

# Analisa Clustering Pada Penerima Pupuk Subsidi Menggunakan Algoritma K-Means

**Diterima:**

10 Mei 2023

**Revisi:**

10 Juli 2023

**Terbit:**

1 Agustus 2023

**<sup>1\*</sup>Darina Zalfa, <sup>2</sup>Arie Nugroho, <sup>3</sup>Dwi Harini**

*<sup>1-3</sup>Universitas Nusantara PGRI Kediri*

**Abstrak**—Pupuk Subsidi yang diatur dalam RDKK bisa didapatkan oleh petani melalui retail atau ecer yang khusus ditujukan dari distributor langsung dan diatur kebijakannya oleh pemerintah daerah setempat. Toko Lancar Jaya merupakan usaha yang bergerak di bidang retail atau ecer pupuk subsidi, Menurut informasi produktifitas tani terkendala pembagian pupuk subsidi yang belum merata, salah satu penyebabnya adalah luas lahan setiap kelompok yang berbeda dan belum maksimal dalam proses validasi data kelompok tani yang belum diupdate sehingga berakibat pada pembagian pupuk subsidi belum merata, hal tersebut juga menyebabkan kekurangan ketersediaan pupuk. Penggunaan metode Clustering merupakan solusi untuk pengelompokan data petani yang diimplementasikan dengan metode Algoritma K-means untuk memberikan hasil spesifik dalam penyajian informasi. Hasil dari penelitian ini adalah terbentuknya model pengklasteran data petani dalam penerimaan pupuk subsidi sesuai luas lahan.

**Kata Kunci**— RDKK, Data Mining, Algoritma K-means Clustering

**Abstract**— *Subsidized fertilizers regulated in the RDKK can be obtained by farmers through retail or e-tails specifically aimed at direct distributors and policies regulated by the local regional government. Toko Lancar Jaya is a business engaged in the retail or retail sector of subsidized fertilizers. According to information on farmer productivity, it is constrained by the uneven distribution of subsidized fertilizers. One of the reasons is that the land area of each group is different and has not been maximized in the process of validating farmer group data which has not been updated so that resulting in the uneven distribution of subsidized fertilizers, this also causes a shortage of fertilizer availability. The use of the Clustering method is a solution for grouping farmer data which is implemented with the K-means Algorithm method to provide specific results in presenting information. The result of this study is the formation of a data clustering model for farmers in receiving subsidized fertilizer according to land area*

**Keywords**— RDKK, Data Mining, Algoritma K-means Clustering

This is an open access article under the CC BY-SA License.



---

## **Penulis Korespondensi:**

Darina Zalfa

Sistem Informasi

Universitas Nusantara PGRI Kediri

---

## I. PENDAHULUAN

Pupuk subsidi merupakan salah satu bentuk program pemerintah dalam meningkatkan produktivitas pertanian, meningkatkan produksi pangan serta ketahanan pangan berkelanjutan. Pupuk subsidi di tujukan pada petani yang diatur dalam RDKK (Rencana Definitif Kebutuhan Kelompok) atau sistem kelompok tani, yang biasanya beranggotakan petani-petani sekala kecil dan menengah. Pupuk Subsidi yang diatur dalam RDKK bisa didapatkan oleh petani melalui retail atau ecer yang khusus ditujukan dari distributor langsung dan diatur kebijakannya oleh pemerintah daerah setempat.

Toko Lancar Jaya merupakan usaha yang bergerak di bidang retail atau ecer pupuk subsidi, dengan kode ecer RT0000035424, dari distributor PT.DIMAR ANTIKA NUGROHO (DIARA) yang beralamatkan di Desa Ngepeh Kecamatan Loceret Kabupaten Nganjuk. Sesuai dengan Rencana Definitif Kebutuhan Kelompok pupuk bersubsidi, di desa Ngepeh terbagi menjadi 12 kelompok tani yang mengecer pupuk subsidi di Toko Lancar Jaya.

Berdasarkan informasi yang didapatkan dari pihak pengelola Toko Lancar Jaya, produktifitas padi, jagung, dan kedelai yang dikategorikan menjadi komoditas unggul di Desa Ngepeh terkendala pembagian pupuk subsidi yang belum merata. Salah satu penyebabnya adalah luas lahan setiap kelompok yang berbeda dan belum maksimal dalam proses validasi data kelompok tani yang belum diupdate sehingga berakibat pada pembagian pupuk subsidi belum merata, hal tersebut juga menyebabkan kekurangan ketersediaanpupuk.

Dari permasalahan tersebut peneliti ingin membantu pemilik Toko Lancar Jaya untuk menganalisa data RDKK untuk dilakukan pengelompokan petani di desa Ngepeh kecamatan Loceret Kabupaten Nganjuk, hal ini bertujuan untuk pemerataan penerimaan pupuk sebsidi berdasarkan luas lahan yang dimiliki setiap petani. Dari tujuan tersebut peneliti memilih menggunakan metode *clustering* yang diimplementasikan untuk memberikan hasil spesifik dalam penyajian informasi. Peneliti menggunakan Metode Algoritma *K-means* yang mana akan dibentuk 3 klaster dengan kategori penerimaan pupuk subsidi tertinggi, sedang, rendah, dalam Penggunaan metode *K-means* peneliti menggunakan tiga perhitungan jarak *Euclidean Distance*, *Manhattan Distance*, dan *Chebychev Distance*

## II. METODE

Metode penelitian ini menggunakan metode *Action reserch*. Metode *action reserch* memiliki empat tahapan dalam medel Action reserch, yaitu *Action Plan*, *Action Taking*, *Evaluation*, *Learning*.

Metode pengumpulan data dalam penelitian ini melalui Study literatur yang mana peneliti mengumpulkan berbagai informasi terkait masalah penelitian melalui tinjauan literature, sehingga peneliti benar-benar mengerti isu, konsep, dan variabel yang ada di dalam penelitian, peneliti juga melakukan wawancara dengan Umar Faruq selaku anak dari pemilik Toko, tahap terakhir pengumpulan data yang dilakukan peneliti yaitu Observasi ke Toko Lancar Jaya yang beralamatkan di Desa Ngepeh Kecamatan Loceret Kabupaten Nganjuk Metode Analisa Data pada penelitian ini menyesuaikan dengan CRISP-DM (cross-Industri Standard Proses for Data Mining), langkah-langkah yang dilakukan peneliti berdasarkan CRISP-DM :

#### 1. Business Understanding

Mahami permasalahan yang ada terkait pembagian pupuk subsidi secara merata yang dikelola oleh toko Lancar Jaya. Kemudian memanfaatkan data RDKK pupuk subsidi yang mana data tersebut pernah digukan atau dimanfaatkan untuk analisis. Dari hasil analisa tersebut akan dilakukan pengelompokan data untuk pemerataan pupuk subsidisesuai luas lahan yang dimiliki setiap petani

#### 2. Data Understanding

Pada tahap ini peneliti melakukan pemahaman terhadap data yang dibutuhkan, yang bertujuan untuk diambil data yang relevan dan memiliki keterkaitan dengan tujuan penelitian ini. Data yang digunakan dalam penelitian ini adalah data RDKK Pupuk Bersubsidi Tahun 2022 yang diperoleh dari Toko Lancar Jaya.

#### 3. Data Pre-paration

Peneliti mengolah data dengan tahapan pada KDD (*Knowledge Discovery in Database*) yang terdiri dari pembersihan data, kemudian melakukan *pre-processing* data, selanjutnya pemilihan data, dan terakhir transformation pada data yang sudah melalui tahap pemilihan data. Tools yang digunakan untuk menunjang pengolahan data yaitu microsoft excel.

#### 4. Modeling

Teknik *Modeling* yang dilakukan pada penelitian ini yaitu dengan *Clustering* menggunakan Algoritma *K-means* kemudian dilakukan proses cluster menggunakan tools *Rapidminer* 10.1.

#### 5. Evaluation

Ditahap evaluasi ini akan diketahui apakah tahapan dari modeling dapat menjawab tujuan yang sudah ditetapkan pada fase pertama. Untuk itu akan dilakukan pengelompokan dari hasil pengelompokan tersebut akan terbentuk suatu pola pengelompokan untuk penerimaan pupuk berdasarkan luas lahan.

#### 6. Deployment

Pada tahap deployment ini peneliti akan menyajikan hasil dari penerapan data mining metode

clustering yang menggunakan Algoritma K-means.

Metode Algoritma K-means Clustering pada dasarnya melakukan dua proses, yaitu proses pendeteksian lokasi pusat tiap *cluster* dan proses pencarian anggota dari tiap-tiap cluster. Langkah- langkah dalam perhitungan Algoritma K-means:

1. Tentukan K sebagai jumlah cluster yang ingin dibentuk.
2. Tentukan nilai centroid (titik pusat cluster).
3. Hitung jarak setiap data ke masing-masing centroid.
4. Pengelompokkan nilai sesuai titik terdekat dari nilai iterasi yang didapat.
5. Menentukan centroid baru dengan cara menghitung nilai rata-rata dari data-data yang terletak pada centroid yang sama. Dengan rumus 2.1 pada tahap kedua, selanjutnya proses kembali ke langkah 3 jika posisi centroid baru dengan centroid yang lama tidak sama.

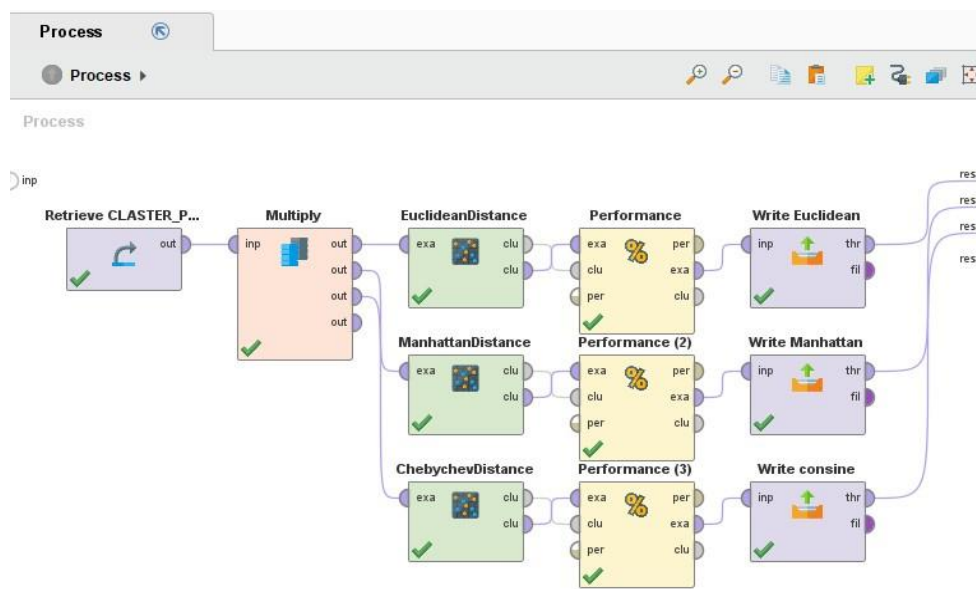
### III. HASIL DAN PEMBAHASAN

Tahap Pertama dari algoritma *k-means* yaitu menentukan pusat *cluster* awal. Menentukan Nilai *Centroid* awal dilakukan secara acak dari data atau objek yang tersedia sebanyak jumlah *cluster k*. *Centroid* awal pada jumlah 3 *cluster*.

Tabel 1. Data Rencana Definitiv Kebutuhan Kelompok

No	Luas Lahan	Urea	NPK	ORGANIK	ZA
1	1.05	279	279	42	2
2	1.8	480	480	72	6
3	2.25	599	599	90	6
4	1.5	599	400	60	4
5	0.6	160	160	24	2
6	1.05	279	279	42	2
7	0.45	119	119	18	0
8	2.25	599	599	90	6
9	2.1	599	560	84	6
10	2.25	599	599	90	6
...	...	...	...	...	...
758	1.05	279	279	42	2

Data yang diperoleh di proses menggunakan tools rapid miner dengan metode K-means menggunakan tiga perhitungan jarak *Euclidean Distance*, *Manhattan Distance*, dan *Chebychev Distance*.



Gambar 1 Proses Data Mining

### 1. Euclidean Distance

Clustering menggunakan algoritma K-Means dengan metode perhitungan jarak Euclidean Distance menghasilkan 3 Cluster dan data yang terkelompok pada masing-masing Cluster, hasil dapat dilihat pada Tabel 2

Tabel 2 Hasil Cluster *Euclidean Distance*

Nominal Value	Absolute Count	Fraction
Cluster 1	392	0.517
Cluster 2	330	0.435
Cluster 3	36	0.047

Pada Tabel 2 dapat dilihat hasil proses perhitungan jarak Euclidean distance yang mana pada cluster 1 terdapat 392 data, pada cluster 2 terdapat 330 data, dan cluster 3 terdapat 36 data.

### 2. Manhattan Distance

Clustering menggunakan algoritma K-Means dengan metode perhitungan jarak Manhattan Distance menghasilkan 3 Cluster dan data yang terkelompok pada masing-masing Cluster, hasil dapat dilihat pada Tabel 3.

Tabel 3 Hasil Cluster *Manhattan Distance*

Nominal Value	Absolute Count	Fraction
Cluster 2	393	0.518
Cluster 1	325	0.429
Cluster 3	40	0.053

Pada Tabel 3 dapat dilihat hasil proses perhitungan jarak *Euclidean distance* yang mana pada cluster 1 terdapat 393 data, pada cluster 0 terdapat 325, dan cluster 2 terdapat 40 data.

### 3. Chebychev Distance

Clustering menggunakan algoritma K-Means dengan metode perhitungan jarak Manhattan Distance menghasilkan 3 Cluster dan data yang terkelompok pada masing-masing Cluster

Tabel 4 Hasil *Cluster Chebychev Distance*

Nominal Value	Absolute Count	Fraction
Cluster 1	394	0.520
Cluster 2	353	0.466
Cluster 3	11	0.015

Pada Tabel 4 dapat dilihat hasil proses perhitungan jarak *Chebychev Distance* yang mana pada cluster 0 terdapat 394 data, pada cluster 1 terdapat 353 data, dan cluster 2 terdapat 11 data.

### 4. Evaluasi

Pada proses evaluasi ini dilihat dari hasil *Davies Bouldin Index* pada proses di *rapid miner*.

Tabel 5. Hasil *Performance Davies Bouldin Index*

Euclidean Distance	Manhattan Distance	Chebychev Distance
<b>0.638</b>	0.644	0.511

Berdasarkan Tabel 5 hasil evaluasi *Davies Bouldin Index* menggunakan 758 data RDKK yang diproses dengan Algoritma K-means menggunakan 3 metode perhitungan jarak yaitu *Euclidean Distance*, *Manhattan Distance*, dan *Chebychev Distance* yang menghasilkan 3 nilai *Davies Bouldin Index* 0.638 untuk metode *Euclidean Distance*, 0.644 untuk metode *Manhattan Distance* dan, 0.511 untuk metode *Chebychev Distance*.

### 5. Analisa Hasil

Metode *Chebychev Distance* memiliki hasil yang cukup baik dalam proses pengklasteran, karena penerapan Metode perhitungan jarak *Chebychev Distance* memiliki nilai *Davies Bouldin Index* 0.511 dari pada metode *Euclidean Distance* yang memiliki nilai *Davies Bouldin Index* 0.638 dan Metode *Manhattan Distance* yang memiliki nilai *Davies Bouldin Index* 0.644. Berdasarkan ketiga hasil tersebut metode

*Chebychev Distance* memiliki hasil yang cukup baik. Evaluasi *Davies Bouldin Index* (DBI) dapat disimpulkan bahwa semakin kecil nilai *Davies Bouldin Index* (DBI) yang diperoleh maka hasil *Cluster* dari metode pengukuran jarak tersebut semakin baik.

#### IV. KESIMPULAN

Berdasarkan hasil perhitungan yang telah dilakukan menggunakan Algoritma *K- Means* dengan 3 metode perhitungan jarak Euclidean Distance, Manhattan Distance, dan Chebychev Distance dan hasil evaluasi *Davies Bouldin Index* menghasilkan nilai DBI yaitu 0.638 pada metode Euclidean Distance, 0.644 pada Manhattan Distance, dan 0.511 pada metode *Chebychev Distance*. Hasil dari evaluasi *Davies bouldin index* tersebut dapat disimpulkan bahwa metode *Chebychev Distance* merupakan metode yang baik digunakan dalam melakukan pengelompokan data petani di desa Ngepeh Kecamatan Loceret Kabupaten Nganjuk berdasarkan luas lahan yang dimiliki setiap petani dengan hasil nilai DBI rendah atau mendekati 0 yaitu 0.511.

#### DAFTAR PUSTAKA

- [1] S. Maysaroh, S. Wulan Purnami, J. Statistika, and J. Arif Rahman Hakim, "ANALISIS KELOMPOK DENGAN METODE KERNEL K-MEANS (STUDI KASUS PENGELOMPOKAN KABUPATEN/KOTA DI INDONESIA BERDASARKAN PENDUDUK DENGAN FAKTOR-FAKTOR RISIKO PENYEBAB PENYAKIT HIPERTENSI)." [Online]. Available: <https://www.researchgate.net/publication/341354022>
- [2] A. I. Warnilah, "ANALISIS ALGORITMA K-MEANS CLUSTERING UNTUK PEMETAAN PRESTASI SISWA STUDI KASUS SMP NEGERI I SUKAHENING," 2016. [Online]. Available: <https://www.researchgate.net/publication/341354022>
- [3] P. Perhitungan Jarak Mohamad Sugeng Pangestu, M. Ayu Fitriani, and M. Sugeng Pangestu, "Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma K- Means Comparison of Euclidean Distance, Manhattan Distance, and Cosine Similarity Calculations on Rice Seed Data Grouping Using the K-Means Algorithm," vol. 19, no. 2, 2022.
- [4] K. Auliasari et al., "Penerapan Algoritma K-Means untuk Segmentasi Konsumen Menggunakan R," 2019.
- [5] S. Kasus, : Hoyweapstore, D. Triyansyah1, and D. Fitriannah2, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing", doi: 10.22441/incomtech.v8i2.4174.
- [6] F. I. Manek, S. Faisal, B. Priyatna, U. Buana, P. Karawang, and I. Abstrak, "Techno Xplore Jurnal Ilmu Komputer dan Teknologi Informasi Penerapan K-Means Clustering untuk Mengelompokkan Pelanggan Berdasarkan Data Penjualan Ayam."
- [7] J. I. Komputer, F. Matematika, I. Pengetahuan, A. Kampus Bina, and W. Pekanbaru, "PENENTUAN LOKASI SUPLIER GAS ELPIJI BERDASARKAN DATA LOKASI PENJUALAN ECERAN GAS ELPIJI MENGGUNAKAN METODE K-MEANS

- (STUDI KASUS KECAMATAN TAMPAN) Beti , Evfi Mahdiyah Mahasiswa Program Studi S1 Sistem Informasi.”
- [8] W. Safira Azis and dan Dedy Atmajaya, “PENGELOMPOKAN MINAT BACA MAHASISWA MENGGUNAKAN METODE K-MEANS,” Jurnal Ilmiah ILKOM, vol. 8, no. 2, 2016.
- [9] K. Auliasari et al., “Penerapan Algoritma K-Means untuk Segmentasi Konsumen Menggunakan R,” 2019.
- [10] M. Benri, H. Metisen, and S. Latipa, “ANALISIS CLUSTERING MENGGUNAKAN METODE K-MEANS DALAM PENGELOMPOKKAN PENJUALAN PRODUK PADA SWALAYAN FADHILA,” 2015.
- [11] R. Kurniawan, S. Defit, and S. Sumijan, “Prediksi Tingkat Kerugian Peternak Akibat Penyakit pada Sapi Menggunakan Algoritma K-Means Clustering,” Jurnal Informasi dan Teknologi, Sep. 2020, doi: 10.37034/jidt.v3i1.87.
- [12] P. Ulil, F. Aulia, and S. Saepudin, “PENERAPAN DATA MINING K-MEANS CLUSTERING UNTUK MENGELOMPOKKAN BERBAGAI JENIS MERK LAPTOP,” 2021.