kepada nilai *K* yang merupakan acuan banyaknya *cluster* yang akan dihasilkan [16]. *K-means* bekerja secara iteratif dimana *centroid* (pusat) digunakan untuk menghitung similaritas antar data dalam *cluster* menggunakan rumus jarak (Euclidean, dsb) [17]. Meski terlihat mudah, metode *K-means* juga memiliki beberapa kelemahan seperti sulitnya menentukan jumlah *cluster* yang cocok (pada data yang abstrak) dan serta hasilnya yang sangat dipengaruhi oleh letak *cluster* awal [15]. Penempatan *centroid* awal yang kurang tepat mudah menyebabkan *K-means* terjebak pada *local optima* [18]. Untuk menghindari berbagai kesalahan ini, *K-means* umumnya digunakan bersama metode evaluasi pendukung lainnya untuk menghasilkan hasil yang maksimal atau dimodifikasi kembali (menjadi algoritma yang berbeda) untuk menutupi kekurangannya.

F. Evaluasi K-Means

Salah satu metode evaluasi yang dapat digunakan pada *K-means* adalah *elbow method* dan *silhouette score*. *Elbow method* merupakan metode untuk mendeteksi *error* pada *K-means* dimana nilai *K* akan dinaikkan sedikit demi sedikit untuk mengurangi *Sum of Square Error* (SSE) sampai mendapatkan hasil yang stabil [18]. Lalu setelah nilai *K* yang terbaik dari *elbow method* dipakai, *silhouette score* dapat digunakan lebih lanjut. *Silhouette score* merupakan nilai rata-rata dari *silhouette coefficient* yang didapat dengan mencari

rata-rata jarak tiap-tiap poin pada data ke *intra-cluster* jika dibandingkan dengan jarak ke *cluster* terdekat lainnya [19]. Nilai *silhouette score* yang mendekati 1 berarti bahwa data sudah cukup berada di *cluster* yang sesuai, sedangkan score 0 melambangkan overlap dengan *cluster* lainnya, dan -1 berarti data berada di *cluster* yang salah [19].

[3] III. METODE DAN PEMBAHASAN

C. Impor Dataset

Proses impor dataset merupakan proses untuk membaca *file* data menjadi *dataframe*. *Dataset* daftar produk makanan/minuman dan *merchant* pada GoFood untuk area Medan, Surabaya, dan DKI Jakarta didapatkan dari Kaggle yang di-publish oleh Reyhan Ariq Syahalam pada tahun 2022. Digunakan *library pandas* untuk mengubah dan mengelola dataset tersebut nantinya.

D. Persiapan Data (Data Preprocessing)

a) Data Cleaning

Data cleaning adalah metode untuk membersihkan struktur data dari berbagai jenis error yang mungkin terjadi, misal karena data yang saling berduplikat atau tidak lengkap [20]. Pada proses ini, data numerik yang diolah adalah data harga makanan, maka proses *data cleaning* yang dilakukan terhadap dataset adalah mengubah tipe data harga makanan (harga normal dan harga diskon) dari string/float menjadi integer (agar bisa dihitung dan disaring berdasarkan harga nantinya).

b) Data Reduction

Data reduction adalah proses untuk mengurangi dimensi dan memperkecil ukuran data [21]. *Data reduction* pada umumnya berfokus akan informasi yang ingin diolah. Dalam hal ini, proses tersebut digunakan untuk menyingkirkan kolom "*display*" dan "*isDiscount*" karena tidak terlalu penting terhadap proses pengolahan data dan pencarian informasi nantinya.

c) Pencarian Titik Koordinat Toko

Proses pencarian titik koordinat toko dilakukan berdasarkan area yang didapatkan. Proses ini dilakukan dengan bantuan *library* Python yaitu *geopy* yang akan mencari nilai *latitude* yang biasa disebut dengan angka garis lintang dan nilai *longitude* yang biasanya disebut dengan angka garis bujur. Pencarian titik koordinat ini dilakukan untuk proses pengolahan data untuk melihat penyebaran toko pada *map*.

E. Pengolahan Data (Data Processing)

a) Mendapatkan Informasi Produk GoFood

Pada tahap ini dilakukan proses pengolahan data untuk mendapatkan informasi terkait dataset yang digunakan. Terdapat beberapa informasi yang akan diperoleh dari dataset yang digunakan seperti *merchant* dengan jumlah makanan/minuman terbanyak, sebaran harga makanan/minuman, *merchant* dengan harga makanan/minuman termahal, area dengan jumlah *merchant* terbanyak, area dengan harga makanan/minuman termurah, dan jenis makanan/minuman rata-rata per-area. *Library*

folium digunakan untuk meletakkan titik koordinat (pada map) yang telah ditentukan pada proses *data preprocessing* sebelumnya.

b) Clustering

Pada tahap ini dilakukan untuk proses pengelompokkan harga produk menjadi beberapa kelompok, digunakan metode pembelajaran mesin secara mandiri (*unsupervised*) yaitu *clustering*. *Clustering* adalah metode pengelompokkan data yang memiliki kemiripan karakteristik tanpa memerlukan label [22]. Dari banyaknya jenis *clustering* yang ada, *K-means* digunakan untuk menganalisis seberapa efektif *clustering* untuk dapat digunakan pada pengelompokkan produk berdasarkan harganya. Tak lupa pula proses evaluasi seberapa efektif proses *clustering* sebelumnya (*K-means*) terhadap dataset yang digunakan, dapat digunakan beberapa metode evaluasi, salah satu yang umum digunakan yaitu metode *silhouette* [23].

[4] IV. HASIL DAN ANALISIS

F. Ringkasan Data

Dataset yang diperoleh dari website kaggle dengan judul dataset “Indonesia food delivery GoFood product list”. *Dataset* tersebut diimpor menjadi *dataframe* melalui *library* *pandas*. Terdapat 9 kolom pada dataset yang digunakan yaitu *merchant_name*, *merchant_area*, *category*, *display*, *product*, *price*, *discount_price*, *isDiscount*, dan *description*. Berikut gambaran tabel kolom pada dataset yang didapatkan:

TABEL 1. DATA MENTAH DATASET

merchant_name	merchant_area	price	description
330 Kopi, Ciledug	jakarta	20000	Sajian Kopi Susu Gula Aren Yang Berbeda ...
330 Kopi, Ciledug	jakarta	22000	Sajian Kopi Susu Gula Aren Yang Berbeda
....
330 Kopi, Ciledug	jakarta	20000	Sajian Susu Coklat Milo Dengan Racikan
Zeger!, Sidoarjo	surabaya	85000	1L Kopi Susu Gula Aren
Zeger!, Sidoarjo	surabaya	85000	1L Salt Caramel Machiato

Data mentah (*raw data*) tersebut terdiri dari 45195 baris dan 9 kolom data. Pada proses *preprocessing* dilakukan perubahan bentuk tipe data kolom “price” dan “discount_price” menjadi bentuk tipe data *integer* dan pada proses ini juga dilakukan penghapusan kolom “display” dan “isDiscount”. Selanjutnya proses pemecahan kolom lokasi, pada proses ini dilakukan pemecahan data dari kolom “merchant_name” menjadi kolom “merchant_name” dan “merchant_location” untuk nantinya diproses agar mendapatkan nilai titik koordinatnya. Pada pemrosesan pencarian titik koordinat juga ditambahkan kolom baru “longitude” dan “latitude”. Sehingga *dataset final* yang akan dilakukan untuk data processing adalah dataset yang terdiri dari 9 kolom yaitu *merchant_name*, *category*, *product*, *price*, *discount_price*, *description*, *area*, *latitude*, dan *longitude*.

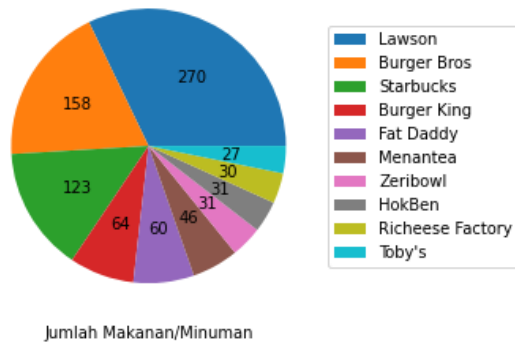
TABEL 2. DATASET FINAL

merchant_name	category	area	latitude	longitude
Banzai!	Cepat saji/Aneka nasi/Jepang	Bogor Utara, Bogor, Jawa Barat	- 6.57171 33	106.7836422
Bunga Bakery	Roti	Cibinong, Bogor, Jawa Barat	- 6.48245 23	106.8016856
....
Yasaka Fried Chicken	Aneka nasi/Ayam & bebek/Cepat saji	Sukolegok, Sidoarjo, Jawa Timur	- 7.37740 0	112.692859
Burger Bros	Jajanan/Barat/Cepat saji/Ayam & bebek	Bintaro, Jakarta Selatan, Jakarta	- 6.27038 79	106.7278611
Geprek Mms	Ayam & bebek/Minuman/Aneka nasi	Gedangan, Sidoarjo, Jawa Timur	- 7.38935 41	112.6932438

G. Pengolahan Informasi Data

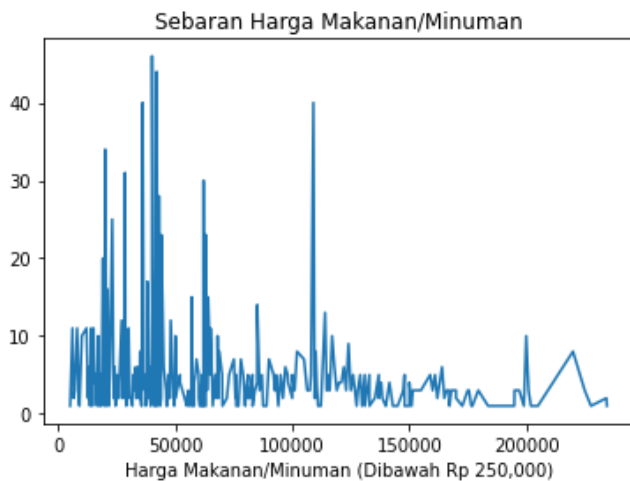
Setelah dilakukan *data preprocessing* maka dilakukan proses *data processing* untuk mendapatkan informasi-informasi dari dataset yang diolah. Informasi yang ingin didapatkan pertama adalah merchant dengan jumlah produk(makanan/minuman) terbanyak. Pada proses ini didapatkan bahwa *merchant* yang memiliki produk terbanyak adalah Lawson dengan total 270 produk dan terendah adalah Geprek Mms sebanyak 1 produk saja. Rata-rata produk setiap *merchant* berada di angka 17.5 produk. Berikut grafik 10 *merchant* dengan jumlah produk terbanyak pada dataset.

10 Toko dengan Jumlah Makanan/Minuman Terbanyak



Gambar 1. 10 Toko dengan jumlah makanan/minuman terbanyak

Informasi selanjutnya yang ingin diperoleh adalah sebaran harga produk (makanan/minuman). Setelah dilakukan proses pengolahan dataset untuk mendapatkan informasi yang dimaksud, didapatkan hasil berupa terdapat 46 produk dengan harga Rp 40.000 dan hanya ada 1 produk dengan harga Rp 21.500. Sebaran tersebut dapat dilihat pada gambar berikut ini:



Gambar 2. Sebaran harga produk

Pada tahapan selanjutnya diperoleh informasi toko dengan harga makanan/minuman termahal dan termurah. Hasil yang didapatkan setelah pengolahan dataset yang ada, didapatkan bahwa *merchant* Wahaha memiliki produk termahal dengan harga Rp 196.500 dan yang termurah adalah Pentol Seafood Korea dengan harga Rp 8.600. Adapun rata-rata produk berada diangka Rp 60.271.45. Berikut 10 *merchant* yang menjual produk termahal dan termurah:



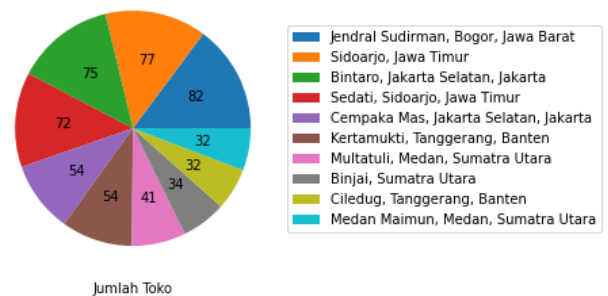
Gambar 3. 10 Toko dengan harga makanan/minuman termahal



Gambar 4. 10 Toko dengan harga makanan/minuman termurah

Lanjut ke proses selanjutnya yaitu mendapatkan informasi terkait area dengan jumlah toko terbanyak. Setelah dilakukan pemrosesan data, didapatkan informasi berupa area yang memiliki *merchant* terbanyak adalah area Jendral Sudirman, Bogor, Jawa Barat dengan 82 *merchant* pada area tersebut dan area yang memiliki *merchant* paling sedikit adalah Waru, Pamekasan, Jawa Timur yang hanya memiliki 1 *merchant* saja, terakhir rata-rata setiap area memiliki 14.68 *merchant*.

10 Area dengan Jumlah Toko Terbanyak



Gambar 5. 10 Area dengan jumlah toko terbanyak

Selanjutnya dilakukan proses untuk mendapatkan informasi terkait area dengan harga makanan/minuman termahal dan termurah dan jenis makanan/minuman rata-rata per area. Setelah dilakukan pemrosesan data, didapatkan informasi berupa area yang harga makanan/minuman termahal adalah Kelapa Gading, Jakarta Utara, Jakarta dengan produk seharga Rp 196.500 dan termurah terdapat pada area Pengantin Ali, Jakarta Timur, Jakarta dengan harga Rp 8.600. Selanjutnya didapatkan informasi tentang jenis makanan/minuman rata-rata untuk setiap area yang menghasilkan 100 baris data yang dapat dilihat pada Tabel 3.



Gambar 6. 10 Area dengan harga makanan/minuman termahal



Gambar 7. 10 Area dengan harga makanan/minuman termurah

TABEL 3. 10 JENIS MAKANAN/MINUMAN RATA-RATA UNTUK SETIAP AREA

No	Area	Kategori
1	Abadi, Medan, Sumatra Utara	Seafood
2	Akses UI, Depok, Jawa Barat	Jajanan/Aneka nasi/Minuman/Jepang
3	Aloha, Sidoarjo, Jawa Timur	Minuman/Kopi/Sate
4	Aren Jaya, Bekasi, Jawa Barat	Korea
5	Bangbarung, Bogor, Jawa Barat	Sweets
....
96	Tunjungan Plaza, Surabaya, Jawa Timur	Minuman
97	Waru, Pamekasan, Jawa Timur	Minuman/Aneka nasi/Jajanan
98	Wiyung, Surabaya, Jawa Timur	Ayam & bebek/Cepat saji
99	Wonoayu, Sidoarjo, Jawa Timur	Cepat saji/Jajanan
100	Yasmin Raya, Bogor, Jawa Barat	Jepang/Aneka nasi

Terakhir adalah proses visualisasi lokasi *merchant* secara langsung pada map. Proses ini menggunakan nilai *latitude* dan *longitude* serta bantuan *library* Python yaitu *folium*. Pada proses ini didapatkan berupa gambaran lokasi-lokasi *merchant* seperti pada gambar 4.8 dan dapat langsung dilihat informasi *merchant* yang bersangkutan seperti pada gambar 7 dan gambar 8.



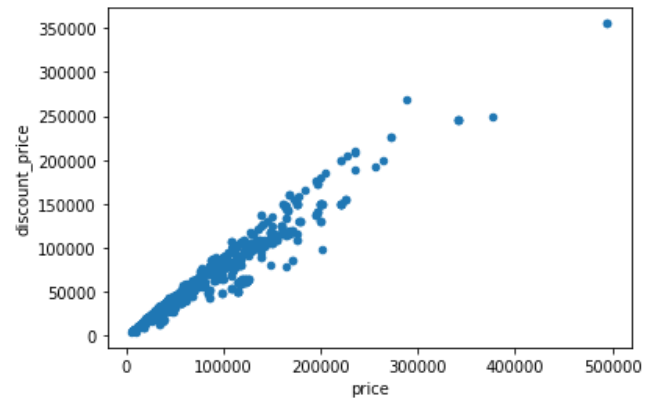
Gambar 8. Visualisasi lokasi *merchant*



Gambar 9. Visualisasi lokasi *merchant*

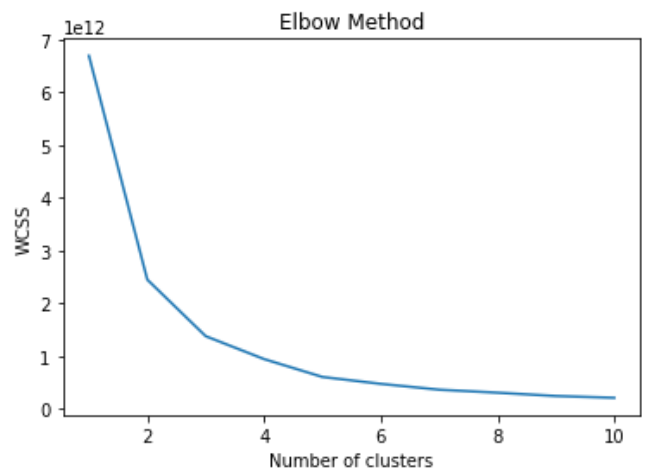
H. Clustering Data Menggunakan Metode K-Means

Pada proses *clustering* harga produk (makanan/minuman) ini menggunakan algoritma *K-means* yang telah disediakan oleh *library* *scikit-learn*. Dengan memanfaatkan *library* yang ada kita hanya perlu memasukkan dataset dan *parameter* yang dibutuhkan untuk proses *clustering K-means*. Proses *clustering* ini menggunakan data harga seluruh produk (kolom "*price*" dan "*discount_price*") dan visualisasi penyebaran data tersebut dapat dilihat pada gambar 9.



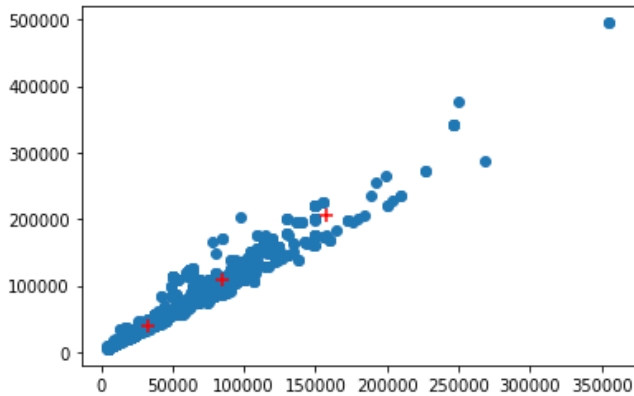
Gambar 10. Visualisasi data harga produk

Sebelum melakukan proses *clustering*, dilakukan proses pencarian nilai *K* dengan *elbow method* sehingga nanti akan didapatkan nilai *K* paling optimal untuk proses *clustering* dengan metode *K-Means*. Pada proses pencarian nilai *K* ini perhitungan ke *centroid* menggunakan metode *within-cluster sum of squares* (WCSS) dan didapatkan nilai optimal *K* yaitu 3.

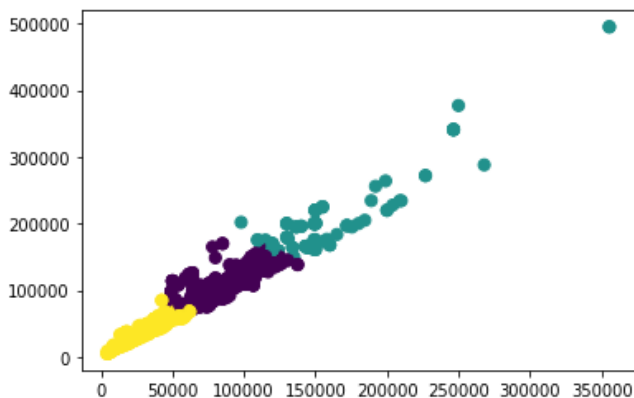


Gambar 11. Visualisasi elbow method

Setelah nilai K didapatkan, dilakukan proses clustering dengan K-Means. Proses pertama adalah menentukan *centroid* untuk *clustering* yang mana *centroid* yang digunakan dapat dilihat pada Gambar 11. Terakhir adalah proses *clustering* dengan dimana akan menghasilkan jumlah cluster sebanyak nilai K dan melakukan visualisasi hasil cluster. Hasil akhir cluster dapat dilihat pada Gambar 12.



Gambar 12. Visualisasi titik centroid



Gambar 13. Visualisasi hasil clustering

Proses *clustering* telah berhasil dilakukan, namun terdapat proses lanjutan yaitu evaluasi hasil *clustering* dengan menggunakan *silhouette coefficient*. Hasil evaluasi clustering ini mendapatkan *silhouette score* sebesar 0.62. Artinya, jarak antar *cluster* dapat dikatakan cukup renggang, namun tidak sampai saling menimpa antar *cluster*.

Hasil *clustering* tersebut pada akhirnya akan dimasukkan ke dalam *dataset* untuk menandakan kelompok produk (makanan/minuman) yang ada pada *dataset* seperti dapat dilihat pada tabel 4. Sebelum itu, dilakukan proses untuk mencari harga maksimum atau tertinggi untuk masing-masing *cluster* dan didapatkan hasil berupa harga produk tertinggi untuk *cluster* 0 adalah Rp 170.001, *cluster* 1 pada harga Rp 495.000, dan *cluster* 2 dengan harga Rp 85.000.

TABEL 4 DATASET DENGAN CLUSTERNYA

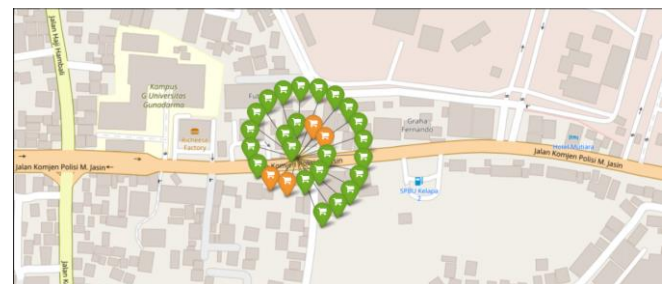
mercha nt_nam e	category	area	latitude	Cluster
Banzai!	Cepat saji/Aneka nasi/Jepang	- 6.571713 3	106.7836 422	0

Bunga Bakery	Roti	- 6.482452 3	106.8016 856	0
....
Yasaka Fried Chicken	Aneka nasi/Ayam & bebek/Cepat saji	- 7.377400	112.6928 59	2
Burger Bros	Jajanan/Barat/ Cepat saji/Ayam & bebek	- 6.270387 9	106.7278 611	2

Dari hasil nilai maksimal masing-masing cluster ditentukan bahwa *merchant* yang menjual produk termasuk *cluster* 2 akan diberi tanda warna hijau, *cluster* 0 berwarna oranye, dan terakhir *cluster* 1 akan diberikan label warna merah. Pewarnaan ini pada hakikatnya dilakukan berdasarkan pola harga dari termurah hingga termahal atau dengan kata lain produk pada *cluster* 2 adalah produk dengan harga murah, *cluster* 0 adalah produk dengan harga terjangkau, dan produk *cluster* 1 adalah produk dengan harga mahal. Contoh pelabelan *merchant* tersebut dapat dilihat pada gambar 13 dan 14.



Gambar 14. Visualisasi penandaan merchant



Gambar 15. Visualisasi penandaan merchant

[5] V. KESIMPULAN

Berdasarkan hasil pembahasan yang telah dijelaskan diatas, maka dapat ditarik beberapa kesimpulan berikut:

- 1) *Dataset* daftar produk makanan/minuman di GoFood berhasil diolah agar dapat menghasilkan informasi setelah beberapa tahapan data preprocessing (*data transformation* dan *data reduction*) serta *clustering* menggunakan algoritma K-Means.

- 2) Proses pengolahan data berhasil untuk menarik informasi seperti *merchant* dengan jumlah makanan/minuman terbanyak, sebaran harga makanan/minuman, *merchant* dengan harga makanan/minuman termahal dan termurah, area dengan jumlah *merchant* terbanyak, area dengan harga makanan/minuman termahal dan termurah, dan terakhir jenis makanan/minuman rata-rata per area.
- 3) Data harga produk makanan/minuman berhasil di-*clustering* dengan algoritma K-Means menjadi 3 buah *cluster*/kategori. Penentuan nilai K pada algoritma K-Means sangat mempengaruhi hasil cluster karena nilai K tersebut adalah jumlah cluster yang akan dibuat. Nilai K didapatkan juga harus melalui proses pencarian K dengan *elbow method* sehingga nilai K yang digunakan adalah nilai terbaik.
- 4) Berdasarkan nilai *silhouette score* yang didapatkan, hasil *cluster* yang terbentuk cukup renggang namun tidak sampai saling tindih dengan cluster lainnya.
- 5) Pemberian label untuk setiap produk berdasarkan *cluster* dapat memberikan gambaran tersendiri akan level produk yang dijualnya antara murah, terjangkau, ataupun mahal. Hal tersebut dipermudah dengan visualisasi langsung pada *map*.

REFERENSI

- [6] L. Erynayati, M. Ayu, D. Geriadi, P. Yuliana, and R. Sawitri, "Pilihan Konsumen Dalam Menggunakan Fitur GoFood : Variabel Kegunaan sebagai Moderator," vol. XXVI, no. 01, pp. 19–34.
- [7] Nurbayti, "Tren Pengguna Aplikasi GoFood di Era Digital," vol. 1, no. 1, pp. 1–10, 2019.
- [8] L. Prapti and Rahoyo, "DAMPAK BISNIS KULINER MELALUI GO FOOD BAGI PERTUMBUHAN EKONOMI DI KOTA SEMARANG Rr . Lulus Prapti NSS , SE , Msi Rahoyo , SE ., MM commerce," vol. 20, no. 2, pp. 120–133, 2018.
- [9] M. M. Ikram, "Keputusan Penggunaan Layanan Gofood Selama Masa Pandemi Covid-19," vol. 9, no. 2, pp. 279–291, 2021.
- [10] R. R. Situmeang, "DAMPAK BISNIS ONLINE DAN LAPANGAN PEKERJAAN TERHADAP PENINGKATAN PENDAPATAN MASYARAKAT (STUDI KASUS JASA BISNIS ONLINE TRANSPORTASI GRAB DI KOTA MEDAN) Pergerakan Ilmu Pengetahuan sangat menentukan berubah atau tidaknya peradaban manusia , dimana manusi," AJIE - Asian J. Innov. Entrep., vol. 03, no. September, pp. 319–335, 2018.
- [11] L. Y. Siregar and M. I. P. Nasution, "PERKEMBANGAN TEKNOLOGI INFORMASI TERHADAP PENINGKATAN BISNIS ONLINE," J. Ilm. Manaj. dan Bisnis, vol. 2, no. 1, pp. 71–75, 2020.
- [12] A. Praditya, "PENGARUH MEDIA SOSIAL DAN KOMUNIKASI BISNIS TERHADAP PERKEMBANGAN BISNIS ONLINE SHOP," vol. 2, no. 1, 2019.
- [13] F. ANGGRAINI, "PENGARUH HARGA, PROMOSI, DAN KUALITAS PELAYANAN TERHADAP LOYALITAS PELANGGAN DIMEDIASI KEPUASAN PELANGGAN PADA KONSUMEN GOJEK," 2020.
- [14] M. L. Lazuardi and I. Sukoco, "Design Thinking David Kelley & Tim Brown: Otak Dibalik Penciptaan Aplikasi Gojek," vol. 02, no. 01, pp. 1–11, 2019.
- [15] B. N. Sari and A. Primajaya, "Penerapan clustering dbscan untuk pertanian padi di kabupaten karawang," J. Inform. dan Komput., vol. 4, no. 1, pp. 28–34, 2019.
- [16] S. Nurdiani, S. Linawati, R. A. Safitri, and E. P. Saputra, "Pengelompokan Perilaku Mahasiswa Pada Perkuliahan E-Learning dengan K-Means Clustering," vol. 19, no. 2, 2019.
- [17] C. Armayani, A. Fauzi, and H. Sembiring, "IMPLEMENTASI DATA MINING PENGELOMPOKAN JUMLAH DATA PRODUKTIVITAS UBINAN TANAMAN PANGAN CLUSTERING DIKAB LANGKAT (STUDI KASUS : BADAN PUSAT STATISTIK LANGKAT)," vol. 5, no. 1, 2021.
- [18] E. Hancer, B. Xue, and M. Zhang, "A survey on feature selection approaches for clustering," Artif. Intell. Rev., no. 0123456789, 2020.
- [19] Z. Nabila, A. R. Isnain, Permata, and Z. Abidin, "ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19," vol. 2, no. 2, pp. 100–108, 2021.
- [20] K. P. Sinaga and M. Yang, "Unsupervised K-Means Clustering Algorithm," pp. 1–13, 2020.
- [21] M. Ahmed, "performance evaluation The k-means Algorithm : A Comprehensive Survey and Performance Evaluation," vol. 9, 2020.
- [22] C. Yuan, "Research on K-Value Selection Method of K-Means Clustering Algorithm," 2019.
- [23] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satot, "Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster," IOP Publ., 2018.
- [24] [19] K. R. Shahapure and C. Nicholas, "Cluster Quality Analysis Using Silhouette Score," 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), 2020, pp. 747–748, doi: 10.1109/DSAA49011.2020.00096.
- [25] Tae, K. H., Roh, Y., Oh, Y. H., Kim, H., & Whang, S. E., "Data cleaning for accurate, fair, and robust models: A big data-AI integration approach", In Proceedings of the 3rd International Workshop on Data Management for End-to-End Machine Learning (pp. 1–4), 2019.
- [26] Li, S., Marsaglia, N., Garth, C., Woodring, J., Clyne, J. and Childs, H., "Data Reduction Techniques for Simulation, Visualization and Data Analysis. Computer Graphics Forum", 37: 422–447. <https://doi.org/10.1111/cgf.13336>, 2018.
- [27] A. Ghosal, A. Nandy, A. K. Das, S. Goswami, and M. Panday, A Short Review on Different Clustering Techniques and Their Applications. Springer Singapore, 2020.
- [28] A. Dockhorn, C. Braune, and R. Kruse, "An Alternating Optimization Approach based on Hierarchical Adaptations of DBSCAN," no. 2, pp. 749–755, 2015.