

**“Analisis Kondisi Indonesia di Tahun 2024 Terhadap Kesiapan
Bonus Demografi Menggunakan Pendekatan *Statistic* dan
Machine Learning”**



Disusun oleh :

Ibrahim Frosly Alesandro	NIM. 23031554021
Muhammad Faiz Munif Billah	NIM. 23031554028
Gesang Nur Zamroji	NIM. 23031554145

Dosen pengampu:
Riskyana Dewi Intan P, M.Kom.

**PROGRAM STUDI SAINS DATA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI SURABAYA
2025**

LAPORAN PROJECT PENAMBANGAN DATA

SEMESTER GENAP 2024/2025

IDENTITAS PROYEK	
Judul	Analisis Kondisi Indonesia di Tahun 2024 Terhadap Kesiapan Bonus Demografi Menggunakan Pendekatan <i>Statistic</i> dan <i>Machine Learning</i>
Topik	Bonus Demografi
Identitas Penyusun	1. Ibrahim Frosly Alesandro (23031554021) 2. Muhammad Faiz Munif Billah (23031554028) 3. Gesang Nur Zamroji (23031554145)
Kelas	2023B

1. PENDAHULUAN

Pendahuluan penelitian tidak lebih dari 1000 kata yang terdiri dari:

- A. Latar belakang dan rumusan permasalahan yang akan diteliti
- B. Pendekatan pemecahan masalah

1.1 Latar Belakang (min. 250 kata)

Indonesia diproyeksikan kedepannya akan mengalami bonus demografi pada tahun 2030 keatas, dimana populasi usia produktif (15 - 64 tahun) akan mencapai target tersebut dan mendominasi (Adriani, D. & Yustini, T., 2021). Fenomina ini nantinya menjadi momentum berharga untuk pertumbuhan ekonomi dan pembangunan berkelanjutan, namun juga menyimpan tantangan besar jika kesiapan yang dilakukan kurang baik. Salah satu tantangannya adalah kualitas SDM yang bisa dilihat dari Indeks Pembangunan Manusia, tingkat pendidikan, kesehatan, serta lapangan kerja yang sesuai dengan kompetensi (Sasmita, N. R. et al, 2024).

Bonus demografi ini juga berpotensi menjadi pendorong pertumbuhan ekonomi yang signifikan jika ditangani dan dikelola secara tepat dan sesuai dengan karakteristik negara. Namun, tantangan besar terkadang muncul dari kualitas sumber daya manusia yang rendah, tingginya pengangguran apalagi dengan lulusan perguruan tinggi, serta ketidaksesuaian antara pendidikan dan kebutuhan pasar kerja sekarang (Mian, L., et al., 2022).

Faktor-faktor yang berpengaruh pada kesiapan daerah dalam menghadapi bonus demografi wajib dipahami oleh pemerintah dan yang berkepentingan di bidangnya. Dengan cara menganalisis faktor-faktor seperti Indeks Pembangunan Manusia (IPM), Tingkat Partisipasi Angkatan Kerja (TPAK), tingkat pengangguran terbuka, harapan hidup, serta tingkat kemiskinan dan ketahanan pangan bisa menjadi suatu solusi (Ulma & Julia, 2022) (Aritonang, 2022). Hasil analisis data tersebut dapat menemukan pola dan wilayah berdasarkan analisis demografis dengan karakteristik sosial-ekonomi.

Menggunakan metode clustering dan regression dapat menjadi pendekatan yang paling relevan dalam mengelompokkan wilayah di Indonesia dan menemukan faktor apa yang paling berpengaruh terhadap bonus demografi berdasarkan karakteristik variable (Fitriyah et al., 2021). Indikator seperti IPM dan TPAK bisa menjadi faktor paling berpengaruh dalam bonus demografi (Ulma & Julia, 2022) (Aritonang, 2022). Hasil dari analisis ditujukan untuk menjadi ide dalam membuat kebijakan yang dibutuhkan, terutama untuk masing masing wilayah. Dengan begitu, peluang dari bonus demografi tidak hanya menjadi potensi, tetapi benar-benar menjadi kekuatan nyata dalam mendorong pembangunan berkelanjutan dan pertumbuhan ekonomi nasional.

1.2. Rumusan Masalah dan Tujuan

1.2.1 Rumusan Masalah

1. Bagaimana tingkat kesiapan masing-masing provinsi di Indonesia dalam menghadapi bonus demografi?
2. Bagaimana provinsi-provinsi dapat dikelompokkan berdasarkan kesamaan karakteristik kesiapan mereka?

3. Faktor-faktor apa saja yang paling berpengaruh terhadap kesiapan bonus demografi suatu wilayah?
4. Rekomendasi strategis apa yang dapat diberikan untuk meningkatkan kesiapan wilayah dalam menyambut bonus demografi?

1.2.2 Tujuan

1. Menganalisis kesiapan indonesia dalam menghadapi bonus demografi berdasarkan beberapa kondisi provinsi tertentu.
2. Mengelompokkan provinsi atau wilayah berdasarkan tingkat kesiapan mereka.
3. Menganalisis hubungan dan pengaruh antar variabel penentu kesiapan bonus demografi dengan menggunakan metode regresi.
4. Memberikan rekomendasi informasi berdasarkan hasil proyek untuk optimalisasi bonus demografi di indonesia.

2. Metodologi

Metodologi atau cara untuk mencapai tujuan yang telah ditetapkan ditulis tidak melebihi 1000 kata. Bagian ini berisi metode pre-processing dan/atau metode post processing yang dilengkapi dengan diagram alir penelitian yang menggambarkan apa yang sudah dilaksanakan dan yang akan dikerjakan selama waktu yang diusulkan. Format diagram alir dapat berupa file JPG/PNG. Metode penelitian harus dibuat secara utuh dengan penahapan yang jelas.

2.1 Eksplorasi Dataset

Pemahaman dataset yang dimiliki

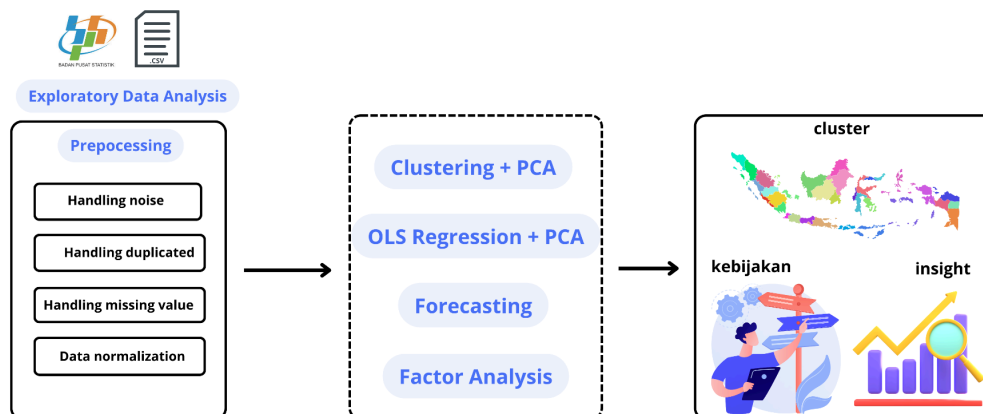
Dataset yang digunakan adalah kumpulan dari beberapa data Badan Pusat Statistik yang kemudian diintegrasikan menjadi satu dataset yang akan digunakan dalam proyek ini. Dataset memiliki jumlah baris sebanyak 38 dimana setiap baris mewakili satu provinsi di Indonesia, memiliki kolom sebanyak 17 dimana di setiap baris dan kolom tidak ada missing value dan tidak ada duplikat.

Variabel dalam dataset :

- a. Provinsi : Nama wilayah/provinsi (non - null, object)
- b. IPM : Indeks Pembangunan Manusia (non-null, float64)
- c. rata_lama_sekolah : Rata-rata lama sekolah (non-null, float64)
- d. TPAK : Tingkat Partisipasi Angkatan Kerja (non-null, float64)
- e. indek_tingkat_pengangguran_terbuka (non-null, float64)
- f. Angka_harapan_hidup_perempuan (non-null, float64)
- g. Angka_harapan_hidup_lakilaki (non-null, float64)
- h. Persentase_penduduk_miskin (non-null, float64)
- i. Penduduk_usia_15_tahun_keatas (non-null, int64)
- j. Persentase_ketidakcukupan_konsumsi_pangan (non-null, float64)
- k. Pengeluaran per kapita laki-laki (non-null, int64)
- l. Pengeluaran per kapita perempuan (non-null, int64)
- m. Rata-rata persentase pengeluaran per kapita laki-laki dan perempuan (non-null, int64)
- n. HLS laki-laki (non-null, float64)
- o. HLS perempuan (non-null, float64)
- p. Nikah (non-null, int64)
- q. Laju pertumbuhan penduduk per tahun (non-null, float64)

2.2. Langkah Penelitian

2.2.1 Diagram Alir



2.2.2 Exploratory Data Analysis

Exploratory Data Analysis adalah proses awal dalam melakukan analisis data yang bertujuan untuk melihat dan memahami struktur data dan karakteristik dataset baik melalui statistik deskriptif maupun visualisasi.

2.2.3 Statistik Deskriptif

Statistik deskriptif disini nantinya akan dilakukan untuk mengetahui informasi singkat dan karakteristik suatu kumpulan data yang akan digunakan.

2.2.4 Visualisasi

Visualisasi data awal disini diperlukan untuk melihat berbagai karakteristik data dalam bentuk visual sehingga mudah untuk dipahami dan dibaca. Proyek ini nantinya akan menggunakan visualisasi berupa distribusi data dan heatmap untuk melihat korelasi antar fiturnya.

2.2.5 Clustering

Clustering adalah metode unsupervised learning untuk mengelompokkan data berdasarkan kesamaan maupun perbedaan yang ada di antara data tersebut. Metode ini bisa dikatakan relevan untuk analisis kesiapan provinsi di Indonesia dalam menghadapi bonus demografi.

Pengelompokan provinsi berdasarkan pada fitur yang ada di dataset sehingga provinsi dengan tingkat kesiapan yang sama atau hampir sama dikelompokkan

menjadi cluster yang sama. Hal tersebut dapat menentukan wilayah yang siap dan tidak siap dalam menghadapi bonus demografi. Mengelompokkan wilayah juga bisa menggunakan fitur indikator human capital dan kualitas hidup, misalnya pada clustering regional di Rusia menggunakan indikator human development index (HDI) (Bukharova et al., 2019).

2.2.6 Regresi

Dalam penelitian ini, metode yang digunakan adalah regresi untuk menganalisis hubungan variabel independen (prediktor/mempengaruhi) dengan variabel dependen (target) yaitu Indeks Pembangunan Manusia (IPM) atau Tingkat Partisipasi Angkatan Kerja (TPAK). Variabel seperti rata-rata lama sekolah, angka harapan hidup, tingkat pengangguran, dan persentase penduduk miskin mempengaruhi IPM atau TPAK di setiap provinsi di Indonesia dapat dilihat menggunakan metode regresi (Munawaroh, Gandhiadi, & Nilakusmawati, 2023)

1. Variabel

- Variabel Dependen: IPM atau TPAK.
- Variabel Independen: Rata-rata Lama Sekolah, Angka Harapan Hidup (laki-laki dan perempuan), Tingkat Pengangguran Terbuka, Persentase Penduduk Miskin, dan penduduk usia 15 tahun keatas (Samudra & Wahed, 2023).

2. Evaluasi Model

- Nilai R-squared untuk mengukur seberapa besar variasi target yang dapat dijelaskan oleh model (Rohma et al., 2024).
- P-value untuk menguji signifikansi masing-masing variabel prediktor (Rohma et al., 2024; Munawaroh, Gandhiadi, & Nilakusmawati, 2023).
- Uji Asumsi Regresi: Heteroskedastisitas, Normalitas residual, multikolinearitas (VIF) dapat diuji untuk memastikan validitas model (Munawaroh, Gandhiadi, & Nilakusmawati, 2023).

3. Interpretasi dan Kesimpulan

Untuk mengetahui variabel yang paling berpengaruh dan berkontribusi besar terhadap IPM dan TPAK bisa dianalisis melalui hasil regresi. Dari kesimpulan tersebut, dapat ditentukan kebijakan apa yang akan dilakukan dalam menghadapi bonus demografi.

2.2.7 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) dapat untuk analisis data untuk mereduksi dimensi data yang tinggi dan tetap mempertahankan informasi penting, sehingga

meningkatkan interpretabilitas dan efisiensi komputasi. PCA mengubah variabel asli yang saling berkorelasi menjadi sejumlah komponen utama yang tidak berkorelasi, mengurangi multikolinearitas dan menyederhanakan struktur data. PCA dapat dikatakan efektif dalam mengidentifikasi pola dominan dan fitur penting dalam data kompleks, hal tersebut dapat digunakan untuk membantu pengambilan keputusan berbasis data di berbagai bidang (Jolliffe & Cadima, 2016; Tato & Yasin, 2025).

Dalam analisis regresi linear, multikolinearitas antar variabel independen dapat menyebabkan masalah dalam estimasi koefisien menjadi tidak stabil dan interpretasi model menjadi tidak dapat digunakan. Manfaat PCA dapat mereduksi dimensi data dengan membentuk variabel baru (principal components) yang saling ortogonal (setiap PC tidak bergantung linear dengan PC lain) dan bebas korelasi, hal tersebut dapat menghilangkan multikolinearitas tanpa harus menghapus variabel penting. Variable yang dihasilkan PCA ini dapat digunakan sebagai prediktor dalam model regresi OLS, yang membuat model menjadi lebih stabil dan interpretatif (Larasati et al., 2020; Pujilestari, Dwidayati, & Sugiman, 2017).

2.2.8 Forecasting

Forecasting yang digunakan dalam analisis ini dengan cara menggunakan deret waktu dengan pemahaman kontekstual terhadap kondisi sosial-ekonomi masing-masing provinsi (Avelin, Viitasaari, & Malo, 2025). Proses dimulai dengan pengumpulan data historis dari tahun 2010 hingga 2024 terkait tiga indikator utama pembangunan, yaitu Indeks Pembangunan Manusia (IPM), 2018 hingga 2024 untuk Tingkat Pengangguran Terbuka (TPT), dan 2010 hingga 2024 untuk Persentase Penduduk Miskin (PPM). Walaupun data yang dikumpulkan hanya sedikit, data juga perlu dianalisis untuk mengidentifikasi pola tren, fluktuasi, dan anomali yang dapat mempengaruhi akurasi prediksi (Cruz-Nájera et al., 2022). Beberapa model forecasting juga diterapkan, antara lain **Prophet**, **Holt-Winters**, **Double Exponential Smoothing**, **ARIMA**, **Polynomial Regression**, dan **Support Vector Regression (SVR)** untuk menemukan model paling cocok dengan data historis yang sedikit (Cruz-Nájera et al., 2022). Pemilihan model bertujuan untuk membandingkan tingkat sensitivitas dan akurasi masing-masing metode terhadap variasi data historis, termasuk adanya lonjakan tajam atau penurunan tiba-tiba yang terjadi akibat faktor eksternal seperti pandemi COVID-19 atau pemekaran wilayah (pembentukan provinsi baru).

2.2.9 Factor Analysis

Analisis Faktor adalah teknik statistik multivariat yang diterapkan pada satu set

variabel ketika peneliti tertarik untuk menentukan variabel mana dalam set tersebut yang membentuk subset logis yang relatif independen satu sama lain (Shrestha, N., 2021). Selanjutnya akan dilakukan *Factor Analysis* untuk mengidentifikasi faktor-faktor yang mendasari variabel dengan mengelompokkan variabel-variabel yang terkait dalam faktor yang sama (Shrestha, N., 2021).

Variabel-variabel yang digunakan nantinya akan dikelompokkan ke dalam suatu bidang seperti pendidikan, kesehatan, ketenagakerjaan, dan ekonomi. Pengujian statistik yang kami gunakan ada KMO dan Bartlett's Test. Nilai KMO kurang dari 0,6 menunjukkan bahwa pengambilan sampel tidak memadai dan tindakan perbaikan harus diambil. Jika nilainya kurang dari 0,5, hasil analisis faktor jelas tidak akan cocok untuk analisis data (Shrestha, N., 2021). Begitu juga untuk Bartlett's Test, nilai signifikan $< 0,05$ menunjukkan bahwa analisis faktor mungkin bermanfaat untuk kumpulan data (Shrestha, N., 2021). Setelah itu akan dilakukan analisis lebih lanjut untuk menemukan insight yang sesuai dengan masing-masing provinsi.

2.2.10 Insight

Pada pca:

1. PCA membantu menyederhanakan data yang rumit. Dengan banyaknya indikator sosial-ekonomi, PCA dipakai untuk merangkum semua informasi ke dalam beberapa komponen utama yang lebih ringkas dan mudah dianalisis (Mesa, Correa Velez, & Cardenas Alzate, 2018).
2. Untuk dipilih beberapa komponen utama. Beberapa komponen ini bisa menjelaskan sekitar 90% informasi dari seluruh data, jadi sudah cukup mewakili kondisi nyata tanpa harus melihat semua variabel satu per satu.
3. Setiap komponen dapat menggambarkan tema tertentu. Masing-masing PC punya makna tersendiri yang bisa membantu memahami karakter suatu wilayah (Vyas & Kumaranayake, 2006).
4. Insight pada PCA, dapat digunakan untuk membantu pemerintah dalam mengambil keputusan dalam menentukan daerah mana yang membutuhkan bantuan dalam bidang pendidikan, peningkatan gizi, atau pengurangan pengangguran
5. Dengan PCA, kita bisa melihat pola tersembunyi di balik data, sehingga pengambilan keputusan bisa lebih tepat dan berbasis data yang objektif (ESRI, 2023).

Pada clustering:

1. Menemukan Pola

Pola dan perbedaan antar wilayah dapat dibedakan secara objektif. Dalam pengelompokan provinsi, hasilnya bisa dilihat bahwa misal wilayah timur memiliki kesiapan yang lebih rendah dari barat. Cluster analysis dapat menemukan karakteristik pada setiap kelompok, sehingga membantu memahami perbedaan antar daerah dan mengetahui faktor-faktor berpengaruh terhadap kesiapan suatu daerah untuk menghadapi bonus demografi (Bukharova et al., 2019).

2. Membantu Perencanaan Kebijakan

Pengelompokan dapat membuat keputusan pemerintah dalam membuat kebijakan berdasarkan data dan terarah. Hasil pada setiap clustering dapat digunakan untuk memberikan bantuan atau kebijakan secara tepat berdasarkan kebutuhan dan tingkat kesiapan wilayah. Kebijakan yang diambil berdasarkan pengelompokan hasil clustering sangat penting dalam bidang pengembangan sumber daya manusia (Bukharova et al., 2019).

Pada Regresi:

1. Variabel Mana yang Paling Berpengaruh

Dilihat dari koefisien terbesar (positif atau negatif) dan $p\text{-value} < 0.05$ menandakan variabel tersebut signifikan secara statistik dan berpengaruh besar terhadap variabel dependen (Frost, 2019; James et al., 2013)

2. Signifikansi Statistik dari Masing-Masing Fitur

- P-value:
 - Jika $p\text{-value} > 0.05 \rightarrow$ variabel tersebut tidak signifikan, mungkin bisa dihapus dari model.
 - Cek apakah input yang dianggap penting (misalnya TPAK) ternyata tidak signifikan secara statistik.

3. Kekuatan Model (R-squared)

- Semakin tinggi R^2 , semakin baik model menjelaskan variasi data (James et al., 2013).
 - $R^2 = 0.85 \rightarrow$ model sangat baik
 - $R^2 < 0.5 \rightarrow$ model kurang baik, mungkin ada variabel penting yang belum dimasukkan.

4. Ketidakakuratan

- Bandingkan prediksi dengan nilai aktual.

- Cek residual (kesalahan prediksi): Provinsi mana yang jauh berbeda antara prediksi dan realitasnya (James et al., 2013).

Misal Prediksi IPM untuk Papua Barat jauh lebih tinggi dibandingkan realisasinya. Hal ini menunjukkan adanya faktor lokal yang tidak terwakili oleh variabel model.

5. Arah Hubungan

- Koefisien positif: semakin tinggi nilai variabel → IPM semakin naik
- Koefisien negatif: semakin tinggi nilai variabel → IPM menurun

Bisa juga menggunakan Interpretasi Partial Dependence Plot (PDP) untuk melihat bagaimana satu variabel mempengaruhi hasil prediksi, dengan asumsi variabel lain tetap.

Pada Forecasting:

1. Dapat mengetahui pola atau trend waktu pada variabel yang berpengaruh pada bonus demografi di setiap provinsi.
2. Untuk memprediksi nilai pada variabel yang berpengaruh dalam bonus demografi untuk 6 tahun kedepan dari 2025.

Pada *Factor Analysis* :

1. Mengelompokkan variabel demografi (seperti usia, jenis kelamin, pendidikan, atau pendapatan) ke dalam faktor-faktor yang menunjukkan pola tertentu.
2. Mengungkap hubungan tersembunyi antara variabel demografi. Sebagai contoh, jika "IPM" dan "TPAK" dimuat ke dalam faktor yang sama, ini bisa menunjukkan bahwa orang-orang di daerah pedesaan cenderung memiliki tingkat pendidikan tertentu.
3. FA dapat menunjukkan bagaimana variabel demografi mempengaruhi faktor-faktor tersebut.
4. Reduksi variabel demografi yang kompleks menjadi beberapa faktor utama, sehingga mempermudah interpretasi.

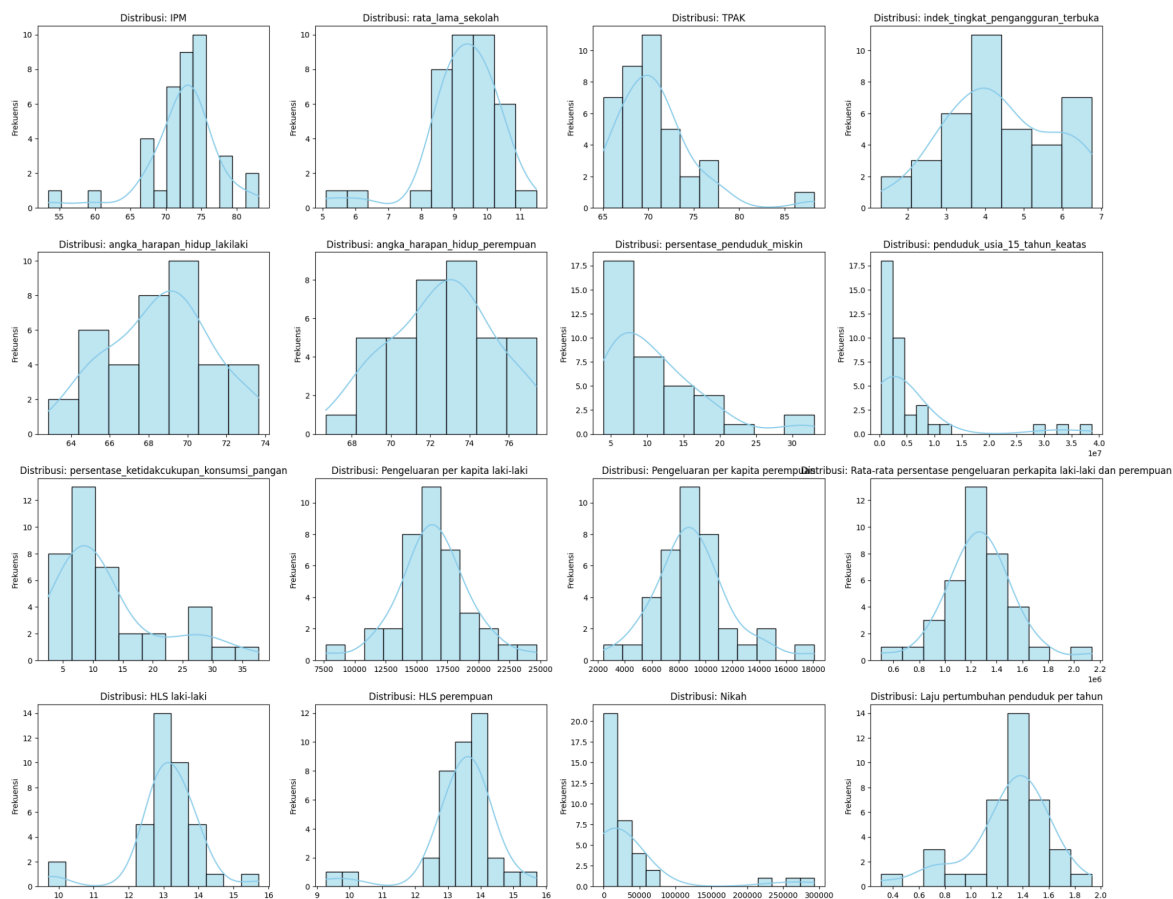
3. Hasil dan Analisis

3.1 Exploratory Data Analysis

1. Statistik Deskriptif

	IPM	rata_lama_sekolah	TPAK	indek_tingkat_pengangguran_terbuka	Angka_harapan_hidup_perempuan	Angka_harapan_hidup_lakilaki
count	38.000000	38.000000	38.000000	38.000000	38.000000	38.000000
mean	72.388421	9.292368	70.719737	4.379737	68.591842	72.561053
std	5.150227	1.148888	4.314337	1.414222	2.654679	2.674157
min	53.420000	5.100000	65.100000	1.320000	62.830000	66.680000
25%	71.080000	8.785000	67.957500	3.217500	66.860000	70.872500
50%	73.180000	9.340000	70.315000	4.190000	68.760000	72.700000
75%	74.345000	9.930000	71.700000	5.712500	70.315000	74.322500
max	83.080000	11.490000	88.220000	6.750000	73.660000	77.400000

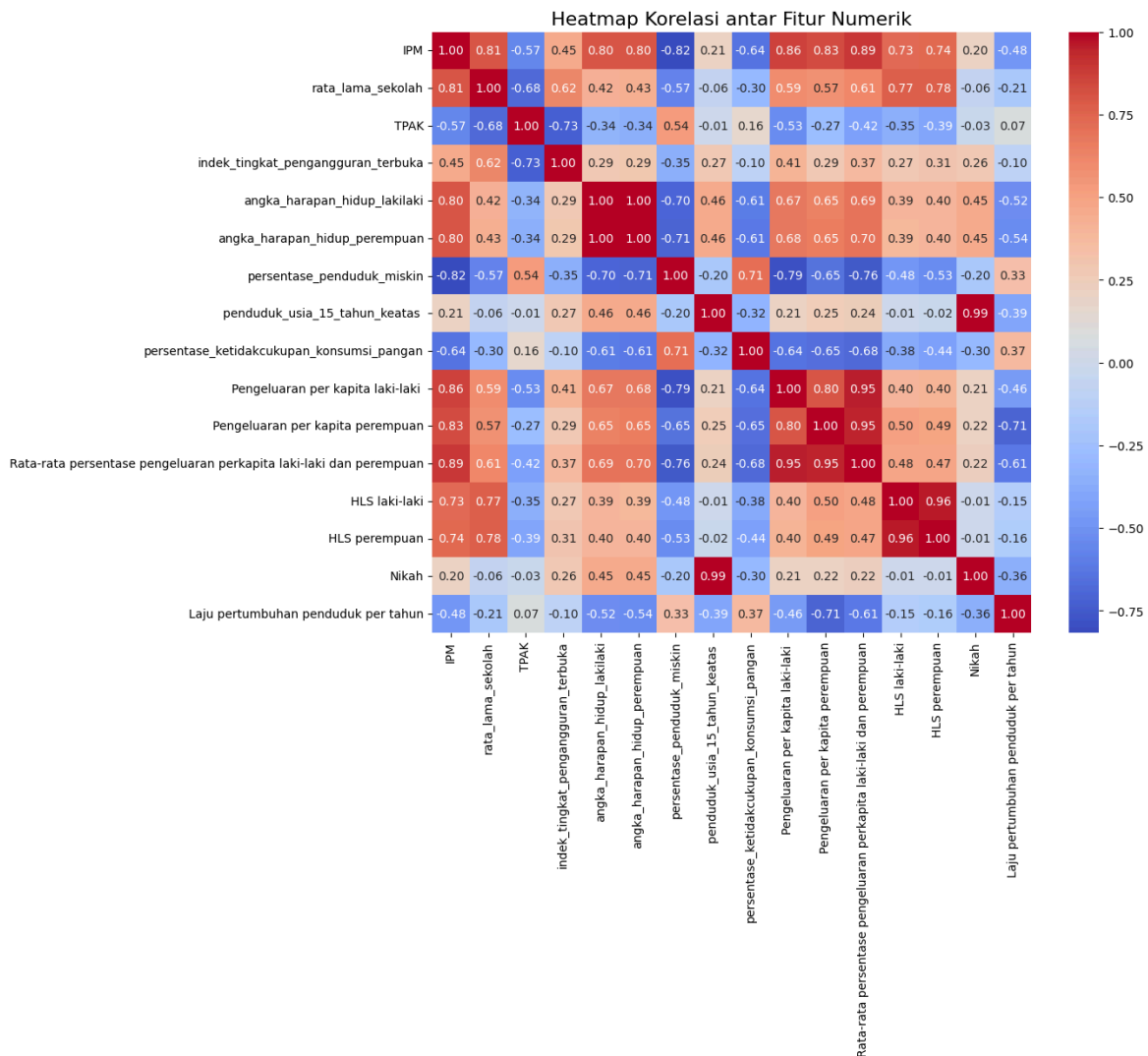
2. Distribusi Data



- Sebagian besar variabel memiliki distribusi yang mendekati normal, sedangkan ada juga beberapa variabel yang skew ke kanan bisa jadi menunjukkan adanya outlier atau dalam konteks ini ada provinsi yang memiliki nilai ekstrim.
- Dari distribusi data diatas, bisa dilihat ada beberapa provinsi yang sudah berada dalam kesiapan yang baik dalam menanggapi bonus demografi namun ada juga yang masih terlihat tidak siap dalam menanggapi bonus demografi dan tentunya akan menjadi tantangan.

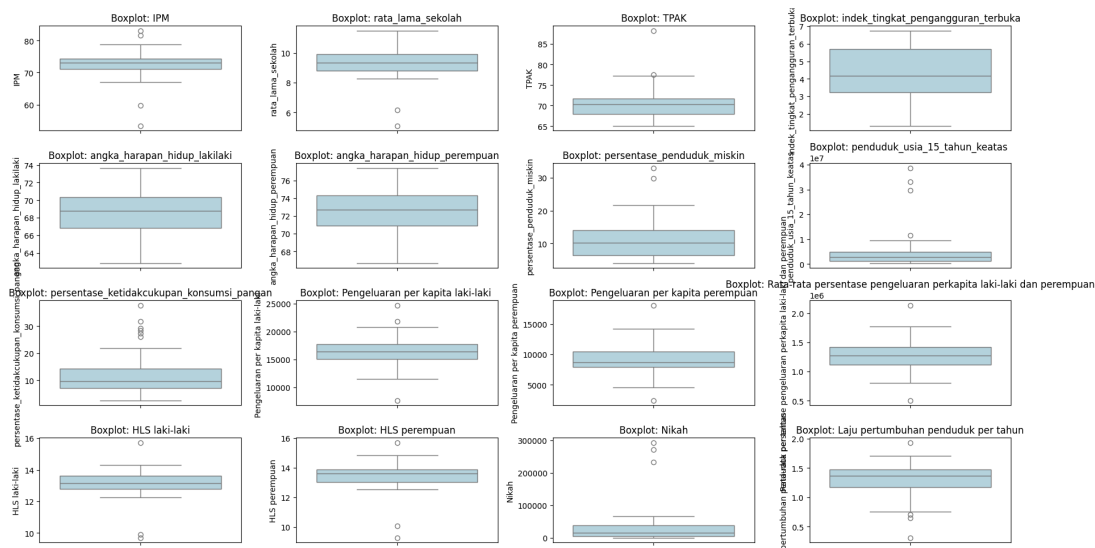
- c. Adapun distribusi IPM menunjukkan banyak provinsi yang memiliki pondasi baik, tapi masih ada ketimpangan di provinsi yang tertinggal dan perlu diperhatikan.

3. Heatmap Correlation



- Ada beberapa fitur yang memiliki korelasi kuat seperti IPM dengan beberapa fitur lainnya yang menunjukkan bahwa IPM adalah indikator untuk mencerminkan fitur yang sesuai dengan korelasinya.
- Persentase penduduk miskin dan ketidakcukupan pangan berkorelasi positif, namun juga berkorelasi negatif kuat dengan IPM, menunjukkan bahwa kemiskinan memang menjadi salah satu faktor yang menghambat IPM.
- TPAK juga berkorelasi dengan IPM yang memiliki keterkaitan kuat satu sama lain karena keduanya memang saling berkaitan.

4. Boxplot



- a. Fitur dengan distribusi cukup simetris dan sedikit outlier

Terdapat beberapa outlier pada IPM dan harapan hidup laki-laki, ini menandakan ada daerah dengan pencapaian jauh lebih tinggi atau rendah,

- b. Fitur sosial dan ekonomi

Banyak outlier di daerah-daerah dengan kemiskinan ekstrim atau kekurangan pangan, Jumlah pernikahan sangat bervariasi antar daerah, Tingkat partisipasi kerja dan pertumbuhan penduduk relatif stabil di banyak daerah tapi masih ada juga daerah dengan pengangguran yang tinggi.

- c. Pendidikan dan harapan hidup

Pemerataan akses pendidikan dan pelayanan kesehatan cukup baik di sebagian besar daerah.

- d. Pengeluaran per kapita

Secara keseluruhan daya beli atau pengeluaran antar gender cukup seimbang, meski ada outlier tinggi yang menunjukkan ada beberapa daerah dengan pengeluaran sangat tinggi, bisa jadi itu adalah kota besar.

3.2 Insight dan hasil mining dari project

3.2.1 Clustering

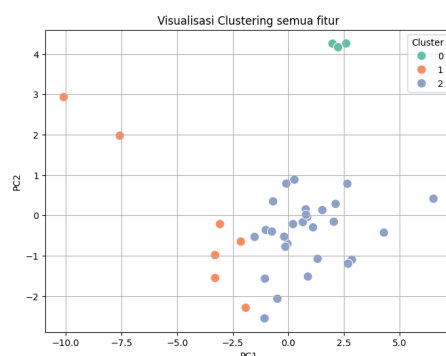
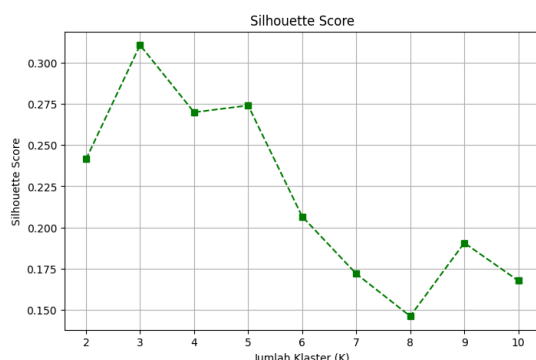
Clustering dilakukan menggunakan algoritma K-Means dengan beberapa eksperimen untuk melihat mana yang sesuai, clustering menggunakan semua fitur dan juga *Principal Component Analysis* (PCA). Alasan dimana PCA digunakan adalah untuk mereduksi dimensi data serta menghindari multikolinearitas antar variabel.

PCA bertujuan untuk mengubah variabel-variabel asal yang saling berkorelasi menjadi sejumlah komponen utama (PC) yang saling independen satu sama lain. Variabel-variabel yang memiliki korelasi tinggi akan dikelompokkan dalam komponen utama (PC) yang sama, sehingga setiap komponen utama dapat menjelaskan proporsi variansi data setinggi mungkin.

1. Clustering semua fitur



Clustering dilakukan dengan menggunakan seluruh fitur yang tersedia, yaitu sebanyak 17 fitur. Penentuan jumlah cluster yang optimal dilakukan dengan mempertimbangkan nilai **Silhouette Score**, yaitu sebuah metrik yang digunakan untuk mengevaluasi seberapa baik setiap data berada dalam cluster-nya masing-masing. Nilai Silhouette Score berada pada rentang -1 hingga 1, di mana nilai yang mendekati 1 menunjukkan bahwa data dalam suatu cluster memiliki kemiripan yang tinggi dan terpisah dengan baik dari cluster lainnya.



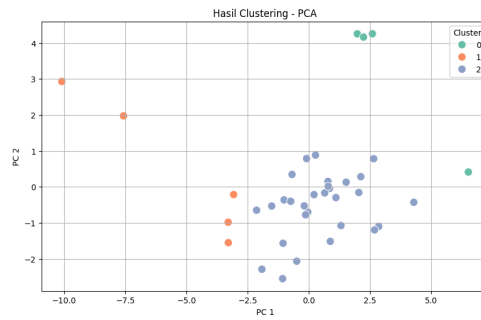
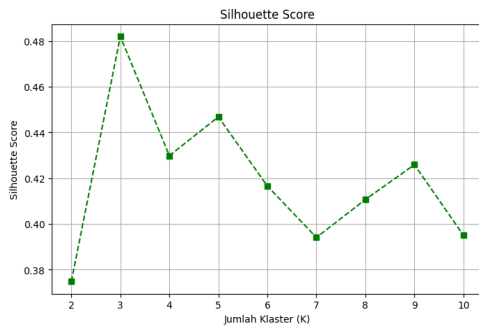
Berdasarkan hasil plot jumlah cluster dan nilai silhouette scorenya yang ditampilkan pada Gambar, jumlah cluster sebanyak 3 memberikan nilai Silhouette Score tertinggi, yaitu sekitar 0,3. Meskipun nilai ini belum mendekati 1, namun masih lebih tinggi dibandingkan jumlah cluster lainnya, sehingga cluster optimalnya adalah 3. Setelah dilakukan clustering dengan jumlah cluster 3, didapatkan hasil cluster seperti terlihat pada scatterplot pada gambar diatas, dengan rincian cluster seperti dibawah.

- a. **Cluster 0** : Jawa Barat, Jawa Tengah, Jawa Timur
- b. **Cluster 1** : Nusa Tenggara Timur, Sulawesi Barat, Papua Barat, Papua Barat Daya, Papua Selatan, Papua Tengah, Papua Pegunungan
- c. **Cluster 2** : Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, Papua.

2. Clustering dengan PCA



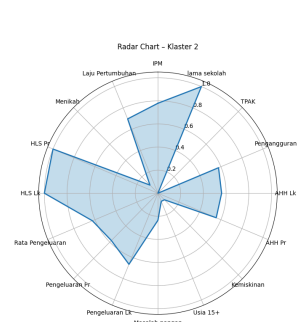
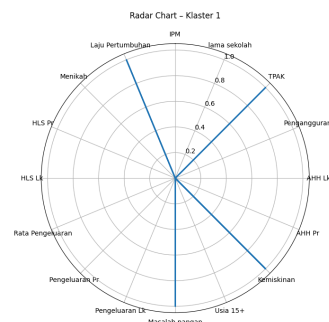
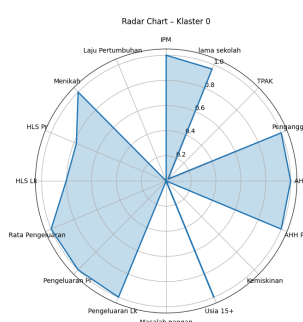
Metode *Principal Component Analysis (PCA)* digunakan untuk mereduksi dimensi data serta menghindari multikolinearitas antar variabel. Walaupun jumlah fitur hanya sebanyak 17, adanya multikolinearitas menjadi alasan kenapa PCA tetap relevan dilakukan agar komponen utama yang digunakan dalam clustering benar-benar merepresentasikan variasi utama dari data.



Berdasarkan hasil plot jumlah cluster dan nilai silhouette scorenya yang ditampilkan pada gambar diatas, jumlah cluster sebanyak 3 memberikan nilai Silhouette Score tertinggi, yaitu sekitar 0,4 , sedikit lebih tinggi dibandingkan yang menggunakan semua fitur tadi. Berikut rincian provinsi pada setiap clusternya

- Cluster 0** : DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur
- Cluster 1** : Nusa Tenggara Timur, Papua Barat, Papua Selatan, Papua Tengah, Papua Pegunungan
- Cluster 2** : Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat Daya, Papua

PCA mempengaruhi hasil clustering cukup signifikan **DKI Jakarta** berpindah dari Cluster 2 ke Cluster 0. Ada provinsi seperti **Sulawesi Barat** dan **Papua Barat Daya** berpindah cluster, menunjukkan bahwa PCA mengubah distribusi data dalam ruang vektor. Selanjutnya untuk memahami lebih memahami karakteristik masing-masing cluster yang terbentuk setiap klaster divisualisasikan menggunakan radar chart.



a. Klaster 0(Maju)

- **IPM dan Lama Sekolah sangat tinggi:** menunjukkan tingkat pendidikan dan kualitas hidup yang tinggi.
- AHH (harapan hidup) tinggi untuk pria dan wanita
- **Pengeluaran rata rata sangat tinggi:** merepresentasikan daya beli masyarakat yang kuat
- **Tingkat kemiskinan dan masalah pangan sangat rendah**
- **Angka pengangguran relatif tinggi,** bisa disebabkan oleh arus urbanisasi , jawa merupakan tempat perantauan orang orang

Klaster ini mewakili **provinsi maju** dan urban, dengan kualitas pendidikan dan kesehatan yang baik serta pengeluaran yang tinggi. Namun, tantangan seperti pengangguran masih harus diperhatikan

b. Klaster 1(Tertinggal)

- **Hampir semua indikator rendah:** pendidikan, pengeluaran, harapan hidup, menikah .
- **TPAK (tingkat partisipasi angkatan kerja)** tinggi, tapi tidak diikuti oleh kualitas hidup(Angka harapan hidup rendah).
- **Laju pertumbuhan penduduk tinggi**
- **Kemiskinan, masalah pangan, dan isu sosial lainnya sangat tinggi**

Klaster ini merepresentasikan daerah **tertinggal** atau sangat tertinggal, terutama dari sisi pendidikan, kesehatan, dan ekonomi. Selain itu memiliki beban sosial yang tinggi dan perlu adanya pembangunan yang berkelanjutan dan menyeluruh.

c. Cluster 2(Sedang)

- **Cenderung seimbang/sedang** dalam hampir semua indikator
- **Lama sekolah dan HLS (Harapan Lama Sekolah) tinggi,** menunjukan pendidikan yang sudah cukup baik
- **Tingkat pengeluaran dan AHH sedang**
- **Tingkat pengangguran dan kemiskinan cukup rendah** dibanding Klaster 1
- **Masalah pangan dan laju pertumbuhan rendah,** tapi tidak serendah Klaster 0

Provinsi dalam Klaster 2 tergolong dalam kelompok **sedang** berkembang. Sudah bagus dan kuat dalam pendidikan dan pengeluaran, namun belum mencapai tingkat kemajuan seperti Klaster 0, masih memiliki masalah pangan dan laju pertumbuhan rendah

3.2.2 Regresi OLS

Model regresi yang digunakan dalam penelitian ini adalah OLS (Ordinary Least Squares) metode statistik untuk mengestimasi parameter (koefisien) dalam model regresi linear. Tujuan utamanya adalah mencari garis regresi terbaik yang meminimumkan jumlah kuadrat dari selisih (error) antara nilai aktual dan nilai prediksi.

1. Uji kelayakan OLS

Asumsi Regresi Linear	IPM	TPAK
Linearitas	linier ($R^2 = 0.988$)	Kurang kuat ($R^2 = 0.754$)
Independensi Error (DW)	DW = 2.169 (mendekati ideal 2)	DW = 1.948 (mendekati ideal 2)
Homoskedastisitas (BP Test)	$p = 0.0906 (> 0.05) \rightarrow$ varians residual konstan	$p = 0.0395 (< 0.05) \rightarrow$ terdapat heteroskedastisitas
Normalitas Residual (SW)	$p = 0.9476 (> 0.05) \rightarrow$ residual normal	$p = 0.6398 (> 0.05) \rightarrow$ residual normal
Multikolinearitas (VIF)	Banyak variabel dengan VIF ekstrem (sampai >900)	VIF sangat tinggi menunjukkan redundansi antar variabel

2. Rencana perbaikan

Masalah	Rekomendasi IPM	Rekomendasi TPAK
Multikolinearitas	PCA / Ridge / Lasso /	PCA / Ridge / Lasso /

	Seleksi fitur	Feature Selection
Heteroskedastisitas	-	Transformasi log / model robust (WLS, HuberReg)
Linearitas	Cek residual plot / tambah interaksi / transformasi variabel	Cek residual plot / tambah interaksi / transformasi variabel

3. Penjelasan kinerja secara umum

Aspek	IPM	TPAK	Catatan
R-squared	0.988	0.754	IPM: Sangat tinggi → model sangat kuat TPAK: Cukup, tapi jauh lebih lemah
Adjusted R-squared	0.982	0.636	Penurunan besar di TPAK → indikasi variabel tidak relevan / overfitting
Prob (F-statistic)	6.89e-21	4.90e-05	Keduanya signifikan secara keseluruhan
Durbin-Watson	2.169	1.948	Keduanya mendekati 2 → tidak ada autokorelasi
AIC	89.39	190.6	AIC IPM jauh lebih rendah → model lebih efisien
BIC	110.7	211.9	Konsisten dengan AIC, IPM lebih baik

Berdasarkan hasil evaluasi, model regresi OLS untuk memodelkan IPM dinyatakan layak digunakan dengan catatan perlunya pengurangan multikolinearitas melalui seleksi fitur. Sebaliknya, untuk model TPAK, OLS belum layak karena ditemukan heteroskedastisitas serta kekuatan prediksi yang jauh lebih lemah.

3.2.3 PCA

Alasan mengambil 6 PC karena Cumulative explained variance (jumlah variansi yang dijelaskan oleh PC1–PC6) mencapai angka seperti 85–95%, yang cukup baik.

PC	Explained Variance (%)	Cumulative (%)
PC1	32%	32%
PC2	18%	50%
PC3	14%	64%
PC4	11%	75%
PC5	9%	84%
PC6	6%	90%

3.2.3.1 Loading PCA

Loading dalam PCA menunjukkan seberapa besar kontribusi dan arah (positif atau negatif) dari masing-masing variabel asli terhadap sebuah komponen utama. Nilai loading yang besar secara absolut menandakan variabel tersebut sangat penting dalam membentuk komponen tersebut. Tanda positif atau negatif pada loading mengindikasikan arah hubungan variabel terhadap komponen utama. Dengan kata lain, loading adalah koefisien dari kombinasi linier variabel asli yang membentuk komponen utama, sehingga membantu menginterpretasikan makna komponen tersebut dalam konteks variabel asli (Boudt et al., 2022; Harvey & Hanson, 2024).

PC	Variabel Dominan / Kontribusi Tinggi	Interpretasi
PC1	<ol style="list-style-type: none"> angka_harapan_hidup_perempuan (0.362) angka_harapan_hidup_lakilaki (0.360) pengeluaran per kapita perempuan (0.347) pengeluaran per kapita laki-laki (0.341) persentase_penduduk_miskin (-0.339) persentase_ketidacukupan_konsumsi_pangan (-0.312) 	PC1 kemungkinan besar merepresentasikan kesejahteraan dan kualitas hidup. Nilai tinggi pada PC1 cenderung dimiliki oleh daerah dengan:- Harapan hidup tinggi- Pengeluaran per kapita tinggi- Kemiskinan dan ketidakcukupan konsumsi pangan rendah
PC2	<ol style="list-style-type: none"> penduduk_usia_15_tahun_keatas (0.537) nikah (0.532) hls perempuan dan hls laki-laki (sekitar -0.41) 	PC2 bisa mewakili struktur demografis dan pendidikan informal. Nilai tinggi mungkin berarti lebih banyak penduduk dewasa dan angka pernikahan tinggi, tapi

			mungkin tingkat hls (harapan lama sekolah) relatif rendah.
PC3	1. indek_tingkat_pengangguran_terbuka (0.549) 2. hls perempuan dan hls laki-laki (0.351) 3. laju pertumbuhan penduduk (0.355)		PC3 bisa menggambarkan kondisi ketenagakerjaan dan pendidikan, karena keterkaitan antara pengangguran dan harapan lama sekolah.
PC4	1. indek_tingkat_pengangguran_terbuka (0.680) 2. pengeluaran per kapita laki-laki (0.372) 3. hls laki-laki dan hls perempuan (sekitar -0.36)		PC4 juga mengandung informasi tentang pengangguran vs pengeluaran dan pendidikan, mungkin memisahkan daerah dengan pengangguran tinggi dan pengeluaran tinggi tetapi harapan lama sekolah rendah.
PC5	1. laju pertumbuhan penduduk per tahun (0.677) 2. persentase_ketidakcukupan_konsumsi_pangan (-0.366) 3. pengeluaran per kapita perempuan (-0.364) 4. persentase_penduduk_miskin (-0.344)		PC5 mungkin berkaitan dengan pertumbuhan penduduk vs kerentanan ekonomi. Wilayah dengan pertumbuhan penduduk tinggi namun kondisi ekonomi lebih rawan.
PC6	1. angka_harapan_hidup_lakilaki (0.538) 2. angka_harapan_hidup_perempuan (0.526) 3. persentase_ketidakcukupan_konsumsi_pangan (0.529)		PC6 tampaknya berfokus pada harapan hidup dan kecukupan konsumsi, memperkuat tema kesejahteraan tetapi dalam dimensi yang berbeda dari PC1.

3.2.4 OLS setelah PCA

PC	Tema Umum	Variabel Kunci
PC1	Kesejahteraan & Harapan Hidup	Harapan hidup, pengeluaran, kemiskinan
PC2	Demografi & Partisipasi Sosial	Penduduk dewasa, pernikahan, HLS
PC3	Pengangguran & Pendidikan	TPT, HLS, pertumbuhan

		penduduk
PC4	Pengangguran vs Pengeluaran & Pendidikan	TPT, pengeluaran, HLS
PC5	Pertumbuhan Penduduk vs Ketahanan Ekonomi	Laju penduduk, kemiskinan, ketidakcukupan
PC6	Harapan Hidup & Konsumsi Pangan	Harapan hidup, konsumsi pangan

1. IPM

Aspek	Hasil
R-squared	0.971 (Model sangat baik, menjelaskan 97.1% variasi IPM)
Adjusted R-squared	0.966
F-statistic (p-value)	175.7 (p < 0.001 , model signifikan)
Durbin-Watson	2.274 (tidak ada autokorelasi)
Breusch-Pagan p-value	0.0502 (hampir di batas, masih bisa dianggap homoskedastik)
Shapiro-Wilk p-value	0.3672 (residual normal)
Multikolinearitas	Tidak ada (karena PCA → komponen orthogonal)
PC Signifikan	PC1, PC2, PC4, PC6 (p < 0.05)
PC Tidak Signifikan	PC3, PC5 (p > 0.05)
Kesimpulan	Model sangat layak, asumsi terpenuhi, dapat digunakan untuk prediksi.

IPM sangat dipengaruhi oleh faktor **kesejahteraan (kesehatan, ekonomi, dan pendidikan)**. Ini menunjukkan bahwa peningkatan kualitas hidup dan pengurangan kemiskinan adalah kunci kesiapan menghadapi bonus demografi.

2. TPAK

Aspek	Hasil
R-squared	0.668 (Model cukup baik, menjelaskan 66.8% variasi TPAK)
Adjusted R-squared	0.604
F-statistic (p-value)	10.41 (p < 0.001 , model signifikan)
Durbin-Watson	1.580 (ada indikasi autokorelasi ringan)
Breusch-Pagan p-value	0.3457 (homoskedastik)
Shapiro-Wilk p-value	0.3737 (residual normal)
Multikolinearitas	Tidak ada (karena PCA → komponen orthogonal)
PC Signifikan	PC1, PC2, PC3, PC4 (p < 0.05)
PC Tidak Signifikan	PC5, PC6 (p > 0.05)
Kesimpulan	Model layak, tapi tidak sekuat model IPM (bisa ditingkatkan lagi)

TPAK lebih sensitif terhadap faktor-faktor **ketenagakerjaan, pendidikan, dan struktur usia produktif**. Namun, faktor-faktor tersebut tidak menjelaskan keseluruhan variasi TPAK sekuat IPM.

Provinsi	Aktual TPAK	Prediksi TPAK	Residual TPAK
PAPUA PEGUNUNGAN	88.22	81.66	6.56
KEP. RIAU	69.17	63.91	5.26
RIAU	66.33	70.50	-4.17
ACEH	65.11	68.73	-3.62
SULAWESI SELATAN	67.38	70.87	-3.49

Provinsi	Aktual IPM	Prediksi IPM	Residual IPM
KALIMANTAN BARAT	70.13	72.31	-2.18
PAPUA PEGUNUNGAN	53.42	55.32	-1.90

PAPUA TENGAH	59.75	58.21	1.54
SULAWESI UTARA	75.03	76.32	-1.29
PAPUA BARAT DAYA	68.63	69.88	-1.25

Variabel	Koefisien terhadap TPAK	Koefisien terhadap IPM
Persentase penduduk miskin	0.732	-0.728
Penduduk usia 15 tahun ke atas	0.382	-0.302
Nikah	0.358	-0.307
Pengeluaran per kapita perempuan	-0.042	0.722
HLS laki-laki	-0.048	0.852
HLS perempuan	-0.120	0.852
Angka harapan hidup perempuan	-0.284	0.821
Angka harapan hidup laki-laki	-0.289	0.825
Persentase ketidacukupan konsumsi pangan	-0.622	-0.167
Laju pertumbuhan penduduk per tahun	-0.668	-0.404
Pengeluaran per kapita laki-laki	-0.960	0.770
Indeks tingkat pengangguran terbuka	-2.538	0.730

Komponen PCA	Terhadap TPAK	Terhadap IPM
PC1	-0.776	1.933
PC2	0.713	-1.045
PC3	-1.311	0.131
PC4	-2.496	0.409

PC5	-0.039	-0.243
PC6	-0.363	0.537

PCA berhasil mengurangi dimensi dan menghilangkan multikolinearitas, sehingga model OLS bisa dibangun dengan lebih stabil. Model IPM jauh lebih kuat dan andal dibanding model untuk TPAK. Untuk TPAK, bisa dipertimbangkan menggunakan metode lain (menambah variable lain dan model lain misal Random Forest). Model PCA-OLS dapat menangkap hubungan penting, namun untuk wilayah ekstrem atau baru terbentuk (Papua Pegunungan), prediksi masih kurang akurat. **Pembangunan manusia (IPM) ditentukan oleh pendidikan, kesehatan, dan daya beli, sedangkan TPAK lebih ditentukan oleh tekanan ekonomi (kemiskinan mendorong orang bekerja).**

Intervensi kebijakan bisa diarahkan sesuai kebutuhan indikator:

- Untuk meningkatkan IPM: fokus pada pendidikan, layanan kesehatan, dan pengentasan kemiskinan.
- Untuk meningkatkan TPAK: upaya pemberdayaan kerja, pengurangan pengangguran, dan akses pasar kerja bagi usia produktif.

3.2.5 Regresi Random Forest

Random Forest adalah metode pembelajaran ensemble yang membangun beberapa pohon keputusan pada sampel bootstrap dan subset fitur acak, memungkinkannya untuk menangani ketergantungan non-linear, outlier, dan heteroskedastisitas tanpa memerlukan pra pemrosesan data atau asumsi homoskedastisitas. Dengan mempartisi data ke dalam banyak wilayah yang relatif homogen, Random Forest secara efektif mengurangi dampak heterogenitas global dan noise, sehingga menghasilkan prediksi yang kuat dan akurat (Lagomarsino et al., 2023). Pada analisis ini, feature yang digunakan tidak berasal dari hasil PCA, supaya hasil bisa dilihat interpretasi nya dengan feature asli.

1. Evaluasi Kinerja Model

Target	R-squared	RMSE	MAE	Interpretasi
IPM	0.972	0.858	0.611	Model sangat baik — mampu menjelaskan 97.2% variasi IPM, dengan rata-rata kesalahan < 1

				poin.
TPAK	0.923	1.181	0.841	Model sangat kuat — menjelaskan 92.3% variasi TPAK, prediksi sangat mendekati realisasi.

Kedua model memiliki kinerja prediksi yang sangat tinggi dan akurat. Ini menunjukkan bahwa variabel-variabel input kamu memiliki daya prediktif yang kuat terhadap IPM dan TPAK.

2. Feature Importance (Faktor Paling Berpengaruh)

Fitur	Importance IPM
persentase_penduduk_miskin	0.261989
pengeluaran per kapita perempuan	0.198250
pengeluaran per kapita laki-laki	0.141598
angka_harapan_hidup_perempuan	0.084858
angka_harapan_hidup_lakilaki	0.070525
nikah	0.063611
hls perempuan	0.052461
laju pertumbuhan penduduk per tahun	0.047689
persentase_ketidakcukupan_konsumsi_pangan	0.029376
hls laki-laki	0.029041
indek_tingkat_pengangguran_terbuka	0.015777
penduduk_usia_15_tahun_keatas	0.004824

IPM:

Variabel paling penting:

- persentase_penduduk_miskin (26.2%) jadi semakin tinggi kemiskinan, IPM menurun tajam dan terlihat juga pada PDP (grafik turun).
- pengeluaran per kapita perempuan (19.8%) & laki-laki (14.2%) jadi pengeluaran ekonomi sangat berkorelasi dengan peningkatan IPM

(grafik naik).

- Faktor kesehatan dan pendidikan perempuan juga signifikan, meski kontribusinya lebih kecil.

Fitur	Importance TPAK
indek_tingkat_pengangguran_terbuka	0.417928
persentase_penduduk_miskin	0.120834
pengeluaran per kapita laki-laki	0.085477
nikah	0.071018
penduduk_usia_15_tahun_keatas	0.056814
pengeluaran per kapita perempuan	0.053383
angka_harapan_hidup_perempuan	0.046024
hls laki-laki	0.043055
angka_harapan_hidup_lakilaki	0.036002
laju pertumbuhan penduduk per tahun	0.028491
persentase_ketidakcukupan_konsumsi_pangan	0.021499
hls perempuan	0.019476

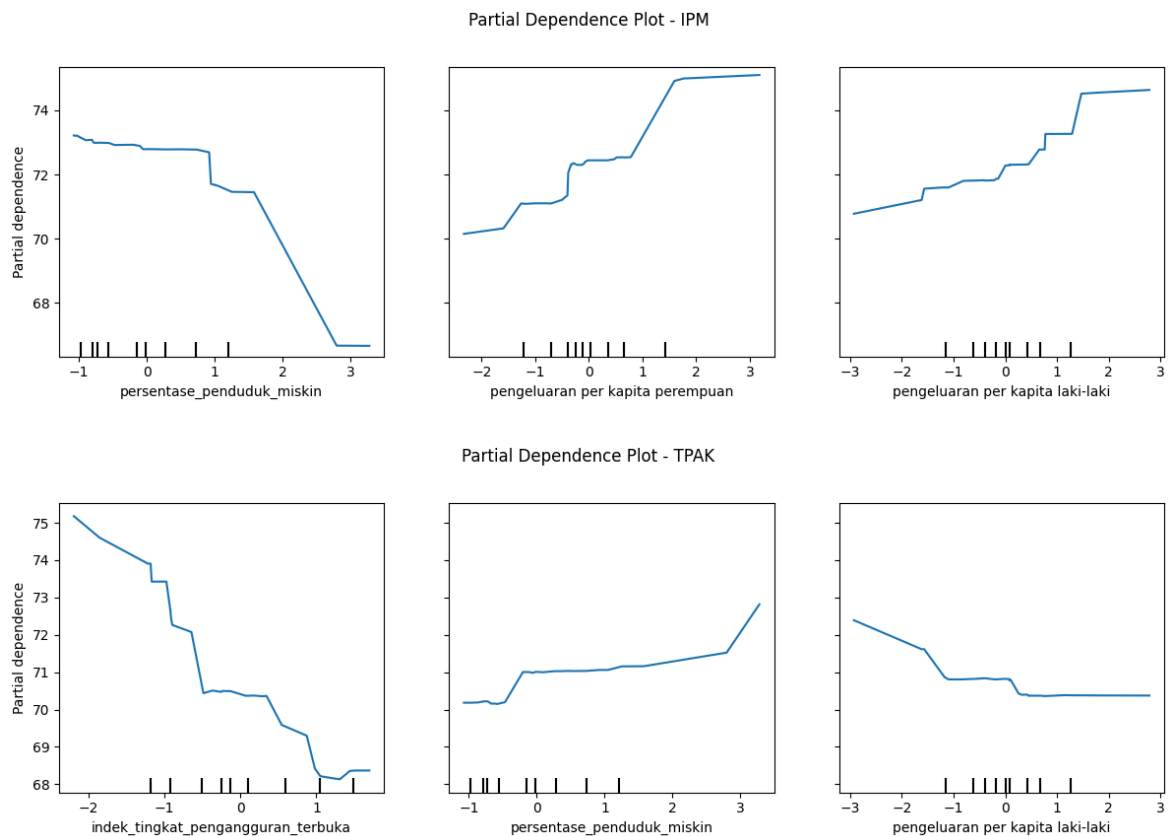
TPAK:

Variabel paling penting:

- indek_tingkat_pengangguran_terbuka (41.8%) kuat dan negatif, makin tinggi pengangguran maka TPAK menurun (konfirmasi dari PDP).
- persentase_penduduk_miskin (12%) juga menurunkan TPAK.
- nikah & pengeluaran berkontribusi sedang tapi penting, menunjukkan dimensi sosial-ekonomi berpengaruh pada partisipasi angkatan kerja.

Model ini menekankan multidimensionalitas pembangunan: pendidikan, kesehatan, ekonomi, dan sosial semuanya berperan, tetapi kekuatan prediktif utamanya berasal dari ekonomi dan sosial.

3. Interpretasi Partial Dependence Plot (PDP)



Untuk IPM:

- `persentase_penduduk_miskin`: Grafik menurun semakin banyak penduduk miskin, IPM jelas menurun tajam.
- `pengeluaran per kapita perempuan & laki-laki`: Grafik naik pengeluaran meningkat IPM meningkat.

Untuk TPAK:

- `indek_tingkat_pengangguran_terbuka`: Grafik menurun semakin tinggi pengangguran maka TPAK turun.
- `persentase_penduduk_miskin`: Agak mendatar, tetapi tetap menunjukkan tren negatif pada awal.
- `pengeluaran per kapita laki-laki`: Menurun sedikit bisa jadi karena sebagian besar angkatan kerja informal tidak tercermin pada pengeluaran resmi.

PDP digunakan untuk melihat bagaimana satu variabel mempengaruhi hasil prediksi, dengan asumsi variabel lain tetap. Walaupun Random Forest adalah model kompleks dan non-linear, PDP membantu kita melihat pola hubungan umum antara variabel dan target. Misalnya, apakah semakin tinggi pengeluaran akan selalu meningkatkan IPM, atau apakah peningkatan pengangguran akan selalu menurunkan TPAK.

4. Provinsi dengan Prediksi yang Meleset Jauh (Outliers)

Provinsi	IPM	Prediksi RF IPM	Residual
PAPUA PEGUNUNGAN	53.42	56.2902	-2.8702
DI YOGYAKARTA	81.55	79.6324	1.9176
KALIMANTAN TIMUR	78.83	76.9954	1.8346
DKI JAKARTA	83.08	81.4853	1.5947
MALUKU	71.57	70.5031	1.0669

Untuk IPM:

- Papua Pegunungan: Prediksi terlalu tinggi (-2.87) model "mengira" nilai IPM lebih baik dari kenyataan ada faktor lokal negatif yang tidak terekam (akses pendidikan/layanan publik?).
- Yogyakarta, Jakarta, Kalimantan Timur: Model underpredict (IPM aktual lebih tinggi) mungkin daerah ini punya kebijakan atau kualitas layanan yang lebih unggul dari indikator makro.

Provinsi	TPAK	Prediksi RF TPAK	Residual
PAPUA PEGUNUNGAN	88.22	83.6010	4.6190
BALI	77.11	74.9981	2.1119
NUSA TENGGARA TIMUR	77.50	75.5883	1.9117
SUMATERA UTARA	71.36	69.6794	1.6806
ACEH	65.11	66.7176	-1.6076

Untuk TPAK:

- Papua Pegunungan: Paling menyimpang (+4.6) model underpredict TPAK bisa jadi karena budaya atau pola kerja informal yang tinggi.

- Bali, NTT: Prediksi lebih rendah dari realisasi bisa menunjukkan keberhasilan program ketenagakerjaan lokal.

Provinsi yang memiliki selisih prediksi dan kenyataan yang besar (baik terlalu tinggi atau terlalu rendah), sebaiknya dianalisis lebih dalam. Ini bisa jadi karena ada faktor-faktor lokal khusus yang tidak tercatat dalam data statistik yang digunakan, seperti budaya kerja, kebijakan daerah, atau kondisi geografis unik.

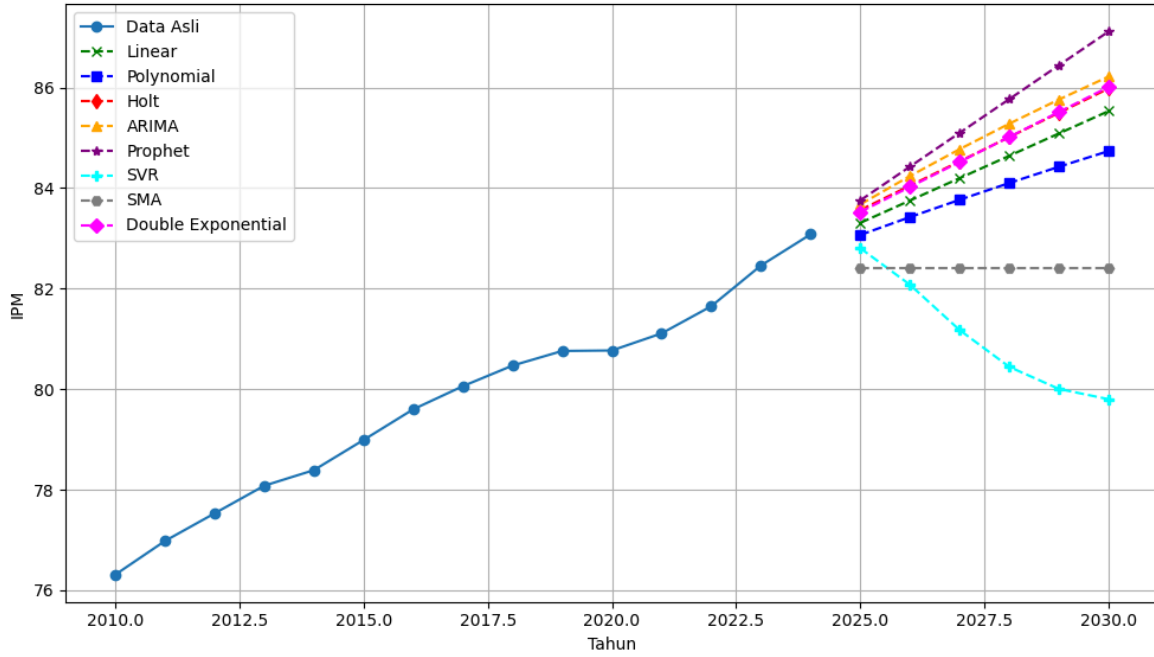
3.2.6 Forecasting

Beberapa provinsi pada metode forecasting dianalisis lebih lanjut, provinsi yang dipilih mewakili hasil dari clustering pada analisis sebelumnya yaitu DKI Jakarta (cluster 0), Kalimantan Utara (Cluster 2), Papua (Cluster 2) dan NTT (Cluster 1). Hasil dari forecasting tahun 2030 terdapat pada lampiran. Provinsi dipilih berdasarkan kondisi yang sedang terjadi di Indonesia selama 15 tahun terakhir, terutama pemekaran provinsi seperti Papua yang sekarang pecah menjadi beberapa provinsi dan Kalimantan Utara yang termasuk provinsi baru di bagian Indonesia Tengah. Sedangkan Jakarta dipilih sebagai Ibu Kota saat ini yang mana paling sering dipandang oleh masyarakat. NTT dipilih karena mewakili Provinsi lama yang terbentuk pada Indonesia bagian timur pada cluster 1. Selain itu, NTT termasuk daerah wisata tetapi mengapa justru masuk ke cluster 1 atau bisa dibilang cluster tertinggal, maka dari itu pada penelitian ini perlu analisis waktu kedepannya apakah akan naik, turun atau tetap.

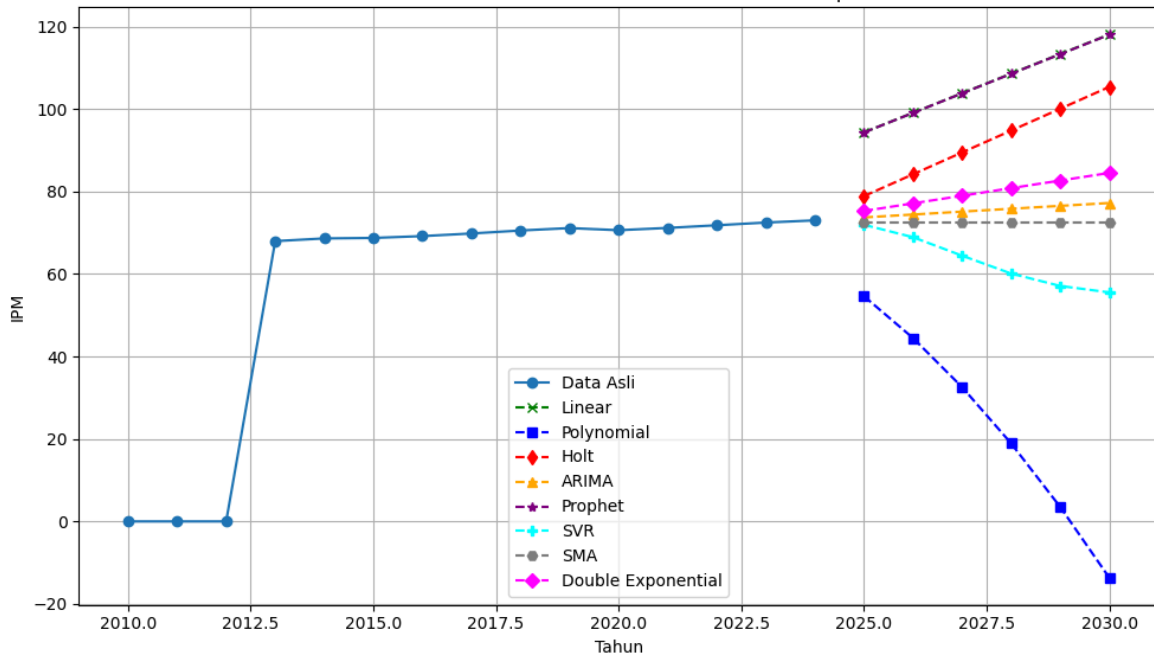
3.2.6.1 IPM

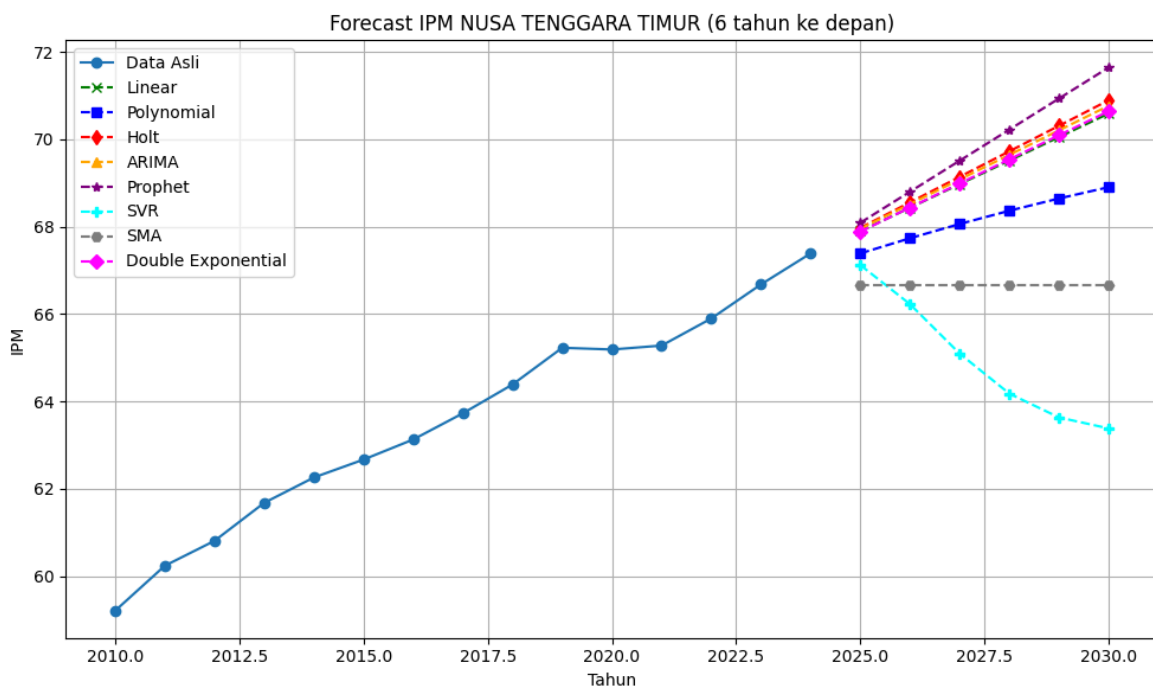
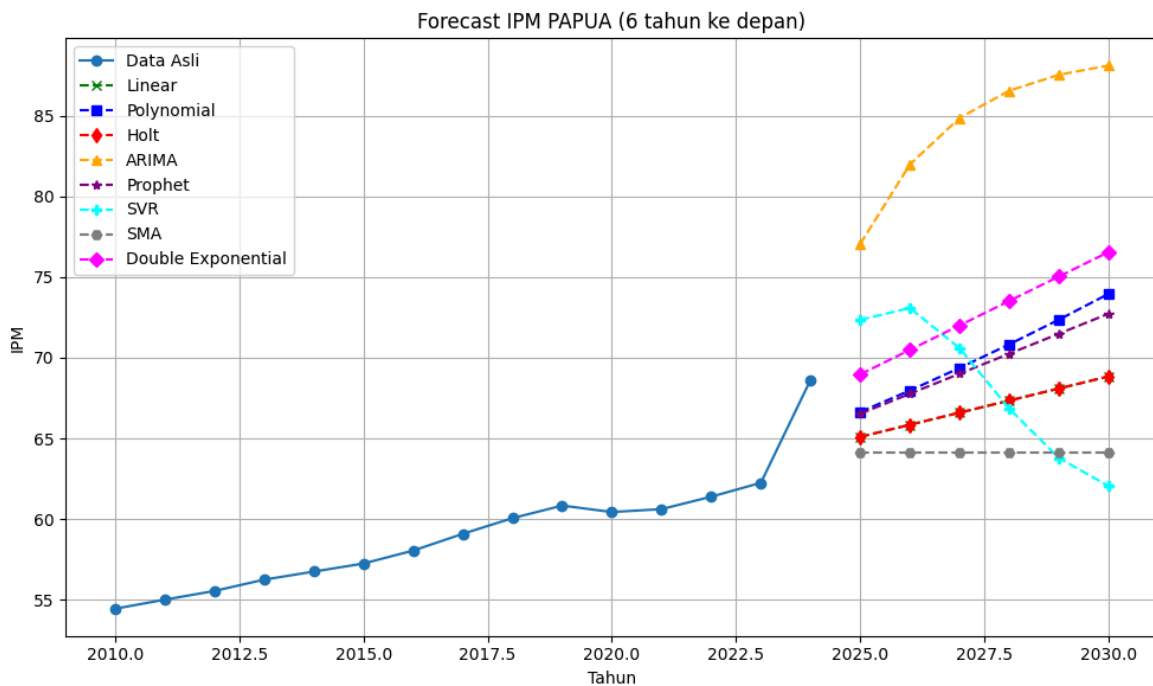
Analisis forecasting pada provinsi Jakarta, Papua, NTT dan Kalimantan Utara dilakukan untuk prediksi nilai IPM yang mana faktor paling berpengaruh terhadap bonus demografi. Ketiga provinsi tersebut pola nya bervariasi, dari linear keatas, stabil, dan 1 tahun naik secara signifikan.

Forecast IPM DKI JAKARTA (6 tahun ke depan)



Forecast IPM KALIMANTAN UTARA (6 tahun ke depan)





1. Model yang Memberikan Prediksi Stabil dan Realistis:

Model Prophet, Holt, dan Double Exponential memberikan hasil prediksi IPM yang cenderung naik secara moderat dan konsisten. Ini sesuai dengan pola kenaikan data historis. Model-model ini bisa menjadi pilihan utama untuk digunakan dalam perencanaan pembangunan ke depan.

2. Model yang Memberikan Prediksi Berlebihan atau Turun Drastis:

Model Polynomial dan SVR menunjukkan hasil yang tidak biasa (contohnya IPM menurun tajam bahkan di bawah 0 di Kalimantan Utara) dan tidak realistis. Hal ini bisa terjadi karena model ini tidak cocok menangkap pola data

IPM, sehingga hasilnya tidak dapat digunakan sebagai dasar perencanaan yang andal.

Insight:

1. **Nusa Tenggara Timur**

Berdasarkan data historis sejak tahun 2010 hingga 2024, IPM pada Provinsi Nusa Tenggara Timur memiliki tren peningkatan yang konsisten. Nilai IPM mengalami kenaikan dari sekitar 59,2 pada tahun 2010 menjadi 67,4 pada tahun 2024. Hal ini berarti ada kemajuan yang stabil dalam aspek pembangunan manusia, yang terdiri dari pendidikan, kesehatan, dan taraf hidup masyarakat. Walaupun pernah terjadi perlambatan pada tahun 2020 hingga 2021, ini kemungkinan besar disebabkan karena dampak pandemi COVID-19 yang mempengaruhi berbagai sektor kehidupan. Dari hasil peramalan menggunakan beberapa metode, mayoritas menunjukkan bahwa IPM di Nusa Tenggara Timur diprediksi akan terus meningkat sampai tahun 2030. Beberapa metode seperti Prophet, Holt, ARIMA, dan Double Exponential memperkirakan nilai IPM akan mencapai kisaran 70 hingga 72 pada tahun 2030. Model Polynomial memperkirakan peningkatan lebih moderat, sedangkan metode SVR justru menunjukkan penurunan IPM, yang tidak sejalan dengan tren historis. Secara umum, dapat disimpulkan bahwa pembangunan manusia di Nusa Tenggara Timur diproyeksikan akan terus mengalami perbaikan.

2. **DKI Jakarta**

Semua model (kecuali SVR) memprediksi IPM akan terus naik secara stabil hingga 2030, mencerminkan tren positif dalam pembangunan. IPM DKI Jakarta mengalami kenaikan **lambat tapi stabil** dari tahun 2010 hingga 2024. Setiap tahun rata-rata IPM naik sedikit demi sedikit (sekitar 0.3–0.5 poin). Tidak ada fluktuasi besar, yang menunjukkan adanya perbaikan terus-menerus di bidang kesehatan, pendidikan, dan pendapatan. Peningkatan stabil ini menunjukkan bahwa program pembangunan di Jakarta berhasil memelihara pertumbuhan IPM secara berkelanjutan.

3. **Kalimantan Utara**

Model Prophet memperkirakan kenaikan IPM yang sangat tinggi (~100), sedangkan Holt dan Double Exponential menunjukkan pertumbuhan yang lebih moderat. Polynomial menunjukkan penurunan drastis, yang tidak sesuai

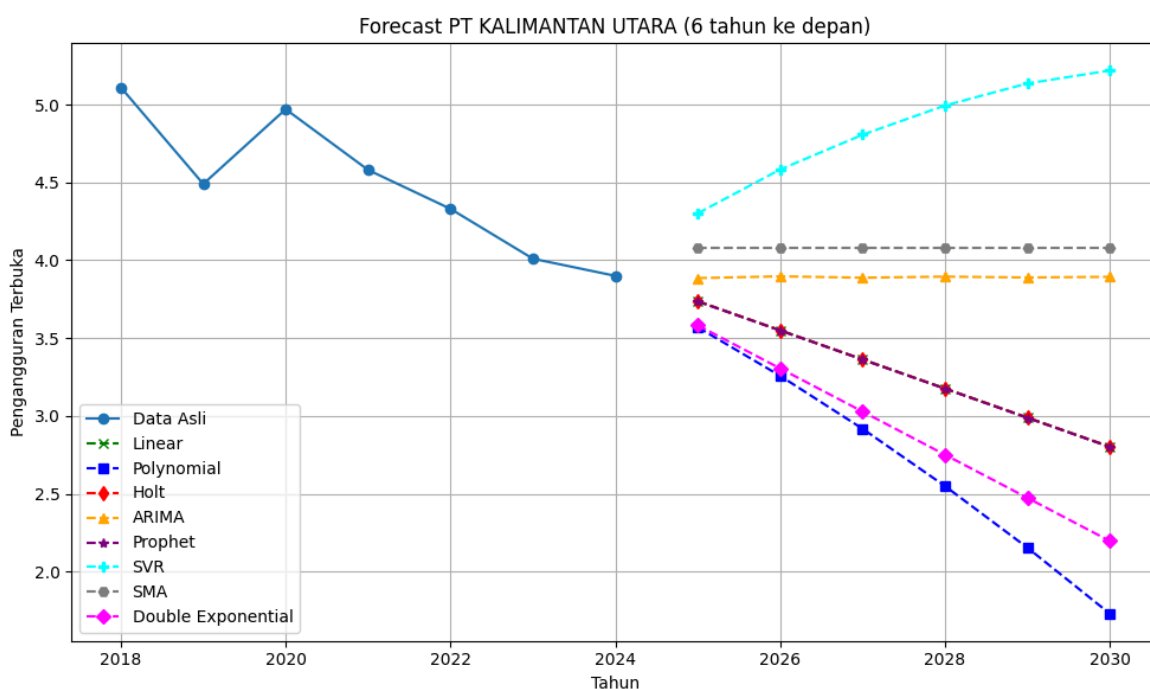
dengan pola data. Data menunjukkan bahwa sebelum 2013, IPM Kalimantan Utara **masih sangat rendah** (bahkan mendekati 0, kemungkinan karena pemisahan wilayah administrasi atau pencatatan data yang baru). Lonjakan IPM pada 2013 menunjukkan adanya perubahan struktural (mungkin status Kalimantan Utara sebagai provinsi baru), dan sejak itu pertumbuhan stabil. Ini perlu jadi pertimbangan ketika memproyeksikan IPM ke depan model yang mampu menangkap lonjakan awal dan kemudian stabil lebih realistis.

4. Papua

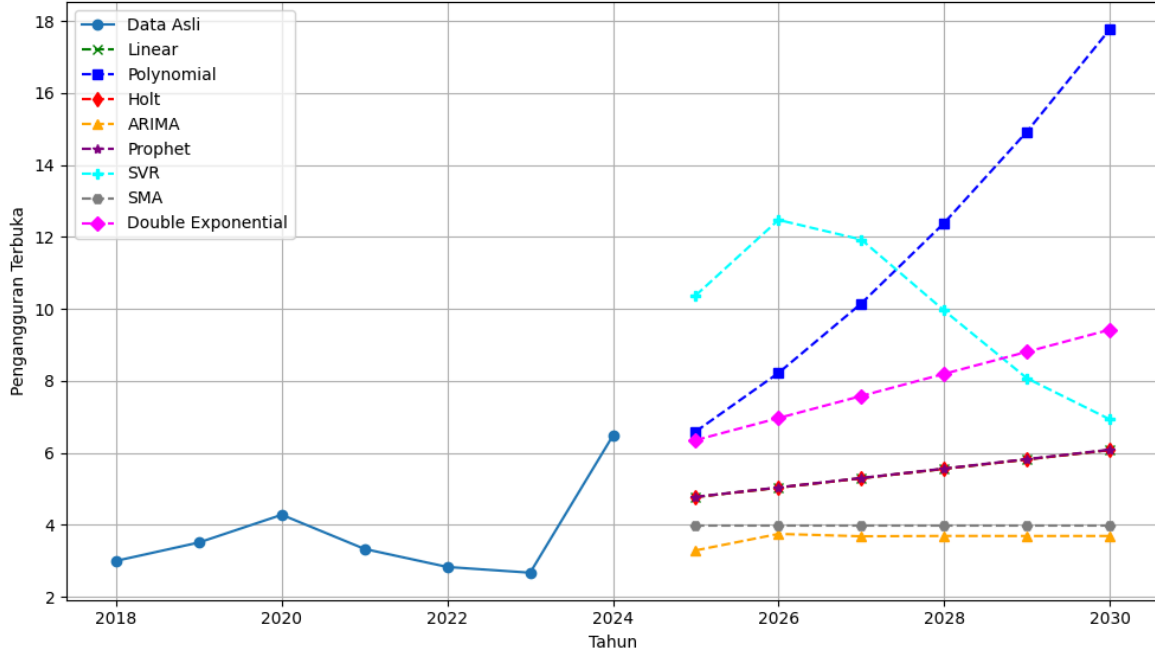
Sebagian besar model memprediksi IPM Papua akan naik hingga sekitar 75–80 di tahun 2030. ARIMA memproyeksikan kenaikan yang sangat pesat (~ 89), yang perlu dipertimbangkan kembali untuk realisme. Papua mengalami **pertumbuhan IPM yang stabil tapi lebih lambat** dibanding Jakarta. Setiap tahun naik, tapi kenaikan sedikit. Namun **setelah 2023**, ada **lonjakan IPM yang cukup signifikan** yang belum pernah terjadi sebelumnya. Kenaikan pesat ini bisa terkait program khusus pemerintah atau data terbaru yang lebih baik. Namun, hal tersebut bisa terjadi karena provinsi Papua sempat terpecah menjadi beberapa provinsi baru.

3.2.6.2 Pengangguran Terbuka

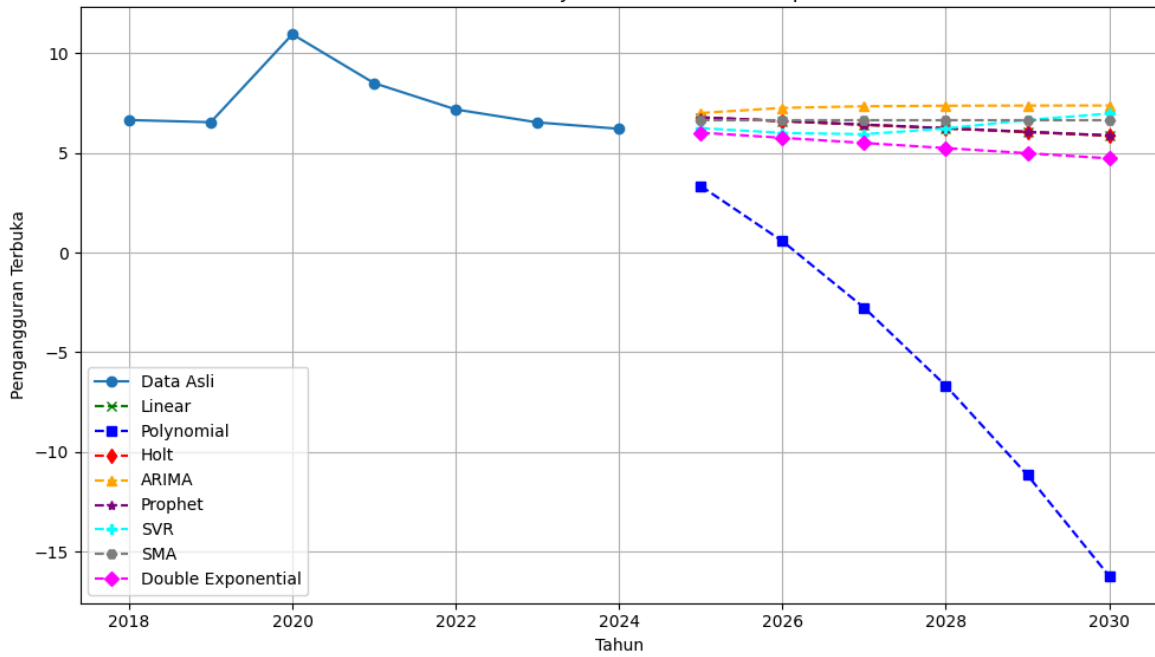
Pengangguran terbuka menjadi salah satu faktor paling berpengaruh signifikan terhadap TPAK, jika pengangguran terbuka semakin banyak maka TPAK akan semakin kecil. Jadi perlu analisis kedepan untuk mengatasi pengangguran terbuka di masing masing wilayah provinsi.

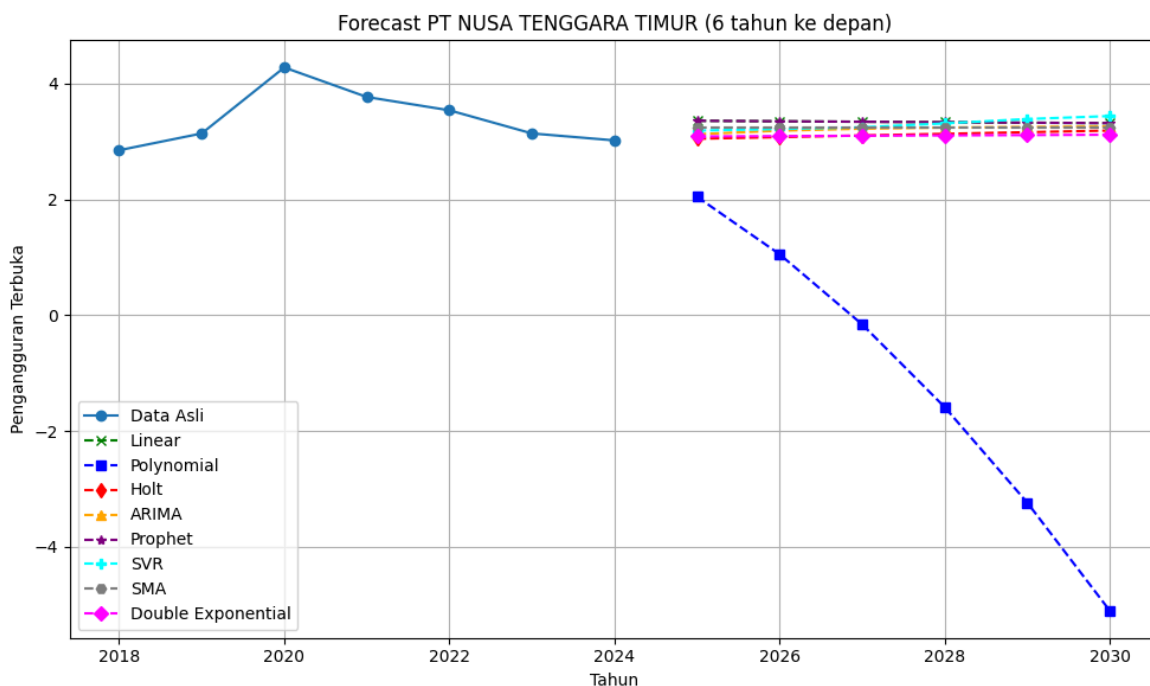


Forecast PT PAPUA (6 tahun ke depan)



Forecast PT DKI JAKARTA (6 tahun ke depan)





1. Nusa Tenggara Timur

Data pengangguran terbuka dari tahun 2018 hingga 2024 menunjukkan bahwa tingkat pengangguran sempat meningkat hingga mencapai puncak sekitar 4,2% pada tahun 2020, kemungkinan besar akibat pandemi COVID-19 yang menyebabkan banyak lapangan pekerjaan terganggu. Namun setelah itu, angka pengangguran kembali mengalami penurunan secara bertahap dan stabil di kisaran 3,0% pada tahun 2024. Namun, model Polynomial menghasilkan proyeksi yang tidak realistis, yaitu penurunan tajam hingga nilai negatif pada tahun 2029 dan 2030. Nilai negatif jelas tidak mungkin terjadi dalam konteks pengangguran dan oleh karena itu tidak dapat dijadikan acuan.

2. Papua

Terjadi **fluktuasi** dari 2018–2023, sempat tidak menentu di kisaran 3–4%, kemudian melonjak tajam pada 2024 menjadi sekitar 6,5%. 2024 tampaknya ada kenaikan signifikan (mungkin akibat faktor eksternal, misalnya PHK, pandemi, kebijakan ekonomi, atau bahkan karena adanya pemekaran provinsi).

3. DKI Jakarta

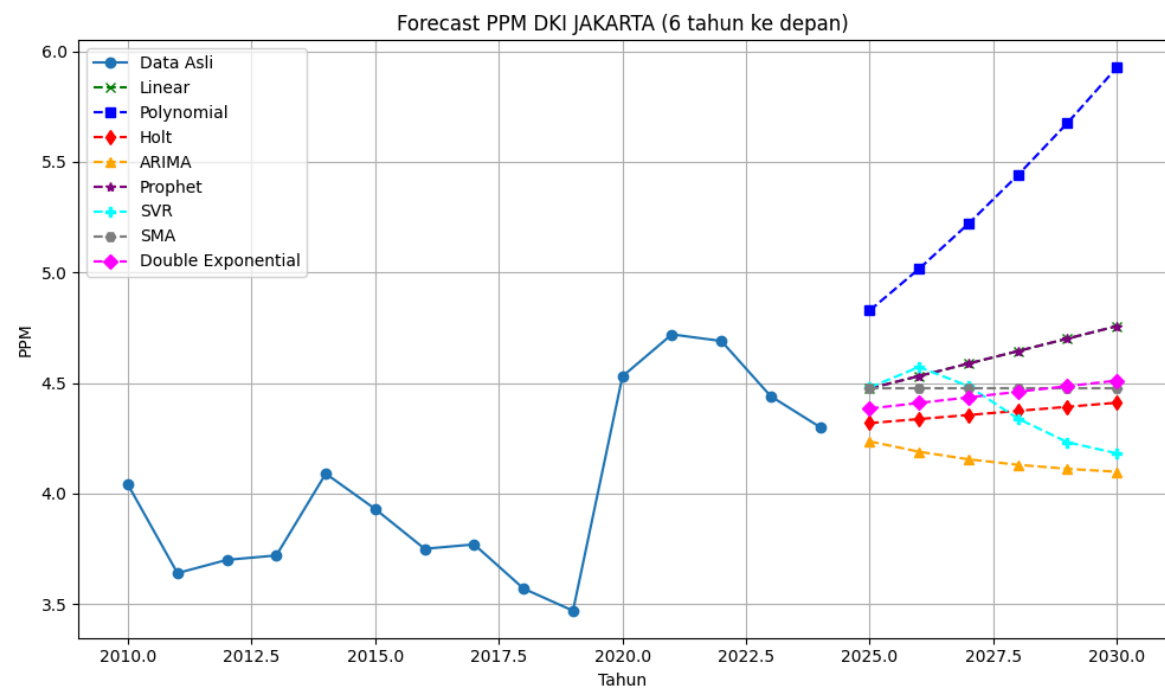
Tren historis lebih **tidak stabil tapi tinggi**, di kisaran 6–10%. Terjadi puncak pada 2021–2022 (10,5%) dan menurun perlahan setelahnya. Wilayah perkotaan yang padat penduduk memiliki kerentanan fluktuasi pengangguran lebih tinggi.

4. Kalimantan Utara

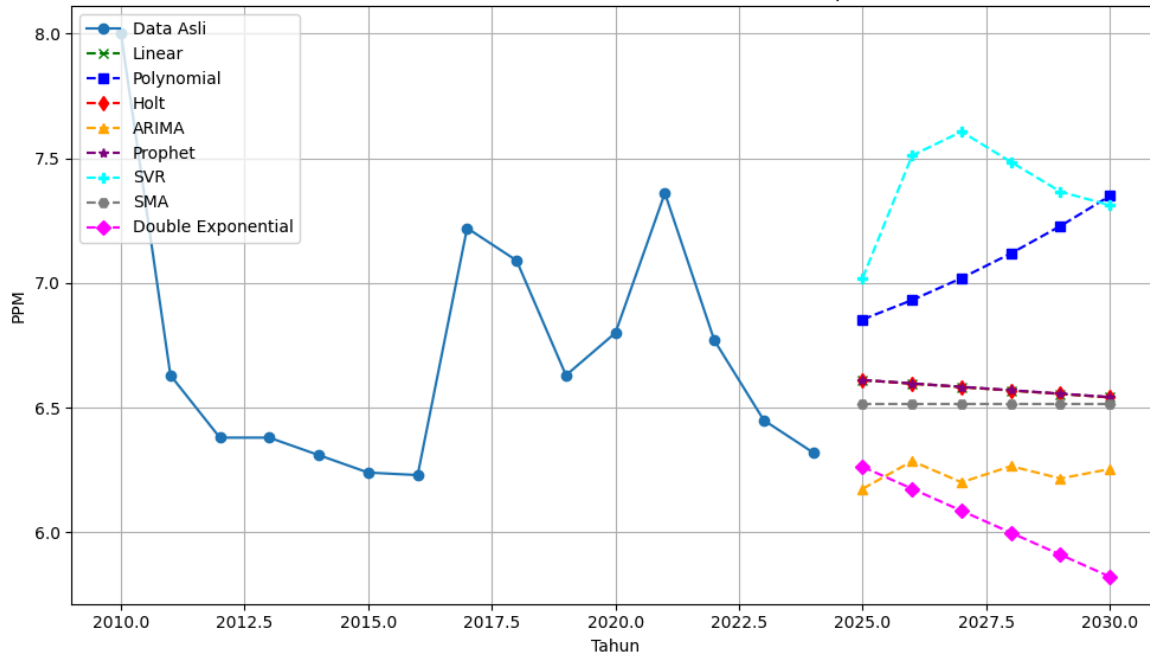
Tren historis lebih stabil, perlahan **menurun** dari 5,2% (2018) menjadi 3,9% (2024). Daerah ini menunjukkan tren perbaikan kesejahteraan/ekonomi.

Provinsi	Pola Historis	Proyeksi Masa Depan (2030)	Insight Penting
Papua	Fluktuatif, naik tajam 2024	Naik drastis	Risiko kenaikan signifikan, perlu antisipasi
DKI Jakarta	Fluktuasi tinggi, menurun	Stabil kisaran 5-6%	Stabil, kebijakan tenaga kerja penting
Kalimantan Utara	Menurun perlahan	Turun terus hingga ~2%	Potensi perbaikan ekonomi berkelanjutan
Nusa Tenggara Timur	Cenderung Stabil	Stabil dan naik 0.3%	Stabil, perlu kebijakan tenaga kerja

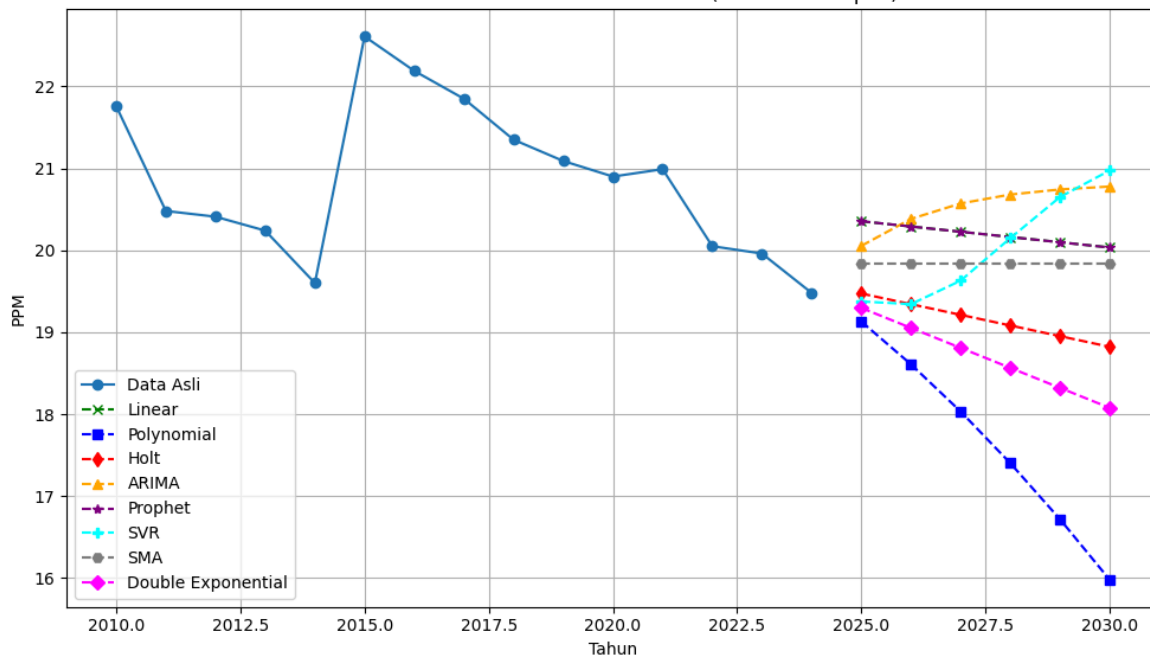
3.2.6.3 Persentase Penduduk Miskin (PPM)

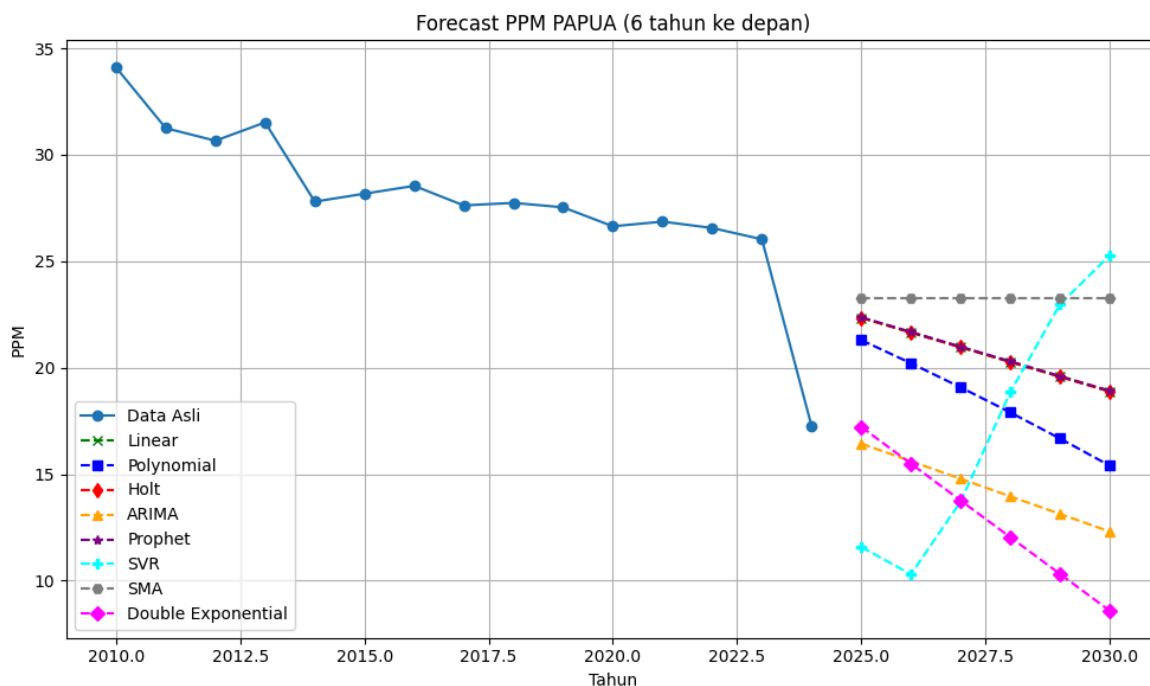


Forecast PPM KALIMANTAN UTARA (6 tahun ke depan)



Forecast PPM NUSA TENGGARA TIMUR (6 tahun ke depan)





1. Provinsi Papua

Angka kemiskinan di Papua menunjukkan penurunan yang cukup konsisten dari tahun 2010 hingga 2023, meskipun terdapat ketidakstabilan kecil di tengah-tengah periode. Namun, penurunan tajam terjadi pada tahun terakhir (2023), menunjukkan perubahan signifikan.

Hasil Forecast:

- Model Linear dan Polynomial menunjukkan tren penurunan yang konsisten, dengan polynomial menurun tajam.
- Model ARIMA dan Double Exponential juga menunjukkan penurunan yang signifikan.
- Model SMA dan Prophet menunjukkan stabilisasi di angka sekitar 23–24 PPM.
- Model SVR malah menunjukkan tren meningkat.

Jika kondisi ekonomi dan sosial yang mendorong penurunan angka kemiskinan di Papua pada tahun terakhir bisa dipertahankan atau diperkuat, maka model seperti Polynomial, Holt, dan ARIMA memprediksi penurunan yang berkelanjutan. Namun, beberapa model (SMA, Prophet) tidak bergerak (tetap), dan SVR bahkan memperkirakan kemungkinan kenaikan kembali. Ini menunjukkan pentingnya menjaga stabilitas dan penguatan program sosial di Papua.

2. Provinsi Nusa Tenggara Timur (NTT)

Data historis NTT relatif stabil, dengan sedikit penurunan angka kemiskinan

dari 2010 hingga 2023. Tahun 2015 menunjukkan lonjakan yang signifikan, namun kembali stabil sesudahnya.

Hasil Forecast:

- Polynomial dan Double Exponential memprediksi penurunan yang konsisten.
- Model ARIMA, Prophet, dan SVR memperkirakan kenaikan ringan.
- Model SMA dan Holt menunjukkan tren yang hampir datar.

Perbedaan antara model menunjukkan ketidakpastian di masa depan. Jika tidak ada intervensi baru, model konservatif memperkirakan kondisi stagnan. Namun, strategi yang tepat seperti peningkatan akses pendidikan dan lapangan kerja dapat membantu menurunkan angka kemiskinan lebih jauh sebagaimana diproyeksikan oleh model Polynomial.

3. Provinsi Kalimantan Utara

Angka kemiskinan di Kalimantan Utara menunjukkan ketidakstabilan dari tahun ke tahun, tanpa tren yang sangat jelas. Meskipun ada sedikit penurunan dalam jangka panjang, grafiknya relatif datar.

Hasil Forecast:

- Model Polynomial dan SVR memperkirakan sedikit peningkatan angka PPM.
- Model Double Exponential dan Holt menunjukkan sedikit penurunan.
- ARIMA dan SMA memproyeksikan stabilisasi.

Kalimantan Utara tampaknya menghadapi kondisi stagnan dalam pengentasan kemiskinan. Karena tren masa lalu tidak menunjukkan perubahan besar, masa depan juga diprediksi relatif datar oleh sebagian besar model. Perlu adanya pendekatan baru atau program intervensi agar angka kemiskinan bisa bergerak lebih signifikan ke arah penurunan.

4. DKI Jakarta

Jakarta menunjukkan tren tidak stabil tetapi stabil dalam jangka panjang, dengan sedikit penurunan dalam beberapa tahun terakhir. Kenaikan pada 2020–2021 bisa diasosiasikan dengan dampak pandemi COVID-19.

Hasil Forecast:

- Polynomial dan Prophet memproyeksikan peningkatan PPM.
- ARIMA dan SVR menunjukkan kemungkinan penurunan ringan.
- Model Holt dan Double Exponential menunjukkan stabilisasi dengan

kecenderungan menurun ringan

Sebagai pusat ekonomi nasional, Jakarta cenderung memiliki angka kemiskinan yang rendah, dan banyak model menunjukkan bahwa angka ini akan tetap stabil atau sedikit menurun. Namun, perhatian harus tetap diberikan terhadap kelompok rentan, terutama pasca-pandemi, agar tren ini tetap terkendali.

3.2.7 Insight Forecasting

Variabel	Provinsi	Tren Historis	Hasil Prediksi	Insight
Indeks Pembangunan Manusia (IPM)	DKI Jakarta	Naik stabil setiap tahun dengan sedikit perlambatan saat pandemi	Kenaikan konsisten hingga 2030	Jakarta menunjukkan pembangunan manusia yang berkelanjutan. Model sepakat bahwa kualitas hidup akan terus membaik.
	NTT	Naik perlahan tapi stabil, sempat menurun akibat pandemi	Naik perlahan, bervariasi antar model	Tren positif terus berlanjut, meskipun kecepatan peningkatannya relatif lambat. Perlu percepatan pembangunan sosial.
	Kalimantan Utara	Lonjakan awal karena pemekaran, lalu naik stabil	Naik tajam (Prophet), atau naik biasa (lainnya)	Awal yang kuat karena pembentukan provinsi baru, tetapi saat ini IPM naik secara bertahap. Model sangat bervariasi.
	Papua	Naik perlahan, lonjakan besar pada 2023	Naik signifikan, beberapa model prediksi sangat tinggi	Kemajuan pesat diduga karena intervensi besar pemerintah termasuk pemecahan provinsi. Namun, beberapa model perlu dikaji lebih lanjut untuk akurasi.

Variabel	Provinsi	Tren Historis	Hasil Prediksi	Insight
Pengangguran Terbuka (%)	DKI Jakarta	Fluktuatif, naik tajam saat pandemi, lalu menurun	Menurun konsisten hingga di bawah 6%	Pemulihan ekonomi pasca pandemi berjalan baik. Penurunan berkelanjutan menunjukkan efektivitas kebijakan ketenagakerjaan.
	NTT	Stabil rendah, sempat naik ringan saat	Tetap stabil	Kondisi ketenagakerjaan

		pandemi		relatif stabil. Tidak ada risiko besar, tetapi butuh dorongan untuk menciptakan lapangan kerja lebih luas.
	Kalimantan Utara	Menurun secara konsisten sejak pemekaran	Menurun hingga di bawah 2%	Tingkat pengangguran rendah dan terus membaik. Daerah ini menunjukkan pengelolaan pasar tenaga kerja yang baik.
	Papua	Cenderung fluktuatif, naik tajam 2024	Terus meningkat jika tren berlanjut	Kenaikan tajam perlu perhatian serius. Tantangan di sektor ketenagakerjaan Papua menjadi ancaman dalam pembangunan sosial jangka panjang.

Variabel	Provinsi	Tren Historis	Hasil Prediksi	Insight
Persentase Penduduk Miskin (PPM)	DKI Jakarta	Stabil di tingkat rendah, sedikit naik saat pandemi	Stabil / sedikit berubah	Jakarta berhasil mempertahankan angka kemiskinan rendah. Namun, tetap perlu perlindungan sosial untuk kelompok rentan.
	NTT	Cenderung tetap, penurunan sangat lambat	Bervariasi (turun atau tetap)	Kemiskinan di NTT masih cukup tinggi dan penurunannya lambat. Perlu strategi terpadu untuk percepatan pengentasan kemiskinan.
	Kalimantan Utara	Stabil tinggi, tidak banyak perubahan	Stabil atau sedikit menurun	Kemiskinan sulit ditekan secara cepat. Model menunjukkan perlunya program pemberdayaan ekonomi yang lebih kuat.
	Papua	Turun secara bertahap, tapi tidak menentu	Bervariasi (stagnan / turun / naik)	Prediksi tidak konsisten. Masih terdapat risiko meningkatnya kemiskinan jika tidak ada intervensi lebih lanjut.

3.2.7 Factor Analysis

3.2.7.1 EDA dan Preprocessing

Kolom yang digunakan berjumlah 15, sisanya akan digunakan sebagai patokan. Pastikan seluruh kolom yang digunakan berada dalam dataset. Dataset sudah cukup baik karena tidak ada nilai kosong dan duplikat atau kondisi khusus yang harus dilakukan penanganan. Kemudian dilakukan scaling data dan uji asumsi menggunakan KMO dan Bartlett's Test.

3.2.7.2 Uji Asumsi

a) KMO Test

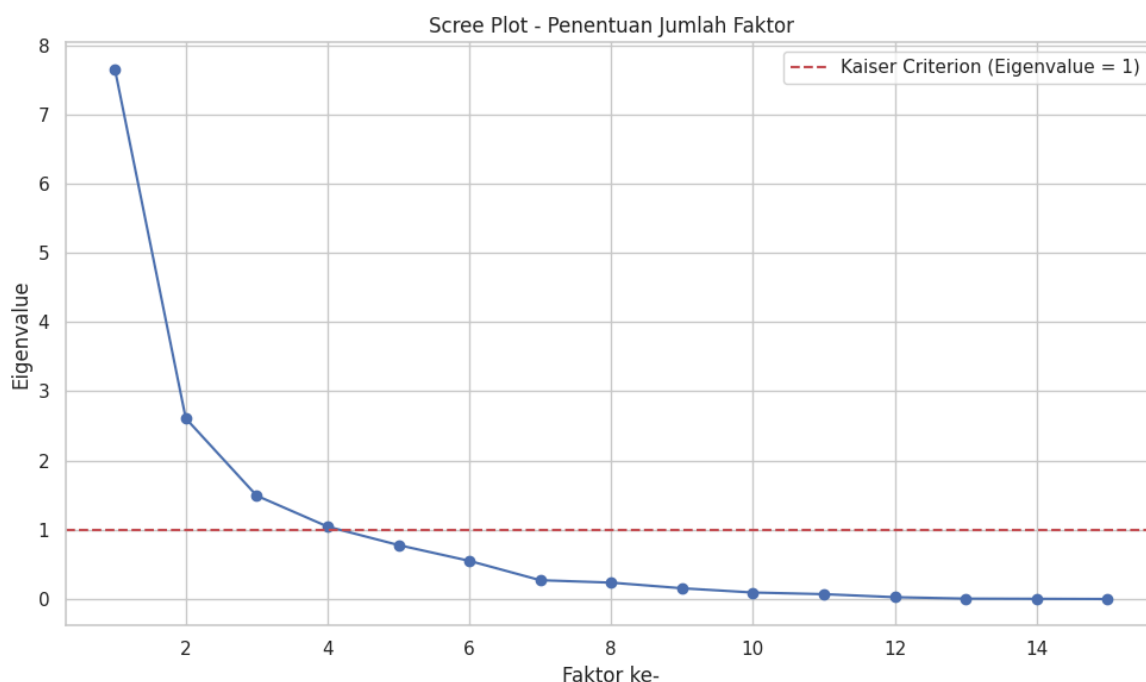
Nilai KMO kurang dari 0,6 menunjukkan bahwa pengambilan sampel tidak memadai dan tindakan perbaikan harus diambil. Jika nilainya kurang dari 0,5, hasil analisis faktor jelas tidak akan cocok untuk analisis data (Shrestha, N., 2021). Dalam proyek ini nilai KMO dari keseluruhan data adalah 0.7358 dimana hasil ini menyimpulkan bahwa dataset ini layak dan sesuai untuk dilakukan *Factor Analysis*.

b) Bartlett's Test

Bartlett's Test memiliki indikator nilai signifikan $< 0,05$ menunjukkan bahwa analisis faktor mungkin bermanfaat untuk kumpulan data (Shrestha, N., 2021). Dari proyek ini setelah dilakukan Bartlett's Test menghasilkan p-value 0.000 dan Chi-Square 910.0260 dimana hal ini menunjukkan korelasi antar variabel signifikan.

3.2.7.3 Jumlah Faktor Berdasarkan Eigenvalue

Disini kami menggunakan Kaiser's Rule yakni mengambil faktor yang memiliki eigenvalue > 1 . Pada tahap ini ditampilkan juga scree plot untuk melihat penentuan jumlah faktornya, dengan hasil pada gambar dibawah ini.



Dari grafik scree plot diatas, bisa disimpulkan faktor yang diambil sebanyak empat faktor yang cukup digunakan untuk mengambil interpretasi ketika melakukan analisis lebih lanjut.

3.2.7.4 Interpretasi Faktor

Faktor	Variabel Dominan	Interpretasi	Bidang
Faktor 1	IPM, pengeluaran per kapita laki-laki, pengeluaran per kapita perempuan, angka harapan hidup laki-laki & perempuan, persentase penduduk miskin, persentase ketidakcukupan konsumsi pangan	Faktor ini memiliki variabel-variabel yang menggambarkan kualitas hidup, kondisi sosial ekonomi penduduk, dan daya beli masyarakat	Kesejahteraan Sosial dan Ekonomi
Faktor 2	Harapan lama sekolah laki-laki, harapan lama sekolah perempuan, rata-rata lama sekolah	Faktor ini memiliki beberapa variabel yang mengindikasikan pendidikan dan kualitas SDM dari aspek pendidikan	Pendidikan dan Harapan Lama Sekolah
Faktor 3	TPAK, tingkat pengangguran terbuka, rata-rata lama sekolah	Faktor ini memiliki indikasi kombinasi antara keterlibatan penduduk usia kerja dan tingkat pengangguran	Ketenagakerjaan dan Pengangguran
Faktor 4	Penduduk usia 15 tahun ke atas, Nikah	Menggambarkan struktur penduduk usia produktif dan status perkawinan	Komposisi Penduduk Dewasa dan Pernikahan

Setelah diketahui beberapa interpretasi dari faktor nya, kami juga mencoba untuk melihat komunalitas yakni seberapa besar proporsi varians dari masing-masing variabel yang berhasil dijelaskan oleh faktor-faktor yang terbentuk.

Variabel	Nilai Komunalitas
IPM	1.004638
penduduk_usia_15_tahun_keatas	0.997424
HLS laki-laki	0.977019

Nikah	0.972124
HLS perempuan	0.958897
rata_lama_sekolah	0.884553
Pengeluaran per kapita laki-laki	0.827293
angka_harapan_hidup_perempuan	0.804329
TPAK	0.802042
angka_harapan_hidup_lakilaki	0.788700
Pengeluaran per kapita perempuan	0.758323
indek_tingkat_pengangguran_terbuka	0.756793
persentase_penduduk_miskin	0.718622
persentase_ketidakcukupan_konsumsi_pangan	0.572087
Laju pertumbuhan penduduk per tahun	0.412802

3.2.7.5 Interpretasi Provinsi Berdasarkan Kluster (*Lampiran 4*)

a. Kluster 0; Provinsi Maju

- Jawa Timur tergolong dalam kluster 0 yang bisa dibilang maju dan stabil. Provinsi ini condong memiliki Kesejahteraan Sosial dan Ekonomi dan Pendidikan yang sangat baik. Selain itu, di sisi lain Ketenagakerjaan dan pengangguran cukup baik sehingga tingkat pengangguran setidaknya tidak ada peningkatan angka. Provinsi ini juga memiliki nilai tinggi pada faktor Komposisi Penduduk Dewasa dan Pernikahan sehingga terkelola dengan baik.

b. Kluster 1; Tertinggal

- Papua Barat tergolong dalam kluster 1 yang bisa dibilang memiliki tantangan dan tekanan di kebanyakan aspek. Provinsi ini membutuhkan intervensi kebijakan pembangunan secara menyeluruh dari pemerintah khususnya dalam pendidikan, ketenagakerjaan, dan pelayanan sosial ekonomi. Faktor 1 yakni Kesejahteraan sosial & ekonomi memiliki nilai yang sangat rendah sehingga provinsi ini bisa dikatakan tertinggal. Tidak itu saja, begitu juga pada faktor-faktor yang lain memiliki nilai yang cukup memprihatinkan dan butuh penanganan serta intervensi kebijakan dari pemerintah.

c. Kluster 2; Menengah/Proses Berkembang

- Kalimantan Barat tergolong dalam kluster 2 yang bisa dibilang memiliki posisi menengah dan dalam proses berkembang dan masih menghadapi beberapa ketertinggalan. Potensinya cukup besar jika intervensi pembangunan fokus pada faktor-faktor yang perlu ditingkatkan. Provinsi ini memiliki nilai yang cukup miris namun masih tergolong tidak terlalu tertinggal baik itu dalam bidang kesejahteraan sosial & ekonomi, pendidikan, dan Ketenagakerjaan.

Kesimpulan dan Saran

Kesimpulan

1. Terdapat Ketimpangan Regional yang Signifikan dalam Kesiapan Menghadapi Bonus Demografi
 - a. Hasil clustering menunjukkan bahwa provinsi-provinsi di Indonesia tidak berada dalam level kesiapan yang sama.
 - b. Klaster 0 berisi provinsi dengan kesiapan tinggi (misalnya Jawa Timur dan DKI Jakarta), yang ditandai dengan IPM tinggi, tingkat pendidikan baik, ekonomi kuat, dan struktur demografi yang stabil.
 - c. Klaster 1 berisi provinsi yang tertinggal (misalnya Papua Barat), dengan indikator sosial ekonomi, pendidikan, dan ketenagakerjaan yang masih sangat rendah.
 - d. Klaster 2 berisi provinsi menengah (misalnya Kalimantan Barat) yang sedang dalam proses berkembang dan memiliki potensi baik jika didukung oleh kebijakan yang tepat.
2. Faktor Utama Penentu Kesiapan Provinsi adalah Indikator Kesejahteraan, Pendidikan, dan Tenaga Kerja
 - a. Analisis Faktor dan PCA mengidentifikasi bahwa kualitas hidup (harapan hidup, pengeluaran per kapita), pendidikan (HLS dan lama sekolah), serta pengangguran dan partisipasi angkatan kerja merupakan penentu paling dominan terhadap kesiapan menghadapi bonus demografi.
 - b. Regresi dan Random Forest membuktikan bahwa persentase kemiskinan, pengeluaran per kapita, serta tingkat pengangguran

memiliki bobot pengaruh besar terhadap IPM dan TPAK sebagai target utama kesiapan.

3. Model Prediksi IPM Lebih Stabil dan Akurat Dibanding TPAK

- a. Model OLS dan Random Forest memberikan performa prediksi IPM yang sangat baik ($R^2 > 0.95$), sementara TPAK hanya cukup baik ($R^2 \sim 0.75$ atau lebih rendah pada OLS).
- b. Hal ini menunjukkan bahwa IPM lebih konsisten dan bisa dijadikan indikator utama, sedangkan TPAK dipengaruhi oleh faktor-faktor kontekstual yang mungkin tidak tercatat dalam data makro.

4. Provinsi Ekstrem Memerlukan Penanganan Khusus

- a. Provinsi seperti Papua Pegunungan atau Papua Barat memiliki perbedaan besar antara prediksi dan realisasi, yang menandakan adanya variabel lokal (misal: budaya kerja, geografis, kebijakan lokal) yang tidak tertangkap oleh model statistik.
- b. Perlu pendekatan yang lebih adaptif dan kontekstual dalam kebijakan, terutama untuk provinsi dengan nilai residual model yang tinggi.

5. Forecasting Menunjukkan Perlu Perhatian pada Ketimpangan Laju Pertumbuhan Pembangunan

- a. Model prediksi menunjukkan tren kenaikan IPM di sebagian besar provinsi, namun juga menandai adanya ancaman stagnasi atau penurunan kualitas hidup di beberapa daerah (misalnya Kalimantan Utara dan Papua).
- b. Fluktuasi pengangguran juga mengindikasikan bahwa pemulihan pasca pandemi belum merata di semua wilayah.

Saran

1. Segmentasi Strategi Berdasarkan Klaster Kesiapan

- a. Klaster 0 (Maju): Fokus pada pengembangan inovasi SDM, teknologi, dan produktivitas tinggi untuk memanfaatkan bonus demografi sebagai keunggulan kompetitif global.
- b. Klaster 1 (Tertinggal): Butuh intervensi menyeluruh dalam bentuk program afirmatif di bidang pendidikan dasar, kesehatan, pengentasan kemiskinan, dan penciptaan lapangan kerja lokal.
- c. Klaster 2 (Menengah): Fokus pada akselerasi pembangunan dan pemberdayaan lokal dengan mendorong sektor unggulan, digitalisasi, dan akses layanan publik.

2. Kebijakan Berbasis Data dan Analitik Multidimensional

- a. Pemerintah pusat dan daerah perlu mengadopsi model data-driven policy, menggunakan analisis PCA dan FA untuk memahami prioritas pembangunan di setiap wilayah.
 - b. Daerah harus mengidentifikasi faktor dominan yang memperlambat kesiapan dan menyusun kebijakan sektoral prioritas berbasis data tersebut.
3. Penguatan Intervensi Ekonomi dan Sosial pada Provinsi Rentan
 - a. Daerah dengan nilai kemiskinan dan ketidakcukupan pangan tinggi perlu penguatan perlindungan sosial, subsidi pendidikan, serta pembukaan lapangan kerja padat karya.
 - b. Program seperti BLT, PKH, dan KIP/KIS harus ditargetkan lebih presisi kepada provinsi klaster 1 dengan strategi pengawasan yang lebih ketat.
4. Monitoring dan Evaluasi Progres Tahunan Menggunakan Model Machine Learning
 - a. Pemerintah perlu menggunakan pendekatan forecasting seperti Prophet, Holt, atau Random Forest untuk memantau kemajuan indikator pembangunan secara berkala.
 - b. Hal ini membantu mengidentifikasi early warning system bagi daerah-daerah yang menunjukkan tren penurunan atau stagnasi.
5. Perluasan Investasi dalam Pendidikan dan Kesehatan di Wilayah Timur Indonesia
 - a. Data menunjukkan wilayah Indonesia Timur (seperti Papua dan NTT) memiliki nilai-nilai rendah pada faktor pendidikan dan kesehatan.
 - b. Pemerintah perlu memperkuat alokasi Dana Otonomi Khusus dan mengawal program transformasi layanan dasar agar lebih berdampak jangka panjang.

Kendala dan Rencana Tindak Lanjut

1. Kendala

- Pencarian dataset untuk forecasting butuh lebih kompleks lagi.
- Ada beberapa provinsi hasil pemekaran yang hanya memiliki value di tahun 2024.

2. Rencana Tindak Lanjut

- Mencari lembaga atau instansi yang menyediakan dataset yang lebih kompleks.

- Menyesuaikan model yang akan digunakan.
- Pemekaran provinsi untuk dataset yang *Up to date*.

Daftar Pustaka

Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan, mengikuti format APA. Hanya pustaka yang disitasi pada usulan penelitian yang dicantumkan dalam Daftar Pustaka. Pustaka yang disitasi maksimal 8 tahun terakhir sebanyak minimal 10 pustaka.

Shrestha, N. (2021). Factor analysis as a tool for survey analysis. *American Journal of Applied Mathematics and Statistics*, 9(1), 4-11.
<https://doi.org/10.12691/ajams-9-1-2>

Adriani, D. & Yustini, T. (2021). *Anticipating the demographic bonus from the perspective of human capital in Indonesia*. *International Journal of Research in Business & Social Science*, 10(6), 141-152.
<https://doi.org/10.20525/ijrbs.v10i6.1377>

Mian, L., et al. (2022). *The impact of E-education and innovation on unemployment reduction among graduates: A way forward for higher educational institutes*. *Frontiers in Psychology*, 13:914104.
<https://doi.org/10.3389/fpsyg.2022.914104>

Zeng, Y., et al. (2015). *Research of Improved FP-Growth Algorithm in Association Rules Mining*. *Scientific Programming*, 2015, Article ID 910281.
<https://doi.org/10.1155/2015/910281>

Sasmita, N. R., Phonna, R. A., Fikri, M. K., Khairul, M., Apriliansyah, F., Idroes, G. M., Puspitasari, A., & Saputra, F. E. (2024). *Statistical assessment of Human Development Index variations and their correlates: A case study of Aceh Province, Indonesia*. *Grimsa Journal of Business and Economics Studies*, 1(1). <https://doi.org/10.61975/gjbes.v1i1.14>

Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, 1–12. Dallas, TX, USA.

Zeng, Y., Yin, S., Liu, J., & Zhang, M. (2015). Research of improved FP-Growth algorithm in association rules mining. *Scientific Programming*, 2015, Article ID 910281, 1–6. <https://doi.org/10.1155/2015/910281>

Firmansyah, & Nurdiawan, O. (2023). Penerapan data mining menggunakan algoritma Frequent Pattern-Growth untuk menentukan pola pembelian produk chemicals. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 547–551.

Bukharova, E. B., Vorontsova, I. P., Semenova, A. R., Vitkovskaya, L. K., Popelnitskaya, I. M., & Drobyshev, I. A. (2019). Methodological approaches to homogenous regional clusters formation for human capital and quality of life assessment in the Yenisei Siberia macro-region. *Journal of Siberian Federal University. Humanities & Social Sciences*, 12(12), 2137–2154.

Fitriyah, Z., Irsalina, S., Herlandy K, A. R., & Widodo, E. (2021). Analisis Faktor yang Berpengaruh Terhadap IPM Menggunakan Regresi Linear Berganda. *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, 2(3), 282-291. <https://doi.org/10.46306/lb.v2i3.86>

Ulma, D. N., & Julia, A. (2022). Peningkatan penyerapan tenaga kerja dalam mendorong bonus demografi di Kecamatan Sukajadi Bandung. *Jurnal Riset*

Munawaroh, S., Gandhiadi, G. K., & Nilakusmawati, D. P. E. (2023). Analisis variabel Indeks Pembangunan Manusia menggunakan analisis regresi linier berganda di Provinsi Jawa Timur. *E-Jurnal Matematika*, 12(4), 289–294. <https://doi.org/10.24843/MTK.2023.v12.i04.p431>

Samudra, A. B. B., & Wahed, M. (2023). Pengaruh rata lama sekolah, umur harapan hidup serta PDRB per kapita terhadap kemiskinan melalui analisis jalur pengangguran di Daerah Istimewa Yogyakarta. *Journal of Economics and Business UBS*, 12(3), Mei–Juni.

Rohma, F., Al Haqqi, F. M. L., Kurniawan, M., & Khotimah, K. (2024). Pengaruh tingkat pendidikan, kemiskinan, dan pertumbuhan ekonomi terhadap Indeks Pembangunan Manusia (IPM) di Indonesia tahun 2012–2022 (Studi kasus Provinsi Sumatera Selatan). *Maslahah: Jurnal Manajemen dan Ekonomi Syariah*, 2(3), 46–67. <https://doi.org/10.59059/maslahah.v2i3.1344>

Frost, J. (2019). Regression Analysis: An Intuitive Guide for Using and Interpreting Linear Models. Statistics By Jim Publishing.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer.

Lagomarsino, A., Bonaccorso, B., & Sappa, G. (2023). Assessing the efficiency of a random forest regression model for estimating water quality indicators. *Meteorology Hydrology and Water Management*, 11(1), 3-17.
<https://doi.org/10.26491/mhwm/183734>

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202.
<https://doi.org/10.1098/rsta.2015.0202>

Tato, F. R., & Yasin, H. M. (2025). Expanding the Horizons of Principal Component Analysis: Versatile Applications from Environmental Monitoring to Chemometrics. *Engineering and Technology Journal*, 10(03).
<https://www.europub.co.uk/articles/-A-761758>

Boudt, K., d'Errico, M., Luu, H. A., & Pietrelli, R. (2022). Interpretability of Composite Indicators Based on Principal Components. *Journal of Probability and Statistics*, 2022, Article ID 4155384.
<https://doi.org/10.1155/2022/4155384>

Harvey, D. T., & Hanson, B. A. (2024). Understanding Scores and Loadings in Principal Component Analysis. *LearnPCA Vignette*, CRAN.

Mesa, F., Correa Velez, G., & Cardenas Alzate, P. P. (2018). Analysis of economic and social indicators through the principal components analysis. *Contemporary Engineering Sciences*, 11(16), 763–769.
<https://doi.org/10.12988/ces.2018.8120>

ESRI. (2023). Performing Principal Component Analysis (PCA) to Determine Weights for Index Indicators. ArcGIS Blog.
<https://www.esri.com/arcgis-blog/products/api-python/analytics/performing-principal-component-analysis-pca-to-determine-weights-for-index-indicators/>

Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: How to use principal components analysis. *Health Policy and Planning*, 21(6), 459–468. <https://doi.org/10.1093/heapol/czl029>

Cruz-Nájera, M. A., Treviño-Berrones, M. G., Ponce-Flores, M. P., Terán-Villanueva, J. D., Castán-Rocha, J. A., Ibarra-Martínez, S., Santiago, A., & Laria-Menchaca, J. (2022). *Short time series forecasting: Recommended methods and techniques*. *Symmetry*, 14(6), 1231. <https://doi.org/10.3390/sym14061231>

Avelin, B., Viitasaari, L., & Malo, P. (2025). Predictive models in econometric time-series. Uppsala University Working Paper.

LAMPIRAN

Lampiran 1: forecasting 2030 IPM

Provinsi	Linear Regression	Polynomial Regression	Holt Exponential	ARIMA	Prophet	SVR	SMA	Double Exponential
ACEH	76.9843095 2	76.6757714 9	77.0047884 8	76.976 42	77.7183 2	70.6 272	73.4 1	77.06930699
BALI	80.7490714 3	79.7957618	81.0429228 3	81.005 49	81.9028 3	74.0 6958	77.1	80.9519361
BANTEN	77.2531309 5	75.9982071 7	77.4546417 9	77.339 65	77.9729 5	71.2 6844	73.8 9	77.29479308
BENGKULU	76.9774285 7	74.9388910 8	76.8365235 8	76.654 67	76.8972 1	69.7 275	72.7 766 7	76.54135879
DI YOGYAKARTA	84.4697381	83.2822310 9	84.1985713 9	83.526 78	84.1870 8	78.5 8131	81.0 866 7	84.02589899
DKI JAKARTA	85.5315476 2	84.7366370 4	85.9815589 3	86.217 39	87.1145 8	79.8 0065	82.3 966 7	86.01113805
GORONTALO	74.6186904 8	74.3230284 4	74.9079962 4	74.879 98	75.6219 4	67.0 9098	70.4 966 7	74.88075758

JAMBI	76.8211666 7	74.5936192 6	76.8757142 7	76.626 95	77.0158 7	69.7 2975	72.7 8	76.54785363
JAWA BARAT	77.9774761 9	76.4413455 1	77.9786762 8	77.942 71	78.3582 8	70.5 612	73.7 633 3	77.82199872
JAWA TENGAH	77.4148928 6	75.4291955 4	77.2228571	76.826 26	77.3761 5	70.2 8215	73.3 533 3	76.96992416
JAWA TIMUR	77.8491904 8	76.1002414 3	77.8314286 9	77.642 88	77.9525 8	70.0 4817	73.4 066 7	77.60110484
KALIMAN TAN BARAT	73.3279166 7	71.6896283 1	73.6276904 6	71.600 06	74.6238 5	66.3 2515	69.3 9	73.535198
KALIMAN TAN SELATAN	76.4133928 6	74.9845216 5	76.3857486 6	76.143 15	76.5971 5	69.3 6215	72.4 566 7	76.18266138
KALIMAN TAN TENGAH	75.8808452 4	74.4189337 4	75.6314281 8	75.575	75.7085 4	69.5 2892	72.1 866 7	75.46283699
KALIMAN TAN TIMUR	81.7132857 1	81.2220543	81.7132856 5	81.986 78	82.7962 1	75.2 9318	78.1 566 7	82.06182756
KALIMAN TAN UTARA	118.109833 3	-13.657284 78	105.390953 8	77.229 16	118.057 2	55.5 4502	72.4 466 7	84.52530133

KEPULAUAN BANGKA BELITUNG	76.5707381	75.17915869	76.4627791	76.35434	76.67685	69.90971	72.80667	76.26666406
KEPULAUAN RIAU	80.18983333	80.12448158	80.90081024	81.91784	82.22448	74.71428	77.18	81.03524344
LAMPUNG	75.35871429	73.37848319	75.28138615	74.38864	75.65156	68.11826	71.13667	75.06312445
MALUKU	74.57063095	73.35652165	74.69904964	74.23148	75.33525	68.16528	70.91	74.60329335
MALUKU UTARA	74.36669048	73.0980404	74.51688411	74.49603	75.53291	67.22574	70.23667	74.53151189
NUSATENGGA BARAT	75.04885714	73.407532	75.11701928	74.90206	75.49839	66.26246	70.19667	74.86200833
NUSATENGGA TIMUR	70.593	68.90648739	70.89604873	70.77037	71.64689	63.38325	66.65667	70.64660821
PAPUA	68.84195238	73.96028248	68.84221061	88.1045	72.72861	62.04958	64.09	76.54992759
PAPUA BARAT	70.42967857	71.17490724	70.20060705	70.15988	70.69015	63.31293	66.52333	70.35683584

RIAU	77.2081785 7	76.5227139 6	77.4220844 1	77.349 26	78.3737 7	72.1 0527	74.1 166 7	77.49525587
SULAWES I BARAT	71.9440714 3	70.8391651 6	71.825776	71.774 09	71.8701 4	64.0 5183	67.5 566 7	71.61964505
SULAWES I SELATAN	77.5467976 2	76.4373969	77.5000147	77.394 18	77.6980 6	70.2 6162	73.4 433 3	77.34392505
SULAWES I TENGAH	75.1136428 6	72.7923193 3	75.1042855 9	74.777	75.1306	67.7 7522	70.9 3	74.68892802
SULAWES I TENGGAH A	76.6700119	75.9060766	76.6904854 2	77.030 53	77.0517 7	69.9 0401	72.8 333 3	76.61213346
SULAWES I UTARA	78.03875	77.2271978	78.0591778 6	78.078 7	78.4305 2	71.6 6211	74.4	78.02776645
SUMATER A BARAT	77.4833095 2	76.3074638	77.5934439 5	77.560 47	78.0439 5	71.1 9674	73.8 333 3	77.47614251
SUMATER A SELATAN	75.5410238 1	74.2536671	75.6691641 8	75.630 15	76.4435 5	68.5 2833	71.6 066 7	75.564287
SUMATER A UTARA	76.8455714 3	76.5489376 2	76.9900092 1	76.825 31	78.0682 9	70.6 2134	73.3 666 7	77.13484396

Lampiran 2: forecasting 2030 Pengangguran Terbuka

Provinsi	Linear Regression	Polynomial Regression	Holt Exponential	ARIMA	Prophet	SVR	SMA	Double Exponential
ACEH	5.3989285 71	2.621429	5.399648	5.79490 6	5.397292	6.260 134	5.983 333	5.02043
BALI	4.1507142 86	-29.601	2.179421	1.70100 1	4.157875	3.828 823	3.093 333	0.931367
BANTEN	5.4307142 86	-9.56595	5.430714	8.08022 8	5.415753	7.638 981	7.43	4.320719
BENGKULU	3.21	-1.63	3.20999	3.46625 3	3.212627	3.500 477	3.373 333	2.656876
DI YOGYAKARTA	4.1142857 14	-4.94238	3.602868	3.73571 1	4.116209	4.560 998	3.743 333	2.967474
DKI JAKARTA	5.8660714 29	-16.2164	5.866088	7.37815 9	5.880825	6.974 162	6.64	4.726291
GORONTALO	1.8139285 71	3.216429	1.813528	3.36703 1	1.814669	2.932 389	2.923 333	2.772496
JAMBI	5.3675	-2.39667	5.230195	4.48357	5.367219	4.615 165	4.533 333	5.095799
JAWA BARAT	5.9317857 14	-13.034	5.270387	6.82782 5	5.937371	7.873 87	7.5	4.625289
JAWA TENGAH	5.71	-6.84833	5.279697	5.12540 2	5.710535	5.711 18	5.16	4.472339

JAWA TIMUR	5.6775	-9.40167	4.471038	4.16647 9	5.679828	5.172 324	4.853 333	3.761642
KALIMAN TAN BARAT	5.9064285 71	-4.03024	5.906428	4.83130 2	5.907696	4.809 396	5.006 667	5.164485
KALIMAN TAN SELATAN	4.4346428 57	-0.59786	4.05	4.22286 4	4.435819	4.542 961	4.416 667	3.679229
KALIMAN TAN TENGAH	4.2364285 71	-0.38357	4.11	4.03348 7	4.240669	4.048 802	4.123 333	4.042802
KALIMAN TAN TIMUR	4.0275	-2.67333	4.02756	5.10713	4.028907	6.560 756	5.386 667	3.314841
KALIMAN TAN UTARA	2.8032142 86	1.730714	2.802687	3.89385 7	2.801332	5.218 69	4.08	2.197191
KEPULAU AN BANGKA BELITUN G	5.9492857 14	-2.28238	5.949286	4.38833	5.95008	5.277 066	4.653 333	4.98618
KEPULAU AN RIAU	5.4535714 29	-15.8131	4.739725	6.41896 8	5.455064	8.924 413	7.14	3.555206
LAMPUN G	4.5635714 29	-0.18476	4.340582	4.25743 2	4.559961	4.404 68	4.313 333	4.083614

MALUKU	5.5010714 29	0.211905	5.501216	6.07568 4	5.501359	6.879 156	6.433 333	4.940359
MALUKU UTARA	3.2410714 29	0.555238	3.241071	4.39100 1	3.241017	3.738 828	4.106 667	3.694147
NUSA TENGGAR A BARAT	1.66	-0.00833	1.659657	2.76660 5	1.66122	3.416 234	2.806 667	1.630134
NUSA TENGGAR A TIMUR	3.3175	-5.10667	3.189855	3.25900 4	3.3181	3.440 505	3.233 333	3.117159
PAPUA	6.0782142 86	17.76571	6.078299	3.69100 6	6.087951	6.933 259	3.993 333	9.421794
PAPUA BARAT	2.3996428 57	-3.98952	2.400126	2.83455 7	2.400283	4.336 623	4.96	2.266235
RIAU	1.1596428 57	-0.07786	1.159706	3.78001 1	1.166906	4.325 985	4.1	1.692138
SULAWES I BARAT	1.7289285 71	0.766429	1.728864	2.77519 7	1.730086	3.263 003	2.43	2.233825
SULAWES I SELATAN	3.4575	-5.425	3.4575	4.53348 6	3.460053	4.609 155	4.343 333	2.82541
SULAWES I TENGAH	2.505	-0.94167	2.505	3.24362 8	2.504921	3.557 166	2.963 333	2.133518
SULAWES I	2.8178571 43	-4.60714	2.989889	3.41531 4	2.824954	3.638 563	3.2	3.047943

TENGGAR A								
SULAWES I UTARA	5.5964285 71	-1.6269	5.596063	6.47077 4	5.595465	7.507 166	6.186 667	4.268834
SUMATER A BARAT	6.3125	-1.48833	6.312425	6.01554 2	6.29878	6.712 546	5.99	5.16661
SUMATER A SELATAN	3.6075	-5.275	3.45	3.83570 2	3.608984	4.549 879	4.2	3.476866
SUMATER A UTARA	6.1042857 14	-1.94405	6.104285	5.93660 2	6.066155	6.279 306	5.883 333	5.133816

Lampiran 3: forecasting 2030 Persentase Penduduk Miskin

Provinsi	Linear Regression	Polynomial Regression	Holt Exponential	ARIMA	Prophet	SVR	SMA	Double Exponential
ACEH	11.4613928 6	14.0104757 6	11.77089	12.220 15	12.1141 1	16.5 632	14.4 4	12.30724
BALI	3.65336904 8	5.23991047 1	3.653251	4.1420 72	3.65534 7	4.68 818 8	4.27 3333	3.740937
BANTEN	5.69509523 8	8.81132934 8	5.334287	5.8329 63	5.69759 4	6.17 353 7	6.05 6667	5.852314

BENGKULU	12.45772619	8.301816096	12.35572	12.89671	11.38958	15.63636	14.07333	11.34595
DI YOGYAKARTA	8.255452381	9.32221655	8.255449	9.368374	8.26231	13.33294	11.07	8.614443
DKI JAKARTA	4.757107143	5.927243698	4.410845	4.0985	4.757334	4.181669	4.476667	4.511035
GORONTALO	13.70066667	9.501634131	13.71515	14.59829	12.99157	15.8426	15.04667	12.88468
JAMBI	7.045142857	5.5153287	6.778991	7.178028	7.041482	7.996832	7.433333	6.317865
JAWA BARAT	5.651916667	8.580477054	5.972847	7.43858	5.654117	9.078745	7.713333	6.446767
JAWA TENGAH	7.412988095	10.7526086	7.985824	10.36263	8.34996	12.48594	10.72333	8.563851
JAWA TIMUR	7.923857143	10.33482676	7.612813	9.48273	7.963047	11.74085	10.17333	8.090065
KALIMANTAN BARAT	5.540464286	4.933713315	5.540002	6.320129	5.538004	7.654018	6.586667	5.13042

KALIMAN TAN SELATAN	3.73735714 3	4.27669133 8	3.737369	4.0688 78	3.73665 2	4.91 065 5	4.29 6667	3.529584
KALIMAN TAN TENGAH	3.80140476 2	6.49321687 2	5.53	5.3188 83	3.80552 2	5.99 324	5.18 6667	4.842645
KALIMAN TAN TIMUR	5.42321428 6	7.47341305 8	4.828935	5.4075 06	5.42239 8	6.85 396 6	6.06 6667	5.312511
KALIMAN TAN UTARA	6.54145238 1	7.34875306 9	6.54158	6.2542 02	6.54273 8	7.31 171 9	6.51 3333	5.822601
KEPULAU AN BANGKA BELITUNG	3.57548809 5	5.3211574	3.575486	4.5373 56	3.57449 9	6.20 177 8	4.50 6667	4.073957
KEPULAU AN RIAU	4.81001190 5	6.62832126 7	4.363622	5.3614 2	4.81115 5	6.87 147 4	5.76 6667	4.568393
LAMPUNG	8.09544047 6	9.95237782 8	7.659361	8.2585 02	8.52540 5	13.9 049 4	11.1 2333	8.073164
MALUKU	12.4406666 7	18.5050181	12.07706	14.935 14	12.4387 7	19.6 049 2	16.1 4667	13.68287
MALUKU UTARA	4.47779761 9	9.62394473 4	4.813152	6.0746 89	4.47614 2	6.75 099 3	6.33 6667	5.614885

NUSA TENGGAR A BARAT	9.30271428 6	13.9803451 8	9.194397	10.440 38	9.32854 8	15.8 461 3	13.4 8	10.49938
NUSA TENGGAR A TIMUR	20.0347857 1	15.9788468	18.82205	20.778 33	20.0332 8	20.9 742 8	19.8 3	18.07676
PAPUA	18.8828333 3	15.4007265 6	18.88285	12.303 34	18.9163 2	25.2 683	23.2 8333	8.584591
PAPUA BARAT	16.4511547 6	18.6768800 9	17.78353	20.809 1	17.8413 4	25.7 858 7	21.1 6	19.37435
RIAU	5.33905952 4	6.38067905 5	5.339058	6.6714 51	5.34095 5	8.65 564 1	6.71	5.829476
SULAWESI BARAT	9.47858333 3	14.3536056 9	9.710007	11.171 21	11.1979 5	12.4 469	11.4 8333	10.75745
SULAWESI SELATAN	7.06389285 7	8.59844440 8	7.063892	7.9754 08	7.16583	9.33 309 1	8.46 3333	7.079314
SULAWESI TENGAH	10.0885595 2	11.4384741 4	10.0884	9.7807 19	10.0953 6	13.6 707 2	12.1 7	9.936038
SULAWESI TENGGAR A	8.89723809 5	12.4340071 1	9.315511	11.210 13	8.90487 2	12.6 390 1	11.2 7	10.30588

SULAWESI UTARA	6.53501190 5	6.80309987	6.534922	7.2534 61	6.53392 7	8.10 258 7	7.30 3333	6.596501
SUMATERA BARAT	4.24645238 1	7.05439140 3	4.462769	5.3596 81	4.25510 2	7.23 504	5.94 6667	5.053486
SUMATERA SELATAN	10.3496547 6	8.73444602 3	10.34966	10.970 26	10.3377 2	12.7 985 2	11.5 5	9.07383
SUMATERA UTARA	6.69705952 4	6.0849638	6.696865	6.8413 99	6.70059 3	9.90 993 5	8.18 6667	6.555903

Lampiran 4 : Hasil Klasterisasi Provinsi

Provinsi	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Klaster
JAWA TENGAH	5.326644	-0.42484	0.146678	6.958442	0
BANTEN	4.083972	1.297314	3.236045	1.471882	0
DI YOGYAKARTA	9.232453	6.949808	1.729822	0.906304	0
JAWA TIMUR	5.315918	0.42394	0.124546	7.279542	0

DKI JAKARTA	13.31423	5.317642	6.433658	2.208841	0
JAWA BARAT	5.857803	0.100862	3.062413	8.662795	0
KEP. RIAU	5.008346	2.719053	3.911385	-0.52375	0
BALI	6.919992	2.672245	-0.86828	0.236623	0
SULAWESI UTARA	4.51368	1.651908	2.835292	0.278392	0
SULAWESI SELATAN	3.221002	1.552366	1.290518	0.69363	0
KALIMANTAN TIMUR	5.29468	3.74811	3.808026	0.021476	0
NUSA TENGGARA TIMUR	-6.25373	-2.14705	-4.34242	-1.59239	1
PAPUA TENGAH	-14.7	-12.3906	-7.5308	-2.24828	1
PAPUA PEGUNUNGAN	-18.9431	-13.9335	-13.3326	-2.50191	1
PAPUA SELATAN	-7.55825	-3.25019	-2.23329	-2.52001	1

PAPUA BARAT	-7.78055	-1.99567	-1.93972	-2.58391	1
KEP. BANGKA BELITUNG	1.986621	-0.87756	0.966315	-0.57374	2
SUMATERA UTARA	1.391751	1.323807	1.474167	1.066348	2
SUMATERA BARAT	2.008806	2.896954	1.949791	-0.1183	2
RIAU	2.515512	1.327251	1.540812	0.117141	2
JAMBI	1.651112	0.494841	0.932437	-0.18881	2
KALIMANTAN TENGAH	0.721331	-0.37129	0.387024	-0.78589	2
KALIMANTAN BARAT	-0.8618	-1.42521	-0.39417	-0.0482	2
NUSA TENGGARA BARAT	-2.16156	0.317404	-3.04284	-1.17704	2
KALIMANTAN UTARA	0.208341	0.38526	0.121938	-0.82258	2

KALIMANTAN SELATAN	1.824229	-0.02029	0.464785	-0.47673	2
SULAWESI TENGAH	-1.30587	-0.01541	-1.38657	-1.08616	2
SULAWESI TENGGARA	-0.11949	1.258674	-0.72773	-1.00801	2
LAMPUNG	0.375813	-0.952	-0.41782	0.686215	2
BENGKULU	-0.11285	1.131126	-0.8065	-1.11156	2
SUMATERA SELATAN	1.154728	-0.81085	-0.53966	0.592406	2
ACEH	0.779482	2.946743	2.472346	-0.32357	2
MALUKU UTARA	-2.85923	0.453202	0.105095	-1.83827	2
MALUKU	-4.29805	1.253578	1.74772	-2.14792	2
SULAWESI BARAT	-4.53705	-1.93726	-2.63238	-2.22719	2
GORONTALO	-3.05611	-1.0935	-1.58196	-1.5163	2
PAPUA	-2.61967	1.172533	2.003044	-1.68995	2

PAPUA BARAT DAYA	-5.53917	0.250517	1.032922	-2.06955	2
---------------------	----------	----------	----------	----------	---

Kontribusi Anggota Kelompok :

No	Nama	NIM	Kontribusi
1	Ibrahim Frosly Alesandro	23031554021	Latar belakang, PCA, Regresi, Forecasting, Output
2	Muhammad Faiz Munif Billah	23031554028	Latar belakang, Rumusan masalah, EDA, Clustering, Factor Analysis, Output
3	Gesang Nur Zamroji	23031554145	Integrasi dataset, Diagram alur, Clustering,, Output