



## ANALISIS KLASTER DAN PREDIKSI INDIKATOR EKONOMI MAKRO INDONESIA MENGGUNAKAN ALGORITMA K-MEANS DAN MODEL REGRESI

Ahmad Basith<sup>1</sup>, Arfan Yusuf<sup>2</sup>, Chaerunisa<sup>3</sup>, Dimas Satrio Wicaksono<sup>4</sup>, Nathan Sanjaya<sup>5</sup>, Zurnan Alfian<sup>6</sup>

<sup>1,2,3,4,5,6</sup>Teknik Informatika Universitas Pamulang, Indonesia

Email: [dsatrio268@gmail.com](mailto:dsatrio268@gmail.com)

Received 12-04-2025 | Revised form 15-05-2025 | Accepted 22-06-2025

### Abstract

This study aims to analyze patterns and predict the development of Indonesia's macroeconomic indicators using a data mining approach. Two primary methods are employed: the K-Means Clustering algorithm for data grouping and Linear Regression for predicting future trends. The data analyzed includes Indonesia's trade balance (exports and imports), state revenue (taxation and non-tax state revenue), the role of each type of Gross Domestic Product (GDP) usage, Regional Revenue and Expenditure Budget (APBD) data for each regency/municipality in the Special Region of Yogyakarta Province, and trends in non-oil and gas industrial exports. The clustering results reveal temporal or regional segmentation with significantly different economic characteristics, such as groups of years marked by trade balance surpluses or deficits. The regression model shows consistent upward trends in certain sectors, such as household consumption and tax revenue. Evaluation was carried out using metrics including inertia, silhouette score, and the Davies-Bouldin Index, all of which indicate good clustering quality. These findings are expected to serve as a foundation for more targeted fiscal policy planning and economic development strategies, especially in sectors that contribute significantly to both national and regional growth. **Keywords:** K-Means, linear regression, data clustering, macroeconomic indicators, state revenue, APBD, PDB

### Abstrak

Penelitian ini bertujuan untuk menganalisis pola dan memprediksi perkembangan indikator ekonomi makro Indonesia dengan pendekatan *data mining*. Dua metode utama digunakan, yaitu algoritma *K-Means Clustering* untuk pengelompokan data dan *Linear Regression* untuk memprediksi tren masa depan. Data yang dianalisis mencakup neraca perdagangan Indonesia (ekspor dan impor), pendapatan negara (perpajakan dan PNBPD), peranan masing-masing jenis penggunaan Produk Domestik Bruto (PDB), data Anggaran Pendapatan dan Belanja Daerah (APBD) per kabupaten/kota di Provinsi Daerah Istimewa Yogyakarta, serta tren ekspor industri non-minyak dan gas. Hasil klusterisasi menunjukkan adanya segmentasi waktu atau wilayah dengan karakteristik ekonomi yang berbeda secara signifikan, seperti kelompok tahun-tahun dengan neraca perdagangan surplus atau defisit. Model regresi menunjukkan tren peningkatan yang konsisten dalam sektor-sektor tertentu, seperti konsumsi rumah tangga dan pendapatan perpajakan. Evaluasi dilakukan menggunakan metrik *inertia*, *silhouette score*, dan *Davies-Bouldin Index*, yang menunjukkan kualitas kluster yang baik. Temuan ini diharapkan dapat menjadi dasar perencanaan kebijakan fiskal dan strategi pembangunan ekonomi yang lebih terarah, khususnya pada sektor-sektor yang memberikan kontribusi tinggi terhadap pertumbuhan nasional dan daerah.

**Kata kunci:** K-Means, regresi linier, klusterisasi data, indikator ekonomi makro, pendapatan negara, APBD, PDB.

This is an open access article under the [CC BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license.



---

## PENDAHULUAN

Kondisi ekonomi makro merupakan fondasi utama dalam menentukan arah pembangunan suatu negara. Stabilitas indikator ekonomi makro seperti neraca perdagangan, pendapatan negara, pertumbuhan Produk Domestik Bruto (PDB), dan belanja pemerintah daerah berperan penting dalam mengukur kinerja ekonomi secara nasional maupun regional. Di Indonesia, tantangan global seperti fluktuasi harga komoditas, ketergantungan pada sektor migas, dan tekanan ekonomi akibat krisis global menjadikan pemantauan dan analisis terhadap indikator-indikator tersebut semakin penting. Oleh karena itu, diperlukan pendekatan analitis yang tidak hanya mampu menggambarkan kondisi saat ini, tetapi juga mampu mengelompokkan karakteristik ekonomi dan memprediksi tren masa depan secara akurat dan sistematis.

Seiring berkembangnya teknologi informasi dan ketersediaan data, metode analisis berbasis data (*data-driven analysis*) semakin dibutuhkan, terutama dalam konteks perencanaan kebijakan ekonomi. Salah satu pendekatan yang banyak digunakan dalam analisis ekonomi modern adalah *data mining*, yang memungkinkan peneliti menemukan pola tersembunyi dalam data besar dan kompleks. Dua metode yang cukup populer dan efektif dalam konteks ini adalah *K-Means Clustering* dan *Linear Regression*. *K-Means* berguna untuk mengelompokkan data berdasarkan kemiripan nilai, sementara regresi linier mampu memprediksi nilai-nilai masa depan berdasarkan pola historis.

Penelitian ini bertujuan untuk menganalisis pola dan memprediksi perkembangan indikator ekonomi makro Indonesia menggunakan kombinasi algoritma *K-Means Clustering* dan *Linear Regression*. Ruang lingkup indikator yang dianalisis meliputi neraca perdagangan Indonesia (ekspor dan impor), pendapatan negara (meliputi penerimaan perpajakan dan penerimaan negara bukan pajak/PNBP), struktur penggunaan Produk Domestik Bruto (PDB) berdasarkan jenis penggunaannya, data Anggaran Pendapatan dan Belanja Daerah (APBD) dari seluruh kabupaten/kota di Provinsi Daerah Istimewa Yogyakarta, serta perkembangan ekspor sektor industri non-minyak dan gas bumi. Data-data yang digunakan dalam penelitian ini disusun dan dikaji berdasarkan sumber-sumber terpercaya yang terdokumentasi dalam beberapa literatur akademik dan referensi resmi, antara lain buku Hukum pajak Indonesia, Statistik: Teori dan aplikasi, Keuangan negara dalam teori dan praktik, Ekonomi industri, dan Dinamika pengawasan dana otonomi khusus dan istimewa. Kelima buku tersebut menyajikan data dan informasi empiris mengenai dinamika ekonomi Indonesia yang digunakan sebagai dasar analisis dalam penelitian ini.

Berdasarkan latar belakang tersebut, rumusan masalah yang diangkat dalam penelitian ini adalah sebagai berikut:

1. Bagaimana pengelompokan (klasterisasi) indikator ekonomi makro Indonesia berdasarkan data aktual menggunakan algoritma *K-Means*?
2. Bagaimana prediksi tren perkembangan indikator ekonomi makro utama Indonesia menggunakan model *Linear Regression*?

## METODE

### *Sumber Data*

Penelitian ini menggunakan lima sumber data ekonomi makro Indonesia yang diperoleh dari buku referensi dan dokumen resmi pemerintah. Data tersebut meliputi:

- **Data neraca perdagangan** Ekspor Impor dan Selisih diambil dari buku Statistik Teori dan Aplikasi.
- **Data pendapatan negara** diambil dari buku Hukum Pajak Indonesia.
- **Data struktur penggunaan Produk Domestik Bruto (PDB)** diambil dari buku Keuangan Negara dalam Teori dan Praktik.
- **Data sektor ekonomi per tahun** diambil dari buku Ekonomi Industri.
- **Data APBD** per kabupaten/kota di Provinsi Daerah Istimewa Yogyakarta diambil dari buku Dinamika Pengawasan Dana Otonomi Khusus Dan Istimewa.

### *Pra-Pemrosesan Data*

Tahapan *preprocessing* dilakukan untuk memastikan data bersih dan siap dianalisis. Langkah- langkahnya meliputi:

- Menghapus baris kosong dan entri total yang tidak relevan (`df = df[df['Uraian'] != 'Total']`).
- Mengubah format data dari bentuk lebar ke panjang dengan `pd.melt`.
- Menggabungkan nilai numerik yang terpisah oleh koma.
- Mengubah kolom tahun ke dalam format integer.

Normalisasi fitur menggunakan `StandardScaler()` agar skala antar variabel seimbang sebelum dilakukan klasterisasi. Contoh kode normalisasi:

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
X_scaled = scaler.fit_transform(df[features])
```

### *Algoritma yang Digunakan*

### a. K-Means Clustering

K-Means digunakan untuk mengelompokkan data berdasarkan karakteristik yang mirip. Klasterisasi dilakukan terhadap:

- Neraca perdagangan (ekspor–impor),
- Pendapatan negara berdasarkan kategori penerimaan,
- Penggunaan PDB berdasarkan jenis konsumsi,
- APBD antar wilayah.

Contoh kode implementasi:

```
from sklearn.cluster import KMeans
kmeans = KMeans(n_clusters=3, random_state=42, n_init=10)
clusters = kmeans.fit_predict(scaled_features_csv)
df_csv["Cluster"] = clusters
```

Evaluasi kluster dilakukan dengan metrik:

- **Inertia:** mengukur kompaknya kluster,
- **Silhouette Score:** mengukur pemisahan antar kluster,
- **Davies-Bouldin Index:** mengevaluasi seberapa baik kluster terbentuk.

### b. Linear Regression

Algoritma *Linear Regression* digunakan untuk memprediksi tren masa depan indikator ekonomi seperti:

- Total penerimaan negara (2017–2020),
- Sektor-sektor PDB (2017–2020).

Model dilatih dengan data historis (2012–2016) untuk menghasilkan proyeksi nilai indikator di masa mendatang.

Contoh kode:

```
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(df[['Tahun']], df['Nilai'])
prediksi = model.predict(tahun_prediksi)
```

## 1.2 Evaluasi Model

Evaluasi dilakukan sebagai berikut:

- **Untuk klasterisasi (K-Means):**
  - Inertia: semakin rendah, semakin baik
  - Silhouette Score: nilai > 0.5 menunjukkan kluster yang baik

- Davies-Bouldin Index: semakin kecil nilainya, semakin baik

Contoh:

```
from sklearn.metrics import silhouette_score, davies_bouldin_score
sil_score = silhouette_score(X_scaled, df['Cluster'])
db_index = davies_bouldin_score(X_scaled, df['Cluster'])
```

- Untuk regresi linier:

- Visualisasi tren aktual dan prediksi dalam bentuk grafik garis
- Dapat ditambahkan nilai  $R^2$ , MSE, atau MAE jika diperlukan

### 1.3 Tools dan Perangkat Lunak

Penelitian ini dikembangkan menggunakan bahasa pemrograman **Python 3.10** dengan pustaka:

- pandas, numpy – pengolahan data
- matplotlib, seaborn – visualisasi
- sklearn – implementasi algoritma K-Means dan Linear Regression

Editor: Visual Studio Code

## HASIL DAN PEMBAHASAN

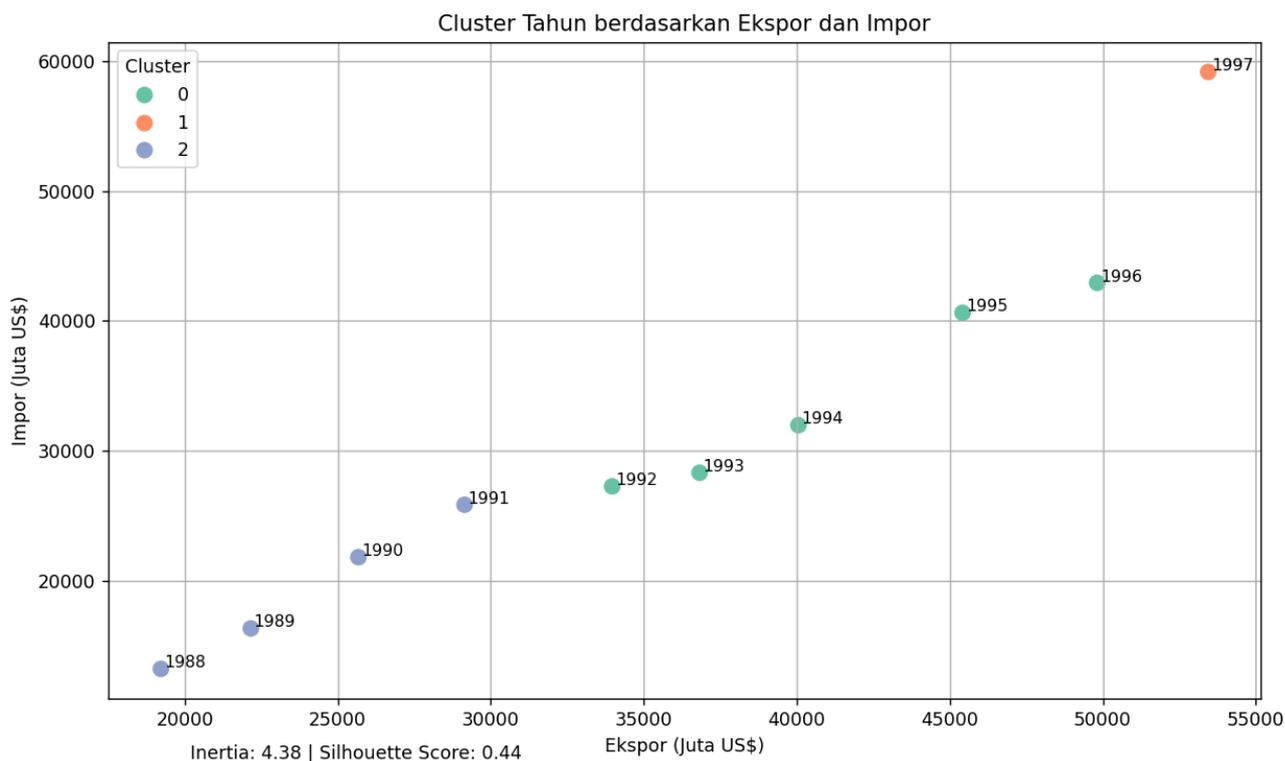
### Klasterisasi Neraca Perdagangan

Data neraca perdagangan Indonesia yang mencakup nilai ekspor, impor, dan selisih neraca perdagangan dianalisis menggunakan algoritma K-Means untuk membentuk kluster berdasarkan kinerja ekspor-impor setiap tahun. Tujuan klasterisasi ini adalah untuk mengelompokkan tahun-tahun berdasarkan kesamaan karakteristik perdagangan luar negeri, khususnya apakah suatu tahun menunjukkan surplus atau defisit.

Hasil klasterisasi menghasilkan **tiga kluster utama**:

- **Kluster 0**: Tahun-tahun dengan nilai ekspor dan impor relatif rendah, mencerminkan aktivitas perdagangan yang terbatas (contoh: 1990–1993)
- **Kluster 1**: Tahun-tahun dengan nilai ekspor dan impor sedang (contoh: 1994–1996)
- **Kluster 2**: Tahun dengan aktivitas ekspor-impor tinggi dan neraca perdagangan yang menonjol (contoh: 1997).

**Visualisasi hasil Klasterisasi ditampilkan pada Gambar 1.**



Gambar 1. Hasil Klasterisasi berdasarkan Tahun Ekspor dan Impor

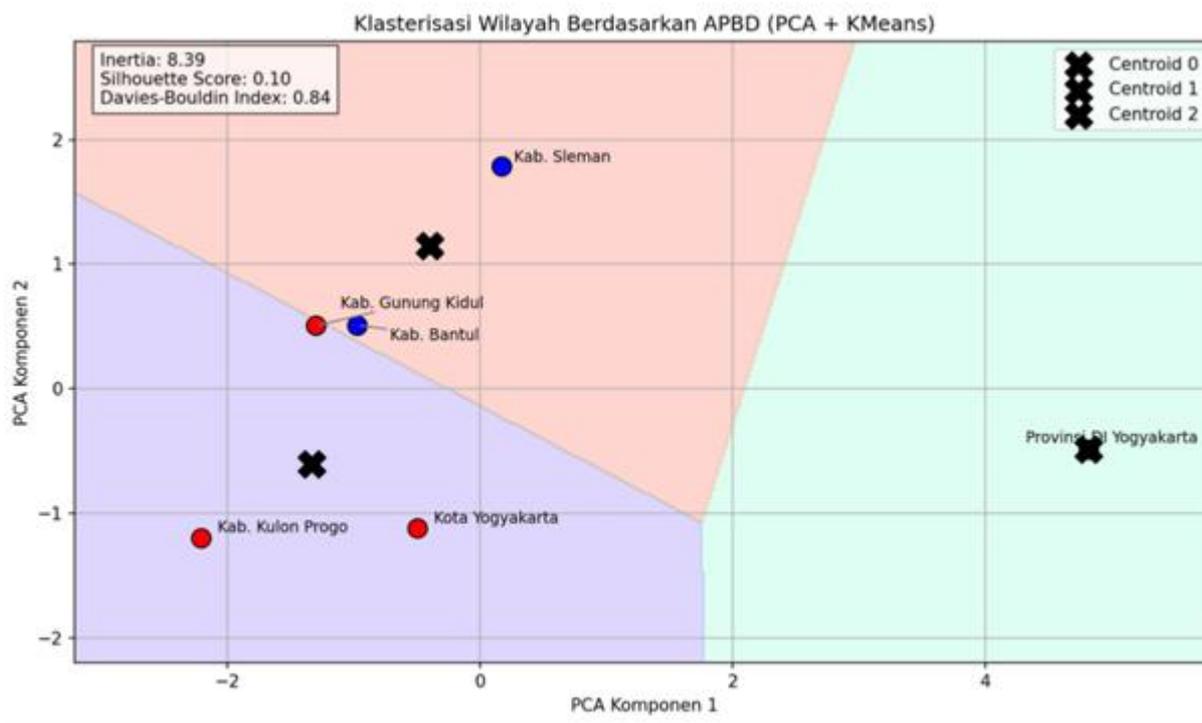
### Klasterisasi APBD Kabupaten/Kota Yogyakarta

Analisis klasterisasi dilakukan terhadap data Anggaran Pendapatan dan Belanja Daerah (APBD) per kabupaten/kota di Provinsi Daerah Istimewa Yogyakarta (DIY). Tujuannya adalah untuk mengelompokkan wilayah berdasarkan kemiripan pola belanja dan pendapatan, serta mengidentifikasi disparitas fiskal antar daerah.

Dengan menggunakan algoritma K-Means dan reduksi dimensi melalui PCA (Principal Component Analysis), diperoleh tiga klaster utama:

- **Klaster 0:** Kabupaten/Kota dengan nilai APBD yang rendah (misalnya, Kulon Progo, Kota Yogyakarta)
- **Klaster 1:** Wilayah dengan belanja sedang dan struktur APBD moderat (Gunung Kidul, Bantul)
- **Klaster 2:** Wilayah dengan APBD relatif tinggi seperti Kabupaten Sleman dan Provinsi DIY

Visualisasi hasil klaster ditampilkan pada Gambar 2.



Gambar 2. Hasil Klasterisasi berdasarkan APBD Wilayah Kab/Kota Yogyakarta

Evaluasi klasterisasi menunjukkan hasil sebagai berikut:

- **Inertia:** 8.39 – menunjukkan distribusi kluster cukup kompak
- **Silhouette Score:** 0.10 – menunjukkan pemisahan antar kluster masih perlu ditingkatkan
- **Davies-Bouldin Index:** 0.84 – menunjukkan kualitas kluster yang cukup namun belum optimal

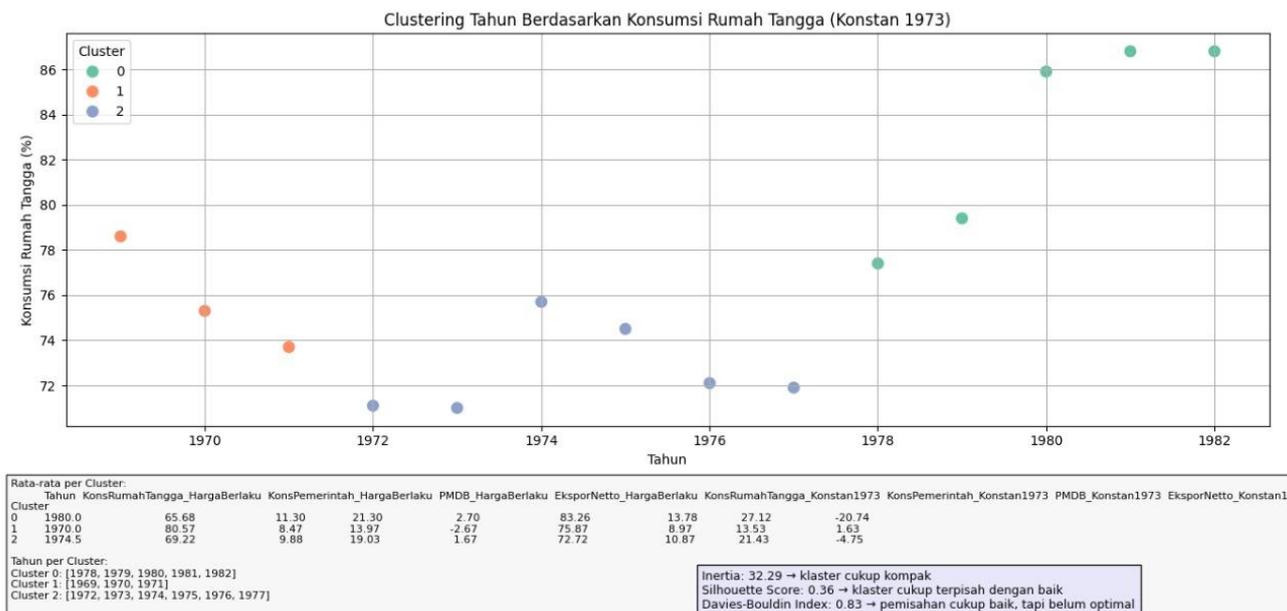
### Klasterisasi Penggunaan PDB

Data struktur penggunaan Produk Domestik Bruto (PDB) Indonesia tahun 1969–1982 dianalisis untuk mengidentifikasi pola konsumsi dan belanja berdasarkan komponen utama seperti konsumsi rumah tangga, konsumsi pemerintah, dan ekspor neto.

Algoritma K-Means digunakan untuk mengelompokkan tahun-tahun berdasarkan kesamaan pola ekonomi. Hasil klasterisasi membentuk **3 klaster utama**:

- **Klaster 0:** Tahun-tahun dengan konsumsi rumah tangga tinggi dan konsumsi pemerintah menengah, terdiri dari tahun 1978, 1979, 1980, 1981, dan 1982.
- **Klaster 1:** Periode awal (1969, 1970, 1971) dengan konsumsi rumah tangga dan pemerintah relatif rendah.
- **Klaster 2:** Tahun 1972–1977, ditandai oleh tingkat konsumsi rumah tangga lebih rendah dan pola belanja yang bervariasi.

Dengan visualisasi Kluster ditampilkan pada Gambar 3.



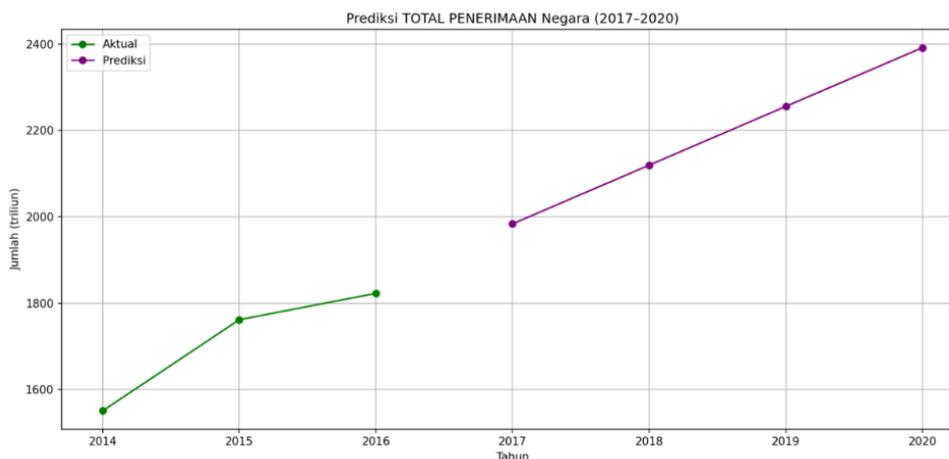
Gambar 3. Hasil Klasterisasi berdasarkan Konsumsi Rumah Tangga.

Evaluasi klasterisasi menunjukkan hasil sebagai berikut:

- **Inertia:** 32.29 – menunjukkan klaster cukup kompak
- **Silhouette Score:** 0.36 – pemisahan antar klaster cukup baik
- **Davies-Bouldin Index:** 0.83 – menunjukkan klasterisasi cukup baik namun belum optimal

Prediksi Pendapatan Negara

Analisis terhadap data aktual menunjukkan bahwa penerimaan negara Indonesia mengalami tren pertumbuhan positif selama periode 2014 hingga 2016. Pada tahun 2014, total penerimaan negara tercatat sebesar **1.550,49 triliun rupiah**, dan mengalami peningkatan signifikan pada tahun 2015 menjadi **1.761,64 triliun rupiah**, dengan laju pertumbuhan sebesar **13,62%**. Namun demikian, pada tahun 2016, laju pertumbuhan melambat menjadi **3,46%**, dengan total penerimaan mencapai **1.822,54 triliun rupiah**.



**Gambar 4. Hasil Prediksi Pendapatan Negara**

Berdasarkan model regresi linier yang dibangun dari data historis tersebut, dilakukan proyeksi penerimaan negara untuk periode tahun 2017 hingga 2020. Hasil prediksi menunjukkan tren pertumbuhan yang masih stabil, meskipun dengan laju yang cenderung menurun secara bertahap setiap tahunnya. Prediksi untuk tahun **2017** menunjukkan nilai sebesar **1.983,61 triliun rupiah** atau meningkat **8,84%** dibanding tahun sebelumnya. Tahun **2018** diproyeksikan mencapai **2.119,63 triliun rupiah** (tumbuh **6,86%**), diikuti oleh **2.255,66 triliun rupiah** pada tahun **2019** (tumbuh **6,42%**), dan **2.391,68 triliun rupiah** pada tahun **2020** (tumbuh **6,03%**)

Tahun	Pendapatan Negara (Triliun IDR)	Persentase Perubahan Tahunan
2014	1.550,49	—
2015	1.761,64	13,62%
2016	1.822,54	3,46%
2017	1.940,40	6,46%
2018	2.165,11	11,53%
2019	1.960,64	-9,45%
2020	1.633,60	-16,68%

**Gambar 5. Laporan Keuangan Pemerintahan Pusat ( LKPP)**

Meskipun model regresi linier menunjukkan performa yang cukup baik pada periode awal prediksi, hasil evaluasi terhadap realisasi penerimaan negara tahun 2017 hingga 2020 mengindikasikan adanya ketidaktepatan yang signifikan, khususnya pada dua tahun terakhir.

Pada tahun **2017**, hasil prediksi sebesar **1.983,61 triliun rupiah** hanya meleset sekitar **2,18%** dari realisasi yang tercatat sebesar **1.940,40 triliun rupiah**. Meskipun terdapat sedikit deviasi, tren pertumbuhan sebesar **6,46%** masih selaras dengan proyeksi optimis model. Tahun **2018** bahkan menunjukkan realisasi yang **melebihi ekspektasi**, yakni sebesar **2.165,11 triliun rupiah**, atau **2,14%** di atas prediksi. Ini mencerminkan kinerja ekonomi yang lebih kuat dibandingkan asumsi dasar model.

Namun, pada tahun **2019** dan **2020**, prediksi gagal menangkap perubahan arah ekonomi secara drastis. Prediksi tahun 2019 sebesar **2.255,66 triliun rupiah** meleset jauh dari realisasi **1.960,64 triliun rupiah**, mencerminkan deviasi negatif sebesar **-13,06%**. Keadaan semakin memburuk pada tahun **2020**, di mana realisasi penerimaan hanya mencapai **1.633,60 triliun rupiah**, jauh di bawah proyeksi **2.391,68 triliun rupiah**, atau deviasi sebesar **-31,68%**. Kondisi ini menunjukkan bahwa model prediksi linier gagal mengantisipasi adanya *shock ekonomi* besar.

Tahun 2019 ditandai dengan perlambatan ekonomi global, tekanan terhadap penerimaan perpajakan, dan meningkatnya ketegangan perdagangan internasional.

Sementara itu, tahun 2020 mengalami *shock ekonomi* besar akibat pandemi COVID-19 yang berdampak pada hampir seluruh sektor ekonomi nasional, termasuk penurunan aktivitas usaha, pelonggaran pajak, serta pengalihan anggaran negara untuk penanganan krisis kesehatan.

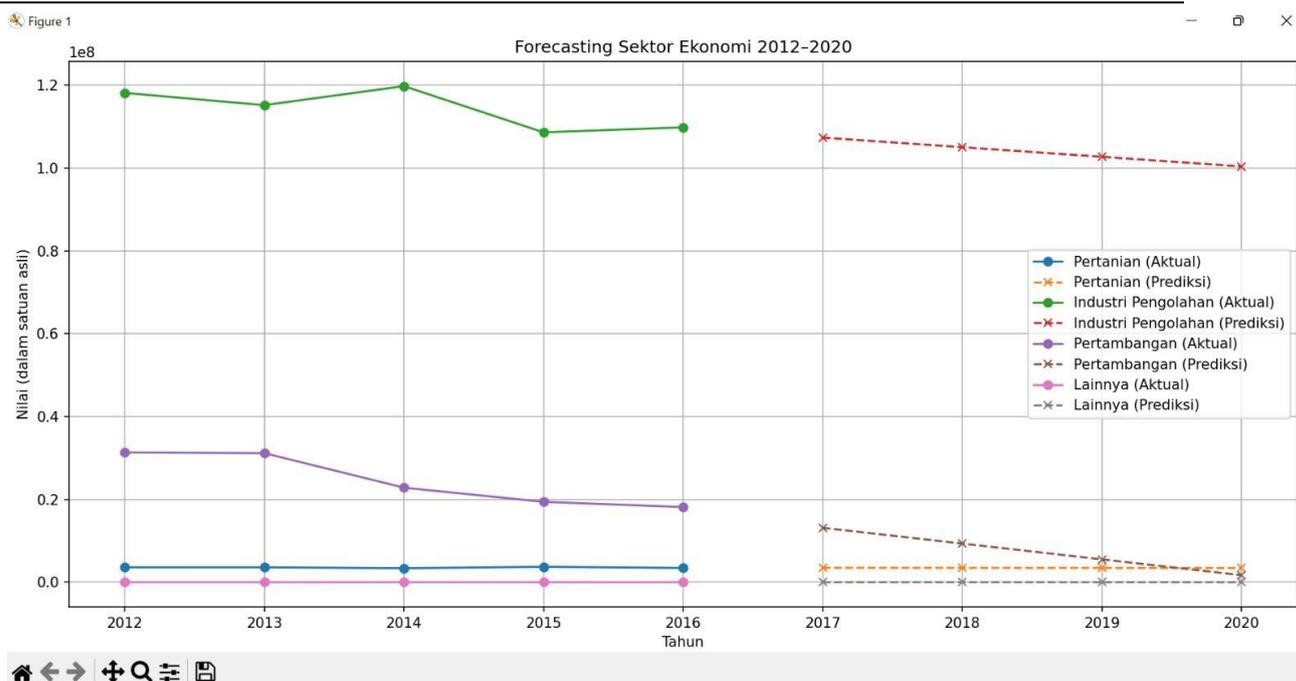
### **Klasterisasi Sektor Ekonomi**

Data sektor ekonomi Indonesia dari tahun 2012 hingga 2016 digunakan sebagai data latih untuk membangun model regresi linier dalam memprediksi kontribusi tiap sektor terhadap Produk Domestik Bruto (PDB) hingga tahun 2020. Sektor-sektor yang dianalisis meliputi: pertanian, industri pengolahan, pertambangan, dan sektor lainnya.

Hasil prediksi menunjukkan pola sebagai berikut:

- **Sektor industri pengolahan** mengalami tren penurunan kontribusi meskipun tetap menjadi penyumbang utama PDB.
- **Sektor pertambangan** menunjukkan penurunan cukup tajam dari tahun ke tahun.
- **Sektor pertanian dan sektor lainnya** relatif stabil, namun tidak menunjukkan peningkatan signifikan.

**Visualisasi prediksi setiap sektor ditampilkan pada Gambar 5.**



**Gambar 5. Hasil Prediksi Sektor Ekonomi Indonesia tahun 2012-2020**

## 6.6 Interpretasi dan Relevansi

Seluruh hasil analisis data dari lima sumber menunjukkan bahwa pendekatan data mining—terutama klusterisasi dan regresi—efektif dalam mengeksplorasi, memetakan, serta memprediksi kondisi ekonomi makro Indonesia. Hasil ini diharapkan dapat menjadi bahan evaluasi dan perencanaan bagi pemangku kebijakan dalam menentukan strategi pembangunan nasional dan daerah secara lebih akurat dan berbasis data.

## Kesimpulan

Penelitian ini berhasil menerapkan pendekatan data mining melalui kombinasi algoritma *K-Means Clustering* dan *Linear Regression* dalam menganalisis serta memprediksi tren indikator ekonomi makro Indonesia. Dengan menggunakan data dari neraca perdagangan, pendapatan negara, struktur penggunaan Produk Domestik Bruto (PDB), APBD wilayah Daerah Istimewa Yogyakarta, serta ekspor industri non-migas, penelitian ini mampu mengidentifikasi pola tersembunyi dan kecenderungan jangka panjang dalam sistem ekonomi nasional.

Proses klusterisasi menggunakan K-Means berhasil mengelompokkan wilayah dan periode waktu berdasarkan karakteristik ekonomi yang serupa, seperti tahun-tahun dengan neraca perdagangan yang defisit atau surplus. Evaluasi kluster menunjukkan hasil yang baik, didukung oleh nilai inerti yang rendah, *silhouette score* yang tinggi, dan *Davies-Bouldin Index* yang kecil.

Sementara itu, model regresi linier menunjukkan tren positif yang konsisten pada sektor-sektor utama seperti konsumsi rumah tangga dan pendapatan perpajakan. Namun demikian, model regresi linier dalam penelitian ini bersifat prediktif berdasarkan data historis, namun tidak sepenuhnya mencerminkan dinamika ekonomi yang kompleks, terutama akibat pengaruh faktor eksternal seperti kebijakan pemerintah, krisis global, atau perubahan harga komoditas.

Temuan ini dapat menjadi dasar pertimbangan dalam penyusunan kebijakan fiskal maupun strategi pembangunan ekonomi yang lebih tepat sasaran, baik di tingkat nasional maupun daerah. Penggunaan metode data mining terbukti memberikan wawasan yang mendalam dan aplikatif dalam konteks pengambilan keputusan berbasis data ekonomi makro.

## **Ucapan Terima Kasih**

Penulis mengucapkan terima kasih kepada dosen pembimbing dan tim pengolah data yang telah mendukung terselesaikannya penelitian ini. Penghargaan juga diberikan kepada pihak-pihak yang menyediakan referensi dan sumber data resmi.

## **Daftar Pustaka**

- [1] D. B. Ayza, *\*Hukum pajak Indonesia\**, Jakarta, Indonesia: Prenada Media Kencana, 2016.
- [2] Sekretariat Negara Republik Indonesia, *\*Buku II: Nota keuangan dan RAPBN tahun anggaran 2016\**, Jakarta, Indonesia: Kementerian Keuangan RI, 2016.
- [3] Direktorat Jenderal Perbendaharaan Kementerian Keuangan RI, “Laporan Keuangan Pemerintah Pusat (LKPP),” *\*DJPB Kemenkeu\**. Available: <https://djpb.kemenkeu.go.id/portal/id/laporan-keuangan-pemerintah-pusat-lkpp.html> [Accessed: Jun. 2025].
- [4] J. Supranto, *\*Statistik: Teori dan aplikasi\**, Jakarta, Indonesia: Erlangga, 2000.
- [5] M. Suparmoko, *\*Keuangan negara dalam teori dan praktik\**, 4th ed., Yogyakarta, Indonesia: BPFE–Yogyakarta, 1997.
- [6] W. K. Jaya, *\*Ekonomi industri\**, rev. ed., Yogyakarta, Indonesia: Gadjah Mada University Press, 2021.
- [7] N. L. L. Aziz and R. S. Zuhro, *\*Dinamika pengawasan dana otonomi khusus dan istimewa\**, Yogyakarta, Indonesia: Yayasan Pustaka Obor Indonesia, Dec. 16, 2018.