

PENERAPAN *DATA MINING* PADA JUMLAH PENDUDUK MISKIN DI INDONESIA

Nugroho Arif Sudibyo¹, Ardymulya Iswardani², Kartika Sari³, Siti Suprihatiningsih⁴
^{1,2} Universitas Duta Bangsa Surakarta
³ Universitas Padjadjaran
⁴ STKIP Pamane Talino, Landak
nugroho_arif@udb.ac.id¹, ardymulya@udb.ac.id², kartika20003@mail.unpad.ac.id³,
s.suprihatiningsih@stkippamanetalino.ac.id⁴

Abstract:

The purpose of this study is to apply data mining to the number of poor people in Indonesia with data from 2006 to March 2020. Cluster analysis will be carried out using the k-means method and then predictions on the number of poor people in Indonesia. The results showed that using the k-means cluster analysis method obtained two clusters. Furthermore, based on predictions that have been carried out with double exponential smoothing with an Alpha value of 1.094 and a Gamma of 0.539, it is predicted that in the next period there will be 27.4405 million poor people in Indonesia.

Keywords: Cluster, K-means, Prediction, Poor.

Abstrak :

Tujuan penelitian ini adalah menerapkan data mining pada jumlah penduduk miskin di Indonesia dengan data 2006 sampai Maret 2020. Akan dilakukan analisis cluster dengan metode k-means kemudian dilakukan prediksi pada jumlah penduduk miskin di Indonesia. Hasil penelitian menunjukkan bahwa dengan analisis *cluster* metode k-means diperoleh dua *cluster*. Selanjutnya, berdasarkan prediksi yang telah dilakukan dengan double exponential smoothing dengan nilai Alpha 1,094 dan Gamma 0,539, diprediksi periode selanjutnya terdapat 27,4405 juta jiwa penduduk miskin di Indonesia.

Kata Kunci : Cluster, K-means, Prediksi, Miskin.

Received: November 30, 2020 / Accepted: December 15, 2020 / Published Online: December 31, 2020

PENDAHULUAN

Setelah satu tahun berlalu pandemi COVID-19 mempengaruhi berbagai aspek kehidupan manusia, salah satunya adalah perekonomian (Sudiby, Iswardani, Septyanto, & Wicaksono, 2020). Perekonomian yang cenderung menurun mengakibatkan jumlah penduduk miskin yang meningkat (Megasari, Amar, & Idris, 2015). Jumlah penduduk miskin di Indonesia tren nya cenderung menurun akan tetapi Maret 2020 mengalami peningkatan (BPS, 2019).

Data mining sebenarnya bukanlah topik yang baru dalam penelitian. *Data mining* digunakan untuk memperbaiki teknik-teknik sebelumnya sehingga bisa mengangani berbagai masalah yang sering dijumpai (Maulana & Fajrin, 2018). Pada penelitian ini akan digunakan analisis cluster dan prediksi pada jumlah penduduk miskin di Indonesia.

Analisis cluster adalah teknik multivariat dengan tujuan utama mengelompokkan objek berdasarkan karakteristik yang dimiliki (Windarto, 2017). Saat ini, analisis cluster telah diterapkan di berbagai bidang sebagai contoh kemiripan DAS (Aytaç, 2020), serangan DOS (Iswardani & Riadi, 2016), pengelompokan mata kuliah (Rustam & Annur, 2019) dan masih banyak lagi. Salah satu analisis cluster yang banyak digunakan adalah K-means (Ramadani, Ambarita, & Pardede, 2019).

K-means juga dapat didefinisikan sebagai metode *clustering* yang termasuk dalam pendekatan partisi (Yang, Long, Ma, & Sun, 2020). Algoritma K-Means adalah model *centroid* (Wang, Da Cunha, Ritou, & Furet, 2019). Mode *centroid* adalah model yang menggunakan *centroid* untuk membuat cluster. *Centroid* adalah titik tengah sebuah cluster (Pietrzykowski, 2017).

Setelah di *cluster* jumlah penduduk miskin di Indonesia akan dilakukan prediksi jumlah penduduk miskin di Indonesia selanjutnya. Prediksi merupakan salah satu topik dalam *data mining* yang sering dilakukan (Eska, 2018). Akan digunakan beberapa metode prediksi dalam penelitian ini yaitu *moving average*, *single exponential smoothing* dan *double exponential smoothing* (Aden & Supriyanti, 2020).

Beberapa penelitian terkait dengan *data mining* pada tingkat kemiskinan di Indonesia adalah sebagai berikut. Penerapan *data mining* untuk clustering data penduduk miskin pada kota Jambi menggunakan algoritma k-means (Sunia, Kurniabudi, & Jusia, 2019). Hasil dari penelitian tersebut terdapat lima *cluster* yang dihasilkan. Selanjutnya, Metode k-means untuk pengelompokan masyarakat miskin menghasilkan tiga *cluster* dalam penelitiannya (Ramadani et al., 2019).

Penerapan *data mining* untuk *clustering* data penduduk miskin menggunakan algoritma hard c-means menghasilkan tiga *cluster* dalam penelitian tersebut. Selanjutnya, implementasi data mining dalam penentuan tingkat kemiskinan Kabupaten Bone Bolango, Provinsi Gorontalo menggunakan fuzzy c-means juga menghasilkan tiga cluster (Kumalasari, Kurniawan, Wiranto, & Kahar, 2020).

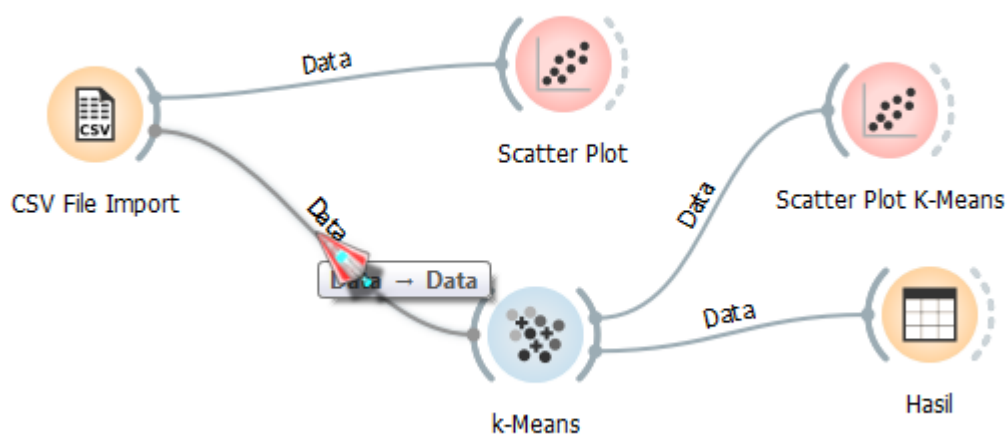
Tujuan penelitian ini adalah menerapkan *data mining* pada jumlah penduduk miskin di Indonesia dengan data 2006 sampai Maret 2020. Akan dilakukan analisis *cluster* dengan metode k-means

kemudian dilakukan prediksi pada jumlah penduduk miskin di Indonesia. Penelitian ini dilakukan dengan bantuan *software* Orange dan Minitab.

METODE PENELITIAN

Berikut adalah metode penelitian dalam penelitian ini.

1. Mengumpulkan data sekunder yaitu data tingkat kemiskinan di Indonesia dari Maret 2006 sampai dengan Maret 2020.
2. Analisis *cluster* metode k-means dengan bantuan *software* Orange. Gambar 1 merupakan *workflow* yang digunakan untuk analisis *cluster* metode k-means.



Gambar 1. Workflow yang digunakan untuk analisis *cluster* metode k-means

3. Prediksi tingkat kemiskinan di Indonesia pada periode selanjutnya. Model yang digunakan dalam prediksi adalah *moving average*, *single exponential smoothing* dan *double exponential smoothing*.

HASIL DAN PEMBAHASAN

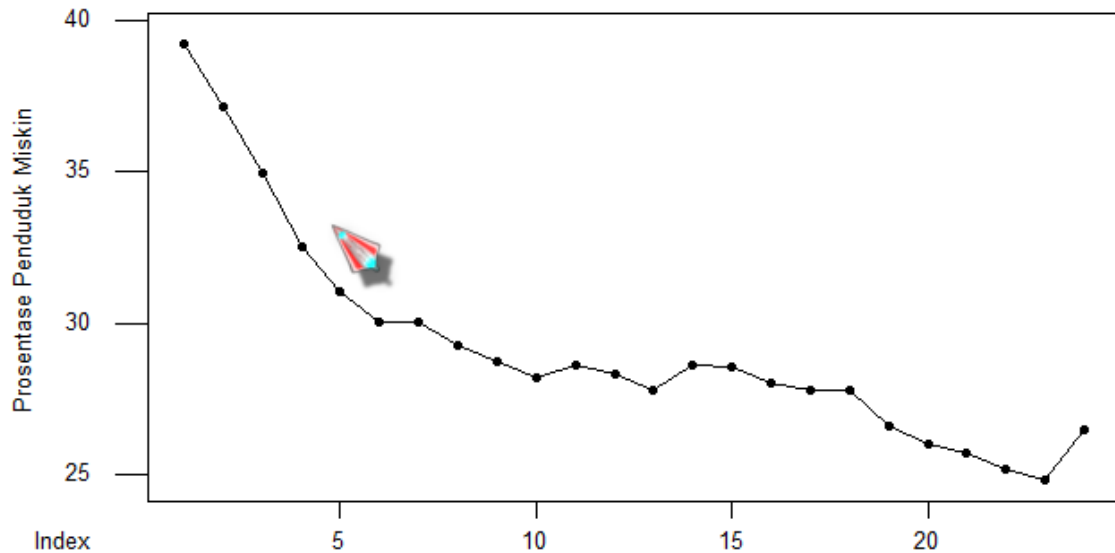
Data tingkat kemiskinan di Indonesia dari Maret 2006 sampai dengan Maret 2020 diambil dari Badan Pusat Statistik (BPS, 2019) merupakan jumlah penduduk miskin (juta). Analisis deskriptif yang sudah dilakukan tersaji pada Tabel 1.

Tabel 1. Analisis Deskriptif Statistik

Variabel	Mean	Std.Deviation	N
----------	------	---------------	---

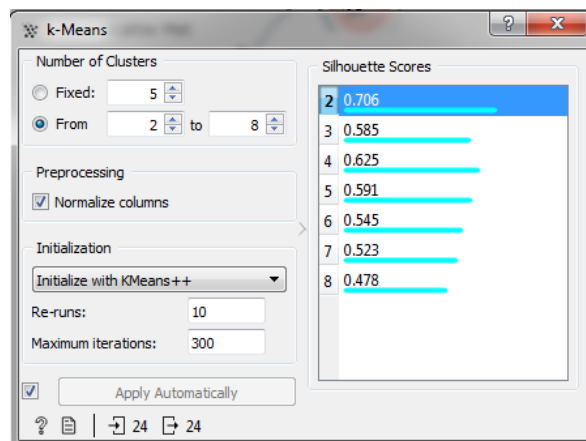
Tingkat kemiskinan	29,206	3,606	24
--------------------	--------	-------	----

Berikut adalah *time series plot* tingkat kemiskinan di Indonesia dari Maret 2006 sampai dengan Maret 2020 tersaji pada Gambar 2.



Gambar 2. *Time series plot* tingkat kemiskinan

Dari Gambar 2 terlihat bahwa tingkat kemiskinan mempunyai tren menurun tetapi pada awal tahun 2020 meningkat. Selanjutnya, akan dilakukan analisis *cluster* pada data tingkat kemiskinan di Indonesia dari Maret 2006 sampai dengan Maret 2020. Data terlebih dahulu dimasukkan dalam *software* Orange. Dengan bantuan *software* Orange diperoleh hasil sebagai berikut.



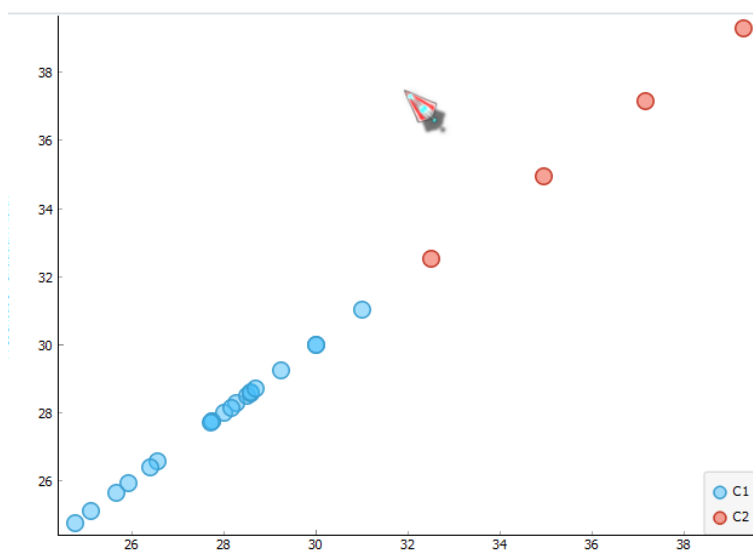
Gambar 3. Analisis cluster metode k-means

Dari Sihouette Score terlihat bahwa *cluster* yang direkomendasikan sebanyak dua *cluster*, sehingga data tersebut akan dibagi menjadi dua *cluster*. Berikut adalah hasil analisi *cluster* metode k-means disajikan pada Gambar 4.

	Cluster	Silhouette	Presentase Penduduk Miskin
1	C2	0.675414	39.3
2	C2	0.689839	37.17
3	C2	0.667074	34.96
4	C2	0.504601	32.53
5	C1	0.600999	31.02
6	C1	0.671922	30.02
7	C1	0.672408	30.01
8	C1	0.700648	29.25
9	C1	0.714108	28.71
10	C1	0.720412	28.17
11	C1	0.716212	28.6
12	C1	0.719668	28.28
13	C1	0.721282	27.73
14	C1	0.716372	28.59
15	C1	0.717414	28.51
16	C1	0.721055	28.01
17	C1	0.721384	27.76
18	C1	0.721394	27.77
19	C1	0.715338	26.58
20	C1	0.710324	25.95
21	C1	0.707537	25.67
22	C1	0.701504	25.14
23	C1	0.697055	24.79
24	C1	0.71427	26.42

Gambar 4. Hasil cluster dengan metode k-means beserta Sihouette Score

Dari Gambar 4 terlihat bahwa *cluster* pertama adalah tahun 2006 sampai tahun 2010 dengan jumlah diatas 31 juta penduduk. Sementara *cluster* kedua dengan jumlah dibawah 31 juta penduduk. Untuk lebih jelasnya berikut merupakan visualisasi hasil *cluster* dengan metode k-means pada Gambar 5.



Gambar 5. Visualisasi hasil *cluster* dengan metode *k-means*

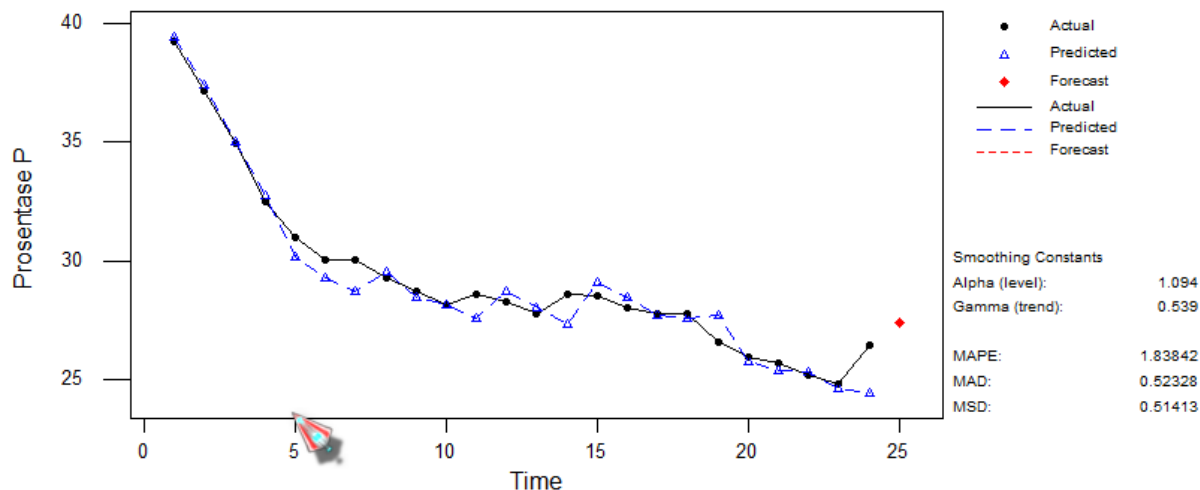
Hasil menunjukkan bahwa tingkat kemiskinan di Indonesia dapat dibagi menjadi dua *cluster*. Hal tersebut mengindikasikan bahwa setelah tahun 2010 masuk pada *cluster* kedua, dengan jumlah penduduk miskin dibawah 32 juta jiwa. Hal tersebut mengindikasikan bahwa tingkat kemiskinan di Indonesia cenderung menurun (Megasari et al., 2015; Soleh, 2015). Akan tetapi, pandemi COVID-19 menyebabkan Maret 2020 jumlah penduduk miskin mengalami kenaikan. Situasi ini tidak hanya dihadapi Indonesia tetapi juga dunia (Sheth, 2020). Selanjutnya diperlukan adanya prediksi untuk jumlah penduduk miskin di Indonesia untuk periode selanjutnya pada Tabel 2.

Tabel 2. Tabel prediksi dengan eror

Metode Prediksi	Prediksi	MAPE	MAD	MSD
<i>Moving average</i> dengan <i>length</i> 3	25,45	4,17282	1,20698	2,77349
<i>Moving average</i> dengan <i>length</i> 6	25,7583	5,16226	1,42722	3,24223
<i>Single exponential smoothing</i>	27,4405	2,24508	0,67409	0,72124
<i>Double exponential smoothing</i>	27,3933	1,83842	0,52328	0,51413

Berdasarkan pada Tabel 2, dari ke empat model prediksi nilai eror (MAPE, MAD dan MSD) yang terkecil adalah model *double exponential smoothing* (lihat Gambar 6). Dengan model tersebut, diprediksi bahwa pada periode selanjutnya sebesar 27,4405 juta jiwa penduduk miskin di Indonesia. Apabila dibandingkan dengan periode sebelumnya sebesar 26,42 juta jiwa berarti terjadi

kenaikan lebih dari 1 juta jiwa. Hal tersebut berakibat pada penduduk miskin di Indonesia yang cenderung terus meningkat (Nurwati, 2008).



Gambar 6. Double exponential smoothing dengan nilai Alpha 1,094 dan Gamma 0,539.

Dari analisis *cluster* dan prediksi yang telah dilakukan, jumlah penduduk miskin di Indonesia cenderung akan meningkat pada periode selanjutnya. Kenaikan atau penurunan jumlah penduduk miskin tidak menjadi refleksi dari keadaan nyata di Indonesia (Martiyani Ramdani, 2015). Akan tetapi perlu menjadi perhatian bersama baik pemerintah maupun masyarakat. Karena jumlah penduduk miskin salah satu indikator keberhasilan pembangunan di suatu negara (Zahra, Fatin A, Afuwu, & Auliyah R, 2019). Tidak bisa dipungkiri bahwa naiknya jumlah penduduk miskin salah satunya disebabkan pandemi COVID 19 dari Desember 2019 sampai sekarang (Chairani, 2020).

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa dengan analisis *cluster* metode k-means diperoleh dua *cluster*. Selanjutnya, berdasarkan prediksi yang telah dilakukan dengan *double exponential smoothing* dengan nilai Alpha 1,094 dan Gamma 0,539, diprediksi periode selanjutnya terdapat 27,4405 juta jiwa penduduk miskin di Indonesia.

DAFTAR PUSTAKA

- Aden, & Supriyanti, A. (2020). PREDIKSI JUMLAH CALON PESERTA DIDIK BARU MENGGUNAKAN METODE DOUBLE EXPONENTIAL SMOOTHING DARI BROWN (Study Kasus : SD Islam Al-Musyarrofah Jakarta). *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 1(1), 56–62. <https://doi.org/10.46306/lb.v1i1>
- Aytaç, E. (2020). Unsupervised learning approach in defining the similarity of catchments: Hydrological response unit based k-means clustering, a demonstration on Western Black Sea Region of Turkey. *International Soil and Water Conservation Research*, 8(3), 321–331. <https://doi.org/10.1016/j.iswcr.2020.05.002>
- BPS. (2019). Profil Kemiskinan di Indonesia. *Berita Resmi Statistik*, (56), 1–12. Retrieved from <https://www.bps.go.id/pressrelease/2020/01/15/1743/persentase-penduduk-miskin-september-2019-turun-menjadi-9-22-persen.html>
- Chairani, I. (2020). Dampak Pandemi Covid-19 Dalam Perspektif Gender Di Indonesia. *Jurnal Kependudukan Indonesia, Edisi Khusus*(Juli), 39–42.
- Eska, J. (2018). *Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5. 2*. <https://doi.org/10.31227/osf.io/x6svc>
- Iswardani, A., & Riadi, I. (2016). Denial of service log analysis using density K-means method. *Journal of Theoretical and Applied Information Technology*, 83(2), 299–302.
- Kumalasari, I., Kurniawan, A., Wiranto, A., & Kahar, R. N. (2020). KOMMAS : Jurnal Pengabdian Kepada Masyarakat Universitas Pamulang EDUKASI INTERNET UNTUK WIRAUSAHA PARA SANTRI PONDOK KOMMAS : Jurnal Pengabdian Kepada Masyarakat Universitas Pamulang. *KOMMAS: Jurnal Pengabdian Kepada Masyarakat*, 1, 104–108.
- Martiyani Ramdani. (2015). Determinan Kemiskinan Di Indonesia Tahun 1982-2012. *Economics Development Analysis Journal*, 4(1), 58–64. <https://doi.org/10.15294/edaj.v4i1.14803>
- Maulana, A., & Fajrin, A. A. (2018). Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Transaksi Penjualan Spare Part Motor. *Klik - Kumpulan Jurnal Ilmu Komputer*, 5(1), 27. <https://doi.org/10.20527/klik.v5i1.100>
- Megasari, H., Amar, S., & Idris, I. (2015). Analisis Perekonomian Dan Kemiskinan Di Indonesia. *Jurnal Kajian Ekonomi*, 3(6), 103629.
- Nurwati, N. (2008). Kemiskinan : Model Pengukuran, Permasalahan dan Alternatif Kebijakan. *Jurnal Kependudukan Padjadjaran*, 10(1), 245387.
- Pietrzykowski, M. (2017). Local regression algorithms based on centroid clustering methods. *Procedia Computer Science*, 112, 2363–2371. <https://doi.org/10.1016/j.procs.2017.08.210>
- Ramadani, S., Ambarita, I., & Pardede, A. M. H. (2019). Metode K-Means Untuk Pengelompokan

- Masyarakat Miskin Dengan Menggunakan Jarak Kedekatan Manhattan City Dan Euclidean (Studi Kasus Kota Binjai). *Information System Development (ISD)*, 04(2), 15–29.
- Rustam, S., & Annur, H. (2019). Akademik Data Mining (Adm) K-Means Dan K-Means K-Nn Untuk Mengelompokan Kelas Mata Kuliah Kosentrasi Mahasiswa Semester Akhir. *ILKOM Jurnal Ilmiah*, 11(3), 260–268. <https://doi.org/10.33096/ilkom.v11i3.487.260-268>
- Sheth, J. (2020). Impact of Covid-19 on consumer behavior: Will the old habits return or die? *Journal of Business Research*, 117, 280–283. <https://doi.org/10.1016/j.jbusres.2020.05.059>
- Soleh, A. (2015). Pertumbuhan Ekonomi Dan Kemiskinan Di Indonesia. *EKOMBIS REVIEW: Jurnal Ilmiah Ekonomi Dan Bisnis*, 2(2), 197–209. <https://doi.org/10.37676/ekombis.v2i2.15>
- Sudibyo, N. A., Iswardani, A., Septyanto, A. W., & Wicaksono, T. G. (2020). Prediksi Inflasi Di Indonesia Menggunakan Metode Moving Average, Single Exponential Smoothing Dan Double Exponential Smoothing. *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 1(2), 123–129. <https://doi.org/10.46306/lb.v1i2.25>
- Sunia, D., Kurniabudi, & Jusia, P. A. (2019). Penerapan Data Mining untuk Clustering Data Penduduk Miskin Menggunakan Algoritma K-Means. *Jurnal Ilmiah Mahasiswa Teknik Informatika, Vol 1 No 2(2016)*, 121–134.
- Wang, Z., Da Cunha, C., Ritou, M., & Furet, B. (2019). Comparison of K-means and GMM methods for contextual clustering in HSM. *Procedia Manufacturing*, 28, 154–159. <https://doi.org/10.1016/j.promfg.2018.12.025>
- Windarto, A. P. (2017). Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method. *International Journal of Artificial Intelligence Research*, 1(2), 26. <https://doi.org/10.29099/ijair.v1i2.17>
- Yang, W., Long, H., Ma, L., & Sun, H. (2020). Research on clustering method based on weighted distance density and k-means. *Procedia Computer Science*, 166, 507–511. <https://doi.org/10.1016/j.procs.2020.02.056>
- Zahra, A., Fatin A, A., Afuwu, H., & Auliyah R, R. (2019). Struktur Kemiskinan Indonesia: Berapa Besar Pengaruh Kesehatan, Pendidikan dan Kelayakan Hunian? *Jurnal Inovasi Ekonomi*, 4(02), 67–74. <https://doi.org/10.22219/jiko.v4i2.9856>