

KLASTERISASI DATA : ANALISIS KINERJA K-MEANS PADA SEKTOR PAJAK, EKSPOR, PERIKANAN, MODAL, DAN SUMBER DAYA

Achmad S.W.A.Nurba¹⁾, Dharma Fathahillah²⁾, Muhamad Shafly Pratama³⁾, Muhammad Rivaldi Bachtiar⁴⁾, Muhammad Chesta Adabi Putra⁵⁾, Sartika Lina Mulani Sitio⁶⁾
Prodi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang

Correspondence author: S.L.M. Sitio, dosen00847@unpam.ac.id, Tangerang Selatan, Indonesia

Abstract

This study aims to cluster Indonesian economic data patterns from five sectors: tax, export, fisheries, capital markets, and resources, using the K-Means algorithm. Data were obtained from BPS, the Ministry of Finance, the Ministry of Marine Affairs and Fisheries, the Financial Services Authority (OJK), and UN Comtrade. Pre-processing was carried out through data cleaning and normalization. The optimal number of clusters was determined using the elbow and silhouette methods. Clustering evaluation used *Inertia*, *Silhouette score*, and the *Davies-Bouldin Index*. The results show variations in cluster patterns in each sector, with the fisheries and capital markets sectors providing the best results (high *silhouette scores*). Visualization using PCA supports cluster interpretation. These findings demonstrate that K-Means is effective in economic data analysis and helps support more adaptive and data-driven policies.

Keywords: *analysis, indonesian economic data, k-means algorithm, clustering*

Abstrak

Penelitian ini bertujuan mengelompokkan pola data ekonomi Indonesia dari lima sektor: pajak, ekspor, perikanan, pasar modal, dan sumber daya, menggunakan algoritma K-Means. Data diperoleh dari BPS, Kementerian Keuangan, KKP, OJK, dan UN Comtrade. Pra-pemrosesan dilakukan melalui pembersihan dan normalisasi data. Jumlah kluster optimal ditentukan menggunakan metode *elbow* dan *silhouette*. Evaluasi klasterisasi menggunakan *Inertia*, *Silhouette score*, dan *Davies-Bouldin Index*. Hasil menunjukkan variasi pola kluster di tiap sektor, dengan sektor perikanan dan pasar modal memberikan hasil terbaik (*silhouette score* tinggi). Visualisasi menggunakan PCA mendukung interpretasi kluster. Temuan ini menunjukkan bahwa K-Means efektif dalam analisis data ekonomi dan bermanfaat untuk mendukung kebijakan yang lebih adaptif dan berbasis data.

Kata Kunci: *analisis, data histori ekonomi, algoritma k-means, Clustering*

A. PENDAHULUAN

Transformasi digital telah mendorong pemanfaatan data dalam mendukung pengambilan keputusan yang lebih tepat sasaran, termasuk dalam perumusan kebijakan ekonomi nasional (Martino, 2023). Indonesia sebagai negara berkembang dengan potensi ekonomi yang luas dan kompleks membutuhkan pendekatan analisis data yang mampu mengungkap pola tersembunyi di berbagai sektor strategis (Hilabi et al., 2025).

Salah satu metode yang efektif untuk mengelompokkan karakteristik data ekonomi yang beragam adalah klusterisasi, khususnya menggunakan algoritma K-Means. Metode ini bekerja dengan membagi data ke dalam beberapa klaster berdasarkan kemiripan sifat, sehingga memudahkan analisis dan penyusunan strategi kebijakan yang lebih terfokus (Reyhan et al., 2024).

Konsep klusterisasi data telah menjadi pilar dalam bidang *data mining*, *machine learning*, dan statistika multivariat selama beberapa dekade. Tujuan utama klusterisasi adalah mengorganisir objek data ke dalam kelompok-kelompok (klaster) sedemikian rupa sehingga anggota dalam satu klaster memiliki tingkat kemiripan yang tinggi, sementara anggota dari klaster yang berbeda memiliki tingkat kemiripan yang rendah (Ali & Masyufah, 2021).

K-Means adalah salah satu algoritma klusterisasi berbasis partisi yang paling sering digunakan dan diakui karena efisiensinya dalam mengelola data berukuran besar. Algoritma ini bekerja secara iteratif untuk mempartisi n observasi menjadi k klaster yang telah ditentukan sebelumnya. Setiap observasi akan ditetapkan ke klaster yang *centroid*-nya (pusat rata-rata klaster) memiliki jarak terdekat (Ahmed et al., 2020).

Proses K-Means dimulai dengan pemilihan k *centroid* awal, yang dapat dilakukan secara acak atau menggunakan metode inisialisasi yang lebih canggih

seperti K-Means++. Setelah inisialisasi, algoritma melalui dua langkah utama yang berulang: (Chen, 2024)

1. Langkah Penugasan (*Assignment Step*): Setiap titik data ditugaskan ke klaster yang *centroid*-nya paling dekat berdasarkan ukuran jarak *Euclidean* atau ukuran kemiripan lainnya.
2. Langkah Pembaruan (*Update Step*): *Centroid* setiap klaster dihitung ulang sebagai rata-rata dari semua titik data yang ditugaskan ke klaster tersebut. Proses iterasi ini berlanjut hingga *centroid* tidak lagi berpindah secara signifikan antar iterasi, atau hingga kriteria konvergensi tertentu terpenuhi (misalnya, jumlah iterasi maksimum).

Keunggulan K-Means terletak pada kecepatan komputasinya, menjadikannya pilihan ideal untuk *dataset* yang besar. Namun, K-Means rentan terhadap inisialisasi *centroid* awal dan cenderung membentuk klaster berbentuk bola.

Evaluasi kualitas hasil klusterisasi merupakan tahapan esensial untuk memastikan validitas dan interpretasi yang akurat dari kelompok yang terbentuk. Beberapa metrik evaluasi yang umum digunakan meliputi:

1. *Inertia*: Dikenal juga sebagai *within-cluster sum of squares* (WCSS), *inertia* mengukur seberapa padat klaster yang terbentuk. Nilai *inertia* adalah jumlah kuadrat jarak antara setiap titik data dan *centroid* klaster tempatnya berada. Semakin rendah nilai *inertia*, semakin kompak dan rapat klaster tersebut. Tujuan utama algoritma K-Means adalah meminimalkan nilai *inertia*.
2. *Silhouette score*: Metrik ini mengukur seberapa konsisten setiap titik data berada dalam klaster yang ditugaskan dibandingkan dengan klaster tetangga terdekatnya. Nilai *silhouette score* berkisar antara -1 hingga 1 . Nilai mendekati 1 menunjukkan klusterisasi yang sangat baik, di mana titik data sangat cocok dengan klaster sendiri dan jauh dari klaster lain. Nilai sekitar 0

menunjukkan tumpang tindih antar klaster, sementara nilai negatif menunjukkan bahwa titik data mungkin salah dikelompokkan.

3. *Davies-Bouldin Index*: Metrik ini mengukur rasio antara rata-rata jarak antar titik dalam klaster dan jarak antar *centroid* klaster. Nilai *Davies-Bouldin Index* yang lebih rendah mengindikasikan klasterisasi yang lebih baik, dengan klaster yang kompak di dalamnya dan terpisah dengan baik satu sama lain. Nilai 0 adalah yang terbaik, dan nilai di bawah 1 umumnya menunjukkan klasterisasi yang valid.
4. *Calinski-Harabasz Index*: Juga dikenal sebagai *Variance Ratio Criterion*, metrik ini mengukur rasio antara varians antar klaster dan varians dalam klaster. Nilai yang lebih tinggi menunjukkan klaster yang lebih baik.

Klasterisasi telah menjadi alat analisis yang sangat berharga dalam berbagai disiplin ilmu ekonomi, seperti segmentasi pasar, kinerja perusahaan, dan identifikasi pola perdagangan internasional (Chen, 2024; Hidayah, 2018).

1. Segmentasi Pasar dan Konsumen: Mengelompokkan konsumen berdasarkan perilaku pembelian, preferensi, atau demografi untuk strategi pemasaran yang lebih bertarget.
2. Analisis Kinerja Perusahaan: Pengelompokan perusahaan berdasarkan metrik keuangan (profitabilitas, likuiditas, solvabilitas) atau operasional untuk benchmarking atau identifikasi risiko.
3. Pengembangan Regional: Mengidentifikasi wilayah geografis dengan karakteristik sosio-ekonomi serupa (misalnya, tingkat pendapatan, pendidikan, atau industri dominan) untuk perumusan kebijakan pembangunan daerah.
4. Studi Perdagangan Internasional: Mengelompokkan negara berdasarkan pola ekspor-impor, struktur tarif, atau

perjanjian perdagangan untuk analisis hubungan bilateral atau multilateral.

5. Analisis Pasar Modal: Mengklaster saham berdasarkan volatilitas, kapitalisasi pasar, sektor industri, atau korelasi harga untuk tujuan diversifikasi portofolio dan manajemen risiko.
6. Sektor Sumber Daya dan Lingkungan: Mengelompokkan daerah berdasarkan tingkat pencemaran, konsumsi energi, atau produksi sumber daya untuk kebijakan keberlanjutan.
7. Sektor Pajak dan Fiskal: Mengelompokkan wajib pajak berdasarkan pola pembayaran atau profil pendapatan untuk analisis kepatuhan atau perumusan kebijakan pajak yang lebih adil.

Studi-studi ini menunjukkan fleksibilitas dan potensi klasterisasi dalam mengungkap struktur data ekonomi yang kompleks, yang pada gilirannya dapat memberikan *insight actionable* bagi pembuat kebijakan dan pelaku bisnis

Dalam konteks tersebut, lima sektor ekonomi utama yaitu pajak, ekspor, perikanan, pasar modal, dan sumber daya alam, menjadi objek kajian penting pada penelitian ini. Masing-masing memiliki peran vital dalam pertumbuhan dan stabilitas ekonomi Indonesia, namun perbedaan karakteristik antar tahun atau wilayah memerlukan segmentasi yang cermat. Penerapan algoritma K-Means diharapkan dapat mengelompokkan data secara objektif untuk mengungkap pola kinerja dan daya saing sektor-sektor tersebut.

Penelitian ini menggunakan data resmi dari BPS, Kementerian Keuangan, KKP, OJK, dan ESDM, dengan rentang waktu antara tahun 2000 hingga 2022. Evaluasi kualitas klaster dilakukan menggunakan *Inertia*, *Silhouette score*, dan *Davies-Bouldin Index*.

Penelitian ini bertujuan mengelompokkan pola data ekonomi Indonesia dari lima sektor: pajak, ekspor,

perikanan, pasar modal, dan sumber daya, menggunakan algoritma K-Means, serta memberikan interpretasi mendalam terhadap klaster-klaster yang terbentuk pada masing-masing sektor ekonomi, dengan tujuan menghasilkan *insight* yang dapat mendukung perumusan kebijakan ekonomi berbasis data yang lebih adaptif dan tepat sasaran di Indonesia.

B. METODE PENELITIAN

Metodologi penelitian ini dirancang secara sistematis untuk mengaplikasikan algoritma K-Means pada berbagai *dataset* ekonomi Indonesia dan melakukan evaluasi komprehensif terhadap hasil klusterisasinya.

Sumber Data

Penelitian ini menggunakan *dataset* dari lima sektor ekonomi utama di Indonesia, yang dikumpulkan dari sumber-sumber resmi terkemuka untuk menjamin validitas dan reliabilitas data. Rentang waktu data bervariasi tergantung pada ketersediaan data pada masing-masing sumber:

1. Data Penerimaan Pajak dan Pendapatan Negara

Data ini mencakup agregasi penerimaan pajak dari berbagai jenis (misalnya, Pajak Penghasilan, Pajak Pertambahan Nilai) dan komponen pendapatan negara non-pajak. Data historis dari tahun 2000-2022 digunakan untuk mengidentifikasi tren dan segmen kinerja fiskal. Sumber data utama adalah publikasi Kementerian Keuangan Republik Indonesia dan Badan Pusat Statistik (BPS).

Tabel 1. Data Penerimaan Pajak dan Pendapatan Negara

Tahun	Pendapatan Negara (dalam miliar rupiah)	Penerimaan Perpajakan (dalam miliar rupiah)	%
2000	205.334,5	115.912,5	56,5
2001	301.077,7	185.540,9	61,6
2002	298.605,3	210.383,8	70,5

Tahun	Pendapatan Negara (dalam miliar rupiah)	Penerimaan Perpajakan (dalam miliar rupiah)	%
2003	341.396,1	241.626,8	70,8
2004	403.769,6	279.207,5	69,2
2005	443.786,7	319.440,5	72,0
2006	659.115,2	425.053,0	64,5
2007	694.087,9	490.010,9	70,6
2008	781.354,1	591.978,4	75,7
2009	848.763,2	619.922,2	73,04
2010	995.271,5	723.306,7	72,67
2011	1.210.599,7	873.873,9	72,19
2012	1.338.109,6	980.518,1	73,28
2013	1.438.891,1	1.077.306,7	74,87
2014	1.550.490,8	1.146.865,8	73,96
2015	1.508.020,3	1.240.418,8	82,25
2016	1.786.225,0	1.539.166,2	86,16

2. Data Sektor Sumber Daya Alam

Data ini mencakup nilai ekspor sektor sumber daya alam dari Indonesia pada periode krusial tahun 1998 hingga 2000, yang meliputi komoditas utama seperti batu bara, minyak bumi, gas alam, dan hasil perkebunan strategis seperti kelapa sawit. Tujuannya adalah untuk mengidentifikasi pola klusterisasi dalam dinamika ekspor selama masa transisi ekonomi pasca-krisis Asia. Hasil klusterisasi menunjukkan adanya pemisahan yang cukup jelas antar tahun, berdasarkan perbedaan signifikan dalam performa ekspor tiap tahunnya. Nilai evaluasi seperti *silhouette score* dan *Davies-Bouldin Index* menunjukkan hasil klusterisasi yang cukup baik untuk skala *dataset* kecil. Sumber data berasal dari UN Comtrade, BPS, serta laporan resmi dari Kementerian ESDM dan Kementerian Perdagangan Republik Indonesia.

Tabel 2. Tabel Data Sektor Sumber Daya Alam

Tahun	Total Pendapatan (Miliar USD)	Bagian Pemerintah (Miliar USD)	Dari Lepas Pantai (Miliar USD)
1998	11.71	7.45	2.61
1999	14.23	9.14	3.75
2000	12.21	7.69	3
	Total Produksi Gas (TSCF)	Produksi Gas Offshore	% Offshore
1998	2.21	0.77	35

Tahun	Produksi Timah (metrik ton)	Perkiraan Produksi Lepas Pantai (Metrik ton)	Gas dijual (MMSCF)	
1999	2.61	1.07	41	
2000	2.38	0.93	39	
1998	48887	34221	394.74	
1999	47753	33427	503.86	
2000	53959	37771	542.37	
Tahun	Gas (MMSCF)	Lift	Gas Ekspor (MMSCF)	Gas Loss (MMSCF)
1998	122.83	1.69	133.49	
1999	122.53	1818.08	122.65	
2000	131.95	1.58	99.62	
Tahun	Gas (MMSCF)	Gas Diproses (MMSCF)	Pemakaian Sendiri (MMSCF)	
1998	6.07	305.05		
1999	9.05	212.04		
2000	10.47	219.31		

3. Data Sektor Perikanan

Data ini mencakup nilai ekspor produk perikanan Indonesia ke Tiongkok dan total impor produk perikanan Tiongkok dari dunia, dalam periode 2001 hingga 2018. Data ini digunakan untuk mengukur daya saing ekspor Indonesia menggunakan pendekatan NRCA (*Normalized Revealed Comparative Advantage*) berdasarkan perbandingan nilai ekspor antar negara dan total pasar global. Hasil analisis mencerminkan tren daya saing Indonesia di pasar Tiongkok dari waktu ke waktu, termasuk fluktuasi nilai ekspor yang terjadi. Sumber data berasal dari UN Comtrade, *International Trade Centre* (ITC), dan Badan Pusat Statistik (BPS).

Tabel 3. Data Sektor Perikanan

Tahun	Impor Cina dari Indonesia	Perikanan	Impor Cina dari Dunia	Perikanan
2001	3.887.887	23.942	243.552.881	1.330.929
2002	4.508.435	17.518	295.170.104	1.564.594
2003	5.746.971	24.710	412.759.796	1.864.899
2004	7.215.671	34.166	561.228.748	2.339.944
2005	8.436.960	40.652	659.952.762	2.879.068
2006	9.605.743	43.348	791.460.868	3.155.136
2007	12.464.207	40.077	956.115.448	3.442.820
2008	14.322.935	49.989	1.132.562.200	3.648.126
2009	13.663.783	42.284	1.005.555.200	3.604.804
2010	20.795.189	67.829	1.396.001.001	4.365.460

Tahun	Impor Cina dari Indonesia	Perikanan	Impor Cina dari Dunia	Perikanan
2011	31.337.084	88.130	1.743.394.904	5.587.562
2012	31.935.986	156.851	1.818.199.200	5.488.977
2013	31.424.279	211.440	1.949.992.300	5.993.493
2014	24.485.247	221.402	1.959.234.625	6.583.348
2015	19.886.914	278.161	1.679.564.325	6.328.008
2016	21.414.036	320.170	1.587.290.688	6.917.637
2017	28.574.306	344.368	1.843.792.939	8.070.777
2018	34.154.685	543.402	2.134.987.265	11.605.738

4. Data Perkembangan Pasar Modal

Dataset ini melibatkan informasi terkait kinerja perusahaan yang terdaftar di Bursa Efek Indonesia (BEI), termasuk harga saham, volume transaksi, kapitalisasi pasar, dan kinerja sektor industri. Data juga mencakup total kapitalisasi pasar dana investasi. Periode data dari tahun 2000-2022. Sumber data utama adalah Otoritas Jasa Keuangan (OJK) dan BEI.

Tabel 4. Data Perkembangan Pasar Modal

Tahun	Perusahaan Saham	Kapitalisasi Dana
1977	1	263.260
1978	1	330.260
1979	4	7.058.116
1980	6	14.588.116
1981	8	19.788.128
1982	13	39.948.205
1983	23	48.055.208
1984	24	57.498.184
1985	24	57.827.872
1986	24	58.349.872
1987	24	58.569.311
1988	24	72.844.043
1989	56	432.8339.874
1990	122	1.779.936.594
1991	139	3.729.841.279
1992	153	6.253.196.082
1993	172	9.787.393.323
1994	217	23.854.393.821
1995	229	31.477.012.164

5. Data Ekspor Alas Kaki ke Negara Mitra Dagang

Data ini merupakan hasil perbandingan nilai ekspor industri alas kaki dari Indonesia dan Vietnam ke sejumlah negara mitra strategis, seperti Amerika Serikat, Jepang, Tiongkok, Jerman, serta negara-negara Eropa lainnya. Analisis difokuskan pada perbandingan pangsa ekspor antara

kedua negara untuk memahami posisi daya saing masing-masing di pasar internasional. Nilai ekspor telah dinormalisasi (*scaled*) agar skala antar negara menjadi sebanding, memungkinkan analisis komparatif yang objektif. Data ini dianalisis menggunakan metode klusterisasi untuk mengelompokkan negara tujuan ekspor berdasarkan kesamaan karakteristik volume ekspor dan pangsa pasar antara Indonesia dan Vietnam. Sumber data diperoleh dari publikasi resmi seperti Badan Pusat Statistik (BPS), UN Comtrade, dan Kementerian Perdagangan Republik Indonesia.

Tabel 5. Data Ekspor Alas Kaki ke Negara Mitra

Rank	Negara	Nilai Ekspor		Share		Pertumbuhan Ekspor 2012-2013	
1	Amerika Serikat	1077	3057	26.8	26,5	16	22
2	Jerman	261	1194	6.8	10.3	3	12
3	Perancis	69	643	1.8	5.6	8	6
4	Inggris	221	564	5.7	4.9	-3	6
5	Italy	138	300	3.6	2.6	-6	13
6	Jepang	216	481	5	4.2	23	6
7	Russia	75	272	1.9	2.4	25	13
8	Belanda	194	264	5	2.3	0	5
9	Belgia	2976	602	7.7	5.2	-2	21
10	Korea Selatan	123	377	3.2	3.3	57	24
11	Tiongkok	377	485	3.7	4.2	15	12

Pra-Pemrosesan Data (Preprocessing)

Kualitas data sangat mempengaruhi hasil klusterisasi. Oleh karena itu, setiap *dataset* melalui tahapan pra-pemrosesan yang ketat:

1. Pembersihan Data: Identifikasi dan penanganan nilai-nilai yang hilang (*missing values*) dilakukan. Strategi yang digunakan dapat berupa imputasi (misalnya, menggunakan nilai rata-rata, median, atau regresi) atau penghapusan baris/kolom yang memiliki terlalu banyak nilai hilang. Penanganan outliers juga dipertimbangkan; pada data ekonomi, outliers seringkali merupakan titik data yang signifikan dan tidak selalu dihapus melainkan

dianalisis secara terpisah atau ditransformasi.

2. Normalisasi Data: Fitur-fitur dalam *dataset* ekonomi sering memiliki skala yang sangat berbeda, yang dapat membiaskan perhitungan jarak K-Means. Oleh karena itu, normalisasi (*scaling*) diterapkan untuk memastikan bahwa semua fitur berkontribusi secara proporsional terhadap perhitungan jarak. Metode yang umum digunakan adalah Min-Max Scaling (menskalakan nilai ke rentang 0–1) atau Standardization (*Z-score scaling*, mengubah data menjadi rata-rata 0 dan standar deviasi 1).

Implementasi Algoritma K-Means

Setelah pra-pemrosesan, algoritma K-Means diterapkan pada setiap *dataset* secara independen. Penentuan jumlah kluster optimal (*k*) adalah langkah krusial. Dalam penelitian ini, *k* optimal ditentukan dengan kombinasi metode:

- a. Metode *Elbow*: Mengidentifikasi titik "siku" pada grafik *inertia* terhadap jumlah kluster yang berbeda. Titik ini menunjukkan *k* di mana penambahan kluster tidak lagi menghasilkan penurunan *inertia* yang signifikan, mengindikasikan *k* yang paling efisien.
- b. *Silhouette Analysis*: Memilih nilai *k* yang menghasilkan *silhouette score* rata-rata tertinggi. Ini memberikan pandangan tentang seberapa baik titik data cocok dengan kluster mereka sendiri dan seberapa jelas kluster terpisah satu sama lain.

Nilai *k* yang digunakan dalam analisis masing-masing sektor akan disebutkan secara spesifik pada bagian hasil. Proses K-Means diimplementasikan menggunakan pustaka *scikit-learn* di Python, memastikan efisiensi komputasi dan konsistensi hasil.

Evaluasi Klusterisasi

Kualitas klusterisasi dievaluasi menggunakan tiga metrik utama yang telah dijelaskan di Tinjauan Pustaka:

- Inertia*: Nilai *inertia* akhir akan dilaporkan untuk setiap model klasterisasi.
- Silhouette score*: Nilai rata-rata *silhouette score* akan dilaporkan untuk setiap model klasterisasi.
- Davies-Bouldin Index*: Nilai *Davies-Bouldin Index* akan dilaporkan untuk setiap model klasterisasi.

Interpretasi dari nilai-nilai metrik ini akan menjadi dasar objektif untuk menilai kualitas pembentukan klaster pada setiap sektor.

Hasil evaluasi ini konsisten dengan temuan sebelumnya yang menyatakan bahwa klasterisasi berbasis metrik *silhouette* dan DBI sangat efektif untuk segmentasi ekonomi berbasis waktu (Hasan, 2024).

Visualisasi Hasil

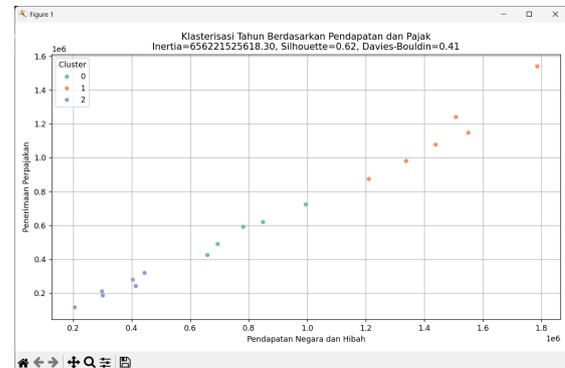
Visualisasi adalah kunci untuk memahami struktur dan distribusi klaster. Untuk *dataset* dengan dimensi tinggi (lebih dari dua fitur), *Principal Component Analysis* (PCA) digunakan untuk mereduksi dimensi data menjadi dua komponen utama (Komponen Utama 1 dan Komponen Utama 2) yang menangkap varians data terbesar. Hasil klasterisasi kemudian diplot dalam ruang dua dimensi ini, dengan setiap titik data diberi warna sesuai dengan klaster keanggotaannya. Jika *dataset* secara intrinsik sudah memiliki dua fitur utama atau kurang, plot langsung antara fitur-fitur tersebut dilakukan. Visualisasi ini membantu dalam interpretasi visual mengenai pemisahan dan karakteristik klaster.

C. HASIL DAN PEMBAHASAN

Hasil Penerapan data mining Klasterisasi Penerimaan Pajak dan Pendapatan Negara

Penerapan algoritma K-Means pada data penerimaan pajak dan pendapatan negara Indonesia dari tahun 2000 hingga

2022 menghasilkan klasterisasi yang jelas, dengan jumlah klaster optimal $k=3$. 454545, *Inertia*: 656,221,525,618.30, *Silhouette score*: 0.62, dan *Davies-Bouldin Index*: 0.41



Gambar 1. Grafik Scatterplot klasterisasi penerimaan pajak dan pendapatan Negara

```
=== Evaluasi Model Klasterisasi ===
Inertia           : 656221525618.30
Silhouette Score  : 0.62
Davies-Bouldin Index : 0.41
```

Gambar 2. Terminal Hasil Program Python

Dengan *silhouette score* 0.62 dan *Davies-Bouldin Index* 0.41, hasil ini mengindikasikan kualitas klasterisasi yang baik, ditandai dengan separasi yang memadai antar klaster dan kekompakan internal klaster. Visualisasi klaster (Gambar 1) secara efektif memisahkan tahun-tahun berdasarkan besaran "Penerimaan Perpajakan" dan "Pendapatan Negara dan Hibah". Tiga kelompok utama teridentifikasi:

- Klaster 0 (Biru): Mewakili periode awal pengamatan (tahun-tahun di awal 2000-an) di mana penerimaan pajak dan pendapatan negara relatif rendah. Periode ini mungkin mencerminkan tahap awal pemulihan ekonomi atau kebijakan fiskal yang lebih konservatif.
- Klaster 1 (Oranye): Menunjukkan periode menengah dengan peningkatan signifikan dalam penerimaan pajak dan pendapatan negara. Tahun-tahun dalam klaster ini mungkin bertepatan dengan pertumbuhan ekonomi yang stabil dan

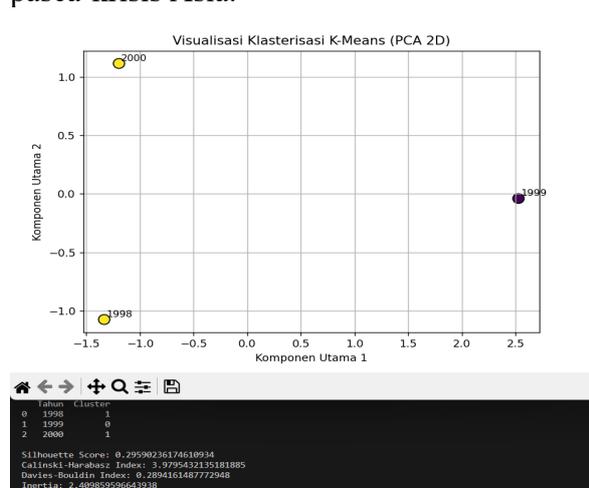
reformasi perpajakan yang mulai memberikan dampak.

3. **Klaster 2 (Merah):** Terletak di bagian kanan atas plot, klaster ini secara jelas merepresentasikan tahun-tahun dengan penerimaan pajak dan pendapatan negara yang sangat tinggi. Periode ini kemungkinan besar didorong oleh pertumbuhan ekonomi yang pesat, harga komoditas yang tinggi, dan efektivitas kebijakan penerimaan negara.

Pola klasterisasi ini sangat relevan untuk analisis fiskal, memungkinkan pemerintah untuk mengidentifikasi faktor-faktor pendorong atau penghambat kinerja penerimaan negara pada berbagai periode waktu. Misalnya, korelasi antara klaster pendapatan tinggi dengan kebijakan ekonomi tertentu dapat memberikan *insight* untuk perumusan kebijakan fiskal di masa depan.

Klasterisasi Sumber Daya Alam

Klasterisasi pada data ekspor sumber daya alam berdasarkan tahun (1998, 1999, dan 2000), yang mencakup pola dan besaran ekspor, menghasilkan segmentasi menarik yang mencerminkan dinamika perdagangan pada masa transisi ekonomi pasca-krisis Asia.



Gambar 3. Grafik Scatterplot terhadap data ekspor sumber daya alam (PCA 2D)

Proses klasterisasi dilakukan menggunakan algoritma K-Means dan divisualisasikan dalam ruang 2 dimensi menggunakan metode PCA (*Principal Component Analysis*). Setiap titik dalam visualisasi merepresentasikan satu tahun, dengan posisinya mencerminkan karakteristik ekspor pada tahun tersebut. Nilai *Inertia*: 2.41, *Silhouette score*: 0.296, *Davies-Bouldin Index*: 0.29, dan *Calinski-Harabasz Index*: 3.98

Nilai *silhouette score* sebesar 0.296 menunjukkan bahwa pemisahan antar klaster cukup baik, meskipun ada sedikit tumpang tindih antar data. Angka ini masih dalam rentang wajar untuk *dataset* kecil dan kompleks seperti ini. *Davies-Bouldin Index* yang rendah (0.29) menunjukkan bahwa klaster-klaster yang terbentuk cukup kompak dan saling berjauhan, yang merupakan indikasi klasterisasi yang baik. Sementara itu, *Calinski-Harabasz Index* sebesar 3.98 mengindikasikan bahwa struktur klaster yang terbentuk masih layak dan memiliki validitas internal, walaupun tidak sekuat hasil pada *dataset* yang lebih besar atau lebih homogen. Nilai *inertia* yang cukup rendah (2.41) juga mengindikasikan bahwa masing-masing data point tidak terlalu jauh dari pusat klasternya, menandakan kepadatan klaster yang cukup baik.

Berdasarkan hasil klasterisasi dan visualisasi (Gambar 3), terdapat dua klaster utama yang terbentuk:

1. **Klaster 0 (Tahun 1999):**

Tahun 1999 muncul sebagai satu-satunya anggota klaster ini, menunjukkan bahwa pola ekspor pada tahun tersebut berbeda signifikan dibandingkan dengan tahun-tahun lainnya. Hal ini mungkin terkait dengan stabilisasi ekonomi pasca krisis 1998, di mana Indonesia mulai menunjukkan pemulihan perdagangan atau penerapan kebijakan ekspor baru yang berpengaruh signifikan terhadap mitra dagang dan struktur ekspor nasional.

2. Klaster 1 (Tahun 1998 dan 2000):

Tahun 1998 dan 2000 berada dalam satu klaster, menandakan bahwa meskipun tahun 1998 merupakan masa krisis ekonomi besar dan 2000 adalah masa pemulihan awal, pola ekspor yang terjadi memiliki kemiripan, baik dalam hal negara tujuan utama, volume ekspor, maupun komposisi produk ekspor. Kemungkinan besar, pada tahun 2000 Indonesia masih meneruskan atau memulihkan pola perdagangan yang mirip dengan masa sebelum krisis, sehingga pola ekspornya menunjukkan kedekatan dengan tahun 1998.

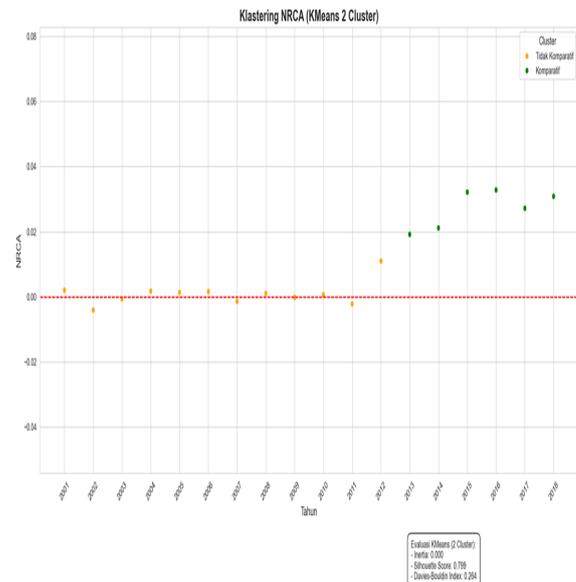
Visualisasi pada gambar 3 memperlihatkan dengan jelas bagaimana data tahun 1999 secara signifikan terpisah dari dua tahun lainnya berdasarkan dua komponen utama. Ini dapat diinterpretasikan sebagai representasi dari pergeseran pola perdagangan yang cukup besar, misalnya akibat perubahan mitra dagang dominan, pergeseran struktur industri ekspor, atau transisi kebijakan perdagangan luar negeri yang terjadi di akhir dekade 1990-an.

Hasil klasterisasi ini memberikan wawasan penting bagi analisis makroekonomi dan perumusan kebijakan perdagangan. Klasterisasi berdasarkan tahun memungkinkan identifikasi momen penting dalam sejarah perdagangan Indonesia, serta menjadi dasar bagi pembuat kebijakan untuk memahami perubahan struktur ekspor dari waktu ke waktu. Temuan ini juga dapat digunakan dalam merumuskan strategi diplomasi ekonomi, peningkatan daya saing produk ekspor, dan diversifikasi pasar tujuan ekspor untuk mengantisipasi ketergantungan terhadap negara atau sektor tertentu.

Klasterisasi Sektor Perikanan

Penerapan algoritma K-Means pada data *Normalized Revealed Comparative Advantage* (NRCA) bertujuan untuk

mengelompokkan tahun-tahun berdasarkan daya saing sektor perikanan Indonesia dalam periode 2001–2018. Hasil klasterisasi optimal menunjukkan $k = 2$ klaster, nilai *Inertia*: 0.000, *Silhouette score*: 0.799, dan *Davies-Bouldin Index*: 0.264



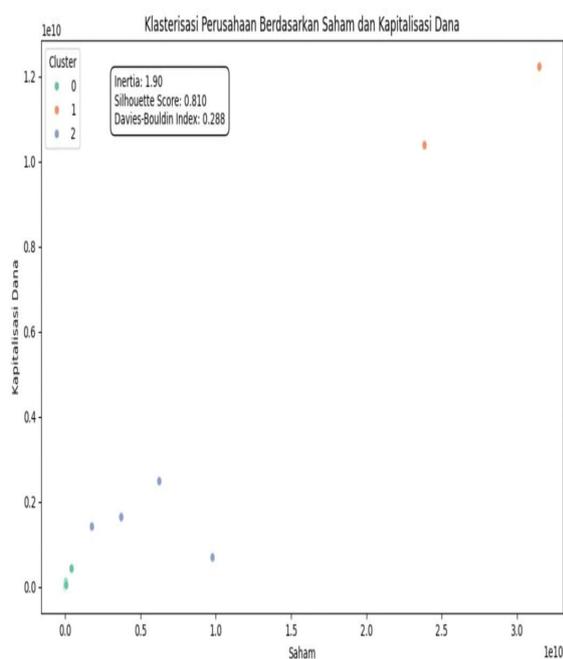
Gambar 4. Grafik Scatter plot time series hasil klasterisasi K-Means terhadap NRCA (2001–2018).

Nilai *silhouette score* yang tinggi dan *Davies-Bouldin Index* yang rendah menunjukkan pemisahan klaster yang sangat baik dan jelas. Klaster terbagi menjadi:

1. Klaster Tidak Komparatif (warna oranye): Periode 2001–2012 umumnya memiliki nilai NRCA mendekati atau di bawah nol, menandakan bahwa sektor perikanan pada periode tersebut belum menunjukkan keunggulan komparatif secara global.
2. Klaster Komparatif (warna hijau): Periode 2013–2018 menunjukkan peningkatan signifikan dalam nilai NRCA, yang konsisten berada di atas nol. Hal ini mencerminkan bahwa sektor perikanan Indonesia mulai memiliki daya saing internasional dan kontribusi ekspor yang lebih kuat.

Hasil ini menunjukkan adanya perubahan struktural dalam daya saing sektor perikanan Indonesia pasca 2012, yang dapat disebabkan oleh peningkatan investasi, kebijakan maritim yang lebih strategis, atau penguatan infrastruktur ekspor. Klasterisasi ini dapat menjadi dasar penting bagi pemerintah dalam merumuskan kebijakan perikanan berorientasi ekspor dan penguatan posisi Indonesia di pasar global.

Klasterisasi Perkembangan Pasar Modal



Gambar 5. Grafik Scatter plot klasterisasi perusahaan berdasarkan saham dan kapitalisasi dana menggunakan K-Means (3 klaster).

Analisis klaster pada data saham dan kapitalisasi dana perusahaan di pasar modal Indonesia memberikan *insight* yang sangat tajam mengenai profil perusahaan dan segmentasi investor. Menghasilkan nilai *Inertia*: 1.90, *Silhouette score*: 0.810, dan *Davies-Bouldin Index*: 0.288 .

Kualitas klasterisasi ini sangat tinggi, sebagaimana ditunjukkan oleh *silhouette score* yang luar biasa (0.810) dan *Davies-Bouldin Index* yang sangat rendah (0.288). Ini mengindikasikan klaster yang sangat

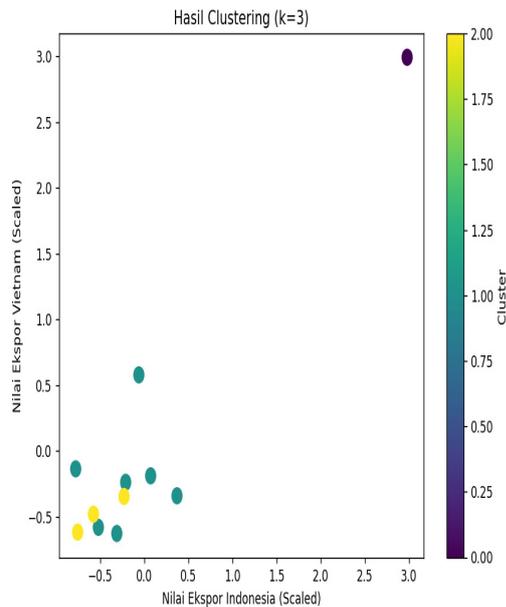
kompak dan terpisah dengan jelas. Visualisasi gambar di atas secara definitif memisahkan perusahaan berdasarkan "Saham" dan "Kapitalisasi Dana" menjadi tiga klaster utama:

1. **Klaster 1 (Oranye):** Klaster ini menonjolkan perusahaan dengan nilai saham dan kapitalisasi dana yang sangat tinggi. Ini kemungkinan besar adalah kelompok perusahaan *blue-chip* atau emiten dengan kapitalisasi pasar terbesar di BEI, yang memiliki stabilitas tinggi dan sering menjadi pilihan utama investor institusional.
2. **Klaster 0 (Biru):** Mewakili kelompok perusahaan dengan kapitalisasi pasar menengah, mungkin dengan volume transaksi yang moderat. Perusahaan-perusahaan ini mungkin masih dalam tahap pertumbuhan atau memiliki pangsa pasar yang signifikan di sektor spesifik.
3. **Klaster 2 (Hijau):** Berisi perusahaan dengan kapitalisasi pasar dan nilai saham yang relatif lebih rendah. Kelompok ini bisa mencakup perusahaan-perusahaan kecil, *startup*, atau perusahaan yang sedang berkembang, yang mungkin menarik bagi investor dengan profil risiko lebih tinggi.

Identifikasi klaster ini sangat bermanfaat bagi investor dalam diversifikasi portofolio berdasarkan profil risiko dan target pengembalian, serta bagi regulator (OJK) untuk memantau stabilitas pasar, mengidentifikasi anomali, dan merumuskan kebijakan yang mendukung pertumbuhan segmen pasar modal yang berbeda.

Klasterisasi Data Ekspor Alas Kaki Antara Indonesia dan Vietnam

Penerapan klasterisasi pada data ekspor alas kaki bertujuan untuk mengelompokkan negara-negara tujuan ekspor berdasarkan kesamaan karakteristik nilai ekspor Indonesia dan Vietnam. Dari hasil analisis, didapatkan bahwa klasterisasi optimal terbentuk pada $k = 3$.



Gambar 6. Grafik Scatter plot klasterisasi ekspor Indonesia dan Vietnam (data diskalakan), menggunakan K-Means (3 klaster).

```
Evaluasi untuk k=3:  
Inertia : 7.00 (lebih rendah lebih baik)  
Silhouette Score : 0.35  
Davies-Bouldin Index : 0.59 (lebih rendah lebih baik)  
Calinski-Harabasz Index : 27.43 (lebih tinggi lebih baik)  
Negara Cluster  
0 Amerika Serikat 0  
1 Jerman 1  
2 Perancis 1  
3 Inggris 1  
4 Italy 1  
5 Jepang 2  
6 Russia 2  
7 Belanda 1  
8 Belgia 1  
9 Korea Selatan 2  
10 Tiongkok 1
```

Gambar 7. Output terminal klasterisasi K-Means (k=3) yang menampilkan pembagian negara dan metrik evaluasi model.

Hasil Evaluasi Klasterisasi:

1. *Inertia*: 7.00 (Semakin rendah semakin baik, menunjukkan sebaran data dalam klaster cukup rapat)
2. *Silhouette score*: 0.35 (Nilai sedang, menunjukkan pemisahan antar klaster

cukup baik namun masih bisa ditingkatkan)

3. *Davies-Bouldin Index*: 0.59 (Cukup baik, semakin kecil menunjukkan klaster semakin terpisah)
4. *Calinski-Harabasz Index*: 27.43 (Nilai yang baik, semakin tinggi menunjukkan klaster lebih terpisah dengan variansi yang kecil dalam klaster)

Interpretasi Klaster:

Berdasarkan hasil klasterisasi dan visualisasi scatter plot, berikut adalah segmentasi negara tujuan ekspor alas kaki:

1. **Klaster 0 (Negara Utama dengan Nilai Ekspor Tertinggi): Amerika Serikat**
Negara ini secara signifikan memiliki nilai ekspor yang jauh lebih tinggi dibanding negara lain, menjadi pasar utama dan prioritas dalam ekspor alas kaki Indonesia.
2. **Klaster 1 (Negara Pasar Menengah dan Stabil):**
Jerman, Perancis, Inggris, Italy, Jepang, Belanda, Belgia, Tiongkok
Negara-negara dalam klaster ini memiliki nilai ekspor yang cukup stabil dan cukup besar, baik dari sisi Indonesia maupun Vietnam, tetapi tidak sebesar Amerika Serikat. Klaster ini menjadi target penting untuk mempertahankan dan meningkatkan volume ekspor.
3. **Klaster 2 (Negara Potensial Berkembang):**
Russia, Korea Selatan
Negara-negara ini memiliki nilai ekspor relatif lebih kecil, namun pertumbuhan ekspornya signifikan dan berpotensi menjadi pasar ekspor yang lebih besar di masa depan.

Implikasi Kebijakan:

Hasil klasterisasi ini dapat dimanfaatkan oleh pelaku industri dan pemerintah dalam: Menentukan prioritas negara tujuan ekspor, dan merancang strategi penetrasi pasar yang sesuai untuk masing-masing klaster.

D. PENUTUP

Penelitian ini telah berhasil menerapkan algoritma K-Means untuk mengelompokkan data ekonomi dari lima sektor strategis di Indonesia: penerimaan pajak, kinerja ekspor, sektor perikanan, perkembangan pasar modal, dan sektor sumber daya alam. Penerapan K-Means terbukti efektif dalam mengungkap struktur dan pola tersembunyi yang bermakna dalam berbagai dataset ekonomi makro. Evaluasi menggunakan metrik *inertia*, *silhouette score*, dan *Davies-Bouldin Index* mengkonfirmasi variasi kualitas klusterisasi antar sektor, dengan kinerja yang sangat baik pada sektor penerimaan pajak dan pasar modal, dan kualitas yang memadai pada sektor ekspor.

Pola-pola yang teridentifikasi melalui klusterisasi memberikan *insight* yang berharga, antara lain: Identifikasi periode-periode dengan tingkat penerimaan pajak dan pendapatan negara yang berbeda, mencerminkan evolusi kinerja fiskal. Pengelompokan perusahaan di pasar modal berdasarkan kapitalisasi dan kinerja saham, esensial untuk segmentasi investasi dan regulasi. Analisis keunggulan komparatif Indonesia dalam sektor ekspor, menunjukkan area kekuatan dan potensi pengembangan. Segmentasi negara mitra dagang berdasarkan pola ekspor, mendukung strategi perdagangan bilateral dan multilateral. Identifikasi potensi klusterisasi di sektor perikanan dan sumber daya alam untuk pengelolaan dan pengembangan yang lebih terarah.

Temuan penelitian ini secara kuat mendukung gagasan bahwa analisis klusterisasi merupakan alat yang ampuh untuk memahami dinamika ekonomi dan mendukung pengambilan keputusan berbasis data yang lebih informatif di Indonesia.

Berdasarkan hasil dan keterbatasan yang ditemukan dalam penelitian ini, beberapa saran untuk penelitian selanjutnya yaitu agar melakukan eksplorasi Algoritma

Klusterisasi Alternatif: Untuk mengatasi keterbatasan K-Means (seperti sensitivitas terhadap outliers atau bentuk kluster non-bola), disarankan untuk mencoba algoritma klusterisasi lain seperti DBSCAN (yang efektif untuk kluster berbasis densitas) atau Gaussian Mixture Model (GMM) yang memungkinkan keanggotaan probabilistik kluster. Perbandingan hasil dari berbagai algoritma dapat memberikan validasi dan perspektif yang lebih komprehensif.

Mengintegrasikan lebih banyak variabel ekonomi yang relevan (misalnya, inflasi, suku bunga, tingkat pengangguran, PDB per kapita, kebijakan moneter, atau indikator sosial) dapat memperkaya model klusterisasi dan menghasilkan kluster yang lebih nuansa serta informatif.

Untuk dataset berbasis waktu (seperti data pajak, ekspor, atau pasar modal), penerapan metode klusterisasi deret waktu atau analisis tren evolusi kluster dari waktu ke waktu dapat memberikan *insight* dinamis tentang bagaimana segmen-segmen ekonomi berubah seiring berjalannya waktu.

Jika memungkinkan, validasi hasil klusterisasi dengan data eksternal yang independen atau melalui wawancara dengan pakar ekonomi dan sektoral dapat meningkatkan kepercayaan terhadap temuan. Analisis kualitatif lebih lanjut tentang karakteristik spesifik masing-masing kluster dapat memberikan konteks dan makna yang lebih dalam.

Melakukan pencarian hiperparameter yang lebih sistematis (misalnya, grid search dengan validasi silang atau randomized search) untuk menentukan jumlah kluster optimal (k) dan parameter algoritma lainnya dapat lebih lanjut meningkatkan kinerja dan stabilitas klusterisasi.

Penelitian selanjutnya dapat berfokus pada bagaimana hasil klusterisasi dapat secara langsung menginformasikan dan memprediksi dampak dari kebijakan ekonomi tertentu di setiap segmen kluster.

E. DAFTAR PUSTAKA

- Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The K-Means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics*, 9(8), 1295. <https://doi.org/10.3390/electronics9081295>
- Ali, A., & Masyfufah, L. (2021). Klasterisasi Pasien BPJS Dengan Metode K-Means Clustering Guna Menunjang Program Jaminan Kesehatan Nasional di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo. *Jurnal Wiyata: Penelitian Sains Dan Kesehatan*, 8(1), 8–22. <https://doi.org/10.56710/wiyata.v8i1.427>
- Chen, Q. (2024). Application of K-Means Algorithm in Marketing. *Advances in Economics, Management and Political Sciences*, 71, 178–184. <https://doi.org/10.54254/2754-1169/71/20241485>
- Goldstein, I., Jiang, W., & Karolyi, G. A. (2019). To FinTech and beyond. *The Review of Financial Studies*, 32(5), 1647–1661.
- Hasan, Y. (2024). Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster K-Means dan DBSCAN. *Kakifikom: Kumpulan Artikel Kaya Ilmiah Fakultas Ilmu Komputer*, 6(1), 60–74. <https://ejournal.ust.ac.id/index.php/KA-KIFIKOM/article/view/3938>
- Hidayah, A. (2018). Implementing Data Clustering to Identify Capital Allocation for Small and Medium Sized Enterprises (SMEs). *ASEAN Marketing Journal*, 10(1), 66–74. <https://doi.org/10.21002/amj.v10i1.10627>
- Hilabi, S. S., Savina, S., & Khairunisa, S. (2025). Pemanfaatan Data Analitik dalam Big Data: Studi Kasus Implementasi di Pemerintahan. *JATISI: Jurnal Teknik Informatika Dan Sistem Informasi*, 12(1), 378–390. <https://doi.org/10.35957/jatisi.v12i1.10916>
- Martino. (2023). Pemanfaatan Kecerdasan Buatan Dalam Siklus Kebijakan Publik: Antara Peluang dan Tantangan. *Prosiding Seminar Nasional Amikom Surakarta*, 54–68. <https://ojs.amikomsolo.ac.id/index.php/semnasa/article/view/129>
- Reyhan, M., Ahmad, D. R., Ramadhan, N. A., Hidayat, R., & Kusumasari, I. R. (2024). Penggunaan Data Analisis dan Big Data dalam Strategi Pengambilan Keputusan Keuangan. *Jurnal Akuntansi, Manajemen Dan Perencanaan Kebijakan*, 2(2), 1–9. <https://doi.org/10.47134/jampk.v2i2.540>
-