

## Penerapan Pengelompokan Produktivitas Hasil Pertanian Menggunakan Algoritma K-Means

Putri Trisnawati<sup>1\*</sup>, Ade Irma Purnamasari<sup>2</sup>

<sup>1,2</sup>Program Studi Manajemen Informatika, STMIK IKMI Cirebon

\*putritrisnawati76@gmail.com

### Abstrak

Sejak dahulu, Indonesia selalu kaya dengan hasil dari pertanian seperti padi, kedelai, jagung, kacang tanah, ketela pohon dan ubi jalar. Selain itu, ada juga hasil dari pertanian yang disebut sebagai hasil pertanian tanaman perdagangan yaitu teh, kopi, kelapa, kina, cengkeh, tebu, karet dan yang lainnya. Sektor pertanian pada tahun 2021 tumbuh 1,84% dan berkontribusi terhadap perekonomian nasional sebesar 13,28%. Kemudian pada 2022, sektor pertanian menunjukkan konsistensi dengan pertumbuhan positif 1,37% dan berkontribusi 12,98% terhadap perekonomian nasional. Maka diperlukan pengelompokan produktivitas hasil pertanian dengan menggunakan metode clustering k-means dikelompokkan data hasil jenis panen paling tinggi dan rendah menurut Kecamatan di Bojonegoro, sehingga dapat diketahui jenis hasil pertanian yang paling produktif, dan kurang produktif. Metode yang digunakan dalam penelitian ini adalah analisis cluster K-Means dengan menentukan terlebih dahulu jumlah kelompok yang akan dibentuk. Dalam penelitian ini data yang digunakan adalah data sekunder tentang hasil pertanian yang berasal dari Satu Data Bojonegoro. Algoritma k-means dalam hal ini akan digunakan untuk pengelompokan hasil pertanian dari tahun 2016– 2022 di Kecamatan Bojonegoro berdasarkan hasil tanaman pangan. Adapun tanaman pangan yang dimaksud adalah padi, bawang merah, kedelai, cabai besar, jagung, dan sebagainya. Dari hasil pengelompokan, hasil pertanian berdasarkan tahun hasil produksi, nantinya akan diketahui jenis panen terbaik dan kecamatan apa yang akan menghasilkan tanaman pangan terproduktif sehingga distribusi tanaman pangan di Kecamatan Bojonegoro dapat dikontrol. Produktifitas hasil pertanian dapat digunakan sebagai strategi untuk meningkatkan hasil-hasil pertanian.

**Kata kunci :** Data Mining, K-means, Produktivitas Hasil Pertanian

### Abstract

Since ancient times, Indonesia has always been rich in agricultural products such as rice, soybeans, corn, peanuts, cassava, and sweet potatoes. In addition, there are also products from agriculture that are referred to as trade crop agricultural products, namely tea, coffee, coconut, quinine, cloves, sugar cane, rubber, and others. The agricultural sector in 2021 will grow by 1.84% and contribute 13.28% to the national economy. Then in 2022, the agricultural sector will show consistency with a positive growth of 1.37% and contribute 12.98% to the national economy. Then it is necessary to group the productivity of agricultural products using the k-means clustering method to group data on the highest and lowest yield types according to the District in Bojonegoro so that the types of agricultural products that are most productive and less productive can be identified. The method used in this study is K-Means cluster analysis by first determining the number of groups to be formed. In this study, the data used is secondary data on agricultural products originating from One Bojonegoro Data. The food crops in question are rice, shallots, soybeans, large chilies, corn, and so on. From the results of grouping agricultural products based on the year of production, the best types of crops will be known, and which districts will produce the most productive food crops so that the distribution of food crops in Bojonegoro District can be controlled. Productivity grouping of agricultural products can be used as a strategy to increase agricultural yields.

**Keywords :** *Data Mining, K-means, Productivity of Agricultural Products.*

## 1. Pendahuluan

Indonesia menjadi salah satu negara agraris dengan mayoritas penduduknya yang bekerja di bidang pertanian. Sektor pertanian ini sangat penting dalam membantu perekonomian nasional di Indonesia, hal ini terlihat dari mayoritas penduduknya yang tinggal dan bekerja dibidang ini. Padi menjadi bahan makanan pokok sehari-hari yang dibutuhkan setiap keluarga sebelum diolah menjadi nasi. Sedangkan bahan pangan lainnya seperti singkong, jagung, padi, kedelai, dan kacang tanah merupakan lauk pelengkap bagi setiap keluarga [1]

Sektor pertanian sendiri merupakan salah satu kekuatan pembangunan ekonomi di Indonesia. Hal ini disebabkan terbukanya penyerapan tenaga kerja di sektor pertanian dan tingginya kontribusi devisa yang dihasilkan. Tanaman pangan bisa menjadi komoditas yang dapat diandalkan sebagai sentra bisnis yang sangat menguntungkan. Terlebih produk-produk dari tanaman pangan sangat ramai peminatnya, baik dipasar dalam negeri ataupun luar negeri.

Selain itu, harga jual yang tinggi juga membuat tanaman perkebunan menjadi salah satu penyumbang devisa negara yang tidak sedikit [2] Salah satu wilayah produktivitas hasil pertanian adalah di Kecamatan Bojonegoro. Kabupaten Bojonegoro yang saat ini lebih dikenal sebagai daerah penghasil migas (minyak dan gas bumi) terbesar di Indonesia, sebenarnya memiliki

potensi sektor pertanian yang tak kalah besarnya dengan setor migas. Apalagi sumber daya migas yang bersifat tidak bisa diperbarui dan pasti akan habis, hal ini tentu saja menjadikan sektor pertanian lebih berpeluang dan menjanjikan keberlanjutan perekonomian daerah. Pemerintah Kabupaten Bojonegoro bertekad menjadikan Bojonegoro sebagai lumbung pangan nasional, sebagai daerah penghasil, pendistribusian, pengolahan, dan perdagangan pangan dalam arti luas dalam rangka peningkatan kesejahteraan masyarakat, peningkatan usaha produktif pengolahan hasil pertanian, serta penyediaan stok pangan secara berkelanjutan melalui program program yang mendukung tekad tersebut [3]

Pendekatan yang umum digunakan dalam analisis klaster adalah dengan menggunakan algoritma K-means. Keunggulan menggunakan Algoritma K-means adalah dapat mengelompokan objek besar dan outlier dengan sangat cepat, sehingga mempercepat proses pengelompokan. Algoritma K-means dalam penelitian ini akan digunakan untuk pengelompokan kecamatan di Bojonegoro berdasarkan hasil tanaman pangan. Adapun tanaman pangan yang dimaksud adalah padi, jagung, ubi kayu, kacang tanag, cabai merah besar, dan sebagainya.

Berdasarkan penelitian sebelumnya dari [4] produktivitas merupakan kemampuan lahan untuk

menghasilkan produksi tanaman tertentu. Lahan yang produktif ialah lahan yang dapat menghasilkan produksi yang baik dan menguntungkan bagi petani. Jika hasil pertanian tidak menguntungkan bagi para petani berarti lahan tersebut tidak produktif kurangnya kualitas tanah yang baik dan perlu pengolahan, perawatan yang lebih maksimal lagi. Kurangnya informasi mengenai kualitas lahan dan perkembangan tentang subur tidaknya tanah tersebut bisa menjadi pengaruh bagi hasil produksi pertanian tersebut yang nantinya akan merugikan petani.

Dalam penelitian ini data yang digunakan adalah data sekunder tentang hasil pertanian yang berasal dari Satu Data Bojonegoro, yang bertujuan untuk mengelompokan produktivitas hasil panen terbaik menurut jenis tanaman di Kecamatan Bojonegoro, sehingga dapat diketahui jenis tanaman hasil pertanian yang paling produktif, dan kurang produktif. Algoritma K-means dalam penelitian ini akan digunakan untuk pengelompokan kecamatan di Bojonegoro berdasarkan hasil tanaman pangan. Adapun tanaman pangan yang dimaksud adalah padi, jagung, ubi kayu, kacang tanag, cabai merah besar, dan sebagainya. Dari hasil pengelompokan kecamatan berdasarkan hasil produksi tanaman pangan, nantinya akan diketahui jenis tanaman pangan terproduktif sehingga distribusi tanaman pangan di Kecamatan Bojonegoro dapat dikontrol

## **2. Tinjauan Pustaka**

### **2.1. Penelitian Terkait**

Adapun penelitian sebelumnya sebagai pendukung penelitian kali ini, antara lain:

1. Budi Prasajo dan Emy Haryanti dalam judul “Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest”. Penelitian ini menggunakan metode Random Forest untuk memprediksi kelayakan dan tidak layaknya seorang nasabah dalam pemberian kredit pinjaman. Hasil akurasi dari algoritma tersebut adalah sebesar 83% dan termasuk ke dalam very good model. Hasil akurasi didapat dari pembagian dataset menjadi 80% sebagai data latih dan 20% sebagai data uji. Penelitian dilakukan untuk menerapkan metode klasifikasi dengan menggunakan algoritma random forest agar dapat menganalisa hasil terbaik pada setiap kreditur[8].

2. Aditya Yudha Perdana dan kawan-kawan dalam penelitiannya yang berjudul “Prediksi Stunting Pada Balita Dengan Algoritma Random Forest”. Tujuan dari penelitian ini adalah untuk memprediksi keadaan stunting pada balita di Kecamatan Pitu Kabupaten Ngawi serta untuk mengembangkan sebuah sistem perangkat lunak berbasis website dan android yang dapat melakukan pengukuran keadaan stunting balita. Hasil akurasi yang didapat dari penelitian ini adalah sebesar 97,89%[9].

3. Penerapan metode datamining dengan model Random Forest dan C4.5 digunakan pada penelitian ini untuk menghasilkan nilai akurasi yang dapat digunakan untuk memprediksi prestasi akademik mahasiswa. Pada penelitian ini dilakukan pengujian menggunakan cross validation dan confusion matrix sehingga menghasilkan nilai akurasi, precision dan recall. Hasil akurasi untuk penelitian ini didapat 92,4% nilai tertinggi dari random forest dibandingkan dengan decision tree C4.5. Penggalan data dilakukan pada penelitian untuk mendapatkan informasi yang baru dan bermanfaat untuk kepentingan perguruan tinggi tersebut dalam memprediksi prestasi akademik mahasiswa, sehingga dapat menyelesaikan masalah-masalah yang timbul dan berdampak pada hasil belajar mahasiswa[10]

## 2.2. Landasan Teori

### 1. Data Mining

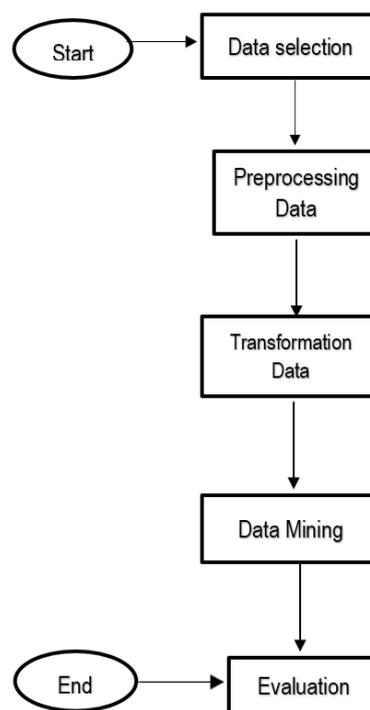
Data mining ialah istilah yang telah digunakan untuk proses otomatis yang menggunakan matematika, kecerdasan buatan, statistik, dan machine learning untuk mengidentifikasi dan mengekstrasi informasi pengetahuan yang tersembunyi dalam database [5] Data mining digunakan untuk mengekstrak informasi berharga dari kumpulan data kemudian menyajikannya dalam format yang mudah dipahami oleh semua orang [6]. Data mining adalah proses yang

bertujuan untuk menemukan nilai tambah berupa informasi yang sebelumnya tidak diketahui dari basis data. Informasi yang didapatkan dengan cara mengekstraksi dan mengidentifikasi pola yang penting atau menarik dari data yang terdapat dalam database [7]

### 2. Knowledge Discovery in Database

KDD merupakan metode yang dapat kita pakai untuk menghasilkan pengetahuan dari database yang sudah ada [8]. Pengetahuan yang diperoleh dapat digunakan untuk basis pengetahuan (knowledge base) yang dipakai dalam keperluan mengambil keputusan [9].

Pada penelitian ini tahapan KDD meliputi : Data Selection, Preprocessing Data, Transformation Data, Data Mining, dan Evaluation, terlihat pada flowchart berikut [10].



Gambar 1. Tahap KDD

### 3. Clustering

Clustering adalah proses pengelompokan objek menurut informasi yang diperoleh dari data yang menggambarkan hubungan antar objek dengan prinsip untuk memaksimalkan kesamaan antar anggota suatu kelas dan meminimumkan kesamaan antar objek kelas/cluster [11]. Tujuan klastering adalah untuk mengidentifikasi sekelompok data dari populasi data untuk menghasilkan atribut dari data itu sendiri [12]

### 4. Algoritma K-Means

Algoritma K-Means merupakan salah satu algoritma klastering yang paling sederhana dan mudah dipahami. Algoritma K-Means yaitu algoritma pengelompokan berdasarkan nilai titik centroid. Titik centroid dari suatu klaster ialah titik pusat klaster [12]. Algoritma K-Means yaitu teknik Clustering berbasis jarak yang membagi data menjadi beberapa cluster dan algoritma ini hanya bekerja pada atribut numerik atau angka [13]

### 5. RapidMiner

RapidMiner yaitu software yang dapat digunakan oleh siapa saja dan bersifat terbuka (open source) atau umum. RapidMiner ini menjadi sebuah solusi untuk menganalisa terhadap data processing atau database [14]. RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi [15]

### 6. Produktivitas pertanian

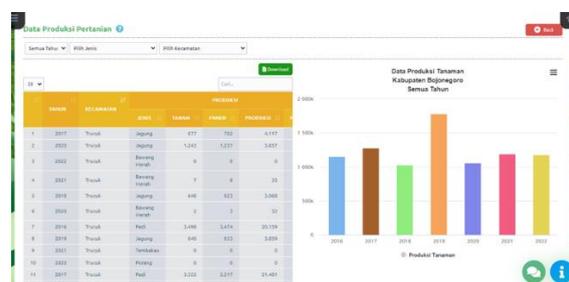
Produktivitas Pertanian yaitu kemampuan sekumpulan sumber daya ekonomi untuk

menghasilkan produk hasil pertanian yang diinginkan. Sumber ekonomi atau sering disebut faktor produksi meliputi tanah, modal, teknologi, tenaga kerja dan bahan baku [4]

## 3. Metode Penelitian

### 3.1. Teknik Pengumpulan Data

Data yang diambil adalah data sekunder data yang diperoleh dari sumber buku-buku dan jurnal ilmiah. Data ini di peroleh dari Dinas Ketahanan Pangan dan Pertanian yang diambil dari web resmi Kecamatan Bojonegoro. Data yang berisi hasil produktivitas pertanian yang di peroleh dari tahun 2016-2022 berdasarkan kecamatan yang ada di Bojonegoro. Data ini berjumlah 952 data dari 6 tahun terakhir. Atribut yang ada didalamnya yaitu tahun hasil produksi, nama kecamatan, jenis tanam, hasil panen, hasil produksi, dan provitas. Berikut adalah link website yang berisi data hasil produktivitas pertanian. <https://data.bojonegorokab.go.id/dinas-ketahanan-pangan.html@detail=produksi-pertanian>. Teknik pengumpulan data yaitu studi dokumentasi yang berisi data hasil produktivitas pertanian dari sumber website resmi Dinas Pertanian Bojonegoro.



Gambar 1 Data Produktivitas Hasil Pertanian

### 3.2. Lokasi Penelitian

Penelitian ini berlokasi di kecamatan Bojonegoro. Kabupaten Bojonegoro yang saat ini dikenal sebagai daerah penghasil migas (minyak dan gas bumi) terbesar di Indonesia.

## 4. Hasil dan Pembahasan

Berikut ini merupakan hasil dari pengelompokan produktivitas hasil pertanian berdasarkan jenis tanam di Kecamatan Bojonegoro dengan menggunakan algoritma K-Means yang diproses dari tools RapidMiner serta tahapan KDD nya adalah sebagai berikut.

### 1. Data Selection

Proses penyelesaian pengelompokan produktivitas hasil pertanian yang pertama dilakukan yaitu pemilihan data. Data yang digunakan adalah data produktivitas hasil pertanian berdasarkan kecamatan yang diperoleh dari tahun 2016-2022 yang bersumber dari situs resmi Dinas Pertanian Kecamatan Bojonegoro, dengan keseluruhan jumlah data yaitu 952 dan jumlah atributnya adalah 7. Atribut yang digunakan dalam pengelompokan produktivitas hasil pertanian itu adalah 3 atribut yaitu nama kecamatan, jenis tanaman dan hasil produksi yang akan diproses melalui tools RapidMiner.

Tabel 1 Hasil Data Selection

Kecamatan	Jenis	Produksi
Balen	Padi	19736,19

Kecamatan	Jenis	Produksi
Balen	Jagung	44
Bojonegoro	Padi	6218,25
Bojonegoro	Jagung	25,69
Bojonegoro	Kedelai	5,03
Bubulan	Padi	4512,66
--	--	--
Balen	Padi	52914,98
Balen	Kedelai	3599,16
Balen	Tembakau	26,33
Balen	Bawang Merah	99
Balen	Jagung	61,08
Balen	Cabai Besar	22,2
Bojonegoro	Padi	11019,92
Bojonegoro	Bawang Merah	5,5

### 2. Preprocessing Data

Tahap preprocessing dilakukan pembersihan data dimana tidak ada missing value yang dapat mendukung proses data mining.

Name	Type	Missing	Statistics
Id Kecamatan	Polynomial	0	Least Bojonegoro (9)
Cluster cluster	Nominal	0	Least cluster_4 (1)
Jenis	Integer	0	Min 1
Produksi	Integer	0	Min 1

Gambar 2 Pengecekan Missing Value

### 3. Transformation Data

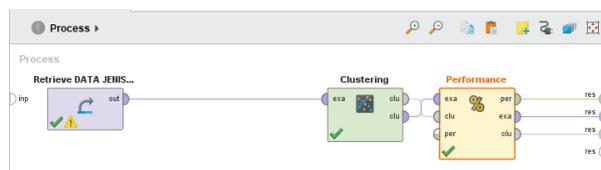
Tahap berikutnya adalah tahap transformasi data, dari 3 atribut yang digunakan, dikelompokan berdasarkan jenis tanamannya, dan setelah diolah data yang awalnya berjumlah 952 kemudian menjadi 512. Disini data yang berjumlah atau hasilnya 0 dihapus dan diminimalisir agar bisa digunakan sesuai

kebutuhan. Data ini berdasarkan jenisnya, yang mana diinisialisasi terlebih dahulu untuk menyesuaikan pada tools RapidMiner nya, data sebagai berikut

Tabel 2 Inisialisasi Jenis Tanaman

Jenis	Inisialisasi
Padi	1
Ubi Kayu	2
Cabai Besar	3
Jagung	4
Kedelai	5
Bawang Merah	6
Cabai Rawit	7
Tembakau	8

#### 4. Data Mining



Gambar 3 Susunan Operator RapidMiner

Algoritma k-means menentukan satu set k cluster dan menugaskan setiap contoh untuk tepat satu cluster. Cluster terdiri dari contoh yang sama. Persamaan antara contoh didasarkan pada ukuran setiap jarak. Satu cluster dalam algoritma k-means ditentukan oleh posisi pusat dalam ruang n-dimensi dari n atribut dari ExampleSet.

#### 5. Evaluasi

Dari proses hasil pengelompokan produktifitas pertanian berdasarkan jenis yang diperoleh dari

Implementasi Agoritma K-Means menggunakan DBI, menghasilkan DBI terbaik yaitu di  $k_9 = 0,196$ .

Tabel 3 Hasil DBI

Cluster	DBI
2	0,245
3	0,240
4	0,247
5	0,235
6	0,237
7	0,237
8	0,221
<b>9</b>	<b>0,196</b>
10	0,203

Berdasarkan hasil dari tabel DBI diatas telah dilakukan k sebanyak 10 kali untuk mengetahui dimana letak k optimum tersebut. Dibawah ini hasil cluster model pada cluster 3 menghasilkan item terbanyak yaitu 205 items dan yang memiliki item paling sedikit itu ada di cluster 4 yaitu 1 items.

### Cluster Model

```
Cluster 0: 96 items
Cluster 1: 20 items
Cluster 2: 30 items
Cluster 3: 205 items
Cluster 4: 1 items
Cluster 5: 25 items
Cluster 6: 49 items
Cluster 7: 3 items
Cluster 8: 83 items
Total number of items: 512
```

Gambar 4 Cluster Model DBI

Setelah mengetahui hasil cluster menggunakan DBI selanjutnya hasil klasterisasi data pengelompokan produktifitas pertanian berdasarkan jenis disajikan dengan



- [4] N. M. Manaor Hara Pardede Akim, "Penerapan Data Mining Pada Daerah Potensi Hasil Pertanian Yang Produktif Menggunakan Metode Algoritma K-Means Di Langkat," *Jurnal Informatika Kaputama*, vol. 6, no. 3, 2022.
- [5] P. Alkhairi dan A. P. Windarto, *Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara*. 2019. [Daring]. Tersedia pada: <https://seminar-id.com/semnas-sainteks2019.html>
- [6] B. Ginting dan F. Riandari, "Implementasi Metode K-Means Clustering Dalam Pengelompokan Bibit Tanaman Kopi Arabika," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 3, no. 2, 2020.
- [7] Sri Marliska Hutabarat dan Anita Sindar, "Data Mining Penjualan Suku Cadang Sepeda Motor Menggunakan Algoritma K-Means," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 2, no. 2, 2019.
- [8] B. Prasojo and E. Haryatmi, "Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random," vol. 02, pp. 79–89, 2021.
- [9] A. Y. Perdana, R. Latuconsina, A. Dinimaharawati, and U. Telkom, "Prediksi Stunting Pada Balita Dengan Algoritma Random," vol. 8, no. 5, pp. 6650–6656, 2021.
- [10] S. Linawati and S. Nurdiani, "Prediksi Prestasi Akademik Mahasiswa Menggunakan," vol. VIII, no. 1, pp. 47–52, 2020.
- [11] D. Ariyanto, "Data Mining Menggunakan Algoritma K-Means untuk Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut," *Jurnal Sistim Informasi dan Teknologi*, hlm. 13–18, Feb 2022, doi: 10.37034/jsisfotek.v4i1.117.
- [9] G. Gustientiedina, M. H. Adiya, dan Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 1, hlm. 17–24, Apr 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [10] D. Amelia, T. N. Padilah, dan A. Jamaludin, "Optimasi Algoritma K-Means Menggunakan Metode Elbow dalam Pengelompokan Penyakit Demam Berdarah Dengue (DBD) di Jawa Barat," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 11, hlm. 207–215, 2022, doi: 10.5281/zenodo.6831380.
- [11] Darmansah dan Ni Wayan Wardani, "Analisa Penyebab Kerusakan Tanaman Cabai Menggunakan Metode K-Means," *Jurnal Teknik Informatika dan Sistem Informasi*, hlm. 126–134, 2020.
- [12] A. Rohmatullah, "Klasterisasi Data Pertanian Di Kabupaten Lamongan Menggunakan Algoritma K-Means Dan Fuzzy C Means," *Jurnal Ilmiah Teknosains*, no. 2, hlm. 1, 2019.
- [13] Zulfa Nabila, A. Rahman Isnain, dan Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 2, hlm. 100, 2021, [Daring]. Tersedia pada: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [14] Y. Ratna Sari, A. Sudewa, D. Ayu Lestari, dan T. Ika Jaya, "Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan RapidMiner," 2020.
- [15] Y. F. S. Y. Damanik, S. Sumarno, I. Gunawan, D. Hartama, dan I. O. Kirana, "Penerapan Data Mining Untuk Pengelompokan Penyebaran Covid-19 Di Sumatera Utara Menggunakan Algoritma K-Means," *Jurnal Ilmu Komputer dan Informatika*, vol. 1, no. 2, Nov 2021, doi: 10.54082/jiki.13.