

Klasterisasi Inflasi Makanan Dari 90 Kota Di Indonesia Tahun 2020-2023 Menggunakan Algoritma K-Means

Food Inflation Clustering from 90 Cities in Indonesia

2020-2023 Using the K-Means Algorithm

Slamet

Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa
slamet0133@gmail.com

Abstract

This study focuses on clustering food inflation from 90 cities in Indonesia using the K-Means clustering method because it only has inflation values. The food inflation data used is monthly data from January 2020 to November 2023. According to the results of processing the inflation data, it produces 5 clusters with details of cluster 0 indicating that the cluster is in the very low category, cluster 1 shows data in the low category, cluster 2 is the cluster in the very low category. moderate, clusters 3 and 4 respectively show the high and very high cluster categories. Cluster 0 has the highest frequency of data appearance with 1403 data.

Keywords: *Inflation, Clustering, K-Means*

Abstrak

Studi ini berfokus pada klasterisasi inflasi makanan dari 90 kota di Indonesia dengan menggunakan metode K-Means clustering karena hanya memiliki nilai inflasi saja. Data inflasi makanan yang digunakan merupakan data bulanan dari bulan Januari 2020 hingga November 2023. Menurut hasil pengolahan data inflasi menghasilkan 5 cluster dengan rincian klaster 0 menunjukkan bahwa klaster dengan kategori sangat rendah, klaster 1 menunjukkan data dengan kategori rendah, klaster 2 merupakan klaster dengan kategori sedang, klaster 3 dan 4 secara berturut-turut menunjukkan kategori klaster tinggi dan sangat tinggi. Klaster 0 memiliki rekuensi kemunculan data terbanyak dengan jumlah 1403 data.

Kata kunci: Inflasi, Klasterisasi, K-Means

Pendahuluan

Salah satu hal yang dapat mempengaruhi perkembangan perekonomian suatu daerah adalah inflasi. Inflasi memiliki dampak positif dan negatif tergantung dari tinggi atau rendahnya inflasi tersebut [1]. Dilihat dari sudut pandang ekonomi, inflasi merupakan suatu fenomena keuangan yang dapat mengakibatkan terjadinya gejala ekonomi. Inflasi sendiri biasanya digunakan untuk mengukur stabilitas perekonomian dalam suatu daerah [2]. Menurut Bank Indonesia, inflasi didefinisikan sebagai peningkatan harga barang dan jasa secara umum dan berkelanjutan dalam periode tertentu .

Naik turunnya harga komoditas makanan terjadi begitu cepat. Untuk menghindari terjadinya ketidakseimbangan antara permintaan dan penawaran, maka diperlukanlah sebuah kebijakan stabilisasi harga. Stabilitas harga inilah yang menjadi tujuan utama dari kebijakan-kebijakan ekonomi dari seluruh negara di dunia [2]. Agar dapat menganalisa tingkat inflasi di setiap daerah, maka diperlukanlah sebuah metode analisa yang tepat sehingga menghasilkan informasi yang akurat. Informasi-informasi tersebut nantinya dapat digunakan untuk mengambil sebuah kebijakan.

Salah satu metode analisa yang dapat digunakan pada penelitian ini adalah metode clustering. Clustering merupakan metode analisis yang bertujuan untuk mengelompokkan data menjadi klaster-klaster berdasarkan

kesamaan karakteristik tertentu [3][4]. Clustering termasuk dalam kategori unsupervised learning karena tidak memiliki label pada setiap kelasnya [5][6][7]. Pada penelitian ini, penulis menggunakan algoritma K-Means Clustering dalam melakukan analisa data.

K-Means merupakan metode pengelompokan secara iteratif dengan cara melakukan partisi untuk mengelompokkan sebagian besar objek. K-Means memiliki keunggulan yaitu mampu melakukan klastering pada data dengan ukuran besar secara cepat [8]. K-Means memiliki kesederhanaan dalam melakukan pengelompokan data kedalam k kelompok berdasarkan centroid masing-masing kelompok [7].

Pada penelitian lain yang serupa, Nuringtyas Andadari dan Adi Nugroho pada tahun 2023 dengan judul “Analisis Laju Inflasi di Kota Semarang Menggunakan Metode K-Means Clustering” [9] menghasilkan 3 kategori klaster yaitu klaster 0 adalah klaster dengan tingkat inflasi rendah, klaster 1 dengan tingkat inflasi sedang, dan klaster 2 dengan tingkat inflasi tinggi. Dalam penelitian tersebut digunakan library pandas sebagai tool untuk menganalisa data. Sedangkan pada penelitian ini akan menggunakan PySpark sebagai tool untuk melakukan analisa dan pengolahan data.

PySpark merupakan API python untuk Apache Spark, sebuah kerangka kerja open source yang memiliki kumpulan library untuk memproses data dalam skala besar. Sama halnya dengan pandas, PySpark juga dapat dipelajari guna membuat analisis data dengan skala besar supaya terukur [10][11].

Belum adanya penelitian yang meneliti tingkat inflasi dari kota-kota di Indonesia, maka dari itu penulis melakukan sebuah penelitian dengan judul “Klasterisasi Inflasi Makanan Dari 90 Kota Di Indonesia Tahun 2020-2023 Menggunakan Algoritma K-Means”. Penelitian ini bertujuan untuk memudahkan pembuat kebijakan dalam membaca tingkat inflasi berdasarkan klaster-klaster yang ada agar dapat buat kebijakan yang sesuai. Dengan jumlah data yang banyak, penggunaan PySpark juga diharapkan mampu melakukan pengolahan data secara maksimal.

Metode Penelitian

Dalam melakukan sebuah penelitian, dibutuhkan juga kerangka pemikiran dari penelitian tersebut agar penelitian lebih terarah dan dapat mencapai tujuan yang diharapkan. Kerangka pemikiran sendiri berisi langkah-langkah yang dilakukan oleh penulis dalam melakukan penelitian. Adapun kerangka pemikiran dari penelitian ini adalah sebagai berikut



Gambar 1. Kerangka Pemikiran

Dataset yang digunakan pada penelitian ini berasal dari data BPS dengan URL <https://www.archive.bps.go.id/indicator/3/1890/1/inflasi-2018-100-menurut-kelompok-dan-sub-kelompok-01-makanan-minuman-dan-tembakau.html>, yang merupakan sumber data publik milik pemerintah. Dataset ini berisi tentang data inflasi setiap bulan dari 90 kota di Indonesia selama 4 tahun. Dataset ini memberikan gambaran realistis mengenai pola inflasi yang terjadi setiap bulan di masing-masing kota. Dengan dataset ini, penulis dapat melakukan analisa terhadap tren inflasi yang terjadi di Indonesia pada sektor makanan.

Dalam analisisnya, penulis dapat menggunakan konsep-konsep dari data mining. Data mining merupakan proses penggalian data yang menghasilkan output berupa pengetahuan [12]. Pada tahun 2005, Gartner Group dalam Larose mengemukakan bahwa data mining merupakan suatu proses menemukan keterkaitan antar variabel menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [13]. Proses pengenalan pola tersebut dilakukan secara otomatis menggunakan komputer dan pola yang ditemukan juga harus bermanfaat [14]. Data mining sendiri memiliki beberapa metode yang dapat digunakan seperti klasifikasi, clustering, regresi, seleksi variabel dan analisa [5].

Dari dataset yang ada, penulis menggunakan metode K-Means clustering untuk mengelompokkan nilai inflasi yang ada kedalam klaster-klaster yang akan terbentuk. K-Means merupakan algoritma yang paling umum digunakan karena relatif cepat dan mudah diimplementasikan. Penerapan dari K-Means memiliki beberapa langkah sebagai berikut [3]:

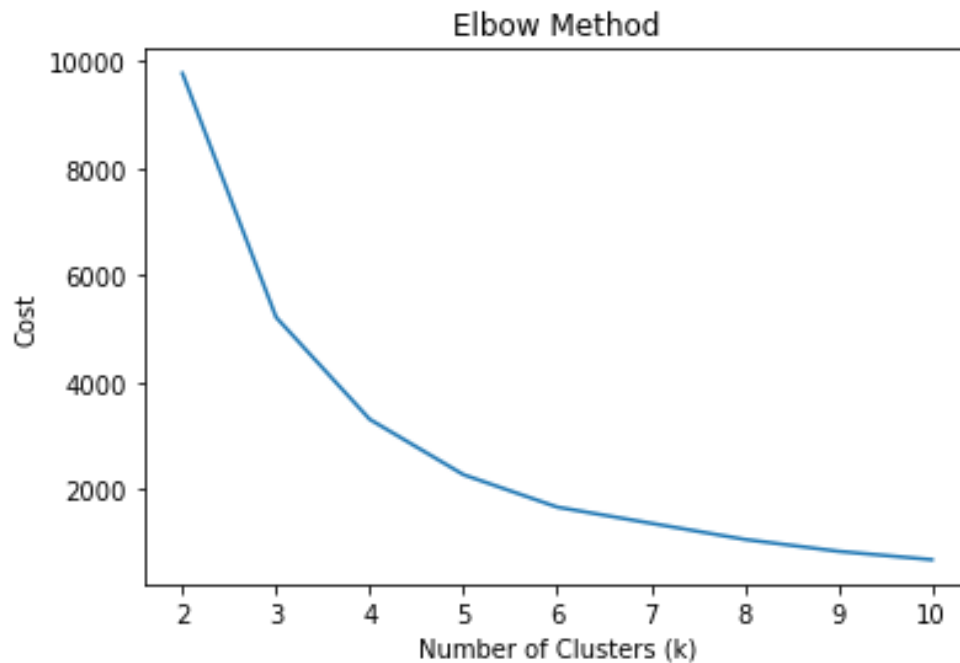
1. Menentukan banyaknya k cluster yang akan dibentuk
2. Inisiasi pusat cluster (centroid) yang ditentukan secara acak
3. Menghitung jarak data dengan centroid. Untuk menghitung semua jarak data dengan pusat cluster dapat digunakan rumus pada teori jarak euclidan dengan rumus:

$$D_{(i,j)} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

4. Menghitung nilai centroid baru dengan cara menghitung rata-rata cluster yang ada
5. Menghitung jarak setiap data dengan centroid baru seperti pada langkah 3
6. Ulangi langkah 3 dan langkah 4 sampai tidak ada perubahan pada anggota cluster.

Hasil dan Pembahasan

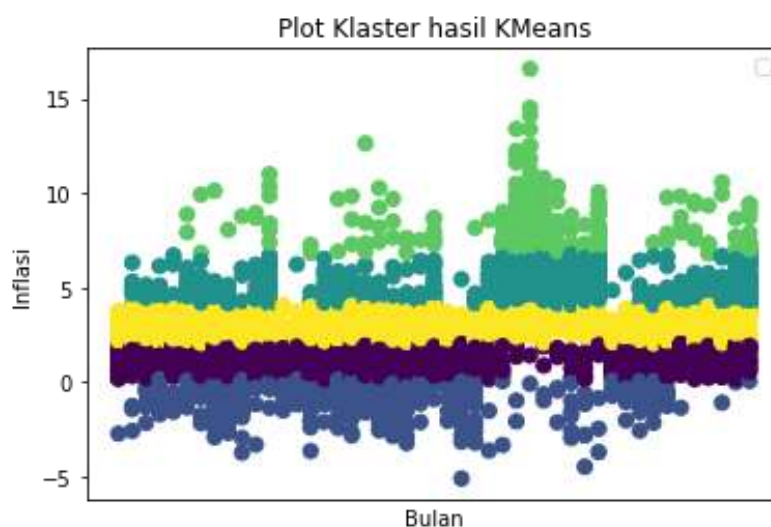
Penelitian ini dimulai dengan menentukan dataset yang akan digunakan untuk klasterisasi inflasi makanan dari 90 kota di Indonesia dengan menggunakan K-Means. Data yang digunakan untuk penelitian ini didapatkan dari website Badan Pusat Statistik, yang dapat diunduh dari URL <https://www.archive.bps.go.id/indicator/3/1917/1/inflasi-tahun-kalender-2018-100-menurut-kelompok-dan-sub-kelompok-01-makanan-minuman-dan-tembakau.html> dengan banyaknya data setelah di transpose sebanyak 4463 record data. Setelah dataset ditemukan, langkah berikutnya adalah preprocessing, atau membersihkan data. Jika data sudah dianggap cukup baik, langkah selanjutnya adalah menentukan pelatihan dan pengujian data menggunakan validasi pembagian, yang membagi 20% data testing dan 80% data training.



Gambar 1. Visualisasi Elbow Method

Tahap selanjutnya setelah preprocessing data adalah mencari nilai k kluster menggunakan elbow method. Gambar 1 menunjukkan bahwa dari visualisasi elbow method menghasilkan nilai k cluster sebanyak 5. Dengan data centroid cluster 0 sampai 4 berturut-turut ditunjukkan pada data dibawah ini

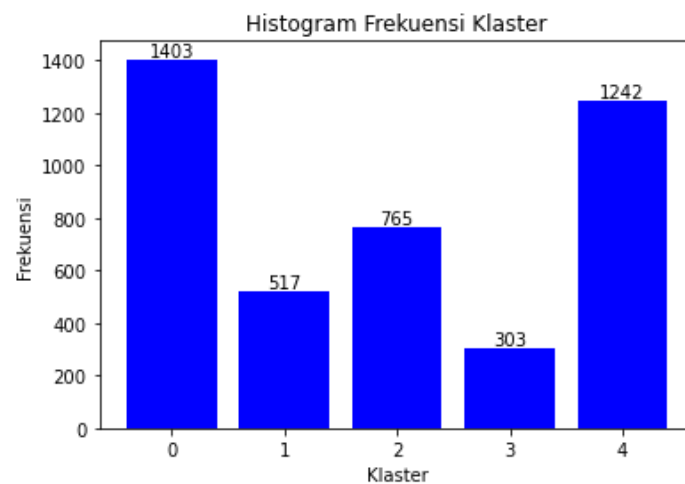
Centroid dari Klaster:
 Klaster 0: [1.13392017]
 Klaster 1: [-0.85524178]
 Klaster 2: [5.1950719]
 Klaster 3: [8.35887789]
 Klaster 4: [2.93700483]



Gambar 2. Visualisasi Hasil Klasterisasi

Gambar 2 merupakan hasil visualisasi dari klassterisasi yang sudah dilakukan. Berdasarkan visualisasi tersebut terdapat 5 klaster dengan warna berbeda. Klaster 0 menunjukkan bahwa klaster dengan kategori sangat rendah, klaster 1 menunjukkan data dengan kategori rendah, klaster 2 merupakan klaster dengan kategori sedang, klaster 3 dan 4 secara berturut-turut menunjukkan kategori klaster tinggi dan sangat tinggi.

Dari klasterisasi yang sudah dilakukan, peneliti membuat histogram untuk menghitung frekuensi setiap klaster. Adapun hasil visualisasi dari histogram tersebut ditunjukkan pada Gambar 3.



Gambar 3. Frekuensi Tiap Klaster

Kesimpulan

Hasil penelitian ini menghasilkan beberapa kesimpulan yaitu klaster yang terbentuk sebanyak 5 klaster dengan rincian klaster 0 menunjukkan bahwa klaster dengan kategori sangat rendah, klaster 1 menunjukkan data dengan kategori rendah, klaster 2 merupakan klaster dengan kategori sedang, klaster 3 dan 4 secara berturut-turut menunjukkan kategori klaster tinggi dan sangat tinggi. Klaster 0 memiliki rekuensi kemunculan data terbanyak dengan jumlah 1403 data. Hasil tersebut menginformasikan kepada kita bahwa inflasi makanan di Indonesia terkategori masih sangat rendah.

Daftar Rujukan

- [1] R. Rahmanta and Y. Maryunianta, "Pengaruh Harga Komoditi Pangan Terhadap Inflasi Di Kota Medan," *J. Agrica*, vol. 13, no. 1, pp. 35–44, 2020, doi: 10.31289/agrica.v13i1.3121.
- [2] E. F. B. Simanungkalit, "Pengaruh Inflasi Terhadap Pertumbuhan Ekonomi Di Indonesia," *J. Manag. Small Medin. Enterp.*, vol. 13, no. 3, pp. 327–340, 2020, doi: 10.35508/jom.v13i3.3311.
- [3] A. A. Maulana, A. W. Rafii, Y. A. Anjelina, and E. Widodo, "Pengelompokan Kecamatan di Kabupaten Bima Berdasarkan Jumlah Produksi dan Luas Panen Bawang Merah Tahun 2021 Menggunakan K-Means Clustering," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 16, no. 1, pp. 442–451, 2023, doi: 10.36456/jstat.vol16.no1.a7032.
- [4] S. Hajar, A. A. Novany, A. P. Windarto, A. Wanto, and E. Irawan, "Penerapan K-Means Clustering pada ekspor minyak kelapa sawit menurut negara tujuan," *Semin. Nas. Teknol. Komput. Sains 2020*, pp. 314–318, 2020.
- [5] J. Hutagalung and F. Sonata, "Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 1187, 2021, doi: 10.30865/mib.v5i3.3113.
- [6] S. N. Br Sembiring, H. Winata, and S. Kusnasari, "Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 1, p. 31, 2022, doi:

10.53513/jursi.v1i1.4784.

- [7] M. R. Muttaqin and M. Defriani, "Algoritma K-Means untuk Pengelompokan Topik Skripsi Mahasiswa," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 121–129, 2020, doi: 10.33096/ilkom.v12i2.542.121-129.
- [8] F. P. Dewi, P. S. Aryni, and Y. Umaidah, "Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 2, pp. 111–121, 2022, doi: 10.14421/jiska.2022.7.2.111-121.
- [9] N. Andadari, "Analisis Laju Inflasi di Kota Semarang Menggunakan Metode K-Means Clustering," vol. 8, no. 4, pp. 1256–1262, 2023, [Online]. Available: [https://repository.uksw.edu/handle/123456789/30611%0Ahttps://repository.uksw.edu/bitstream/123456789/30611/3/T1_67219069_Daftar Pustaka.pdf](https://repository.uksw.edu/handle/123456789/30611%0Ahttps://repository.uksw.edu/bitstream/123456789/30611/3/T1_67219069_Daftar%20Pustaka.pdf).
- [10] P. V. Rao, A. Faheem, A. Begum, and V. Aadi, "Data Behaviour Analysis using Intelligent Big Data Analytics," pp. 426–438.
- [11] H. S. Kudale, M. V Phadnis, P. J. Chittar, K. P. Zarkar, and B. K. Bodhke, "a Review of Data Analysis and Visualization of Olympics Using Pyspark and Dash-Plotly," *Int. Res. J. Mod. Eng. Technol. Sci. www.irjmets.com @International Res. J. Mod. Eng.*, no. 06, pp. 2582–5208, 2093, [Online]. Available: www.irjmets.com.
- [12] Anggada Maulana, "Konsep Dasar Data Mining," *Konsep Data Min.*, vol. 1, pp. 1–16, 2018.
- [13] A. O. P. Dewi, "Big Data di Perpustakaan dengan Memanfaatkan Data Mining," *Anuva J. Kaji. Budaya, Perpustakaan, dan Inf.*, vol. 4, no. 2, pp. 223–230, 2020, doi: 10.14710/anuva.4.2.223-230.
- [14] M. Rizki, D. Devrika, I. H. Umam, and F. S. Lubis, "Aplikasi Data Mining dalam Penentuan Layout Swalayan dengan Menggunakan Metode MBA," *J. Tek. Ind. J. Has. Penelit. dan Karya Ilm. dalam Bid. Tek. Ind.*, vol. 5, no. 2, p. 130, 2020, doi: 10.24014/jti.v5i2.8958.