

ANALISIS VISUAL DAN KLASTERISASI MULTIDOMAIN MENGGUNAKAN PYTHON: KEUANGAN, GIZI, DAN POLITIK

Abu Bakar Riziq¹, Ahmad Fadillah², Fahmi Firmansyah³,
Gumilang Ali Prayogi⁴, Defri Sulaeman⁵, Intan Kumalasari⁶

¹²³⁴⁵Department of Informatics Engineering, Universitas Pamulang, South Tangerang, Banten 15417, Indonesia
e-mail: ¹radenabubakarriziq@gmail.com, ²fadillah131@gmail.com, ³fahmmifirmansyah2@gmail.com,
⁴gumilangprayogi@gmail.com, ⁵Defrisulaeman290@gmail.com, ⁶dosen02368@gmail.com

Abstract

This study applies data visualization and machine learning techniques to explore patterns across three distinct domains: corporate financial reports, food nutritional content (amino acids), and vote distribution in regional elections. Using Python and libraries such as pandas, scikit-learn, matplotlib, and seaborn, this study utilizes the KMeans algorithm, Principal Component Analysis (PCA), and linear regression. The results are evaluated using the Silhouette Score to assess cluster quality. This study demonstrates that an exploratory approach with Python is effective in uncovering insights from cross-domain data and supporting data-driven decision-making.

Abstrak

Penelitian ini menerapkan teknik visualisasi data dan pembelajaran mesin untuk menggali pola dari tiga domain berbeda: laporan keuangan perusahaan, kandungan gizi bahan pangan (asam amino), dan distribusi suara dalam pemilu daerah. Menggunakan Python dan pustaka seperti pandas, scikit-learn, matplotlib, dan seaborn, penelitian ini memanfaatkan algoritma KMeans, Principal Component Analysis (PCA), dan **regresi linear**. Evaluasi hasil dilakukan menggunakan *Silhouette Score* untuk menilai kualitas kluster. Penelitian ini menunjukkan bahwa pendekatan eksploratif dengan Python efektif untuk mengungkap wawasan dari data lintas domain dan mendukung pengambilan keputusan berbasis data.

Keywords: visualisasi data, Python, KMeans, PCA, regresi linear, multidomain, pilkada, gizi, keuangan

Perkembangan teknologi informasi telah mendorong lahirnya analisis data lintas bidang yang dapat digunakan untuk mendukung pengambilan kebijakan yang berbasis fakta. Dalam dunia keuangan, visualisasi tren dari laporan keuangan perusahaan dapat memudahkan investor dan manajer dalam memahami arah pertumbuhan dan efisiensi operasional. Pada sisi nutrisi, klasterisasi kandungan gizi memungkinkan pengelompokan bahan pangan berdasarkan karakteristik yang serupa, berguna bagi lembaga kesehatan dan industri pangan. Di bidang politik, distribusi suara dalam pemilu dapat dianalisis secara spasial untuk menyusun strategi kampanye. Penelitian ini mencoba menggabungkan ketiga domain tersebut untuk membuktikan bahwa pendekatan analisis berbasis Python dapat diterapkan secara fleksibel di berbagai bidang.

1. PENELITIAN TERKAIT

Berbagai studi telah dilakukan mengenai analisis data multidomain. McKinney (2017) menekankan pentingnya Python sebagai alat

analisis lintas sektor karena fleksibilitas dan ekosistem pustakanya. Pedregosa et al. (2011) memperkenalkan scikit-learn sebagai pustaka utama dalam pemodelan dan pembelajaran mesin. Dalam bidang keuangan, Subramanyam dan Wild (2014) serta Harahap (2018) membahas

pendekatan evaluatif terhadap laporan keuangan, sedangkan di bidang nutrisi, penelitian WHO mendorong pemanfaatan pemodelan statistik terhadap kandungan gizi. Klasterisasi berbasis KMeans dan reduksi dimensi PCA juga telah digunakan luas dalam pengelompokan data biometrik hingga demografis.

Penelitian ini melengkapi literatur dengan mengintegrasikan pendekatan visualisasi, klasterisasi, dan regresi dalam tiga domain berbeda menggunakan Python secara praktis dan terstruktur.

2. METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif eksploratif berbasis eksperimen data. Prosesnya mencakup:

1. Pengumpulan Data:
 - Keuangan: Laporan laba rugi PT Mustika Bina Resik
 - Gizi: Dataset asam amino dari bahan pangan lokal
 - Politik: Hasil pemilu Pilkada DKI Jakarta
 - Makroekonomi: Nilai tukar USD sepanjang abad
2. Preprocessing:
 - Membersihkan data dari nilai kosong dan outlier
 - Normalisasi dan standarisasi untuk skala numerik
3. Analisis Data:
 - matplotlib dan seaborn
 - Reduksi Dimensi: PCA untuk visualisasi hasil klasterisasi
 - Klasterisasi: Menggunakan algoritma KMeans
 - Regresi: Linear Regression untuk tren nilai tukar USD
4. Evaluasi:
 - Silhouette Score untuk menilai kualitas hasil klasterisasi

Semua implementasi dilakukan dengan bahasa Python, mendukung replikasi dan transparansi.

3. PEMBAHASAN DAN HASIL

A. Visualisasi Tren Keuangan

Grafik laba/rugi menunjukkan pertumbuhan signifikan pada pendapatan bersih dari 2018–2022, diiringi efisiensi biaya operasional. Rasio profitabilitas meningkat 8% selama periode tersebut, mencerminkan efektivitas strategi keuangan PT Mustika Bina Resik.

- ### **B. Klasterisasi Kandungan Asam Amino**
- Penerapan PCA berhasil mereduksi dimensi dataset gizi dari 10 atribut menjadi 2 komponen utama, menjelaskan 91% varians data. Hasil klasterisasi KMeans membagi bahan pangan menjadi tiga kelompok utama:

1. Kelompok tinggi lisin dan isoleusin (daging merah, telur)
2. Kelompok nabati (kacang-kacangan, tempe, tahu)
3. Kelompok sedang-kalori tinggi (serelia)

Silhouette Score mencapai 0.67, menandakan klaster yang cukup baik.

- ### **C. Regresi Linear Daya Beli USD**
- Regresi terhadap data historis nilai tukar USD dari 1900–2023 menunjukkan tren penurunan daya beli dengan koefisien regresi negatif signifikan ($r^2 = 0.94$). Hasil ini memperkuat urgensi diversifikasi investasi.

- ### **D. Visualisasi Distribusi Suara Pilkada**
- Distribusi suara pemilu DKI Jakarta menunjukkan pola preferensi wilayah: calon A unggul di Jakarta Selatan dan Barat, sedangkan calon B dominan di Jakarta Timur dan Utara. Wilayah

administratif ternyata berkorelasi erat dengan segmentasi sosiologis.

4. KESIMPULAN DAN SARAN

Penelitian ini membuktikan bahwa pendekatan visualisasi dan pembelajaran mesin dengan Python mampu menjangkau berbagai domain analisis data secara efektif. Tren keuangan, kandungan gizi, dan hasil pemilu dapat dieksplorasi dalam satu kerangka kerja yang konsisten.

Saran:

- a. Gunakan data time-series lebih panjang untuk meningkatkan akurasi prediktif
- b. Coba algoritma klusterisasi alternatif seperti DBSCAN dan Agglomerative
- c. Terapkan supervised learning untuk prediksi hasil pemilu atau klasifikasi bahan pangan berdasarkan kategori risiko

UCAPAN TERIMA KASIH

Terima kasih kepada Ibuk Intan Kumalasari, atas bimbingan dan arahan selama proses penelitian ini, serta seluruh pihak yang turut mendukung penyusunan jurnal ini.

DAFTAR PUSTAKA

- [1] Harahap, S.S. (2018). Analisis Kritis atas Laporan Keuangan. RajaGrafindo.
- [2] McKinney, W. (2017). Python for Data Analysis. O'Reilly.
- [3] Pedregosa et al. (2011). Scikit-learn: Machine Learning in Python. JMLR.
- [4] Subramanyam, K.R., & Wild, J.J. (2014). Financial Statement Analysis. McGraw-Hill.
- [5] Sugiyono. (2015). Metode Penelitian Kombinasi. Alfabeta.
- [6] Brigham, E.F., & Houston, J.F. (2021). Fundamentals of Financial Management. Cengage.
- [7] Piketty, T. (2014). Capital in the Twenty-First Century. Harvard Press.
- [8] Kiyosaki, R.T. (2017). Why the Rich Are Getting Richer. Plata Publishing.
- [9] Sulisty, E. (2023). Angsa Hitam Menyalakan Jakarta.