

EKSPLORASI PENGGUNAAN ALGORITMA *K-MEANS* CLUSTERING UNTUK ANALISIS TREND HARGA KOMODITAS WILAYAH INDONESIA TIMUR

Marchel Yusuf Rumlawang Arpi

Program Studi Teknik Informatika, Teknik Informasi, Universitas Tarumanagara
Jl. Letjen S. Parman No.1, Jakarta Barat, DKI Jakarta, Indonesia 11410
e-mail: marchelyusufstudy@gmail.com

ABSTRAK

Penerapan *Machine Learning* telah banyak bermanfaat dalam pengambilan keputusan dan membantu melakukan analisis di berbagai bidang industri. Salah satu algoritma yang paling sering digunakan adalah *K-Means* untuk klusterisasi. Algoritma tersebut digunakan untuk menganalisis tren harga komoditas di Indonesia Timur. Tujuan penelitian ini adalah untuk melihat bagaimana *K-Means* bekerja dalam mengelompokkan setiap komoditas di suatu wilayah ke dalam beberapa kluster, menghitung skor siluet rata-rata yang dihasilkan, dan mengukur sejauh mana setiap sampel dalam kluster tertentu sesuai dengan kumpulan kelompok terdekatnya dibandingkan dengan kumpulan tetangga terdekat. Hasil segmentasi pasar menunjukkan bahwa setiap kelompok dengan perilaku dan karakteristik serupa dapat diperlakukan secara berbeda dalam strategi pemasaran dan penetapan harga terutama untuk komoditas antara beras dan daging sapi serta minyak goreng dan gula yang saling berpengaruh bahwa jika harga salah satu komoditas naik, harga komoditas lain juga akan naik.

Kata kunci: Pembelajaran Mesin, Klusterisasi, *K-Means*, Komoditas, Tren

ABSTRACT

The application of machine learning has been very useful in decision making and analysis in various industrial fields. One of the most frequently used algorithms is K-Means for clustering. This algorithm is used to analyze commodity price trends in Eastern Indonesia. The purpose of this study is to see how K-Means works in grouping each commodity in a region into several clusters, calculating the average silhouette score produced, and measuring the extent to which each sample in a particular cluster matches its closest group compared to its closest neighbors. The market segmentation results show that each group with similar behavior and characteristics can be treated differently in marketing and pricing strategies, especially for commodities such as rice and beef, as well as cooking oil and sugar, which influence each other in such a way that if the price of one commodity rises, the price of other commodities will also rise.

Keywords: Clustering, Commodities, *K-Means*, Machine Learning, Trends.

1. PENDAHULUAN

Komoditas adalah bahan baku yang dapat diperdagangkan dengan nilai yang sama. Harga pangan yang tersedia di wilayah timur Indonesia masih tidak stabil [1]. Ancaman bencana alam seperti banjir, longsor, dan bencana lainnya sering menggagalkan panen sehingga pasokan menjadi mudah busuk dan mengganggu distribusi komoditas pangan, terutama karena produksi terpusat di bagian barat Indonesia, sehingga distribusi memakan waktu lama untuk mencapai wilayah tujuan. Hal ini menyebabkan harga pangan masuk ke kategori harga tinggi [2].

Sektor pertanian sangat penting dalam perekonomian nasional, seperti yang terlihat dari jumlah orang yang tinggal dan bekerja di sektor ini. Indonesia memiliki lahan pertanian yang luas, sehingga ketahanan pangan perlu ditingkatkan untuk pertumbuhan ekonomi yang lebih lanjut. Dengan demikian, ketersediaan makanan sehat dan bergizi dapat dipenuhi oleh negara untuk rakyatnya. Ada juga tantangan dalam menghadapi kenaikan harga pangan akibat permintaan yang tinggi pada hari-hari perayaan atau risiko seperti kekurangan lahan dan penurunan produktivitas pertanian.

Klasterisasi menjadi salah satu metode yang efisien dan cepat yang bekerja pada atribut numerik dengan mengandalkan teknik pengelompokan berbasis jarak terdekat dengan membagi data ke dalam beberapa kluster [3]. Pusat Informasi Harga Pangan Strategis Nasional (PIHPS-Nasional) adalah lembaga pemerintah yang menyediakan data historis dan terkini. Berdasarkan data dari Pusat Informasi Harga Pangan Strategis (PIHPS) Nasional, per Juli 2021, secara rata-rata harga cabai mengalami kenaikan yang cukup signifikan. Kenaikan harga cabai rawit per kg setara dengan harga daging ayam per ekor mencapai Rp.67.500,-/Kg pada akhir Juli 2021 di Provinsi Maluku. Tidak hanya di Maluku, kenaikan harga cabai rawit hampir di seluruh provinsi juga ikut mengalami hal tersebut. Harga rata-rata secara nasional yakni sebesar Rp. 41.244,-/Kg [4].

2. TINJAUAN LITERATUR

Penelitian ini bertujuan untuk mengeksplorasi penerapan algoritma K-Means dalam menganalisis kluster tren harga komoditas di wilayah di Indonesia Timur. Algoritma K-Means digunakan untuk inisialisasi parameter karena sederhana dan bekerja dengan baik pada dataset besar dibandingkan dengan *hierarchical clustering* [5]. Algoritma K-Means bekerja dengan cara secara acak menetapkan nilai kluster (k), di mana nilai tersebut menjadi pusat kluster, yang juga dikenal sebagai *centroid*. Kelemahan algoritma K-Means adalah saat menentukan kluster awal karena hal ini bergantung pada data awal yang diberikan [6]. K-Means adalah salah satu algoritma dalam teknik penambangan data yang dapat mengelompokkan data heterogen karena algoritma pengelompokan hanya mampu mengenali nilai atribut yang homogen. Proses algoritma K-Means berbeda dengan algoritma penambangan data lainnya seperti Algoritma Apriori yang mencari himpunan item yang sering muncul pada model data heterogen, kemudian memangkas dan menghitung berdasarkan jumlah himpunan item K . Pengelompokan mengelompokkan beberapa objek ke dalam k kelas berdasarkan perhitungan jarak mereka ke pusat kluster [7].

Objek kluster dapat dilihat dari jarak objek ke titik pusat terdekat. Setelah mengetahui titik pusat terdekat, objek akan diklasifikasikan sebagai anggota atau bukan dari kategori tersebut. Proses klasterisasi merupakan bagian dari *unsupervised learning* yaitu proses yang membagi data ke dalam kelas atau kelompok berdasarkan tingkat kemiripannya. Beberapa penelitian sebelumnya telah dilakukan, misalnya penelitian menggunakan algoritma K-Means untuk memetakan pasar tradisional berdasarkan harga komoditas pangan dan pengelompokan data ekspor buah berdasarkan negara tujuan [8] [9] [10].

3. METODE PENELITIAN

Data yang dikumpulkan dalam studi ini diperoleh dari situs web PIHPS (<https://www.bi.go.id/hargapangan/TabelHarga/PasarTradisionalDaerah>).

3.1 Pengolahan Data

Dataset yang telah diunduh, disusun ke dalam file *excel* dengan contoh seperti pada Tabel 1. Dengan total 1278 sampel dan ditambahkan 1 fitur, yaitu “*Date*” mulai dari 2020-01-01 hingga 2023-06-30 sehingga terdapat 11 fitur dalam dataset. Setiap provinsi ditetapkan sebagai nama lembar kerja. Penanganan nilai yang hilang dilakukan karena terdapat cukup banyak sel dalam dataset yang tidak memiliki nilai. Penanganan ini menggunakan fungsi *ffill()* untuk mengisi nilai yang hilang dengan nilai terakhir yang ditemukan sebelumnya, yang dikenal sebagai *forward fill*, dan fungsi *bfill()* untuk mengisi nilai yang hilang dengan nilai berikutnya yang ditemukan, yang dikenal sebagai *backward fill*.

Tabel 1. Dataset Komoditas Pangan (*sheet_name* = “Maluku”)

Date	Rice	Chicken Meat	Beef	Chicken Eggs	Shallots	Garlic	Red Chillies	Cayenne Pepper	Cooking Oil	Granulated Sugar
01/01/2020 00:00										
01/02/2020 00:00	11.350	36.250	108.750	23.900	34.650	26.400	29.500	33.650	13.500	13.600
01/03/2020 00:00	11.350	36.250	108.750	23.500	34.650	27.000	30.500	32.600	13.500	13.600
01/04/2020 00:00										
01/05/2020 00:00										
01/06/2020 00:00	11.350	36.250	108.750	23.500	34.150	27.750	31.750	36.250	13.500	13.700
01/07/2020 00:00	11.350	36.000	108.750	23.500	33.250	30.150	35.450	37.950	13.650	13.750
01/08/2020 00:00	11.350	36.000	108.750	23.500	33.250	30.650	35.350	38.150	13.650	13.800
01/09/2020 00:00	11.500	36.000	108.750	23.500	35.250	30.650	38.400	39.400	13.650	13.800
01/10/2020 00:00	11.500	36.000	108.750	23.500	35.250	31.500	38.500	38.400	13.800	13.850

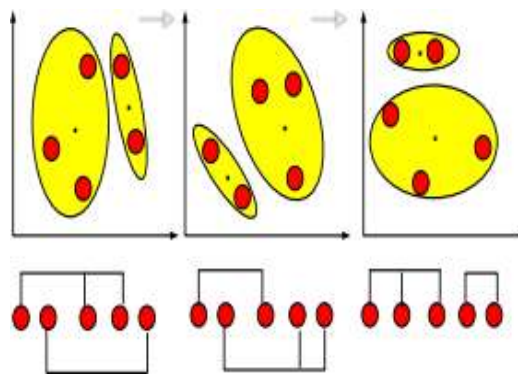
3.2 Pengolahan Data

Setelah itu, dilakukan pengecekan ulang untuk melihat hasil penanganan nilai yang hilang. Namun, fitur “Tanggal” akan dihilangkan untuk melihat korelasi antara fitur-fitur lain, yaitu Beras, Daging Ayam, Daging Sapi, Telur Ayam, Bawang Merah, Bawang Putih, Cabai Merah, Cabai Cayenne, Minyak Goreng, dan Gula Pasir, kemudian diubah formatnya agar dapat dilakukan pengelompokan. Nilai korelasi antar komoditas dapat di lihat seperti pada Tabel 2.

Tabel 2. Tabel Korelasi antar Komoditas (*sheet_name* = “Maluku”)

	Rice	Chicken Meat	Beef	Chicken Eggs	Shallots	Garlic	Red Chillies	Cayenne Pepper	Cooking Oil	Granulated Sugar
Rice	1	0.535392	0.556169	0.529433	0.044769	0.084742	0.279846	0.020478	0.297801	0.00855
Chicken Meat	0.535392	1	0.783095	0.740624	0.024208	-0.098476	0.443287	0.416433	0.735797	-0.070005
Beef	0.556169	0.783095	1	0.804268	0.021021	-0.12801	0.400997	0.451264	0.873997	-0.166296
Chicken Eggs	0.529433	0.740624	0.804268	1	0.203435	-0.052158	0.326175	0.306905	0.659793	0.060064
Shallots	0.044769	0.024208	0.021021	0.203435	1	0.409911	0.068549	-0.131984	0.139119	0.585251
Garlic	0.084742	-0.098476	-0.12801	-0.052158	0.409911	1	-0.197592	-0.27199	-0.056636	0.383203
Red Chillies	0.279846	0.443287	0.400997	0.326175	0.068549	-0.197592	1	0.592875	0.459628	-0.102969
Cayenne Pepper	0.020478	0.416433	0.451264	0.306905	-0.131984	-0.27199	0.592875	1	0.45821	-0.446608
Cooking Oil	0.297801	0.735797	0.873997	0.659793	0.139119	-0.056636	0.459628	0.45821	1	-0.033592
Granulated Sugar	0.00855	-0.070005	-0.166296	0.060064	0.585251	0.383203	-0.102969	-0.446608	-0.033592	1

3.3 K-Means



Gambar 1. Tahapan Metode K-Means

Konsep dalam algoritma ini bertujuan untuk memaksimalkan dan meminimalkan kesamaan antar kelas. Misalnya, terdapat kumpulan objek, proses pertama dapat dikelompokkan menjadi beberapa kumpulan kelas dan kemudian menjadi kumpulan reguler sehingga dapat dihasilkan berdasarkan klasifikasi tertentu dalam kelompoknya didasarkan pada tingkat *similarity*-nya. Penentuan kesamaan antara dua objek dalam klaster dilakukan dengan menghitung jarak antar objek. Dengan begitu, tujuan pengelompokan data dalam meminimalkan fungsi objektif yang didefinisikan dalam proses *clustering*, umumnya akan selalu meminimalkan variasi dalam suatu kluster serta memaksimalkan variasi antara kluster. Alur penelitian dapat di lihat seperti pada Gambar 2.



Gambar 2. Alur Penelitian

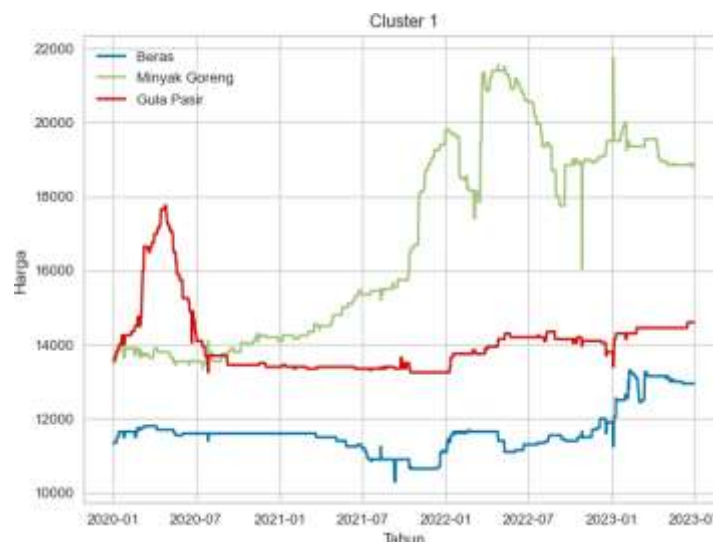
4. HASIL DAN PEMBAHASAN

Masih dengan *sheet_name* yang sama, proses klasterisasi dilakukan dengan memberikan nilai untuk *n_clusters*=5 lalu *cluster_labels* di isi dengan data yang telah ditranspose sebelumnya. Panjang *cluster_labels* yang diketahui adalah 10. Kemudian dibuat *dataframe* baru dengan 2 kolom seperti yang dapat di lihat pada Tabel 3 dengan indeks mengikuti *cluster_labels*.

Tabel 3. Tabel Jumlah Komoditas dan Klasternya

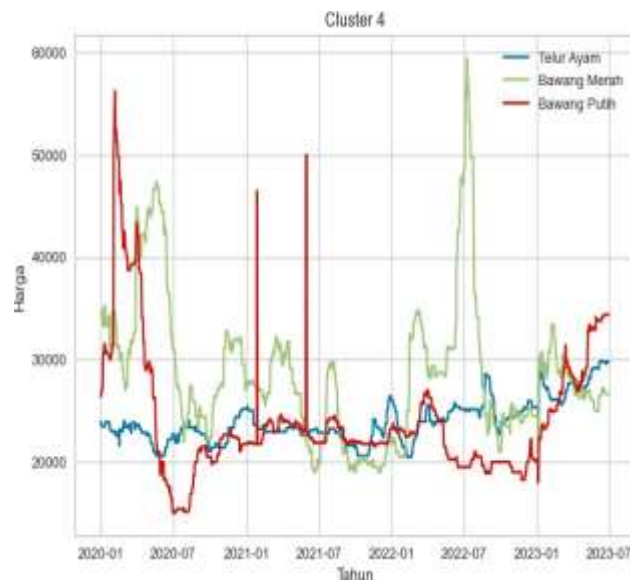
Commodities	Cluster	
0	Beras	0
1	Daging Ayam	4
2	Daging Sapi	1
3	Telur Ayam	3
4	Bawang Merah	3
5	Bawang Putih	3
6	Cabai Merah	2
7	Cabai Rawit	2
8	Minyak Goreng	0
9	Gula Pasir	0

Pergerakan harga yang ditampilkan pada Gambar 3 untuk *cluster* 1 (antara Beras, Minyak Goreng, dan Gula Pasir) menunjukkan lonjakan harga pangan pada tahun 2022 karena kemungkinan oleh gangguan distribusi dan pasokan pada pasar domestik dengan dampak terbesar pada komoditas yang bahan bakunya bergantung pada minyak goreng.



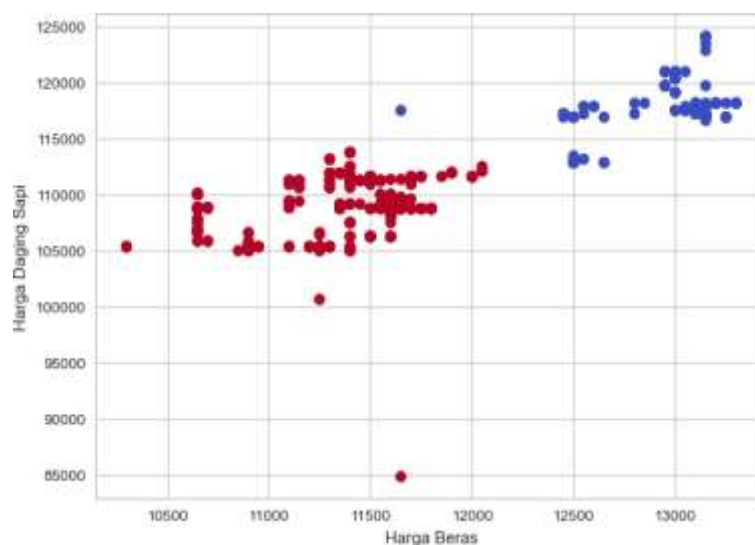
Gambar 3. Tren Harga Komoditas *Cluster* 1

Melompat ke *cluster* 4 yang ditunjukkan melalui Gambar 4 untuk melihat ketiga komoditas hortikultura yaitu bawang merah dan bawang putih ternyata sangat sensitif terhadap gangguan yang sama pada *cluster* 1 (cuaca, impor, distribusi), sedangkan komoditas peternakan seperti telur ayam lebih dipengaruhi oleh biaya input produksi. Lonjakan terbesar di 2022 kembali menegaskan efek gangguan rantai pasok dan inflasi pangan.



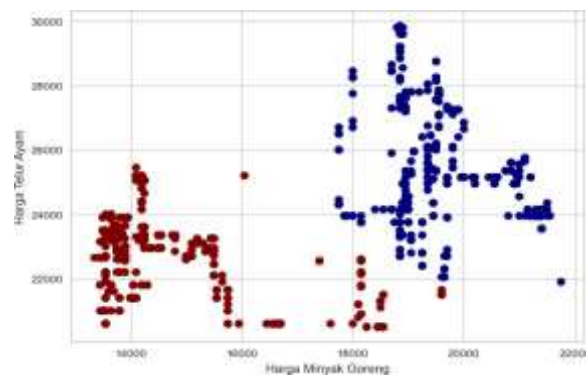
Gambar 4. Tren Harga Komoditas Cluster 4

Nilai keterhubungan antara beras dan daging sapi bersifat positif yang dihasilkan dari tabel korelasi yang sebelumnya sudah ditunjukkan dan dapat di lihat perbandingannya pada Gambar 5 yang menyimpulkan bahwa kenaikan harga beras dari kisaran 10.000–12.000 rupiah/kg berkaitan dengan harga daging sapi yang berada di rentang 85.000–115.000 rupiah/kg (klaster merah), sedangkan pada wilayah dengan harga beras lebih tinggi, sekitar 12.500–13.500 rupiah/kg, harga daging sapi juga meningkat ke kisaran 115.000–122.000 rupiah/kg (klaster biru). Pola ini menegaskan bahwa pergerakan harga kedua komoditas saling terkait dan dipengaruhi oleh faktor biaya logistik, distribusi, serta daya beli wilayah. Dengan demikian, semakin tinggi harga beras di suatu daerah, semakin tinggi pula harga daging sapi yang tercatat, menandakan keterkaitan struktural antara pasar pangan pokok.



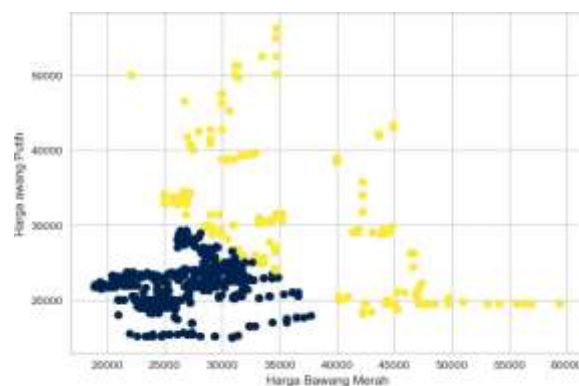
Gambar 5. Perbandingan Harga Beras dan Daging Sapi

Terdapat korelasi positif yang cukup kuat antara minyak goreng dan telur ayam seperti pada Gambar 6 walaupun berbeda klaster namun dapat disimpulkan dengan adanya kenaikan harga minyak goreng dari kisaran 13.000–16.000 rupiah/kg diikuti juga oleh peningkatan harga telur ayam dari sekitar 21.000–24.000 rupiah/kg (klaster merah), sementara pada saat harga minyak goreng meningkat ke kisaran 18.000–21.000 rupiah/kg, harga telur ayam juga naik ke rentang 25.000–29.000 rupiah/kg (klaster biru).



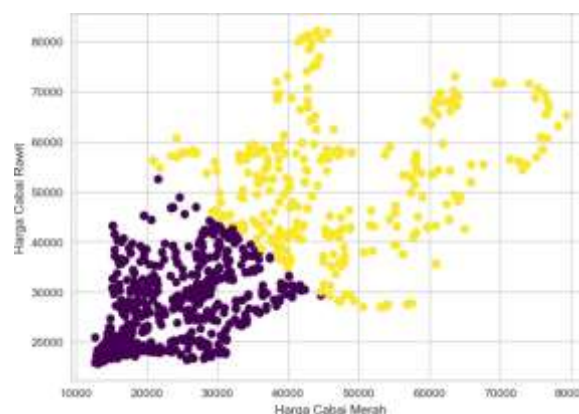
Gambar 6. Perbandingan Harga Minyak Goreng dan Telur Ayam

Sedangkan, korelasi antara harga bawang merah dan harga bawang putih yang ditunjukkan pada Gambar 7 menampilkan korelasi yang ternyata cukup lemah dan tidak linear. Pada kluster harga rendah (biru tua), harga bawang merah berada di kisaran 20.000–35.000 rupiah/kg dan harga bawang putih relatif stabil pada 15.000–25.000 rupiah/kg. Namun pada kluster harga tinggi (kuning), harga bawang merah mengalami lonjakan hingga 60.000 rupiah/kg dan bawang putih mencapai 50.000 rupiah/kg, hal ini dapat terjadi karena kemungkinan dominasi impor yang mengakibatkan kenaikan harga secara ekstrem.



Gambar 7. Perbandingan Harga Bawang Merah dan Bawang

Selain dari bahan pokok yang telah ditunjukkan pada gambar-gambar sebelumnya, terdapat perbandingan harga antar cabai merah dan cabai rawit yang berada dalam satu kluster yang sama. Harga cabai merah dan cabai rawit menunjukkan korelasi positif yang paling kuat, di mana kenaikan harga salah satunya akan diikuti oleh peningkatan harga komoditas lainnya. Kedua jenis cabai cenderung bergerak searah dengan satu kelompok pada rentang 20.000–30.000 rupiah/kg dan kelompok lain pada 40.000–70.000 rupiah/kg



Gambar 8. Perbandingan Harga Cabai Merah dan Cabai Rawit

5. KESIMPULAN

Perbandingan yang ditunjukkan memberikan pemahaman bahwa terdapat perilaku yang tidak jauh berbeda antar komoditas yang diukur. Komoditas dengan nilai korelasi lemah yang ditandai oleh nilai koefisien yang cenderung mendekati nol menunjukkan bahwa perubahan harga pada satu komoditas memiliki dampak kecil terhadap komoditas lainnya. Hal ini dapat dibuktikan dengan tabel korelasi yang telah ditunjukkan dengan contoh komoditas seperti beras dan bawang merah, bawang putih dan gula, atau telur ayam dan bawang putih tidak berada dalam satu klaster yang sama, sehingga dapat diambil kesimpulan bahwa perubahan harga antar komoditas memiliki pengaruh kecil terhadap komoditas lainnya.

Namun, nilai koefisien korelasi antar komoditas yang berkisar antara 0.5 dan 0.6 menggambarkan adanya hubungan korelasi positif. Dengan begitu, perubahan harga pada suatu komoditas akan memiliki pengaruh signifikan pada komoditas lain secara langsung. Contohnya, terdapat dalam satu klaster dengan nilai koefisien positif yang paling kuat yaitu cabai merah dan cabai rawit. Penelitian ini masih memerlukan perbaikan dan pengembangan dengan metode yang lebih lanjut. Misalnya, pada data yang digunakan dapat ditambahkan agar hasil terlihat variatif dan menerapkan teknik penanganan untuk *outlier*, sehingga mengurangi bias pada data. Selain itu, untuk penelitian selanjutnya disarankan untuk menggunakan metode *silhouette score* sebagai evaluasi untuk hasil yang sudah diperoleh.

DAFTAR PUSAKA

- [1] D. Ariestiyanti dan V. Adrison, "Revitalisasi Pasar Dan Stabilisasi Harga Komoditas Pangan," *Buletin Ilmiah Litbang Perdagangan (BILP)*, vol. 2, no. 14, pp. 261-282, Dec 2020.
- [2] B. Marpaung, H. Siregar dan L. Anggraeni, "Analysis of El Niño Impact and the Price of Food Commodities on Inflation," *Jurnal Ekonomi Indonesia*, vol. 1, no. 8, pp. 21-35, Aug 2018.
- [3] M. H. Siregar, "Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan)," *Jurnal Teknologi dan Open Source (JTOS)*, vol. 2, no. 1, pp. 83-91, Dec 2018.
- [4] Y. Prayoga, H. S. Tambunan dan I. Parlina, "Penerapan Clustering Pada Laju Inflasi Kota Di Indonesia Dengan Algoritma K-Means," *BRAHMANA: Jurnal Penerapan Kecerdasan Buatan*, vol. 1, no. 1, pp. 24-30, 2019.
- [5] J. Nasir, "Penerapan Data Mining Clustering Dalam Mengelompokkan Buku Dengan Metode K-Means," *Jurnal SIMETRIS*, vol. 2, no. 11, 2020.
- [6] T. Siburian, M. Safii dan I. Parlina, "Penerapan Algoritma K-Means Clustering untuk Pengelompokan Harga Eceran Beras di Pasar Tradisional Berdasarkan Wilayah Kota," dalam *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 2019.
- [7] R. Muliono dan Z. Sembiring, "Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen," *Journal Computer Engineering System and Science*, vol. 2, no. 4, Juli 2019.
- [8] A. Praja, C. Lubis dan D. E. Herwindiati, "Deteksi Penyakit Diabetes Dengan Metode Fuzzy C-Means Clustering Dan K-Means Clustering," *Computatio : Journal of Computer Science and Information Systems*, vol. 1, no. 1, Apr 2017.
- [9] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Islam Indragiri)," *SISTEMASI*, vol. 3, no. 7, p. 238, 2018.
- [10] "Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 2, no. 1, p. 151, 202.