
PENERAPAN ALGORITMA *K-MEANS CLUSTERING* UNTUK ANALISIS POLA DATA EKONOMI HISTORIS

Abed Neco¹⁾, Firman Aziz Saputra²⁾, Nazar Fadhil Abdullah³⁾, Rizky Ramadhani⁴⁾,
Testarina Tatiana Hermansyah⁵⁾, 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

Historical economic and financial data are available in vast volumes, yet extracting non-trivial insights hidden within them remains a significant challenge, primarily due to the reliance on traditional, hypothesis-driven analysis methods. In the Indonesian context, the comprehensive application of clustering techniques to uncover objective data narratives remains unexplored, mainly raising the urgency of developing a data-driven approach. This study aims to address this gap by demonstrating the capabilities and flexibility of the K-Means algorithm as a robust exploratory analysis method. The study employs a comparative case study approach on five independent datasets purposefully selected to cover diverse domains and periods: bank merger trends (1971–1988), critical macroeconomic indicators (1992–2003), state-owned bank financial performance (2004–2014), bird's nest exports (2017–2021), and comparable economic data from the United States (1930–1955). Methodologically, each dataset was rigorously pre-processed before being clustered using the K-Means algorithm, with the quality of the results quantitatively evaluated using the Silhouette Score, Davies-Bouldin Index, and Inertia metrics. The results demonstrate powerful clustering performance, with three of the five case studies achieving Silhouette Scores above 0.70, indicating dense and well-defined data segmentation. Key findings demonstrate that the formed clusters successfully map historical periods objectively; for example, the algorithm automatically isolates the extreme anomaly of the 1998 monetary crisis as a unique cluster, identifies the peak of the banking merger era as a phase of intense consolidation, and groups state-owned banks into distinct strategic segments based on their capital and profitability profiles. This study confirms that K-Means is an effective exploratory analysis method, capable of transforming complex historical data into structured insights to support more informed and evidence-based policy formulation.

Keywords: *analysis, economic data, k-means algorithm, Clustering*

Abstrak

Meskipun data ekonomi dan keuangan historis tersedia dalam volume yang sangat besar, upaya untuk mengekstrak wawasan non-trivial yang tersembunyi di dalamnya tetap menjadi tantangan signifikan, terutama karena ketergantungan pada metode analisis tradisional yang bersifat hypothesis-driven. Dalam konteks Indonesia, aplikasi teknik *Clustering* secara komprehensif untuk menangkap

narasi data yang objektif masih belum banyak dieksplorasi, sehingga memunculkan urgensi untuk mengembangkan pendekatan berbasis data. Penelitian ini bertujuan untuk menjawab kesenjangan tersebut dengan mendemonstrasikan kapabilitas dan fleksibilitas algoritma K-Means sebagai metode analisis eksplorasi yang tangguh. Untuk mencapai tujuan ini, penelitian menerapkan pendekatan studi kasus komparatif pada lima *dataset* independen yang sengaja dipilih guna mencakup domain dan periode waktu yang beragam: tren merger bank (1971-1988), indikator makroekonomi kritis (1992-2003), kinerja keuangan bank BUMN (2004-2014), ekspor komoditas sarang burung walet (2017-2021), dan data ekonomi perbandingan dari Amerika Serikat (1930-1955). Secara metodologis, setiap *dataset* diproses secara ketat melalui pra-pemrosesan sebelum dikelompokkan menggunakan K-Means, dengan kualitas hasil dievaluasi secara kuantitatif melalui metrik *Silhouette score*, Davies-Bouldin Index, dan *Inertia*. Hasil penelitian menunjukkan kinerja klusterisasi yang sangat kuat, di mana tiga dari lima studi kasus mencapai *Silhouette score* di atas 0.70, yang mengindikasikan segmentasi data yang padat dan terdefinisi dengan baik. Temuan utama menunjukkan bahwa kluster yang terbentuk berhasil memetakan periode-periode historis secara objektif; sebagai contoh, algoritma ini secara otomatis mengisolasi anomali ekstrem krisis moneter 1998 sebagai sebuah kluster unik, mengidentifikasi puncak era merger perbankan sebagai fase konsolidasi yang intens, serta mengelompokkan bank-bank BUMN ke dalam segmen strategis yang berbeda berdasarkan profil modal dan profitabilitasnya. Studi ini mengonfirmasi bahwa K-Means adalah metode analisis eksplorasi yang efektif, mampu mentransformasi data historis yang kompleks menjadi wawasan terstruktur untuk mendukung perumusan kebijakan yang lebih informatif dan berbasis bukti.

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

A. PENDAHULUAN

Di era ekonomi global yang serba terhubung dan didorong oleh transformasi digital, dunia dihadapkan pada volume data yang belum pernah terjadi sebelumnya. Lembaga pemerintah, bank sentral, dan organisasi riset kini memiliki akses ke triliunan titik data yang mencakup segala hal, mulai dari indikator makroekonomi harian, statistik perdagangan internasional, hingga laporan keuangan korporasi yang terperinci. Banjir data (*data deluge*) ini, alih-alih secara otomatis memberikan kejelasan, justru sering kali menghadirkan tantangan baru untuk mencari cara mengekstrak wawasan yang bermakna dan dapat ditindaklanjuti dari lautan informasi yang begitu luas dan kompleks. Fenomena

ini, yang sering disebut sebagai era "*Big Data*", telah mendorong pergeseran paradigma dalam analisis ekonomi (Memon, 2021).

Secara tradisional, analisis ekonomi sangat bergantung pada metode statistik dan ekonometrika konvensional, seperti analisis regresi atau model deret waktu. Metode-metode ini sangat berharga dan telah terbukti efektif untuk menguji hipotesis yang telah ditentukan sebelumnya. Namun, keterbatasan utamanya adalah sifatnya yang *hypothesis-driven*. Analisis harus terlebih dahulu memiliki dugaan atau teori tentang hubungan antar variabel. Pendekatan ini kurang ideal untuk menemukan pola-pola yang tidak terduga, hubungan non-linear yang rumit, atau struktur tersembunyi yang

tidak pernah terpikirkan sebelumnya oleh para analis, sebuah keterbatasan yang diakui dalam literatur ekonomi modern (Varian, 2018). Akibatnya, banyak wawasan berharga yang mungkin terkubur di dalam data dan tidak pernah terungkap.

Kesenjangan inilah yang coba dijawab oleh disiplin ilmu *data mining*. Berbeda dengan pendekatan konvensional, *data mining* menawarkan paradigma analisis yang bersifat eksplorasi. Tujuannya bukan untuk mengonfirmasi hipotesis, melainkan untuk menemukan pola secara otomatis dari data itu sendiri (Tan et al., 2016). Dengan memanfaatkan kekuatan komputasi modern, *data mining* memungkinkan kita untuk menyisir *dataset* masif guna mengidentifikasi tren, anomali, dan pengelompokan yang signifikan, yang pada akhirnya dapat menjadi dasar untuk perumusan hipotesis dan kebijakan yang lebih mendalam.

Dalam dunia *data mining*, klusterisasi termasuk metode yang fundamental dan sangat kuat untuk menemukan pola tersembunyi. *Clustering* adalah proses mengelompokkan sekumpulan objek dalam hal ini, bisa berupa periode waktu, negara, atau perusahaan sedemikian rupa sehingga objek dalam kelompok yang sama disebut *Cluster* lebih mirip satu sama lain daripada objek di kelompok lain (Sinaga & Yang, 2020). Keindahan dari teknik ini adalah sifatnya yang *unsupervised* (tanpa pengawasan), artinya algoritma belajar dari data itu sendiri tanpa memerlukan label atau kategori yang telah ditentukan sebelumnya. Dalam konteks ekonomi, ini berarti kita dapat menemukan pengelompokan alami dari periode-periode historis misalnya, mengidentifikasi "periode stabilitas tinggi," "periode gejolak krisis," dan "periode pemulihan" secara objektif berdasarkan data.

Di antara berbagai algoritma *Clustering*, K-Means menonjol sebagai salah satu metode yang paling populer dan banyak digunakan. Popularitasnya bukan tanpa alasan. K-Means dikenal karena

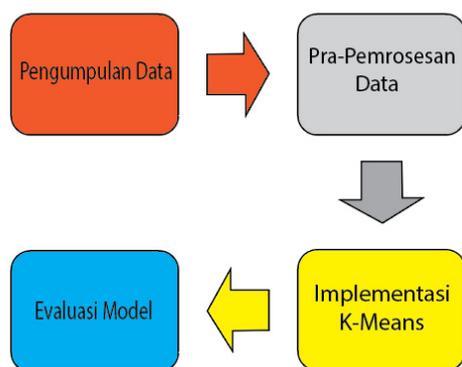
kesederhanaan konseptualnya, efisiensi komputasinya yang tinggi pada *dataset* besar, dan kemudahan dalam menginterpretasikan hasil klasternya (Efron & Hastie, 2021). Algoritma ini bekerja dengan mempartisi data ke dalam k jumlah klaster yang telah ditentukan, di mana setiap titik data menjadi milik klaster dengan rata-rata (mean) terdekat. Karena karakteristik inilah, K-Means menjadi alat yang sangat efektif untuk analisis eksplorasi awal guna mendapatkan pemahaman pertama tentang struktur data.

Meskipun potensi K-Means sangat besar, aplikasinya untuk membongkar narasi ekonomi historis di Indonesia khususnya dengan pendekatan studi kasus yang beragam masih relatif terbatas. Banyak analisis masih terfokus pada domain tunggal atau periode waktu yang sempit. Oleh karena itu, penelitian ini bertujuan untuk mendemonstrasikan fleksibilitas dan kekuatan interpretatif dari algoritma K-Means dengan menerapkannya pada lima *dataset* ekonomi dan keuangan yang berbeda secara fundamental. Studi ini akan menunjukkan bagaimana K-Means dapat mengubah data historis yang mentah dan kompleks menjadi wawasan terstruktur, seperti mengidentifikasi dampak krisis moneter 1998, memetakan siklus pasar komoditas, atau memahami dinamika restrukturisasi perbankan, yang semuanya krusial untuk perumusan kebijakan yang lebih informatif.

B. METODE PENELITIAN

Metodologi penelitian ini dirancang sebagai sebuah proses penemuan pengetahuan (*knowledge discovery*) yang terstruktur (Géron, 2022). Proses ini diawali dengan tahap fundamental yaitu pengumpulan data, di mana *dataset* historis yang beragam dari domain ekonomi dan keuangan dikurasi secara saksama. Data yang terkumpul kemudian memasuki tahap persiapan atau pra-pemrosesan, sebuah

langkah kritis yang mencakup pembersihan dari *noise* dan anomali serta normalisasi skala data untuk menjamin integritas analisis. Pada tahap inti, yaitu pemodelan, algoritma K-Means diaplikasikan sebagai alat utama untuk melakukan partisi data. Tujuannya adalah untuk secara otomatis mengidentifikasi pengelompokan alami dalam data mengungkap pola, tren, dan periode-periode unik yang tidak terlihat pada pandangan pertama [X, Y]. Akhirnya, kualitas dan kekokohan dari setiap model klaster yang terbentuk diuji melalui serangkaian metrik evaluasi kuantitatif, yang menjadi dasar untuk validasi sebelum melangkah ke tahap interpretasi hasil. Tahapan penelitian yang dilakukan dapat dilihat Gambar. 1.



Gambar 1. Tahapan Penelitian

Pengumpulan Data

Data yang digunakan dalam penelitian ini bersumber dari berbagai publikasi resmi dan arsip statistik yang mencakup lima studi kasus independen, yaitu:

1. Perkembangan Merger Bank di Indonesia (1971–1988): Data tahunan yang aslinya berasal dari Departemen Keuangan diperoleh dari buku 'Tinjauan Ekonomi Atas Dampak Paket Deregulasi Tahun 1988 Pada Sistem Keuangan Indonesia' karya Nasution (Nasution, 1990).
2. Indikator Makroekonomi PMDN, Bunga Kredit, dan Kurs di Indonesia (1992–2003): Data bulanan mengenai PMDN, Bunga Kredit (%), dan Kurs bersumber

dari publikasi Bank Indonesia (BI) dan Badan Koordinasi Penanaman Modal (BKPM), yang dikutip dari buku "Pendekatan Populer dan Praktis Ekonometrika untuk Analisis Ekonomi dan Keuangan" karya Nachrowi dan Usman (Djalal, 2006).

3. Kinerja Keuangan Bank BUMN (2004–2014): Data untuk analisis ini didasarkan pada laporan keuangan tahunan (Annual Report and Financial Statement) masing-masing bank, yang telah disarikan dan disajikan dalam buku "Valuasi Bisnis Teknologi Informasi" karya Abdurrahman (Abdurrahman, 2019).
4. Ekspor Sarang Burung Walet (2017–2021): Data volume ekspor merupakan data publikasi dari Badan Pusat Statistik (BPS) yang dirujuk dari buku "Rahasia Sukses Budi Daya Burung Walet" karya Tribowo (Vebriansyah, 2017).
5. Konsumsi Tembakau di AS (1930–1955): Data historis untuk studi kasus pembandingan ini bersumber dari arsip Modern Business Statistics, dikutip dari buku "Statistika" karya Sudjana (Organization, 2021).

Pra-pemrosesan Data

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

1. Pembersihan data: Menghapus atau memperbaiki data yang tidak konsisten, duplikat, atau mengandung kesalahan agar kualitas data meningkat.
2. Penanganan nilai yang hilang (*missing values*): Menangani data yang tidak lengkap, misalnya dengan mengisi nilai yang hilang menggunakan rata-rata, median, modus, atau menghapus baris/kolom yang terlalu banyak nilai kosongnya.
3. Normalisasi fitur numerik: Menyamakan skala variabel numerik agar tidak ada fitur yang mendominasi karena skala yang berbeda.

4. Penggunaan *StandardScaler*: Teknik normalisasi yang mengubah data sehingga memiliki rata-rata 0 dan standar deviasi 1, memastikan setiap fitur memiliki kontribusi yang seimbang dalam proses analisis, terutama dalam algoritma berbasis jarak seperti K-Means.

Implementasi 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:

1. Metode Elbow: Teknik ini digunakan untuk menentukan jumlah kluster optimal (nilai k) dengan cara memplot grafik antara jumlah kluster dan nilai *Inertia* (total jarak dalam kluster). Titik "siku" atau perubahan drastis terakhir pada grafik menunjukkan bahwa penambahan jumlah kluster setelah titik tersebut tidak lagi memberikan pengurangan *Inertia* yang signifikan. Titik ini dianggap sebagai jumlah kluster yang paling efisien.
2. *Silhouette Analysis*: Metode ini memilih nilai k yang menghasilkan nilai rata-rata *silhouette score* tertinggi. Skor ini mencerminkan seberapa baik tiap data cocok dengan klusternya sendiri dibandingkan dengan kluster lain. Nilai yang tinggi menunjukkan pemisahan kluster yang jelas dan kohesi yang baik dalam masing-masing kluster.

Evaluasi Model

Kualitas kluster dievaluasi menggunakan tiga metrik:

1. *Inertia*: Mengukur kepadatan kluster. Nilai yang lebih rendah lebih baik.
2. *Silhouette score*: Mengukur seberapa baik data terpisah antar kluster (Ahmed et al., 2020). Nilai mendekati +1 menunjukkan pengelompokan yang baik.
3. Davies-Bouldin Index: Mengukur rasio antara sebaran intra-kluster dengan jarak

antar-kluster (Azmin, 2022). Nilai yang lebih rendah lebih baik yang lebih efektif dan efisien.

C. HASIL DAN PEMBAHASAN

Hasil Penerapan *data mining* menggunakan algoritma *K-Means Clustering* dalam analisis data ekonomi historis dapat memberikan wawasan yang sangat berharga bagi para analis dan pengambil kebijakan untuk memahami pola-pola fundamental dari dinamika pasar. Langkah-langkah penerapan algoritma K-Means dalam penelitian ini dapat diuraikan sebagai berikut:

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data deret waktu (*time series*) historis dari berbagai indikator ekonomi dan keuangan. Data ini dikumpulkan dari sumber-sumber yang kredibel seperti publikasi resmi pemerintah, laporan keuangan, dan arsip statistik, yang kemudian disusun menjadi lima studi kasus independen seperti yang telah dijelaskan sebelumnya.

Tabel 1. *Dataset* Perkembangan Merger Bank di Indonesia

Tahun	Jumlah Bank	Jumlah Bank yang Merger	Jumlah Bank Hasil Merger
1971	128	3	1
1972	117	11	5
1973	109	17	8
1974	99	19	9
1975	92	10	5
1976	86	12	6
1977	85	6	3
1978	79	9	4
1979	77	8	4
1980	75	4	2
1981	71	8	4
1982	70	2	1
1983	70	2	1
1984	70	2	1
1985	69	0	0
1986	69	6	2
1987	67	2	1
1988	66	2	1

Tabel 2. *Dataset* PMDN, Bunga Kredit, dan Kurs di Indonesia

Tahun	Bulan	PMD (Triliun Rupiah)	Bunga (%)	Kurs (Rupiah)
1993	Februari	1.29	17.70	2,067.00
1993	Maret	3.06	17.65	2,071.00
1993	April	2.38	16.57	2,074.00
1993	Mei	3.27	16.38	2,079.00
1993	Juni	1.35	16.37	2,088.00
1993	Juli	2.50	16.39	2,096.00
1993	Agustus	2.29	15.92	2,102.00
1993	September	5.08	15.61	2,108.00
1993	Oktober	1.66	15.58	2,106.00
1993	November	5.75	15.38	2,106.00
1993	Desember	9.51	15.15	2,110.00
1994	Januari	5.76	14.89	2,122.00
1994	Februari	7.38	14.66	2,136.00
1994	Maret	3.52	14.58	2,144.00
1994	April	3.61	14.48	2,149.00
1994	Mei	3.12	14.10	2,155.00
1994	Juni	5.79	14.06	2,160.00
1994	Juli	5.24	14.11	2,169.00
1994	Agustus	4.95	13.98	2,175.00
1994	September	4.16	13.96	2,181.00
1994	Oktober	5.81	14.11	2,186.00
1994	November	3.50	14.11	2,193.00
1994	Desember	0.43	14.09	2,200.00
1995	Januari	7.38	14.05	2,207.00
1995	Februari	4.47	14.11	2,212.00
1995	Maret	4.51	14.08	2,219.00
1995	April	4.39	14.31	2,227.00
1995	Mei	5.41	14.51	2,236.00
1995	Juni	3.10	14.57	2,246.00
1995	Juli	3.81	14.64	2,256.00
1995	Agustus	9.83	14.71	2,266.00
1995	September	8.98	14.77	2,276.00
1995	Oktober	5.71	14.63	2,285.00
1995	November	8.92	14.82	2,296.00
1995	Desember	11.71	14.91	2,308.00
1996	November	2.95	15.10	2,368.00
1996	Desember	3.70	15.02	2,383.00
1997	Januari	10.40	15.03	2,406.00
1997	Februari	4.17	15.08	2,419.00
1997	Maret	5.89	15.08	2,433.00
1997	April	14.86	14.93	2,440.00
1997	Mei	7.37	14.74	2,450.00
1997	Juni	11.13	14.58	2,599.00
1997	Juli	7.56	16.51	3,035.00
1997	Agustus	9.81	16.68	3,275.00
1997	September	12.66	16.43	3,670.00
1997	Oktober	6.71	16.38	3,648.00
1997	November	5.18	15.37	4,650.00
1997	Desember	9.99	16.14	10,375.00
1998	Januari	9.99	16.14	10,375.00
1998	Februari	3.49	16.32	8,750.00

Tahun	Bulan	PMD (Triliun Rupiah)	Bunga (%)	Kurs (Rupiah)
1998	Maret	5.49	16.95	8,325.00
1998	April	4.25	17.67	7,970.00
1998	Mei	5.41	18.34	10,525.00
1998	Juni	1.24	18.39	14,900.00
1998	Juli	12.40	18.94	13,000.00
1998	Agustus	2.38	20.21	11,075.00
1998	September	10.43	20.53	10,700.00
1998	Oktober	2.41	22.00	7,550.00
1998	November	1.44	22.14	7,300.00
1998	Desember	1.81	19.39	8,025.00
1999	Januari	1.01	22.31	8,950.00
1999	Februari	6.98	22.24	8,730.00
1999	Maret	1.61	22.49	8,685.00
1999	April	0.53	22.38	8,260.00
1999	Mei	0.93	21.93	8,105.00
1999	Juni	8.10	21.45	6,726.00
1999	Juli	1.42	20.59	6,875.00
1999	Agustus	7.93	19.56	7,565.00
1999	September	2.47	18.91	6,900.00
1999	Oktober	2.04	19.14	7,425.00
1999	November	1.59	20.97	7,100.00
1999	Desember	1.43	17.31	7,425.00
2000	Januari	3.01	17.26	7,505.00
2000	Februari	3.17	16.48	7,590.00
2000	Maret	2.33	16.34	7,945.00
2000	April	0.76	16.22	8,620.00
2000	Mei	0.81	15.79	8,735.00
2000	Juni	1.73	15.40	9,003.00
2000	Juli	8.99	15.34	8,290.00
2000	Agustus	1.33	16.19	8,780.00
2000	September	6.00	16.67	9,530.00
2000	November	1.35	16.35	9,595.00
2000	Desember	0.95	16.37	9,450.00
2001	Januari	5.16	16.43	9,835.00
2001	Februari	1.98	16.31	10,400.00
2001	Maret	3.02	16.16	11,675.00
2001	April	1.60	16.21	11,058.00
2001	Mei	3.99	16.17	9,525.00
2001	Juni	2.61	16.26	8,865.00
2001	Juli	4.35	16.44	9,675.00
2001	Agustus	0.80	16.61	10,435.00
2001	September	4.25	16.83	10,430.00
2001	Oktober	2.89	17.11	10,400.00
2001	November	0.67	16.36	10,320.00
2001	Desember	0.86	16.40	10,189.00
2002	Januari	1.96	16.45	9,665.00
2002	Februari	5.57	16.50	9,316.00
2002	Maret	2.17	17.41	8,785.00
2002	Mei	8.57	18.50	8,318.00
2003	Januari	1.03	17.53	8,876.00
2003	Februari	0.47	17.60	8,905.00
2003	Maret	1.03	17.67	8,908.00
2003	April	1.96	17.64	8,675.00
2003	Mei	1.46	17.59	8,279.00

Tahun	Bulan	PMD (Triliun Rupiah)	Bunga (%)	Kurs (Rupiah)
2003	Juni	2.21	17.35	8,285.00
2003	Juli	3.09	16.82	8,505.00
2003	Agustus	1.18	16.49	8,535.00

Tahun	Bulan	PMD (Triliun Rupiah)	Bunga (%)	Kurs (Rupiah)
2003	September	1.95	16.42	8,389.00
2003	Oktober	1.76	16.16	8,495.00

Tabel 3. Dataset Kinerja Keuangan Bank BUMN (BNI)

Tahun	Ekuitas (K)	Buruh (L)	CapexTI (I)	Capex +2.5% opex (I)	Pendapatan (t)	Pendapatan (t-1)	ROE (%)
2006	14.794	2.909	78	3.912	15.044	12.707	22,60
2007	17.220	3.691	19	3.724	14.878	15.044	8,00
2008	15.789	3.299	113	3.471	16.628	14.878	9,00
2009	19.531	3.460	321	4.478	19.447	16.628	16,30
2010	33.446	4.127	277	3.827	18.837	19.447	24,70
2011	38.108	5.042	312	4.060	20.692	18.837	20,06
2012	43.797	5.578	641	4.264	22.705	20.692	19,99
2013	48.017	6.084	888	4.584	26.451	22.705	22,47
2014	61.421	6.781	803	6.298	33.365	26.451	23,64

Tabel 4. Dataset Kinerja Keuangan Bank BUMN (Mandiri)

Tahun	Ekuitas (K)	Buruh (L)	CapexTI (I)	Capex +2.5% opex (I)	Pendapatan (t)	Pendapatan (t-1)	ROE (%)
2004	24.994	2.402	650	3.346	13.581	11.753	22,80
2005	23.290	3.294	337	3.471	11.444	13.581	2,50
2006	26.415	3.079	190	3.325	13.078	11.444	10,00
2007	29.300	4.114	242	3.968	15.732	13.078	15,80
2008	30.640	4.713	488	4.701	19.400	15.732	18,10
2009	35.351	5.043	594	5.183	22.440	19.400	22,06
2010	42.125	6.156	972	7.335	28.767	22.440	24,24
2011	62.889	7.172	992	9.528	35.546	28.767	21,98
2012	76.090	8.567	970	10.872	41.931	35.546	22,60
2013	89.408	10.017	1.048	12.315	49.909	41.931	22,23
2014	105.799	11.565	1.642	14.329	56.882	49.909	25,81

Tabel 5. Dataset Kinerja Keuangan Bank BUMN (BRI)

Tahun	Ekuitas (K)	Buruh (L)	CapexTI (I)	Capex +2.5% opex (I)	Pendapatan (t)	Pendapatan (t-1)	ROE (%)
2006	17.090	4.831	69	2.399	21.071	17.254	33,75
2007	19.832	5.274	44	3.710	23.241	21.071	31,64
2008	22.801	6.329	190	3.316	28.097	23.241	34,50
2009	27.513	6.676	128	4.413	35.334	28.097	35,22
2010	37.030	8.676	155	6.270	44.615	35.334	43,83
2011	50.302	8.701	168	6.018	48.164	44.615	42,49
2012	65.411	9.606	917	7.037	49.610	48.164	38,66
2013	80.541	12.232	734	7.480	59.461	49.610	34,11
2014	98.937	14.112	1.864	8.412	75.122	59.461	31,22

Tabel 6. Dataset Ekspor Sarang Burung Walet

No	Negara	2017	2018	2019	2020	2021
1.	Hongkong	487,8	290,4	644,1	897,2	989,9
2.	Tiongkok	55,5	69,6	129,1	263,5	228,8
3.	Singapura	71,6	90,0	75,4	68,8	80,0
4.	Amerika Serikat	18,2	16,6	47,0	20,4	66,2
5.	Vietnam	624,5	806,1	329,9	27,2	71,3
6.	Kanada	3,4	2,7	2,0	1,8	2,3
7.	Taiwan	8,3	5,5	5,6	12,9	20,6
8.	Thailand	3,9	4,4	1,1	2,4	1,8
9.	Jepang	0,3	1,0	0,1	0,3	0,1
10.	Kamboja	0,6	0,0	0,0	0,0	0,0

Tabel 7. Dataset Konsumsi Tembakau di AS

Tahun	Pemakaian (pounds)
1930	8,8
1931	8,4
1932	7,6
1933	7,8
1934	8,3
1935	8,2
1936	8,8
1937	9,0
1938	8,8
1939	8,8
1940	9,1
1941	9,0
1942	10,7
1943	11,5
1944	11,2
1945	12,5
1946	12,2
1947	12,0
1948	12,1
1949	11,9
1950	12,0
1951	12,5
1952	12,9
1953	12,9
1954	12,2
1955	12,3

2. Pra-pemrosesan Data

Arsitektur komputasi untuk penelitian ini dibangun di atas bahasa pemrograman Python. Pustaka Pandas diandalkan untuk seluruh tugas pra-pemrosesan dan manipulasi data. Sementara itu, kerangka kerja Scikit-learn menyediakan perangkat esensial untuk pemodelan dan evaluasi. Hal ini mencakup penggunaan kelas *StandardScaler* untuk normalisasi data, implementasi

algoritma KMeans untuk proses klasterisasi, serta kalkulasi metrik kuantitatif seperti *silhouette_score* dan *davies_bouldin_score* untuk validasi performa klaster.

Berikut adalah potongan kode yang digunakan dalam proses klasterisasi:

```

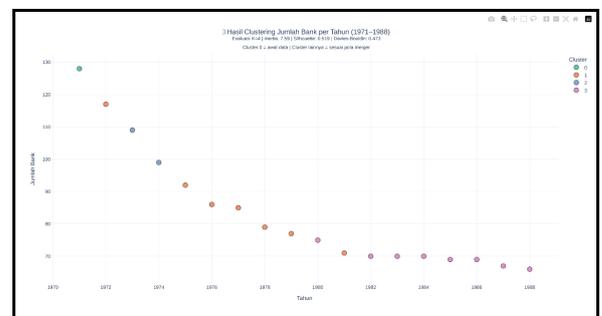
1 import pandas as pd
2 from sklearn.preprocessing import StandardScaler
3 from sklearn.cluster import KMeans
4 from sklearn.metrics import silhouette_score, davies_bouldin_score
    
```

Gambar 2. Potongan Kode dalam Proses Klasterisasi

Implementasi K-Means dan Evaluasi Model

1. Klasterisasi Tren Merger Bank di Indonesia

Penerapan algoritma K-Means pada data Tren Merger bank di Indonesia dari tahun 1970 hingga 1988 menghasilkan klasterisasi yang jelas, dengan jumlah klaster optimal $k=4$, *Inertia*: 7.59, *Silhouette score*: 0.519, *Davies-Bouldin Index*: 0.473



Gambar 3. Grafik Tren Merger Bank

Kombinasi metrik evaluasi ini menunjukkan bahwa hasil klasterisasi memiliki kualitas yang baik dan signifikan secara statistik. *Silhouette score* sebesar 0.519 menandakan klaster cukup jelas dan terpisah, sementara *Davies-Bouldin Index* sebesar 0.473 menunjukkan klaster yang kompak dan terpisah dengan baik. *Inertia* sebesar 7.59 juga mencerminkan keberhasilan algoritma dalam meminimalkan variasi internal. Secara keseluruhan, ketiga metrik ini mengonfirmasi bahwa pembentukan empat

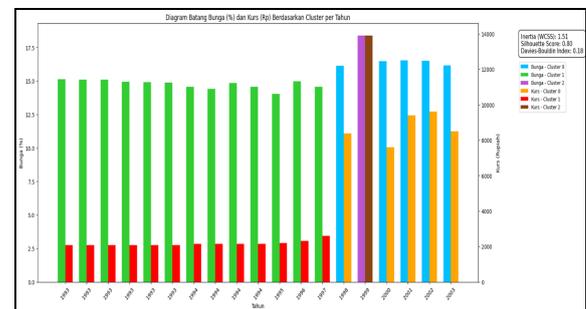
klaster mencerminkan struktur nyata dalam data, bukan sekadar pengelompokan acak. Interpretasi Klaster:

- a. Klaster 0 (Hijau): Klaster ini hanya mencakup tahun 1971, saat jumlah bank berada di titik tertinggi. Karena posisinya yang sangat berbeda dari tahun-tahun lain, algoritma mengelompokkannya secara terpisah. Ini menandai awal era sebelum konsolidasi besar-besaran dimulai. Periode ini dapat digunakan oleh pengambil kebijakan sebagai baseline sebelum reformasi, menjadi acuan dalam mengukur dampak intervensi kebijakan pada sektor perbankan.
- b. Klaster 1 (Oranye): Klaster ini terdiri dari sebagian besar tahun ketika jumlah bank menurun secara konsisten. Ini menunjukkan periode di mana merger dan akuisisi berlangsung aktif dan teratur, menandai fase utama restrukturisasi sektor perbankan. Temuan ini berguna untuk mengevaluasi efektivitas kebijakan merger dan deregulasi selama fase transisi. Pengambil kebijakan dapat menggunakannya sebagai referensi dalam merancang strategi konsolidasi lanjutan.
- c. Klaster 2 (Biru): Hanya mencakup dua tahun yang menunjukkan penurunan paling curam dalam jumlah bank. Klaster ini menggambarkan titik krusial dari proses konsolidasi, ketika perubahan struktural terjadi dengan sangat cepat dan intens. Periode ini dapat dijadikan indikator krisis mikro dalam sistem perbankan. Analisis ini bermanfaat untuk mendesain sistem peringatan dini terhadap ketidakseimbangan sektor keuangan.
- d. Klaster 3 (Ungu): Klaster ini berisi tahun-tahun ketika jumlah bank mulai stabil dan tidak banyak berubah. Ini mencerminkan periode pasca-konsolidasi. Hasil ini dapat digunakan sebagai bukti keberhasilan kebijakan sebelumnya. Pemerintah dan otoritas

keuangan dapat menggunakannya untuk mempertahankan stabilitas dan mencegah kembali ke fase fragmentasi.

2. Klasterisasi PMDN, Bunga Kredit (%), dan Kurs

Penerapan algoritma K-Means pada data Suku Bunga (%) dan Kurs Rupiah di Indonesia dari tahun 1993 hingga 2003 menghasilkan klasterisasi yang sangat jelas dan terdefinisi dengan baik, dengan jumlah klaster optimal $k=3$, *Inertia*: 1.51, *Silhouette score*: 0.80, *Davies-Bouldin Index*: 0.18



Gambar 4. Grafik PMDN, Bunga Kredit (%), dan Kurs

Evaluasi metrik secara kolektif menunjukkan kualitas klasterisasi yang sangat tinggi dan valid secara statistik. Nilai *Silhouette score* sebesar 0.80 mengindikasikan klaster yang padat dan terpisah jelas, dengan tumpang tindih minimal. Hal ini diperkuat oleh *Davies-Bouldin Index* yang rendah (0.18), yang mencerminkan klaster yang kompak dan terpisah baik. Selain itu, *Inertia* yang rendah (1.51) menandakan sebaran data dalam klaster sangat kecil, menunjukkan keseragaman karakteristik anggota klaster. Kombinasi ketiga metrik ini mengonfirmasi bahwa pola yang terbentuk mencerminkan struktur nyata dalam data, bukan kebetulan. Interpretasi Klaster:

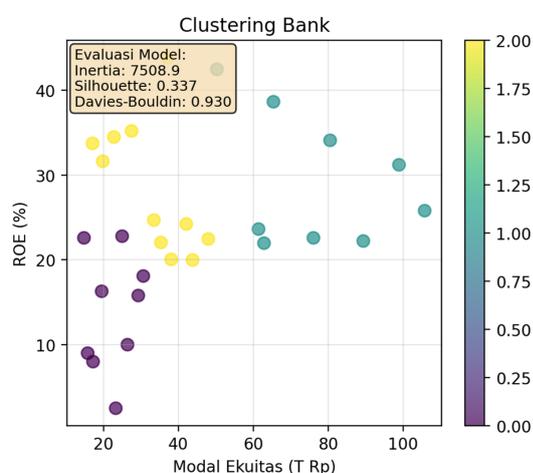
- a. Klaster 1 (Hijau & Merah): mewakili periode stabil sebelum krisis (1993–1997), dengan suku bunga moderat (14–15%) dan nilai tukar stabil di bawah Rp 3.000. Klaster ini mencerminkan kondisi

ekonomi yang solid dan konsisten sebelum guncangan krisis terjadi.

- b. Klaster 2 (Ungu & Cokelat): mengisolasi tahun 1998 sebagai puncak krisis, ditandai oleh lonjakan suku bunga di atas 18% dan pelemahan drastis nilai tukar hingga sekitar Rp 14.000. Ini menunjukkan anomali ekstrem yang berhasil diidentifikasi algoritma. Identifikasi otomatis tahun 1998 sebagai anomali ekstrem ini sangat signifikan, karena temuan ini sejalan dengan berbagai analisis ekonomi makro yang mendalam tentang Krisis Finansial Asia, yang menempatkan Indonesia sebagai salah satu negara yang paling parah terdampak. Algoritma K-Means berhasil menangkap peristiwa bersejarah ini murni dari data numerik.
- c. Klaster 0 (Biru & Oranye): menggambarkan masa pasca-krisis (1999–2003), dengan suku bunga dan kurs yang masih tinggi namun mulai membaik. Klaster ini mencerminkan fase pemulihan yang bergejolak namun bergerak menuju kestabilan.

3. Klasterisasi Kinerja Keuangan Bank BUMN

Penerapan algoritma K-Means terhadap data kinerja keuangan Bank BUMN berhasil menghasilkan pengelompokan yang sangat informatif.



Gambar 5. Grafik Kinerja Keuangan Bank BUMN

Dengan menggunakan Modal Ekuitas dan *Return on Equity* (ROE) sebagai variabel, model ini secara optimal membagi bank ke dalam tiga klaster ($k=3$) yang masing-masing mewakili segmen strategis yang unik dengan nilai *Inertia*: 7508.9, *Silhouette score*: 0.337, dan *Davies-Bouldin Index*: 0.930.

Metrik evaluasi menunjukkan bahwa struktur klaster yang terbentuk cukup baik. *Silhouette score* sebesar 0.337 mengindikasikan bahwa bank-bank dalam satu klaster cenderung mirip satu sama lain dibandingkan dengan bank di klaster lain, meskipun terdapat sedikit tumpang tindih antar-klaster. Nilai *Davies-Bouldin Index* sebesar 0.930 mendukung hal ini, menunjukkan bahwa klaster-klaster yang terbentuk cukup padat dan terpisah dengan jelas. Sementara itu, nilai *Inertia* yang cukup besar, yaitu 7508.9, disebabkan oleh skala data yang besar (modal dalam triliunan rupiah). Nilai ini berfungsi sebagai dasar perbandingan ketika ingin mengevaluasi jumlah klaster yang berbeda di masa depan.

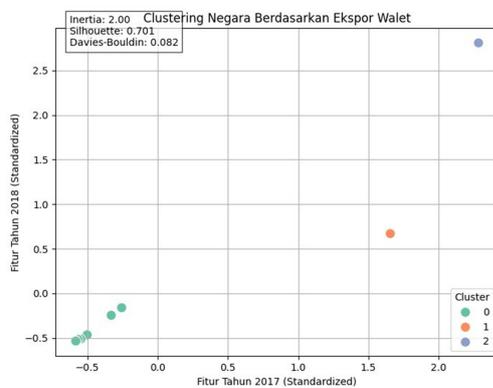
Interpretasi Klaster:

- a. Klaster Ungu (*Cluster 0*): Bank-bank dalam klaster ini memiliki modal ekuitas yang relatif kecil, umumnya di bawah 30 triliun rupiah. *Return on Equity* (ROE) mereka bervariasi antara 2% hingga 23%. Ini menunjukkan bahwa klaster ini terdiri dari bank-bank kecil dengan kinerja yang beragam sebagian mungkin kurang efisien, sementara sebagian lainnya cukup baik untuk ukuran modalnya.
- b. Klaster Kuning (*Cluster 2*): Klaster ini mencakup bank-bank dengan modal ekuitas menengah, sekitar 30 hingga 50 triliun rupiah. ROE mereka tergolong tinggi, sebagian besar di atas 20%, bahkan ada yang mencapai lebih dari 30%. Bank-bank ini sangat efisien dan mampu menghasilkan profit yang besar dibandingkan modal yang mereka miliki, menandakan performa keuangan yang kuat.

c. Klaster Hijau/Tosca (*Cluster 1*): Bank-bank di klaster ini merupakan kelompok dengan modal ekuitas terbesar, yaitu di atas 60 triliun rupiah. Meskipun demikian, ROE mereka berada di tingkat menengah, yaitu antara 21% hingga 38%. Ini merupakan ciri khas bank besar yang cenderung tumbuh lebih stabil dan tidak se-agresif bank menengah dalam menghasilkan keuntungan, namun tetap memiliki performa yang solid.

4. Klasterisasi Ekspor Sarang Burung Walet

Penerapan algoritma *Clustering* pada data ekspor sarang burung walet menghasilkan klasterisasi yang sangat baik dan jelas, dengan jumlah klaster optimal yang ditentukan adalah $k=3$. Hasil ini secara efektif membagi negara-negara eksportir ke dalam tiga kelompok yang sangat berbeda berdasarkan performa ekspor mereka pada tahun 2017 dan 2018. *Inertia*: 2.00, *Silhouette score*: 0.701 dan *Davies-Bouldin Index*: 0.082.



Gambar 6. Grafik Ekspor Sarang Burung Walet

Metrik evaluasi ini memberikan bukti kuat bahwa hasil klasterisasi sangat baik dan signifikan secara statistik. *Silhouette score* sebesar 0.701, yang sangat mendekati 1, menunjukkan pemisahan klaster yang sangat jelas, di mana negara-negara dalam satu kelompok jauh lebih mirip satu sama lain daripada dengan kelompok lain. *Davies-Bouldin Index* yang sangat rendah

(0.082) mendukung hal ini dengan menunjukkan bahwa klaster yang terbentuk sangat padat (kompak) dan terpisah dengan baik. Nilai *Inertia* yang kecil (2.00) juga mengindikasikan bahwa variasi di dalam tiap klaster sangat minim.

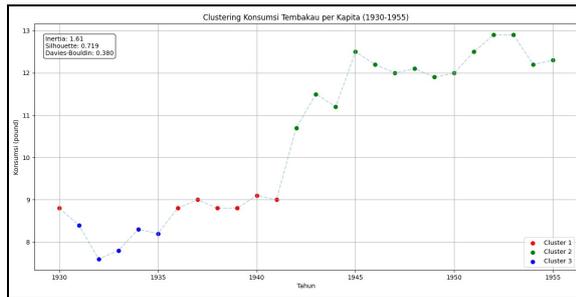
Secara keseluruhan, ketiga metrik ini secara meyakinkan menegaskan bahwa pembagian menjadi tiga kelompok eksportir mencerminkan struktur pasar yang nyata dan bukan hasil pengelompokan acak.

Interpretasi Klaster:

- Klaster 0 (Hijau Tosca): terdiri dari negara-negara dengan nilai ekspor yang rendah pada tahun 2017 dan 2018, yang seluruhnya berada di kuadran kiri bawah. Kelompok ini mewakili eksportir skala kecil dengan volume ekspor yang seragam dan rendah dibandingkan negara lain, membentuk klaster yang cukup padat.
- Klaster 1 (Oranye): hanya berisi satu negara yang berada di kuadran kanan atas, menunjukkan nilai ekspor yang tinggi di kedua tahun. Negara ini dikategorikan sebagai eksportir besar dengan profil yang unik dan berbeda dari negara lain, sehingga membentuk klaster tersendiri.
- Klaster 2 (Biru): juga hanya terdiri dari satu negara, namun posisinya lebih ekstrem di kanan atas, terutama pada sumbu Y tahun 2018. Negara ini merupakan eksportir dominan atau pemimpin pasar, dengan volume ekspor yang jauh melampaui semua negara lain, menunjukkan peningkatan kekuatan ekspor dari tahun ke tahun.

5. Klasterisasi Pemakaian Tembakau Per-Capita di Amerika Serikat

Penerapan algoritma K-Means pada data konsumsi tembakau per kapita dari tahun 1930 hingga 1955 menghasilkan klasterisasi yang sangat baik dan jelas, dengan jumlah klaster optimal $k=3$, *Inertia*: 1.61, *Silhouette score*: 0.719, dan *Davies-Bouldin Index*: 0.380.



Gambar 7. Grafik Konsumsi Tembakau di AS

Metrik evaluasi ini memberikan bukti kuat bahwa hasil klusterisasi sangat baik dan signifikan secara statistik. *Silhouette score* sebesar 0.719 menunjukkan pemisahan klaster yang jelas, di mana tahun-tahun dalam satu kelompok jauh lebih mirip satu sama lain daripada dengan kelompok lain. Davies-Bouldin Index yang rendah (0.380) mendukung hal ini dengan menunjukkan klaster yang padat dan terpisah dengan baik. *Inertia* sebesar 1.61 juga mengindikasikan bahwa variasi dalam tiap klaster sangat kecil. Secara keseluruhan, ketiga metrik ini menegaskan bahwa pembagian tiga periode konsumsi tembakau mencerminkan struktur historis yang nyata dan bukan hasil pengelompokan acak.

Interpretasi Klaster:

- Klaster 3 (Biru):** Klaster ini mencakup periode Depresi Besar, ditandai dengan konsumsi tembakau paling rendah, sekitar 7.6–8.4 pound. Krisis ekonomi parah menurunkan daya beli masyarakat secara signifikan, dan algoritma berhasil mengelompokkan tahun-tahun ini sebagai fase yang berbeda dan tertekan secara ekonomi.
- Klaster 1 (Merah):** Klaster ini menunjukkan fase transisi pasca-Depresi dan pra-Perang Dunia II, dengan konsumsi berada di tingkat menengah (8.8–9.1 pound). Ekonomi mulai pulih, dan konsumsi tembakau perlahan meningkat, menandai pemulihan bertahap dalam perilaku konsumen.
- Klaster 2 (Hijau):** Klaster ini menggambarkan lonjakan konsumsi tembakau selama dan setelah Perang

Dunia II, mencapai puncak hampir 13 pound. Rokok menjadi bagian dari budaya militer dan simbol patriotisme, yang terus berlanjut ke era kemakmuran pasca-perang. Klaster ini menangkap perubahan besar dalam pola konsumsi masyarakat.

D. PENUTUP

Penelitian ini secara komprehensif berhasil membuktikan bahwa algoritma K-Means merupakan metode yang kuat, fleksibel, dan efektif untuk mengekstrak wawasan dari data ekonomi historis yang kompleks. Melalui penerapan pada lima studi kasus yang sangat beragam, studi ini menunjukkan validitas kuantitatif yang tinggi, dengan mayoritas kasus mencapai *Silhouette score* di atas 0.70, yang mengindikasikan pembentukan klaster yang padat dan terdefinisi dengan baik. Secara signifikan, algoritma ini tidak hanya mampu memetakan periode-periode historis penting secara objektif, seperti mengisolasi anomali krisis moneter 1998, tetapi juga menghasilkan wawasan strategis dengan mengelompokkan entitas seperti bank BUMN berdasarkan profil modal dan profitabilitasnya.

Implikasi dari temuan ini bagi pembuat kebijakan dan pelaku ekonomi sangat nyata. Metode ini dapat dimanfaatkan sebagai sistem peringatan dini (*early warning system*) untuk mendeteksi potensi krisis, sebuah area di mana aplikasi *machine learning* telah menunjukkan potensi besar (Goldstein et al., 2019). Selain itu, metode ini dapat membantu kementerian dalam melakukan *benchmarking* kinerja entitas secara objektif, serta merancang kebijakan perdagangan yang lebih tertarget. Untuk pengembangan di masa depan, disarankan agar penelitian selanjutnya mengeksplorasi algoritma *Clustering* yang lebih canggih (seperti DBSCAN atau Hierarchical *Clustering*) untuk menangani struktur data yang lebih rumit, serta mengintegrasikan sumber data alternatif seperti data sentimen

untuk pemahaman yang lebih holistik. Selain itu, hasil klusterisasi ini dapat menjadi fondasi untuk membangun model prediktif yang bertujuan meramalkan probabilitas transisi suatu periode ekonomi dari klaster 'stabil' ke 'krisis'. Pada akhirnya, studi ini menegaskan bahwa pendekatan *data mining* seperti K-Means mampu menjembatani kesenjangan antara data historis dan pengambilan keputusan berbasis bukti yang proaktif.

E. DAFTAR PUSTAKA

- Abdurrahman, L. (2019). *Valuasi Bisnis Teknologi Informasi*. Bandung: Penerbit Informatika.
- Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8), 1295.
- Azmin, S. (2022). *Conditional variational laplace autoencoder based network intrusion detection system*.
- Djalal, N. (2006). *Pendekatan populer dan praktis ekonometrika untuk analisis ekonomi dan keuangan*. Universitas Indonesia Publishing.
- Efron, B., & Hastie, T. (2021). *Computer age statistical inference, student edition: algorithms, evidence, and data science* (Vol. 6). Cambridge University Press.
- Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. “O’Reilly Media, Inc.”
- Goldstein, I., Jiang, W., & Karolyi, G. A. (2019). To FinTech and beyond. *The Review of Financial Studies*, 32(5), 1647–1661.
- Memon, S. (2021). Machine Learning for Economists. *The Pakistan Development Review*, 60(2), 201–211.
- Nasution, A. (1990). Tinjauan ekonomi atas dampak paket deregulasi tahun 1988 pada sistem keuangan Indonesia. (*No Title*).
- Organization, W. H. (2021). *WHO report on the global tobacco epidemic 2021*.
- Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised K-means Clustering algorithm. *IEEE Access*, 8, 80716–80727.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2016). *Introduction to data mining*. Pearson Education India.
- Varian, H. R. (2018). *Artificial intelligence, economics, and industrial organization* (Vol. 24839). National Bureau of Economic Research Cambridge, MA, USA:
- Vebriansyah, R. (2017). *Rahasia Sukses Bisnis Walet dari Pakar dan Praktisi*. Penebar Swadaya Grup.