

ALGORITMA K-MEANS DALAM IMPLEMENTASI BIDANG PEKERJAAN

Akmal Khansa Al Irsyad¹, Priati Assiroj², Besse Hartati³

¹⁻³Program studi Manajemen Teknologi Keimigrasian, Politeknik Imigrasi

¹akmalkhansaalirsyad@gmail.com, ²priati.assiroj@poltekim.ac.id,

³hartatimassaile@yahoo.com

ABSTRACT

Rapid technological advances have affected various fields, especially in data management. The increasing volume of data generated from various sources demands efficient management and analysis methods. Data mining techniques offer a structured approach in processing, classifying, and grouping data to support decision making in various fields. This study is a systematic review of the application of data mining techniques, with a primary focus on the K-Means Clustering algorithm. This study analyzes the trend of data mining applications, especially in data classification and grouping to improve the effectiveness of decision making. Based on a systematic literature review, it was found that the K-Means Clustering algorithm is widely applied in sales analysis, market segmentation, stock optimization, and predictions in the social and health fields. In addition, other algorithms such as Decision Tree, Naïve Bayes, and K-Nearest Neighbor are also commonly used in predictive analysis and data classification. This study provides insight into the effectiveness of various data mining techniques and their future development opportunities.

Keywords: data mining, k-means clustering, classification

ABSTRAK

Perkembangan teknologi yang pesat telah mempengaruhi berbagai bidang, terutama dalam pengelolaan data. Meningkatnya volume data yang dihasilkan dari berbagai sumber menuntut metode pengelolaan dan analisis yang efisien. Teknik penambangan data menawarkan pendekatan terstruktur dalam memproses, mengklasifikasi, dan mengelompokkan data untuk mendukung pengambilan keputusan di berbagai bidang. Penelitian ini merupakan tinjauan sistematis penerapan teknik penambangan data, dengan fokus utama pada algoritma K-Means Clustering. Penelitian ini menganalisis tren penerapan penambangan data, terutama dalam klasifikasi dan pengelompokan data untuk meningkatkan efektivitas pengambilan keputusan. Berdasarkan tinjauan pustaka sistematis, ditemukan bahwa algoritma K-Means Clustering banyak diterapkan dalam analisis penjualan, segmentasi pasar, optimasi stok, dan prediksi di bidang sosial dan kesehatan. Selain itu, algoritma lain seperti Decision Tree, Naïve Bayes, dan K-Nearest Neighbor juga umum digunakan dalam analisis prediktif dan klasifikasi data. Penelitian ini memberikan wawasan tentang efektivitas berbagai teknik penambangan data dan peluang pengembangannya di masa mendatang.

Kata kunci: data mining, cluster k-means, klasifikasi

A. Pendahuluan

Pada era yang telah mengalami kemajuan yang sangat pesat di berbagai bidang khususnya bidang teknologi data dan sains, dunia harus mengikuti pola dan kemajuan yang telah berlangsung dari masa ke masa (Wiryaningrum et al., 2022). Teknologi memberikan manfaat dan kemudahan di setiap bidangnya karena pada dasarnya, segala hal yang ada di dunia yang memiliki ketergantungan dengan teknologi harus diiringi dengan kemampuan manusia itu sendiri agar tidak disalahgunakan. Teknologi yang didasari pada suatu sistem harus dikelola dan disimpan dengan bijak agar menjadi kemudahan bagi manusia dalam penggunaannya pada berbagai bidang pekerjaan (Awailiyah et al., 2024). Data yang ada pada sistem tersebut tentu merupakan data yang sangat beragam dan rumit sehingga perlu dianalisis dan dilakukan pengelolaan data dengan sangat hati-hati. Dalam melakukan pengelolaan data tersebut diperlukan suatu pola atau metode khusus yang telah ada dan digunakan oleh berbagai organisasi atau Lembaga besar.

Perkembangan teknologi tidak hanya di lingkup tertentu saja, Pada

bidang pendidikan, bisnis, sosial dan pemasaran juga sudah menjadi pondasi utama dalam melakukan dan memanfaatkan bidang tersebut. Di zaman yang sudah canggih saat ini, manusia tidak akan lepas dari perkembangan teknologi informasi. Setiap manusia memanfaatkan dan memahami penggunaan dan dampak dari teknologi untuk mempermudah aktivitas pekerjaan, bersosial dan mencari informasi dan kebutuhan hidup (Assiroj, 2022). Dalam penerapan data yang tergolong besar, teknologi berperan aktif untuk pengelolaan data sebagai sistem yang dapat menjadi informasi secara akurat. Dilakukannya data mining menjadi langkah pasti dalam mengelola sumber data yang memiliki cakupan yang sangat luas (Huda Abdaul Irkham, 2020).

Data *mining* merupakan sebuah metode yang cocok, yang dimana merupakan suatu pola atau ilmu dalam pola analisis data dalam pengelompokan dan klasifikasi dari data dengan berbagai teknik dan metode yang ada. Data *mining* ini dapat digunakan dan diaplikasikan di berbagai bidang keilmuan yang ada. Meningkatnya kapasitas data yang dihasilkan oleh sumber yang tersebar,

data mining menjadi suatu tren yang penting yang digunakan oleh berbagai bidang yang ada di dunia. Teknik yang ada pada data mining dapat menciptakan berbagai keputusan dan langkah yang akan dilakukan oleh organisasi atau lembaga berdasarkan analisis yang diperoleh dari data yang ada (L Hartono et al., 2021).

Berbagai macam metode yang digunakan dalam data mining yang meliputi teknik klasifikasi, klusterisasi, asosiasi dan lainnya. Data mining juga memiliki algoritma dalam penerapan tekniknya diantaranya Decision Tree, K-Means, Apriori yang dimana menjadi algoritma yang umum digunakan selama proses pengimplementasian data (Dan et al., 2021).

Penelitian data mining dengan algoritma K-Means menggunakan teknik *clustering* untuk mengelompokkan data. *Clustering* merupakan metode dengan maksud untuk mengelompokkan data berdasarkan karakteristik yang berbeda ke dalam kelompok. K-Means mencakup aspek penting dalam penambahan data yang dijelaskan sebagai algoritma penyaringan. Teknik *Clustering* dilakukan dengan pendekatan partisi

dan hirarki atau sering disebut dengan *partition based clustering* dan *hierarchical clustering* (Sophia & Jasmir, 2024).

Dalam menemukan hasil data mining dengan cara klusterisasi dapat menggunakan algoritma K-Means. K-Means merupakan sebuah algoritma dengan teknik *clustering* atau pengelompokan data dalam menemukan pola dari hasil data mining (Ulfah & Sri Irtwaty, 2022). Algoritma K-Means dikategorikan pada cara pengelompokkan data ke dalam cluster yang diteliti. Kelemahan K-Means terlatak pada sensitifnya pemilihan titik centroid dan tidak dapat mengelompokkan data yang memiliki kluster bulat (Romadhona et al., 2022).

Metode penelitian yang digunakan yaitu meninjau pada *Systematic Literature Review (LSR)*. SLR dimanfaatkan untuk melakukan analisis informasi dari jurnal yang dilakukan penelitian oleh peneliti sebelumnya untuk memperoleh hasil tinjauan yang sistematis dengan berbagai cara dan tahapan (Putri Yuanita et al., 2024).

Melalui tinjauan ini, diharapkan dapat menciptakan wawasan yang lebih luas mengenai sistem data

mining khususnya algoritma K-Means sehingga dapat dimanfaatkan dengan bijak. Penelitian ini juga diharapkan dapat memberikan gambaran mengenai teknik algoritma k-means pada implementasi data mining.

B. Metode Penelitian

Untuk melakukan studi literatur, penelitian ini menerapkan tinjauan literatur dengan metode Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA). PRISMA merupakan pedoman yang digunakan untuk penyusunan laporan systematic review secara transparan dan komprehensif. Proses ini mencakup 5 tahap (Syahfitri et al., 2024).

a. Tahap 1 : Penetapan kriteria kelayakan artikel ditentukan oleh kriteria inklusi (IC), yaitu :

- IC1 : artikel berdasarkan penelitian asli yang telah dilakukan dan ditulis dengan bahasa Indonesia.
- IC2 : artikel yang diterbitkan pada tahun 2020 sampai 2024.
- IC3 : artikel bertujuan untuk menganalisis

teknik dan metode peneliti lain dalam menerapkan data mining.

b. Tahap 2 : Menentukan sumber informasi

- Literatur dapat dilihat melalui basis data daring dalam berbagai koleksi akademik diantaranya, Google Scholar, Crossref, dan Semantic Scholar.
- Penemuan dilakukan pada artikel yang memenuhi syarat untuk IC dan juga dilakukan pencarian lebih lanjut untuk menemukan penelitian yang memiliki kaitan dengan data mining.

c. Tahap 3 : Pemilihan literatur

- Pemilihan kata kunci adalah "Implementasi K-Means", dan "Implementasi Data Mining".
- Untuk meneliti dan menentukan judul, ringkasan dan artikel, kata kunci yang diperoleh dari hasil

pencarian yang selaras dengan persyaratan penelitian akan dievaluasi dengan menerapkan tahap 3 dan 4.

- d. Tahap 4 : Pengumpulan data
Data dikumpulkan melalui pembuatan formular ekstraksi data manual. Sebanyak 57.500 dievaluasi menggunakan kata kunci "Implementasi K-Means", dan "Implementasi Data Mining". dari semua sumber dan kriteria yang memungkinkan. 334 artikel memenuhi persyaratan awal. Setelah penyaringan tambahan, 31 artikel dipilih untuk penelitian ini. Tabel 1 menampilkan data yang dikumpulkan.

- e. Tahap 5 : Pemilihan item data
Data diperoleh dari artikel terpilih yang berisi teknik dan metode yang digunakan untuk melakukan pembahasan dan penerapan dari data mining khususnya *K-Means Clustering*.

Dalam penelitian yang dilakukan, untuk mencari Literature Review melalui *Harzing Publish and*

Perish dengan 3 (tiga) sumber yang telah disebutkan yaitu google scholar, crossref, semantic scholar. Penelitian ini dalam menganalisis jurnal yang di review untuk menentukan penelitian yang akurat dan relevan dengan tema yang dipilih dilakukan dengan memisahkan jurnal mana yang cocok untuk dijadikan sumber studi literatur. Kecocokan data yang diperlukan untuk studi literatur meliputi nama penulis, tahun terbit jurnal, perangkat lunak atau software yang digunakan, bidang pekerjaan dan hasil implementasi terhadap bidang tersebut yang diterapkan melalui algoritma K-Means.

C. Hasil Penelitian Dan Pembahasan

Penelitian ini meninjau segala teknik dan metode yang diterapkan oleh berbagai peneliti pada sumber yang telah dicari untuk proses data mining. Penelitian ini menganalisis studi literatur yang ditinjau dengan melakukan pendataan dari nama penulis, tahun terbit, serta hasil implementasinya. Tabel I menunjukkan bagaimana peneliti melakukan riset terkait pengimplementasian algoritma K-Means pada berbagai bidang. Tabel

tersebut juga menampilkan perangkat lunak atau aplikasi yang digunakan dalam penelitian, dan menampilkan hasil dari implementasi pada bidang seperti kesehatan, pendidikan, pemasaran, keuangan dan investasi, agrikultur, transportasi, dan sosial. Teknik yang dipakai yaitu clustering, dengan menerapkan algoritma K-Means. Penelitian ini menyebutkan perangkat lunak atau *software* yang digunakan secara umum yaitu RapidMiner, disamping menggunakan Matlab, Weka, Python, Orange dan aplikasi Tanagra. Dalam penelitian ini RapidMiner menjadi perangkat yang paling banyak digunakan pada analisis data dengan teknik *clustering* menggunakan algoritma K-Means. Penelitian ini memperlihatkan metode Davies-Bouldin Index (DBI) pada jurnal data pembayaran transaksi dengan python oleh Dodi Alexsander, dkk tahun 2022 yang memperoleh nilai DBI sebesar 0.5021 (Manalu & Gunadi, 2022).



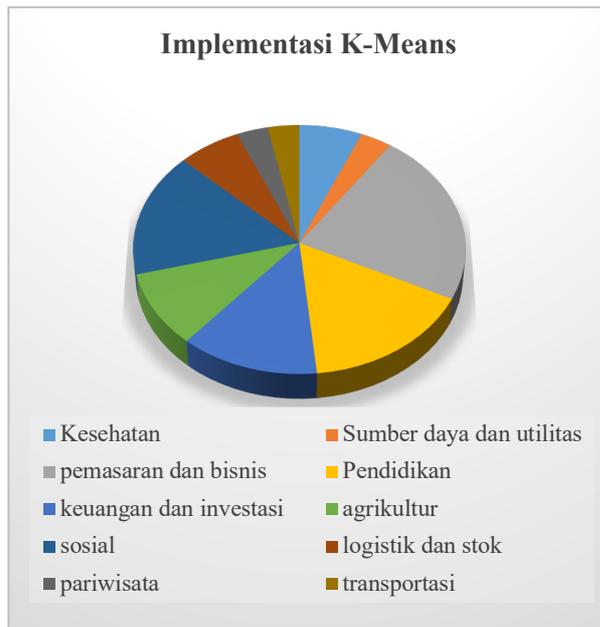
Gambar 1 Flowchart LSR

Tabel 1 Implementasi K-Means dalam bidang pekerjaan

| Perangkat Lunak | Bidang | Hasil |
|-----------------|--------------------------|---|
| MATLAB | Kesehatan | Segmentasi pasien rawat inap berdasarkan pola kunjungan |
| RapidMiner | Kesehatan | Analisis penyebaran Covid-19 |
| WEKA | Sumber Daya dan Utilitas | Pengelompokan pelanggan berdasarkan konsumsi air |
| RapidMiner | Agrikultur | Pengelompokan wilayah potensial produksi daging |
| RapidMiner | Pemasaran | Pengelompokan pelanggan potensial berdasarkan riwayat pembelian |
| Python | Keuangan | Pengelompokan pelanggan berdasarkan pembayaran transaksi |

| | | |
|------------|------------|---|
| RapidMiner | Pendidikan | Klasifikasi buku berdasarkan peminjaman |
|------------|------------|---|

| | | |
|------------|------------------------------|--|
| RapidMiner | Keuangan dan Investasi | Pengelompokan suku bunga deposito berdasarkan profitabilitas |
|------------|------------------------------|--|



Gambar 2 Implementasi K-Means

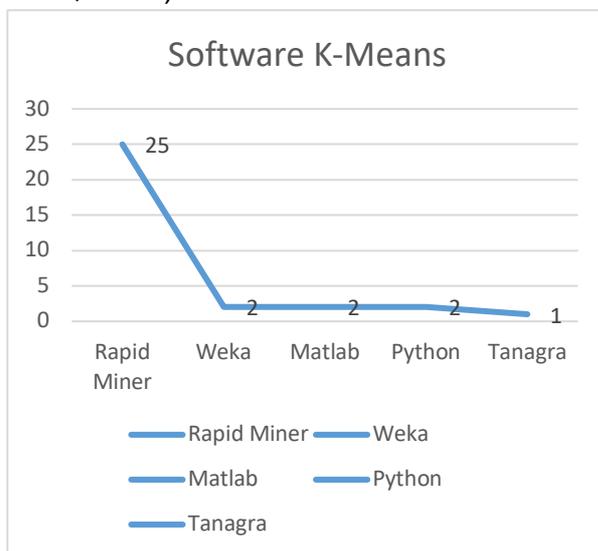
Dari tabel 1, diambil 8 contoh bidang pekerjaan dari 31 jurnal yang membahas pemanfaatan K-Means dan dianalisis sebagai gambaran bahwa K-Means dapat digunakan pada bidang pekerjaan. Berdasarkan gambar 2 dapat dijelaskan bahwa dengan melakukan analisis pada implementasi metode K-Means dengan teknik *clustering* terhadap berbagai bidang dari jurnal ilmiah yang diteliti, algoritma dengan K-Means paling banyak digunakan pada bidang pekerjaan untuk memberikan peran layanan dan sebagai fasilitator penyediaan barang dan jasa. Contohnya pada bidang

pemasaran dan bisnis, dengan target yang difokuskan yaitu segmentasi pelanggan dan analisis pola pembelian. Dari salah satu bidang ini dapat disimpulkan bahwa K-Means menjadi cara akurat dan efektif untuk mengelompokkan data pelanggan dari kebiasaan konsumsi terhadap produk tersebut sehingga dapat meningkatkan strategi pemasaran yang jelas.

Pada bidang keuangan dan investasi juga menganalisis jurnal yang menjelaskan bahwa pengelompokan nasabah berdasarkan risiko kredit serta saham berdasarkan tren harga dapat membantu lembaga keuangan untuk melakukan mitigasi risiko dan melakukan analisis prediksi pasar investasi (Wibowo & Sasongko, 2022). Algoritma K-Means dimanfaatkan juga dalam pengelompokan pasien berdasarkan riwayat penyakit, jenis layanan, dan BPJS pasien. Dengan meminjam penyakit pasien dapat membantu pihak rumah sakit untuk melakukan optimalisasi layanan kesehatan serta mengidentifikasi tren penyakit apa saja yang sering terjadi di suatu wilayah (Widiya, 2022). Dalam mengelompokkan mahasiswa dengan

indeks prestasi akademik dan beasiswa di bidang pendidikan, *K-Means* berperan sehingga dapat membantu instansi pendidikan untuk penentuan keputusan agar kualitas akademik meingkat (Usanto, 2023).

Disamping itu, pengelolaan sumber daya dan logistik untuk memprediksi stok barang, pasokan dan mengelompokan lokasi distribusi suatu barang atau produk dilakukan dengan teknik *clustering*. Sehingga nantinya dapat membantu perusahaan dalam mengoptimalkan operasional dan mengetahui jumlah stok barang (Zafira et al., 2024).



Grafik 3 Software K-Means

Pada penelitian ini, dibuat grafik untuk menampilkan hasil dari implementasi bidang dengan pemanfaatan teknik data mining yaitu *clustering* dengan algoritma *K-Means*. Dilihat dari diagram gambar 2, dapat dilihat bahwa implementasi *K-Means* yaitu pemasaran dan bisnis menjadi bidang yang memiliki persentase hasil implementasi *K-Means* paling tinggi. Diikuti oleh bidang lainnya seperti bidang kesehatan, keuangan dan investasi, dan pendidikan. Hal ini dapat disimpulkan bahwa dengan melakukan analisis terhadap pelanggan, manajemen risiko keuangan, segmentasi produk pemasaran rata rata memanfaatkan *K-Means* untuk mengelompokan data dibanding bidang agrikultur dan transportasi yang tergolong sedikit dalam pengelompokan data.

D. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat ditarik beberapa Kesimpulan dimana pada penelitian sebelumnya telah mengkaji data mining dengan teknik clustering dan dengan algoritma *K-Means*. Proses pengelompokan data menggunakan

teknik clustering dengan algoritma K-Means adalah pola yang digunakan dalam pengelompokan data yang besar seperti pengelompokan data pasien, manajemen pemasaran stok barang dan analisis nilai akademik. K-Means dengan keunggulannya yaitu memungkinkan proses data dengan akurat dan cepat serta memberikan hasil klaster yang dapat dimengerti. Dari perangkat lunak yang digunakan, RapidMiner menjadi pilihan utama untuk pengimplementasian algoritma, selain MATLAB, WEKA, dan Python. Hasil penelitian ini memperlihatkan K-Means sebagai potensi besar bahwa dalam pengimplementasian di berbagai bidang dapat dikembangkan lebih lanjut untuk meningkatkan efektivitas algoritma K-Means untuk melakukan analisis data di masa depan. Perlu disampaikan juga bahwa penelitian ini memiliki keterbatasan dalam pencarian data pada sumber jurnal dari tahun 2020 sampai 2024.

DAFTAR PUSTAKA

- Dan, Ardilla, Y., Manuhutu, A., Ahmad, N., Hasbi, I., Agnes Manuhutu, M., Ridwan, M., Khrisna Wardhani, A., Alim, S., Romli, I., Religia, Y., Tri Octafian, D., Utan Sufandi, U., & Ernawati, I. (2021). *DATA MINING*. www.penerbitwidina.com
- Assiroj, P. (2022). IMPLEMENTASI METODE SEARCH ENGINE OPTIMIZATION (SEO) PADA SITUS WEB IMIGRASI WONOSOBO. *INFOTECH Journal*, 8(1).
<https://doi.org/10.31949/infotech.v8i1.2239>
- Awailiyah, C., Oktaviana, D., & Herlambang, Y. T. (2024). Tantangan dan Peluang Teknologi dalam Dinamika Kehidupan di Era Teknologi. *UPGRADE: Jurnal Pendidikan Teknologi Informasi*, 1(2).
<https://doi.org/10.30812/upgrade.v1i2.3729>
- Huda Abdaul Irkham. (2020). *Perkembangan Teknologi Informasi dan Komunikasi (TIK) Terhadap Kulaitas Pembelajaran*

- Di Sekolah Dasar Irkham Abdaul Huda (Vol. 2).*
- L Hartono, A., Assiroj, P., Priyatna, B., Nurhayati, Fauzi, A., Yuniar Rahman, A., & Shofiah Hilabi, S. (2021). Analysis of Drug Data Mining with Clustering Technique Using K-Means Algorithm. *Journal of Physics: Conference Series*, 1908(1).
<https://doi.org/10.1088/1742-6596/1908/1/012024>
- Manalu, D. A., & Gunadi, G. (2022). IMPLEMENTASI METODE DATA MINING K-MEANS CLUSTERING TERHADAP DATA PEMBAYARAN TRANSAKSI MENGGUNAKAN BAHASA PEMROGRAMAN PYTHON PADA CV DIGITAL DIMENSI. *Infotech: Journal of Technology Information*, 8(1), 43–54.
<https://doi.org/10.37365/jti.v8i1.131>
- Putri Yuanita, S., Luffina, E., Dimentieva, I., Naufal Pribadi, A., Shaunan Syahidan, M., Alief Fauzan Akbar Safa, M., Loekito, J., Nuswantoro PSDKU Kediri, D., Penanggungan No, J., Lor, B., Kota, K., Kediri, K., & Timur, J. (2024). Science, Technology and Management Journal Identifikasi Teknik Data Mining Metode Asosiasi: Systematic Literature Review Info Artikel. *Science Technology and Management Journal*, 4(2), 57.
<https://doi.org/10.53416/stmj.v4i2>
- Romadhona, W., Indarmawan Nugroho, B., & Alim Murtopo, A. (2022). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means. *Jurnal Minfo Polgan*, 11(2), 100–104.
<https://doi.org/10.33395/jmp.v11i2.11797>
- S, U. (2023). Penerapan Data Mining Dengan Mengimplementasikan Algoritma K-Means Dalam Proses Clustering Untuk Pengelompokan Mahasiswa Calon Penerima Beasiswa KIP. *Building of Informatics, Technology and Science (BITS)*, 5(1).
<https://doi.org/10.47065/bits.v5i1.3411>
-

- Sophia, A., & Jasmir. (2024). *Penerapan Data Mining dalam Pengelompokan Uang Kuliah Tunggal (UKT) Menggunakan Metode K-Means Pada Universitas Jambi* (Vol. 9, Issue 1). PROFITABILITAS (Application of Data Mining on Deposit Investment Rate in Indonesia Using K-Means Clustering Method for Profitability Clustering). *ASCARYA*, 2(1). <https://doi.org/10.53754/iscs.v2i1.369>
- Syahfitri, R. I., Ayu, D., Zahroh, S., Sinaga, H., Tanjung, H. Y., Fidorova, Y., & Rahmadhona Tumanggor, S. (2024). *El-Mujtama: Jurnal Pengabdian Masyarakat Analisis Pengaruh Pola Makan terhadap Gastritis Menggunakan Metode PRISMA*. <https://doi.org/10.47467/elmujtama.v4i2.3391>
- Ulfah, M., & Sri Irtwyaty, A. (2022). *Penerapan Data Mining Clustering Menggunakan Metode K-Means Dalam Pengelompokan Buku Perpustakaan Politeknik Negeri Balikpapan*. Widiya, L. (2022). Penerapan Data Mining Pengelompokan Hasil Diagnosa Pasien BPJS Berdasarkan Usia Menggunakan Metode Clustering (Studi Kasus: RSUD Bidadari Binjai) The Application of Data Mining for BPJS Patient Diagnostic Results Grouping by Age Using the Clustering Method (Case Study: RSUD Bidadari Binjai). In *JOURNAL OF INFORMATION AND TECHNOLOGY UNIMOR*. www.kaputama.ac.id
- Wibowo, A., & Sasongko, R. (2022). PENERAPAN DATA MINING PADA SUKU BUNGA INVESTASI DEPOSITO DI INDONESIA MENGGUNAKAN METODE K-MEANS CLUSTERING UNTUK PENGELOMPOKAN Wiryany, D., Natasha, S., & Kurniawan, R. (2022). PERKEMBANGAN TEKNOLOGI INFORMASI DAN KOMUNIKASI TERHADAP PERUBAHAN SISTEM KOMUNIKASI INDONESIA. In *Jurnal Nomosleca, Oktober* (Vol. 8, Issue 2).

Zafira, F., Irawan, B., & Bahtiar, A.
(2024). PENERAPAN DATA
MINING UNTUK ESTIMASI STOK
BARANG DENGAN METODE K-
MEANS CLUSTERING. In *Jurnal
Mahasiswa Teknik Informatika*
(Vol. 8, Issue 1).