

Prediksi Tingkat Pengangguran di Indonesia Berdasarkan Faktor Ekonomi

Nama Mentor: Maura Yufi Septania Putri



MariBelajar

Disusun oleh: Kelompok 14 Bangsawan

Adinda Aditya - 10571797 / Universitas Persada Indonesia YAI

Aji Pangestu - 12000413 / Universitas Nusantara PGRI Kediri

Evan Fulka Bima Maheswara -10556039 / Universitas Pembangunan Nasional Veteran
Jakarta

Sufadlan Nugraha-11246637 / Universitas Pendidikan Indonesia

Data Analyst & AI
Program Studi Independen Angkatan 7
PT. MariBelajar Indonesia Cerdas
Tahun 2024

Daftar Isi

DAFTAR ISI	2
LEMBAR PENGESAHAN.....	3
A. LATAR BELAKANG	4
B. PERUMUSAN MASALAH.....	5
C. SPESIFIKASI KEBUTUHAN.....	5
D. RANCANGAN SOLUSI.....	7
E. HASIL DAN PEMBAHASAN	10
F. KESIMPULAN	25
G. LAMPIRAN.....	26

Lembar Pengesahan

Prediksi Tingkat Pengangguran di Indonesia Berdasarkan Faktor Ekonomi

Disusun oleh:

Adinda Aditya - 10571797 / Universitas Persada Indonesia YAI

Aji Pangestu - 12000413 / Universitas Nusantara PGRI Kediri

Evan Fulka Bima Maheswara -10556039 / Universitas Pembangunan Nasional
Veteran Jakarta

Sufadlan Nugraha-11246637 / Universitas Pendidikan Indonesia

Disetujui oleh:

Maura Yufi Septania Putri

A. Latar Belakang

Tingkat pengangguran adalah salah satu indikator utama dalam menilai kesejahteraan sosial dan kondisi ekonomi suatu negara. Di Indonesia, meskipun ada tren penurunan, masalah pengangguran tetap menjadi tantangan besar. Menurut laporan World Economic Outlook yang diterbitkan oleh International Monetary Fund (IMF) pada April 2024, Indonesia mencatatkan tingkat pengangguran tertinggi di antara negara-negara ASEAN, dengan angka mencapai 5,2%. Posisi ini menunjukkan bahwa Indonesia masih menghadapi kesulitan dalam menyerap tenaga kerja, meskipun ada perbaikan kecil dibandingkan tahun sebelumnya yang tercatat 5,3%. Sementara negara-negara tetangga seperti Thailand dan Singapura berhasil menurunkan tingkat pengangguran mereka secara signifikan, Indonesia tetap berada di posisi teratas.

Badan Pusat Statistik (BPS) pada Februari 2024 melaporkan bahwa Tingkat Pengangguran Terbuka (TPT) Indonesia menurun menjadi 4,82%, yang berarti sekitar 5 dari setiap 100 orang dalam angkatan kerja mengalami pengangguran. Meskipun ada penurunan, pengangguran masih menjadi masalah yang harus diatasi, dengan kelompok lulusan Sekolah Menengah Kejuruan (SMK) menjadi penyumbang terbesar dalam angka pengangguran di Indonesia. Bahkan, berdasarkan data BPS, terdapat peningkatan angka pengangguran pada jenjang pendidikan yang lebih tinggi, seperti lulusan D4, S1, S2, dan S3.

Faktor-faktor ekonomi seperti pertumbuhan ekonomi, inflasi, dan kebijakan pemerintah sering dianggap sebagai penyebab utama tingginya angka pengangguran ini. Pertumbuhan ekonomi yang tidak merata, serta ketidakseimbangan sektor pekerjaan, berkontribusi terhadap kesulitan masyarakat dalam memperoleh pekerjaan yang layak. Selain itu, kurangnya keterampilan yang sesuai dengan kebutuhan pasar kerja juga mempengaruhi tingkat pengangguran, terutama di kalangan lulusan sekolah dan perguruan tinggi.

Dengan tingkat pengangguran yang tinggi ini, pemangku kepentingan di Indonesia perlu memiliki analisis yang lebih mendalam tentang faktor-faktor apa saja yang paling memengaruhi tingkat pengangguran. Oleh karena itu, proyek ini bertujuan untuk mengembangkan model prediksi tingkat pengangguran di Indonesia dengan menggunakan data ekonomi yang relevan, termasuk faktor pendidikan, usia, jenis kelamin, pertumbuhan ekonomi, dan sektor pekerjaan. Melalui pendekatan machine learning, model ini diharapkan dapat memberikan wawasan yang lebih akurat dan terukur tentang faktor-

faktor yang memengaruhi tingkat pengangguran serta memberikan prediksi yang berguna untuk pengambilan keputusan kebijakan.

Model prediksi ini akan membantu para pemangku kepentingan, termasuk pemerintah, sektor swasta, dan lembaga swadaya masyarakat, untuk merumuskan kebijakan yang lebih tepat sasaran dalam mengurangi pengangguran, serta menciptakan program-program yang relevan untuk mengatasi ketidakseimbangan pasar kerja. Selain itu, dengan adanya dashboard interaktif yang menampilkan prediksi pengangguran per provinsi, diharapkan analisis ini dapat diakses dengan mudah oleh publik dan digunakan untuk merencanakan kebijakan berbasis data yang lebih efektif.

B. Perumusan Masalah

- Bagaimana cara mengembangkan model prediksi tingkat pengangguran di Indonesia menggunakan teknik machine learning, berdasarkan faktor-faktor ekonomi dan demografi yang relevan?
- Faktor-faktor ekonomi dan demografi apa saja yang paling signifikan mempengaruhi tingkat pengangguran di Indonesia, dan bagaimana model machine learning dapat mengidentifikasinya?
- Bagaimana merancang dan mengembangkan dashboard interaktif yang menyajikan prediksi tingkat pengangguran secara visual, serta memungkinkan pengguna untuk memahami tren dan distribusi pengangguran berdasarkan faktor-faktor demografi?

C. Spesifikasi Kebutuhan

- **Kebutuhan Data**
Untuk pengembangan model prediksi tingkat pengangguran di Indonesia, data yang digunakan harus mencakup berbagai variabel yang relevan. Dataset utama yang dibutuhkan adalah data tingkat pengangguran per provinsi di Indonesia yang mencakup informasi terkait pendidikan, usia, jenis kelamin, serta sektor pekerjaan. Data ini mencakup periode waktu yang cukup panjang agar model machine learning yang dikembangkan dapat mengidentifikasi tren dan pola jangka panjang. Selain itu, data ekonomi yang memengaruhi tingkat pengangguran, seperti pertumbuhan ekonomi, inflasi, dan PDRB, juga harus dimasukkan dalam dataset. Sumber data utama akan berasal dari Badan Pusat Statistik (BPS) dan lembaga pemerintah lainnya yang menyediakan data terkait ketenagakerjaan dan ekonomi. Untuk memastikan keandalan dan akurasi model, kualitas data yang digunakan harus sangat baik, dengan data yang bersih, lengkap, dan konsisten di seluruh periode waktu yang dianalisis.

- **Kebutuhan Teknologi**

Dalam pengembangan proyek ini, Python akan menjadi bahasa pemrograman utama untuk melakukan analisis data dan pengembangan model machine learning. Python dipilih karena memiliki banyak library yang mendukung pengolahan data, statistik, dan pembuatan model prediksi, seperti pandas dan scikit-learn. Library-library ini akan digunakan untuk mengembangkan model prediksi tingkat pengangguran yang akurat dengan menerapkan teknik machine learning yang sesuai. Untuk visualisasi data dan pembuatan dashboard interaktif, Power BI akan digunakan. Dengan menggunakan Power BI ini, akan memungkinkan kita untuk membuat grafik dan visualisasi yang informatif serta mudah dipahami oleh pengguna. Selain itu untuk penyimpanan data, cloud storage penyimpanan OneDrive di Microsoft Teams akan digunakan untuk menyimpan dataset besar, file dokumen, ataupun laporan yang diperlukan untuk memudahkan akses dan pengolahan data secara efisien.

- **Kebutuhan Fungsional**

Sistem yang dikembangkan harus mampu memprediksi tingkat pengangguran per provinsi di Indonesia dengan akurasi yang tinggi, berdasarkan faktor-faktor ekonomi dan demografi yang relevan. Model machine learning yang dibangun akan menggunakan data historis untuk mempelajari pola-pola yang ada dan menghasilkan prediksi yang dapat diandalkan. Selain itu, sistem ini harus memiliki kemampuan untuk mengevaluasi kinerja model dengan menggunakan metrik evaluasi yang relevan, seperti RMSE, r2 score, ataupun yang lainnya untuk memastikan kualitas prediksi yang dihasilkan. Salah satu elemen penting dari proyek ini adalah pembuatan dashboard interaktif yang menyajikan prediksi tingkat pengangguran serta tren dari waktu ke waktu. Dashboard ini akan memberikan visualisasi yang jelas dan mudah dipahami oleh pemangku kepentingan, dengan grafik tren, peta distribusi pengangguran, dan analisis faktor-faktor yang memengaruhi tingkat pengangguran. Sistem harus memastikan bahwa dashboard dapat diakses dengan mudah oleh pengguna yang memiliki latar belakang non-teknis dan menyediakan fitur interaktif yang memungkinkan pengguna untuk memilih provinsi dan melihat informasi terkait dengan prediksi pengangguran.

- **Kebutuhan Non-Fungsional**

Selain kebutuhan fungsional yang jelas, sistem juga harus memenuhi berbagai kebutuhan non-fungsional yang penting untuk keberhasilan proyek. Pertama, kinerja sistem sangat penting, sehingga proses analisis

data dan prediksi model harus dapat dilakukan dalam waktu yang wajar, memastikan pengguna tidak mengalami kendala saat mengakses informasi atau melakukan analisis. Scalability juga merupakan faktor yang perlu diperhatikan, mengingat adanya kemungkinan penambahan data baru di masa depan. Sistem harus mampu menampung data yang lebih besar dan memperbarui model prediksi seiring waktu tanpa mengorbankan kinerja. Terakhir, user experience sangat penting dalam pembangunan dashboard interaktif. Antarmuka dashboard harus didesain agar mudah digunakan, intuitif, dan menarik bagi pengguna, memastikan pemangku kepentingan dapat mengakses dan memahami informasi dengan mudah meskipun mereka tidak memiliki latar belakang teknis.

D. Rancangan Solusi

Untuk mengatasi masalah tingkat pengangguran yang tinggi di Indonesia, yang telah dijelaskan sebelumnya, proyek ini akan mengembangkan sebuah sistem prediksi berbasis machine learning dan dashboard interaktif. Rancangan solusi ini bertujuan untuk memanfaatkan data ekonomi dan demografi yang relevan untuk memprediksi tingkat pengangguran di tingkat provinsi serta untuk memvisualisasikan hasil prediksi secara efektif, sehingga dapat memberikan wawasan yang berguna bagi pemangku kepentingan dalam pengambilan kebijakan. Berikut tahapan dari Rancangan Solusi yang telah dilakukan:

1. Pengumpulan dan Pembersihan Data

Langkah pertama dalam rancangan solusi ini adalah pengumpulan data yang relevan dan berkualitas. Dataset yang akan digunakan mencakup data tingkat pengangguran per provinsi di Indonesia, dengan variabel-variabel seperti tingkat pendidikan, usia, jenis kelamin, sektor pekerjaan, serta data ekonomi yang mempengaruhi tingkat pengangguran, seperti pertumbuhan ekonomi, inflasi, dan PDRB. Data akan dikumpulkan dari sumber resmi, seperti Badan Pusat Statistik (BPS), yang menyediakan dataset terkait ketenagakerjaan dan indikator ekonomi.

Setelah data terkumpul, langkah berikutnya adalah proses pembersihan dan pra-pemrosesan data. Proses ini meliputi penghilangan data yang hilang, perbaikan kesalahan dalam data, serta konversi variabel ke format yang sesuai untuk analisis. Tujuan utama dari tahap ini adalah untuk memastikan bahwa data yang digunakan berkualitas tinggi, lengkap, dan konsisten di seluruh periode waktu yang

dianalisis, agar model yang dikembangkan dapat memberikan prediksi yang akurat.

2. Pengembangan Model Prediksi

Setelah data siap, langkah selanjutnya adalah mengembangkan model prediksi tingkat pengangguran. Proyek ini akan menggunakan teknik machine learning untuk membangun model yang mampu mengidentifikasi pola dan hubungan antara faktor ekonomi dan demografi dengan tingkat pengangguran. Model ini akan dilatih menggunakan data historis untuk memahami tren jangka panjang dan faktor-faktor utama yang memengaruhi tingkat pengangguran di Indonesia.

Untuk pengembangan model prediksi, Python akan digunakan sebagai bahasa pemrograman utama, dengan memanfaatkan berbagai library machine learning yang ada, seperti scikit-learn dan pandas. Beberapa algoritma machine learning yang dipertimbangkan untuk digunakan antara lain regresi linier, random forest, dan support vector machine (SVM). Setelah model dikembangkan, model akan dievaluasi menggunakan metrik evaluasi yang relevan seperti RMSE (Root Mean Squared Error) atau R^2 score untuk memastikan akurasi prediksi yang tinggi.

3. Evaluasi dan Tuning Model

Setelah model awal dibangun, langkah selanjutnya adalah evaluasi dan tuning model. Evaluasi model bertujuan untuk mengukur akurasi dan ketepatan prediksi yang dihasilkan. Jika model tidak memberikan hasil yang memadai, proses tuning akan dilakukan dengan mengubah hyperparameter atau mencoba algoritma machine learning lainnya.

Proses tuning juga akan mencakup pemilihan fitur (feature selection) yang paling relevan untuk meningkatkan performa model. Misalnya, analisis korelasi antar variabel ekonomi dan demografi dapat membantu menentukan faktor-faktor yang paling berpengaruh terhadap tingkat pengangguran, yang akan digunakan sebagai input untuk model.

4. Pembuatan Dashboard Interaktif

Setelah model prediksi selesai dan memiliki performa yang memadai, solusi selanjutnya adalah membuat dashboard interaktif untuk

menampilkan hasil prediksi tingkat pengangguran serta analisis faktor-faktor yang mempengaruhi pengangguran. Power BI akan digunakan untuk mengembangkan dashboard ini, karena platform ini memungkinkan pembuatan visualisasi data yang informatif dan mudah dipahami oleh berbagai pemangku kepentingan.

Dashboard akan dirancang untuk menampilkan beberapa elemen visual, seperti grafik tren pengangguran per provinsi dari waktu ke waktu, peta distribusi pengangguran per provinsi, serta analisis distribusi pengangguran berdasarkan faktor demografi seperti usia, pendidikan, dan sektor pekerjaan. Salah satu fitur utama dari dashboard ini adalah kemampuannya untuk menampilkan prediksi tingkat pengangguran secara dinamis dan memungkinkan pengguna untuk memfilter data berdasarkan provinsi, waktu, dan faktor-faktor lainnya.

Interaksi dengan dashboard akan dibuat semudah mungkin, dengan desain yang intuitif, sehingga pengguna yang tidak memiliki latar belakang teknis dapat dengan mudah memahami informasi yang disajikan. Selain itu, dashboard juga akan mendukung kemampuan untuk memperbarui data dan model prediksi secara berkala, seiring dengan datangnya data baru.

5. Penyajian dan Pengujian Sistem

Setelah sistem selesai dikembangkan, langkah berikutnya adalah pengujian sistem secara menyeluruh untuk memastikan bahwa seluruh fitur bekerja dengan baik. Uji coba model prediksi akan dilakukan dengan menggunakan data yang tidak digunakan dalam pelatihan model (data uji), untuk menilai sejauh mana model dapat memprediksi tingkat pengangguran yang sebenarnya.

Selain itu, dashboard interaktif akan diuji oleh pengguna akhir untuk memastikan bahwa antarmuka yang disediakan mudah digunakan dan memberikan wawasan yang relevan. Umpan balik dari pemangku kepentingan akan digunakan untuk menyempurnakan dashboard dan meningkatkan fungsionalitas sistem.

6. Penyusunan Laporan dan Presentasi

Pada tahap akhir, hasil dari proyek ini akan disusun dalam bentuk laporan yang menyeluruh. Laporan ini akan mencakup penjelasan tentang data yang digunakan, metodologi yang diterapkan dalam pengembangan model, evaluasi model, serta analisis faktor-faktor yang mempengaruhi tingkat pengangguran di Indonesia. Laporan ini juga

akan menyertakan analisis visual yang ditampilkan dalam dashboard interaktif.

Selain laporan tertulis, hasil proyek ini juga akan dipresentasikan kepada para pemangku kepentingan untuk memberikan wawasan yang lebih mendalam mengenai dinamika pengangguran di Indonesia dan rekomendasi kebijakan berdasarkan hasil prediksi.

E. Hasil dan Pembahasan

Pada bagian Hasil dan Pembahasan ini, akan dijelaskan secara mendalam mengenai solusi yang diusulkan dalam capstone project ini, yaitu model prediksi tingkat pengangguran di Indonesia berbasis machine learning dan dashboard interaktif untuk visualisasi hasil prediksi. Solusi ini dikembangkan untuk membantu para pemangku kepentingan, seperti pemerintah, lembaga pendidikan, dan masyarakat umum, dalam memahami dan mengantisipasi tren pengangguran di berbagai provinsi di Indonesia. Setiap fitur dari solusi ini akan dijabarkan secara rinci, bagaimana fitur tersebut bekerja, serta bagaimana solusi ini dapat efektif dalam menyelesaikan masalah yang ada. Selain itu, solusi ini juga akan dibandingkan dengan solusi yang telah ada sebelumnya, serta dibahas keterbatasan yang mungkin dihadapi dalam implementasi dan penggunaan sistem ini.

1. Hasil solusi yang dikembangkan
 - a. Model Machine Learning (Python)

```

import pandas as pd

df = pd.read_csv('DataUtama.csv')
df

[1]
...

```

	Tahun	Provinsi	Jumlah Penduduk (ribu)	Laju Pertumbuhan Penduduk per Tahun	Persentase Penduduk	Kepadatan Penduduk per km persegi (km2)	Persentase Laki-laki	Persentase Perempuan	Persentase Tenaga Kerja Formal	Persentase Tenaga Kerja Informal	APS 7-12(SD)	APS 13-15(SMP)	APS 16-18(SMA)	APS 19-24(D3/D4/S1)	Pertumbuhan PDB Atas Dasar Harga Konstan 2010	Laju PDB per Kapita Atas Dasar Harga Berlaku (ribu rupiah)	Pengangguran Terbuka
0	2016	Aceh	5096.2	2.01	1.97	88	49.949950	50.050050	42.83	57.17	99.82	97.89	81.82	33.94	3.29	26852.0	
1	2016	Bali	4200.1	1.21	1.62	727	50.347567	49.652433	45.68	54.32	99.35	97.55	81.98	25.36	6.33	46211.0	
2	2016	Banten	12203.1	2.23	4.72	1263	50.980392	49.019608	61.51	38.49	99.43	95.59	67.00	20.74	5.28	42440.0	
3	2016	Bengkulu	1904.8	1.69	0.74	96	50.980392	49.019608	34.08	65.92	99.70	96.96	78.37	28.93	5.28	29076.0	
4	2016	DI Yogyakarta	3720.9	1.18	1.44	1188	49.443883	50.556117	45.09	54.91	99.84	99.62	87.20	49.95	5.05	29553.0	
...
275	2023	Sulawesi Tenggara	2749.0	1.69	0.99	76	50.666009	49.333991	38.16	61.84	99.20	95.00	74.60	31.73	5.35	64088.0	
276	2023	Sulawesi Utara	2681.5	0.82	0.96	185	51.076321	48.923679	41.86	58.14	99.34	95.00	74.55	23.57	5.48	64131.0	
277	2023	Sumatera Barat	5757.2	1.45	2.07	137	50.372208	49.627792	35.81	64.19	99.76	96.79	84.33	38.08	4.62	54327.0	
278	2023	Sumatera Selatan	8743.5	1.17	3.14	101	50.932287	49.067713	37.03	62.97	99.41	95.27	71.71	19.79	5.08	71950.0	
279	2023	Sumatera Utara	15386.6	1.43	5.52	212	50.174390	49.825610	41.48	58.52	99.51	96.76	79.25	28.61	5.01	68306.0	

```

df.info()

[1]
...
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 280 entries, 0 to 279
Data columns (total 17 columns):
 #   Column                                                                 Non-Null Count  Dtype
---  -
 0   Tahun                                                                 280 non-null   int64
 1   Provinsi                                                            280 non-null   object
 2   Jumlah Penduduk (ribu)                                             280 non-null   float64
 3   Laju Pertumbuhan Penduduk per Tahun                             279 non-null   float64
 4   Persentase Penduduk                                               280 non-null   float64
 5   Kepadatan Penduduk per km persegi (km2)                         280 non-null   int64
 6   Persentase Laki-laki                                              280 non-null   float64
 7   Persentase Perempuan                                              280 non-null   float64
 8   Persentase Tenaga Kerja Formal                                    280 non-null   float64
 9   Persentase Tenaga Kerja Informal                                 280 non-null   float64
10   APS 7-12(SD)                                                       280 non-null   float64
11   APS 13-15(SMP)                                                     280 non-null   float64
12   APS 16-18(SMA)                                                     280 non-null   float64
13   APS 19-24(D3/D4/S1)                                                280 non-null   float64
14   Laju Pertumbuhan PDB Atas Dasar Harga Konstan 2010              280 non-null   float64
15   PDB per Kapita Atas Dasar Harga Berlaku (ribu rupiah)            280 non-null   float64
16   Tingkat Pengangguran Terbuka                                     280 non-null   float64
dtypes: float64(14), int64(2), object(1)
memory usage: 37.3+ KB

df.isnull().sum()

[3]
...
Tahun                                0
Provinsi                             0
Jumlah Penduduk (ribu)                0
Laju Pertumbuhan Penduduk per Tahun    1
Persentase Penduduk                  0
Kepadatan Penduduk per km persegi (km2) 0
Persentase Laki-laki                 0
Persentase Perempuan                 0
Persentase Tenaga Kerja Formal        0
Persentase Tenaga Kerja Informal      0
APS 7-12(SD)                         0
APS 13-15(SMP)                       0
APS 16-18(SMA)                       0
APS 19-24(D3/D4/S1)                  0
Laju Pertumbuhan PDB Atas Dasar Harga Konstan 2010 0
PDB per Kapita Atas Dasar Harga Berlaku (ribu rupiah) 0
Tingkat Pengangguran Terbuka          0
dtype: int64

df['Laju Pertumbuhan Penduduk per Tahun'].fillna(df['Laju Pertumbuhan Penduduk per Tahun'].mean(), inplace=True)
df.isnull().sum()

[4]
...
Tahun                                0
Provinsi                             0
Jumlah Penduduk (ribu)                0
Laju Pertumbuhan Penduduk per Tahun    0
Persentase Penduduk                  0
Kepadatan Penduduk per km persegi (km2) 0
Persentase Laki-laki                 0
Persentase Perempuan                 0
Persentase Tenaga Kerja Formal        0
Persentase Tenaga Kerja Informal      0
APS 7-12(SD)                         0
APS 13-15(SMP)                       0
APS 16-18(SMA)                       0
APS 19-24(D3/D4/S1)                  0
Laju Pertumbuhan PDB Atas Dasar Harga Konstan 2010 0
PDB per Kapita Atas Dasar Harga Berlaku (ribu rupiah) 0
Tingkat Pengangguran Terbuka          0
dtype: int64

df.duplicated().sum()

[5]
...
0

```

Pertama-tama, kita akan memasukkan data yang telah siap pakai ke dalam python nya. Kemudian, kita akan lakukan pre-processing pada data tersebut, karena data tersebut hanya

memiliki 1 nilai kosong dan itu di kolom laju pertumbuhan penduduk per tahun. Cara menangani nya, kami mengganti nilai tersebut menggunakan nilai rata-rata.

```
df.describe()
```

	Tahun	Jumlah Penduduk (ribu)	Pertumbuhan Penduduk per Tahun	Persentase Penduduk	Kepadatan Penduduk per km persegi (km2)	Persentase Laki-laki	Persentase Perempuan	Persentase Tenaga Kerja Formal	Persentase Tenaga Kerja Informal	APS 7-12(SD)	APS 13-15(SMP)	APS 16-18(SMA)	APS 19-24(D3/D4/S1)	Laju Pertumbuhan PDB Atas Dasar Harga Konstan 2010	Kap Data Berla
count	280.00000	280.000000	280.000000	280.000000	280.000000	280.000000	280.000000	280.000000	280.000000	280.000000	280.000000	280.000000	280.000000	280.000000	280.000000
mean	2019.50000	15364.586071	1.585914	5.714357	721.742857	50.793826	49.206174	40.995464	59.003536	98.797143	95.245607	74.572179	27.111929	4.372429	64704
std	2.29539	44922.229510	0.600378	16.701990	2631.404537	0.950724	0.950724	10.399441	10.399441	2.800748	3.420248	5.896248	6.438474	3.661074	50195
min	2016.00000	666.300000	0.380000	0.260000	9.000000	48.533196	46.685341	15.570000	28.450000	81.110000	78.860000	62.070000	13.810000	-15.740000	16094
25%	2017.50000	2173.000000	1.210000	0.807500	49.750000	50.193000	48.804295	35.605000	54.315000	99.077500	93.992500	69.932500	23.335000	3.557500	37940
50%	2019.50000	4233.200000	1.500000	1.580000	104.000000	50.738916	49.261084	39.655000	60.345000	99.430000	95.790000	73.765000	25.775000	5.025000	49900
75%	2021.50000	8564.775000	1.802500	3.162500	260.000000	51.195705	49.804295	45.685000	64.395000	99.610000	97.440000	79.040000	30.550000	5.560000	64095
max	2023.00000	278696.200000	4.130000	100.000000	16158.000000	53.314659	51.466804	71.550000	84.430000	99.900000	99.720000	91.170000	51.850000	22.940000	322615

```
df['Provinsi'].unique()
```

```
array(['Aceh', 'Bali', 'Banten', 'Bengkulu', 'DI Yogyakarta', 'DKI Jakarta', 'Gorontalo', 'Indonesia', 'Jambi', 'Jawa Barat', 'Jawa Tengah', 'Jawa Timur', 'Kalimantan Barat', 'Kalimantan Selatan', 'Kalimantan Tengah', 'Kalimantan Timur', 'Kalimantan Utara', 'Kepulauan Bangka Belitung', 'Kepulauan Riau', 'Lampung', 'Maluku', 'Maluku Utara', 'Nusa Tenggara Barat', 'Nusa Tenggara Timur', 'Papua', 'Papua Barat', 'Riau', 'Sulawesi Barat', 'Sulawesi Selatan', 'Sulawesi Tengah', 'Sulawesi Tenggara', 'Sulawesi Utara', 'Sumatera Barat', 'Sumatera Selatan', 'Sumatera Utara'], dtype=object)
```

```
from sklearn.preprocessing import LabelEncoder
# Membuat objek LabelEncoder
label_encoder = LabelEncoder()
# Mengonversi kolom 'Provinsi' menjadi numerik
df['Provinsi_encoded'] = label_encoder.fit_transform(df['Provinsi'])
df
```

	Tahun	Provinsi	Jumlah Penduduk (ribu)	Pertumbuhan Penduduk per Tahun	Persentase Penduduk	Kepadatan Penduduk per km persegi (km2)	Persentase Laki-laki	Persentase Perempuan	Persentase Tenaga Kerja Formal	Persentase Tenaga Kerja Informal	APS 7-12(SD)	APS 13-15(SMP)	APS 16-18(SMA)	APS 19-24(D3/D4/S1)	Laju Pertumbuhan PDB Atas Dasar Harga Konstan 2010	PDB per Kapita Atas Dasar Harga Berla (ribu rupiah)	Peng
0	2016	Aceh	5096.2	2.01	1.97	88	49.949950	50.050050	42.83	57.17	99.82	97.89	81.82	33.94	3.29	26852.0	
1	2016	Bali	4200.1	1.21	1.62	727	50.347567	49.652433	45.68	54.32	99.35	97.55	81.98	25.36	6.33	46211.0	
2	2016	Banten	12203.1	2.23	4.72	1263	50.980392	49.019608	61.51	38.49	99.43	95.59	67.00	20.74	5.28	42440.0	
3	2016	Bengkulu	1904.8	1.69	0.74	96	50.980392	49.019608	34.08	65.92	99.70	96.96	78.37	28.93	5.28	29076.0	
4	2016	DI Yogyakarta	3720.9	1.18	1.44	1188	49.443383	50.556617	45.09	54.91	99.84	99.62	87.20	49.95	5.05	29553.0	
...
275	2023	Sulawesi Tenggara	2749.0	1.69	0.99	76	50.666009	49.333991	38.16	61.84	99.20	95.00	74.60	31.73	5.35	64088.0	
276	2023	Sulawesi Utara	2681.5	0.82	0.96	185	51.076321	48.923679	41.86	58.14	99.34	95.00	74.55	23.57	5.48	64131.0	
277	2023	Sumatera Barat	5757.2	1.45	2.07	137	50.372208	49.627792	35.81	64.19	99.76	96.79	84.33	38.08	4.62	54327.0	
278	2023	Sumatera Selatan	8743.5	1.17	3.14	101	50.932287	49.067713	37.03	62.97	99.41	95.27	71.71	19.79	5.08	71950.0	

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 280 entries, 0 to 279
Data columns (total 18 columns):
#   Column                                Non-Null Count  Dtype
---  --
0   Tahun                                280 non-null    int64
1   Provinsi                             280 non-null    object
2   Jumlah Penduduk (ribu)               280 non-null    float64
3   Laju Pertumbuhan Penduduk per Tahun  280 non-null    float64
4   Persentase Penduduk                 280 non-null    float64
5   Kepadatan Penduduk per km persegi (km2)  280 non-null    int64
6   Persentase Laki-laki                 280 non-null    float64
7   Persentase Perempuan                 280 non-null    float64
8   Persentase Tenaga Kerja Formal       280 non-null    float64
9   Persentase Tenaga Kerja Informal     280 non-null    float64
10  APS 7-12(SD)                         280 non-null    float64
11  APS 13-15(SMP)                       280 non-null    float64
12  APS 16-18(SMA)                       280 non-null    float64
13  APS 19-24(D3/D4/S1)                  280 non-null    float64
14  Laju Pertumbuhan PDB Atas Dasar Harga Konstan 2010  280 non-null    float64
15  PDB per Kapita Atas Dasar Harga Berla (ribu rupiah)  280 non-null    float64
16  Tingkat Pengangguran Terbuka         280 non-null    float64
17  Provinsi_encoded                      280 non-null    int32
dtypes: float64(14), int32(1), int64(2), object(1)
memory usage: 38.4+ KB
```

Selanjutnya, kami melakukan analisa statistik untuk setiap variabel numeriknya untuk menganalisis secara general terlebih dahulu. Dan setelah itu, kami memutuskan untuk melakukan encoding pada kolom Provinsi agar bisa digunakan pada model yang akan dibuat serta dilakukan analisis heatmap juga.

```

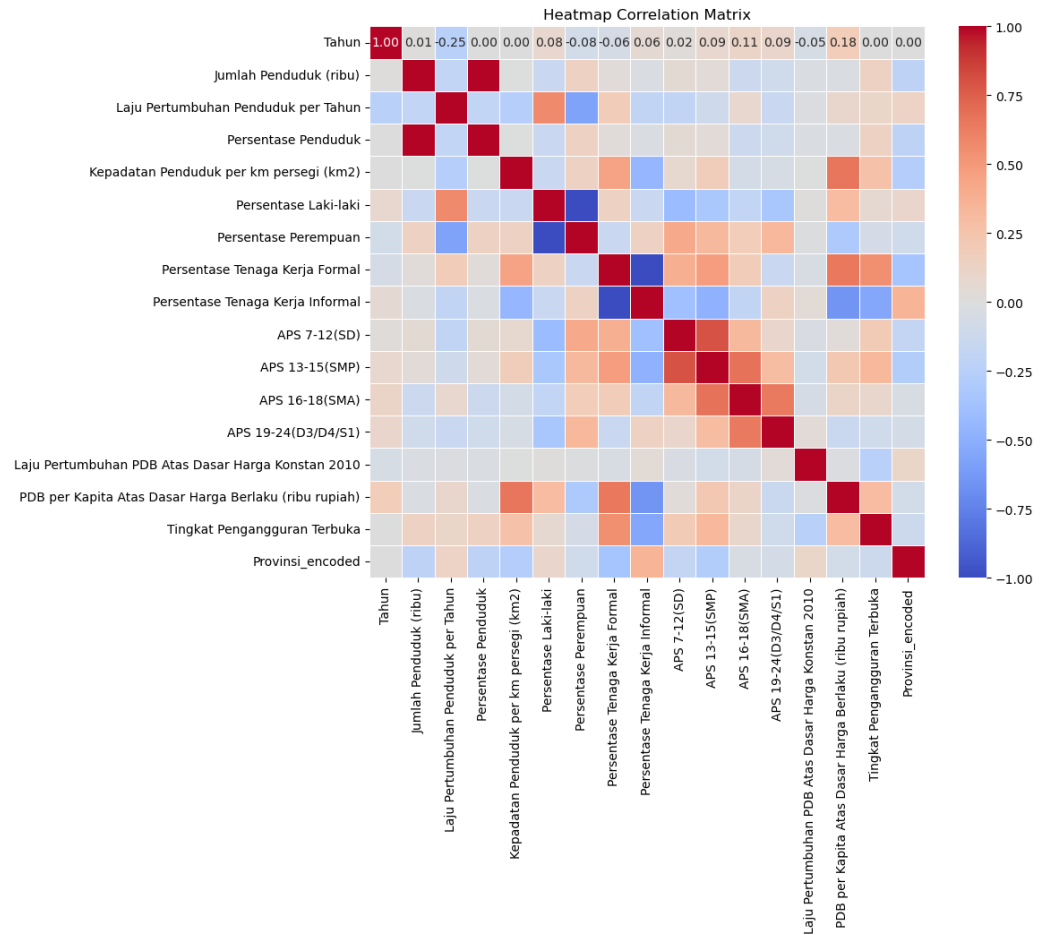
import seaborn as sns
import matplotlib.pyplot as plt

# Menghitung correlation matrix
correlation_matrix = df.select_dtypes(include=['float64', 'int64', 'int32']).corr()

# Membuat heatmap dari correlation matrix
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", fmt=".2f", linewidths=0.5)

# Menampilkan plot
plt.title("Heatmap Correlation Matrix")
plt.show()

```



Berdasarkan matriks korelasi yang dihasilkan, terlihat beberapa hubungan menarik antar variabel dalam dataset. Tingkat Pengangguran Terbuka menunjukkan korelasi negatif yang cukup kuat dengan PDB per Kapita, yang mengindikasikan bahwa provinsi yang lebih tinggi cenderung memiliki tingkat pengangguran yang lebih rendah. Hal ini mungkin mencerminkan hubungan antara kesejahteraan ekonomi dengan kesempatan kerja yang lebih banyak. Sebaliknya, Jumlah Penduduk dan Tingkat Pengangguran Terbuka tidak memiliki korelasi yang signifikan, menunjukkan bahwa ukuran populasi provinsi tidak secara langsung memengaruhi tingkat pengangguran terbuka. Persentase Tenaga Kerja Formal menunjukkan korelasi negatif yang moderat dengan tingkat pengangguran, meskipun hubungan ini tidak sekuat yang ditemukan dengan PDB per Kapita. Variabel seperti Kepadatan Penduduk dan Laju Pertumbuhan PDB memiliki korelasi sedang dengan beberapa variabel lain, yang mungkin mencerminkan dinamika ekonomi dan sosial di masing-masing provinsi. Secara keseluruhan, analisis korelasi ini memberikan petunjuk penting mengenai faktor-faktor yang perlu dipertimbangkan dalam pembuatan model untuk memprediksi tingkat pengangguran terbuka, dengan fokus pada variabel-variabel ekonomi seperti PDB dan tenaga kerja formal.

```

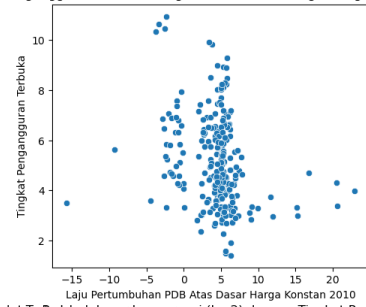
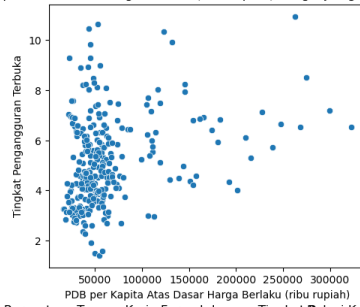
features_to_plot = [
    'PDB per Kapita Atas Dasar Harga Berlaku (ribu rupiah)',
    'Laju Pertumbuhan PDB Atas Dasar Harga Konstan 2010',
    'Persentase Tenaga Kerja Formal',
    'Kepadatan Penduduk per km persegi (km2)'
]

plt.figure(figsize=(12, 10))

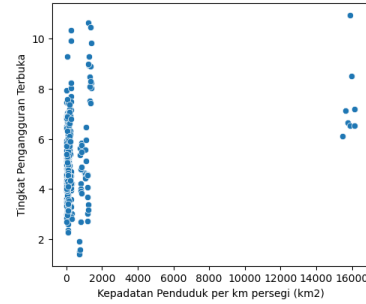
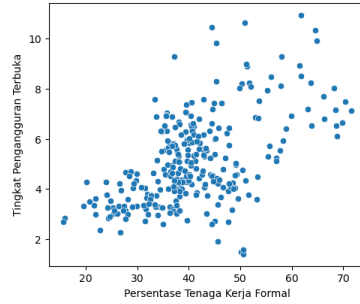
for i, feature in enumerate(features_to_plot, 1):
    plt.subplot(2, 2, i)
    sns.scatterplot(x=df[feature], y=df['Tingkat Pengangguran Terbuka'])
    plt.title(f'Relasi {feature} dengan Tingkat Pengangguran Terbuka')
    plt.xlabel(feature)
    plt.ylabel("Tingkat Pengangguran Terbuka")

```

Relasi PDB per Kapita Atas Dasar Harga Berlaku (ribu rupiah) dengan Tingkat Pengangguran Terbuka



Relasi Persentase Tenaga Kerja Formal dengan Tingkat Pengangguran Terbuka



Berdasarkan visualisasi yang Anda tampilkan, kita dapat menarik beberapa wawasan penting tentang hubungan antara Tingkat Pengangguran Terbuka dan berbagai variabel ekonomi dan sosial. Pertama, hubungan antara PDB per Kapita dan Tingkat Pengangguran Terbuka menunjukkan pola yang negatif: provinsi dengan PDB per kapita yang lebih tinggi cenderung memiliki tingkat pengangguran yang lebih rendah, memperkuat anggapan bahwa peningkatan kesejahteraan ekonomi dapat mengurangi pengangguran. Kedua, Laju Pertumbuhan PDB juga menunjukkan korelasi negatif dengan tingkat pengangguran, dimana provinsi dengan pertumbuhan ekonomi yang lebih pesat memiliki tingkat pengangguran yang lebih rendah. Ketiga, Persentase Tenaga Kerja Formal berbanding terbalik dengan tingkat pengangguran terbuka, meskipun hubungannya tidak sekuat yang ditemukan pada PDB per Kapita, ini menunjukkan bahwa lebih banyak tenaga kerja yang bekerja di sektor formal dapat berkontribusi pada pengurangan pengangguran. Terakhir, Kepadatan Penduduk memiliki korelasi yang lebih kompleks, dengan beberapa provinsi padat penduduknya justru memiliki tingkat pengangguran yang lebih tinggi, yang mungkin disebabkan oleh kurangnya lapangan pekerjaan yang sesuai dengan jumlah penduduk. Secara keseluruhan, visualisasi ini menunjukkan bahwa faktor ekonomi, terutama PDB dan tenaga kerja formal, memiliki pengaruh signifikan terhadap tingkat pengangguran terbuka.

Setelah itu, kami melakukan EDA menggunakan berbagai visualisasi untuk mencari hubungan antar variabel yang ada.

```

from sklearn.preprocessing import LabelEncoder
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.model_selection import train_test_split

X = df[['Tahun', 'Provinsi_encoded', 'Jumlah Penduduk (ribu)', 'Laju Pertumbuhan Penduduk per Tahun',
        'Persentase Penduduk', 'Kepadatan Penduduk per km persegi (km2)', 'Persentase Laki-laki',
        'Persentase Perempuan', 'Persentase Tenaga Kerja Formal', 'Persentase Tenaga Kerja Informal',
        'APS 7-12(SD)', 'APS 13-15(SMP)', 'APS 16-18(SMA)', 'APS 19-24(D3/D4/S1)',
        'Laju Pertumbuhan PDB Atas Dasar Harga Konstan 2010', 'PDB per Kapita Atas Dasar Harga Berlaku (ribu rupiah)']]
y = df['Tingkat Pengangguran Terbuka']

# Membagi data menjadi training dan testing set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Membangun model Random Forest Regressor
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)

# Memprediksi pada data uji
y_pred = rf_model.predict(X_test)

# Menghitung RMSE dan R2
rmse = mean_squared_error(y_test, y_pred, squared=False)
r2 = r2_score(y_test, y_pred)

rmse, r2

```

(0.7777267059238361, 0.7883455934654265)

```

from sklearn.model_selection import GridSearchCV

# Menyiapkan parameter grid untuk Random Forest Regressor
param_grid = {
    'n_estimators': [100, 200, 300],
    'max_depth': [10, 20, None],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4],
    'max_features': ['auto', 'sqrt', 'log2']
}

# Inisialisasi model Random Forest Regressor
rf_model = RandomForestRegressor(random_state=42)

# Melakukan GridSearchCV untuk mencari kombinasi parameter yang optimal
grid_search = GridSearchCV(estimator=rf_model, param_grid=param_grid, cv=5, n_jobs=-1, verbose=2, scoring='neg_mean_squared_error')

# Fit model dengan data training
grid_search.fit(X_train, y_train)

# Menampilkan hasil terbaik dari GridSearchCV
best_params = grid_search.best_params_
best_model = grid_search.best_estimator_

# Memprediksi pada data uji dengan model terbaik
y_pred_optimized = best_model.predict(X_test)

# Menghitung RMSE dan R2 untuk model yang sudah dioptimasi
rmse_optimized = mean_squared_error(y_test, y_pred_optimized, squared=False)
r2_optimized = r2_score(y_test, y_pred_optimized)

best_params, rmse_optimized, r2_optimized

```

Fitting 5 folds for each of 243 candidates, totalling 1215 fits
c:\Users\Evan\miniconda3\envs\myenv\lib\site-packages\sklearn\model_selection_validation.py:425: FitFailedWarning:
405 fits failed out of a total of 1215.

```

... ({'max_depth': 20,
      'max_features': 'sqrt',
      'min_samples_leaf': 1,
      'min_samples_split': 2,
      'n_estimators': 200},
     0.7463582807418664,
     0.8050747941578719)

```

Terakhir, kita masuk ke pembuatan model machine learning yang akan dibuat. Kami menggunakan Random Forest Regression sebagai model utama nya dalam mengembangkan machine learning ini, hal ini dipilih karena model ini dapat memberikan hasil yang lebih baik dan lebih stabil dibandingkan dengan Linear Regression untuk data yang lebih kompleks, seperti data ekonomi dan demografi yang digunakan dalam proyek ini. Linear regression memiliki keterbatasan dalam menangani hubungan non-linier antara variabel-variabel input, sementara Random Forest lebih fleksibel dan mampu menangani interaksi antar fitur secara efektif.

Hasil awal dari model Random Forest menunjukkan performa yang cukup baik, namun tidak optimal. Oleh karena itu, dilakukan optimisasi lebih lanjut menggunakan GridSearchCV.

GridSearchCV bekerja dengan mencari kombinasi parameter terbaik untuk meningkatkan akurasi model. Beberapa parameter yang dioptimalkan antara lain `max_depth`, `n_estimators`, dan `min_samples_split`.

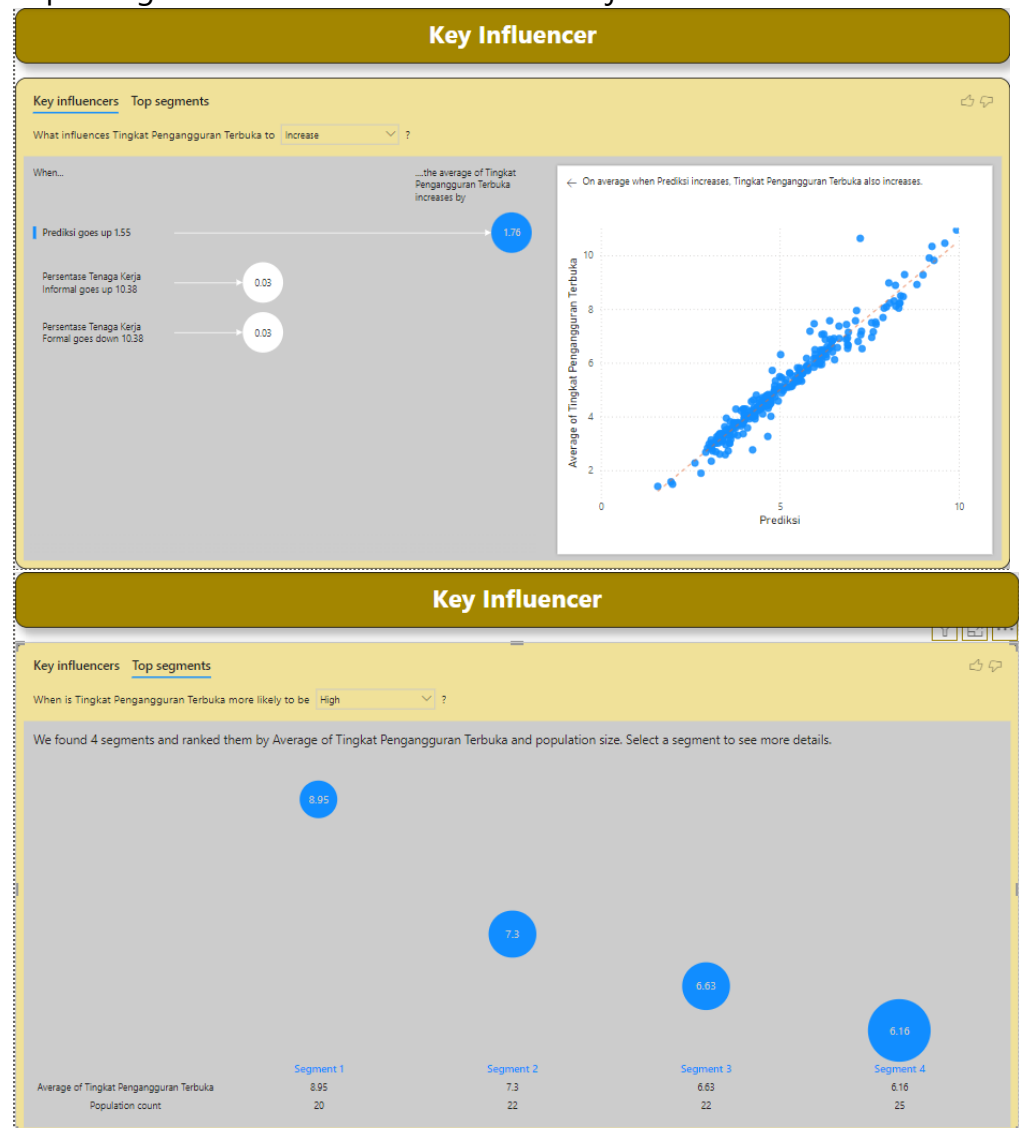
Setelah optimisasi dilakukan, model menunjukkan perbaikan yang signifikan pada RMSE (Root Mean Squared Error) dan R^2 score. Peningkatan ini menunjukkan bahwa model yang telah dioptimalkan memberikan prediksi yang lebih akurat dan lebih dapat diandalkan, yang pada gilirannya dapat membantu pemangku kepentingan dalam membuat kebijakan untuk mengatasi masalah pengangguran di Indonesia.

b. Dashboard Interaktif (Power BI)



Page report yang telah dibuat ini memberikan wawasan yang berguna tentang tingkat pengangguran terbuka dan prediksi tingkat pengangguran di Indonesia berdasarkan faktor ekonomi dari tahun 2016 hingga 2023. Dari visualisasi tersebut, dapat dilihat bahwa tingkat pengangguran terbuka mengalami fluktuasi, dengan penurunan yang signifikan pada tahun 2023, yang menunjukkan adanya perbaikan kondisi ketenagakerjaan secara nasional. Di sisi lain, prediksi tingkat pengangguran menunjukkan tren yang lebih stabil namun sedikit lebih rendah dibandingkan dengan tingkat pengangguran terbuka yang tercatat. Grafik tersebut juga memungkinkan pengguna untuk membandingkan data pengangguran dengan prediksi untuk masing-masing tahun secara visual. Melalui fitur interaktif pada dashboard ini, pengguna dapat memilih provinsi tertentu untuk melihat tren pengangguran di tingkat regional, memberikan gambaran yang lebih terperinci tentang distribusi pengangguran

di Indonesia. Dengan demikian, dashboard ini tidak hanya memberikan informasi yang mudah dipahami, tetapi juga menyediakan alat untuk analisis lebih lanjut bagi pemangku kepentingan dalam merumuskan kebijakan berbasis data.

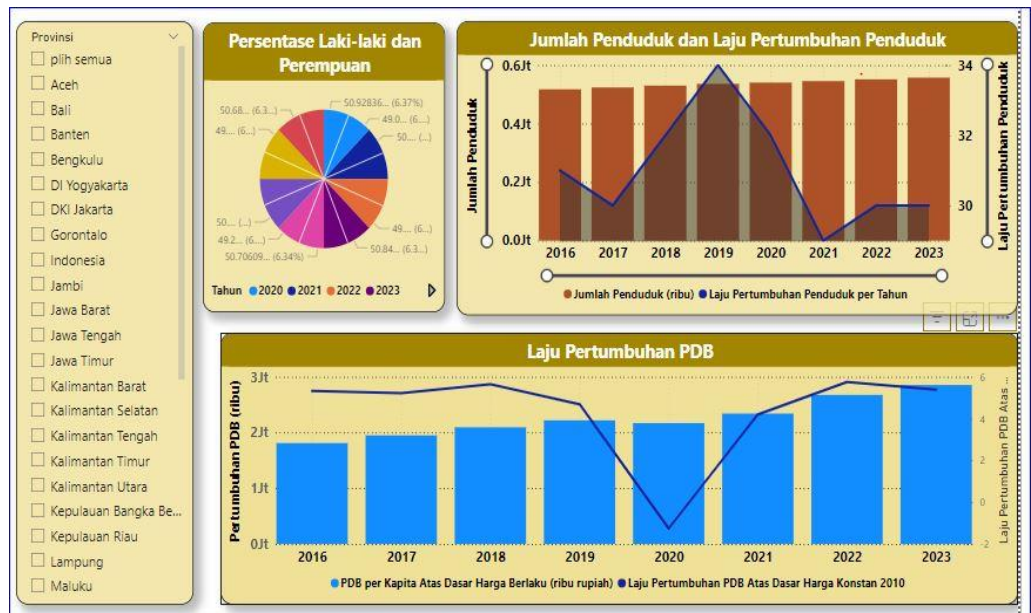


Berdasarkan hasil analisis menggunakan Key Influencer di Power BI, terlihat bahwa prediksi model memiliki pengaruh paling signifikan terhadap Tingkat Pengangguran Terbuka. Ketika nilai prediksi meningkat sebesar 1.55, rata-rata Tingkat Pengangguran Terbuka naik sebesar 1.75. Hal ini menunjukkan bahwa model machine learning yang digunakan mampu memprediksi tren pengangguran dengan akurasi yang kuat. Selain itu, persentase tenaga kerja informal menjadi faktor lain yang mendorong peningkatan pengangguran. Kenaikan sebesar 10.38% dalam tenaga kerja informal berkorelasi dengan peningkatan Tingkat Pengangguran Terbuka. Ini menandakan bahwa tingginya

ketergantungan pada pekerjaan informal berdampak negatif terhadap kestabilan pasar tenaga kerja. Sebaliknya, penurunan tenaga kerja formal sebesar 10.38% juga berperan dalam peningkatan pengangguran, menekankan pentingnya pengembangan lapangan kerja formal di berbagai sektor.

Dari segi segmentasi, data menunjukkan bahwa kelompok dengan prediksi tingkat pengangguran lebih tinggi memiliki rata-rata Tingkat Pengangguran Terbuka yang jauh di atas rata-rata keseluruhan (5.11). Segment 1 mencatat angka tertinggi dengan rata-rata 8.95, lebih tinggi 3.84 unit dari rata-rata nasional, dan mewakili 7.1% dari total data. Segment 2 dan Segment 3 juga menunjukkan tingkat pengangguran yang signifikan, masing-masing berada di angka 7.3 dan 6.63. Sementara itu, Segment 4 memiliki nilai terendah, dengan rata-rata 6.16, meskipun masih lebih tinggi 1.05 unit dari rata-rata keseluruhan. Segmentasi ini mengindikasikan adanya wilayah atau kelompok tertentu yang membutuhkan intervensi lebih lanjut untuk mengatasi masalah pengangguran.

Dengan hasil ini, kita dapat menyimpulkan bahwa peningkatan tenaga kerja informal dan penurunan tenaga kerja formal adalah dua faktor utama yang memengaruhi Tingkat Pengangguran Terbuka. Dengan memahami pola ini, pembuat kebijakan dapat fokus pada peningkatan kualitas lapangan kerja formal, mendorong program pelatihan tenaga kerja, dan merancang kebijakan yang lebih spesifik untuk segmen dengan risiko pengangguran tertinggi. Visualisasi ini tidak hanya menunjukkan hubungan antar variabel, tetapi juga memberikan gambaran komprehensif tentang segmen populasi yang paling terdampak.



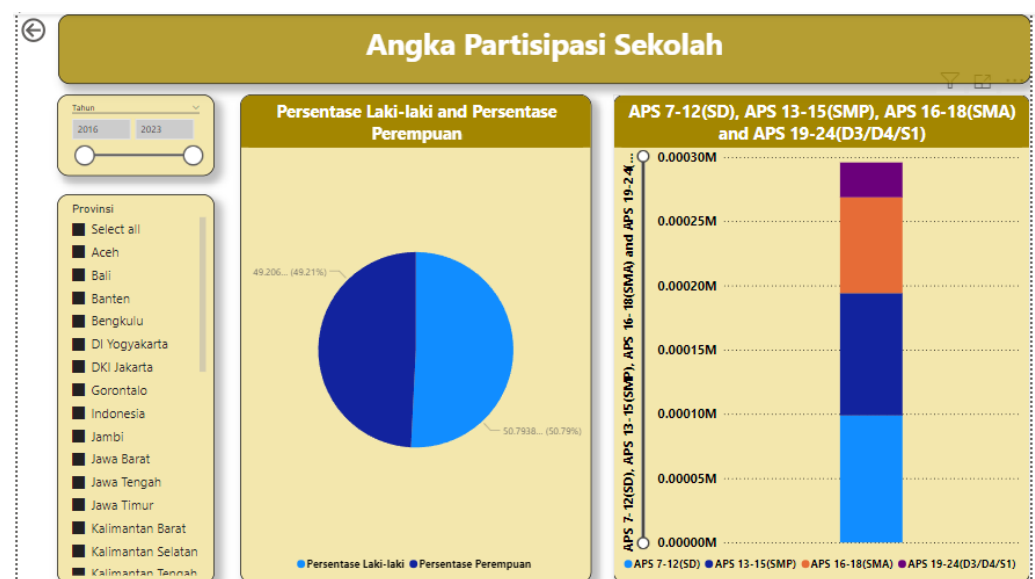
Berdasarkan data yang disajikan pada dashboard di atas, persentase penduduk laki-laki dan perempuan dari tahun 2020 hingga 2023 menunjukkan distribusi yang seimbang, di mana proporsi laki-laki sedikit lebih tinggi dibanding perempuan dengan perbedaan yang tidak signifikan, berkisar antara 50.2% hingga 50.8%. Hal ini mengindikasikan stabilitas distribusi gender dalam populasi yang dapat memengaruhi kebijakan pembangunan di bidang pendidikan, kesehatan, dan ketenagakerjaan.

Sementara itu, jumlah penduduk terus mengalami peningkatan dari tahun 2016 hingga 2023, mencapai hampir 0.6 juta (600 ribu) penduduk pada tahun 2023. Namun, laju pertumbuhan penduduk mengalami fluktuasi yang cukup tajam. Pertumbuhan penduduk tertinggi terjadi pada 2019 dengan tingkat sekitar 34%, tetapi setelah itu mengalami penurunan signifikan hingga titik terendah pada tahun 2021, berada di bawah 30%. Penurunan ini kemungkinan dipengaruhi oleh faktor eksternal seperti pandemi COVID-19, kebijakan pengendalian penduduk, dan tren keluarga kecil yang semakin meningkat. Tahun 2022 dan 2023 menunjukkan sedikit peningkatan, tetapi masih di bawah angka pertumbuhan sebelum 2019.

Di sisi lain, laju pertumbuhan ekonomi yang diukur melalui PDB per kapita menunjukkan tren pemulihan yang positif. PDB per kapita mengalami peningkatan bertahap dari tahun 2016 hingga 2023, dengan lonjakan signifikan pada 2022 dan 2023.

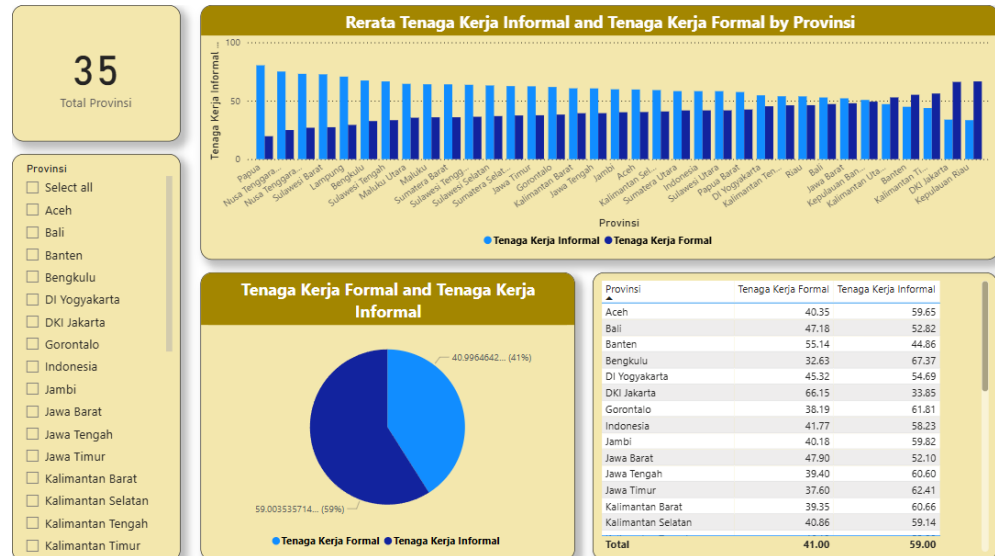
Sementara itu, laju pertumbuhan PDB sempat mengalami penurunan tajam pada 2019, yang kemungkinan disebabkan oleh perlambatan ekonomi global atau faktor internal lainnya. Pemulihan mulai terlihat pada tahun 2020 dan terus berlanjut hingga mencapai tren positif di tahun 2022 dan 2023, didorong oleh kebijakan pemulihan ekonomi dan penguatan investasi.

Secara keseluruhan, data ini menunjukkan tren positif dalam pertumbuhan ekonomi dan jumlah penduduk meskipun laju pertumbuhan penduduk mengalami perlambatan. Pemerintah dapat memanfaatkan momentum pemulihan ekonomi dengan mendorong kebijakan investasi, penguatan infrastruktur, serta peningkatan kualitas hidup masyarakat. Selain itu, evaluasi kebijakan kependudukan perlu dilakukan untuk menjaga keseimbangan pertumbuhan penduduk dan memastikan pembangunan berkelanjutan.



- Perbandingan Gender: Pie Chart menunjukkan persentase perempuan 49,21% dan laki-laki 50,79%.
- Data Per Provinsi: Diagram batang di bagian bawah menampilkan data APS (Angka Partisipasi Sekolah) berdasarkan provinsi. Ini memberikan gambaran tentang seberapa baik akses pendidikan di masing-masing provinsi.
- Tren Waktu: Dengan adanya slider tahun, kita bisa menganalisis perubahan dari tahun ke tahun (2016 hingga 2023) dalam partisipasi sekolah, baik untuk perempuan maupun laki-laki.
- Angka Partisipasi Sekolah (APS) Berdasarkan Jenjang:

1. APS 7-12 (SD) memiliki persentasi 99.00.
2. APS 13-15 (SMP) memiliki persentasi 95.57.
3. APS 16-18 (SMA) memiliki persentasi 72.37.
4. (APS 19-24 (D3/D4/S1) memiliki persentasi 25.34.



Distribusi tenaga kerja informal yang dominan menandakan tantangan ekonomi struktural, di mana sektor informal masih menjadi sandaran utama bagi sebagian besar pekerja. Sebaliknya, tenaga kerja formal hanya mencakup sebagian kecil, meskipun perannya sangat penting dalam mendukung stabilitas ekonomi dan kesejahteraan pekerja.

DKI Jakarta dan Kepulauan Riau tercatat sebagai provinsi dengan proporsi tenaga kerja formal tertinggi, yaitu masing-masing sebesar 66.15% dan 66.59%. Hal ini menunjukkan bahwa wilayah dengan tingkat industrialisasi dan urbanisasi yang tinggi cenderung memiliki porsi tenaga kerja formal lebih besar.

Di sisi lain, Papua dan Nusa Tenggara Timur (NTT) memiliki tingkat tenaga kerja formal paling rendah, masing-masing hanya sebesar 19.64% dan 24.95%, sementara porsi tenaga kerja informal mencapai 80.36% dan 75.05%. Ini menegaskan bahwa wilayah-wilayah tersebut masih sangat bergantung pada sektor informal. Provinsi seperti Nusa Tenggara Barat (NTB), Sulawesi Barat, dan Lampung mencatat dominasi tenaga kerja informal, dengan angka masing-masing mencapai 73.07%, 72.65%, dan 70.68%. Keterbatasan lapangan kerja formal di daerah ini memaksa sebagian besar penduduk untuk bekerja di sektor

informal, yang cenderung kurang stabil dan minim perlindungan sosial.

Wilayah Jawa dan Sumatera menunjukkan perbandingan yang lebih seimbang antara tenaga kerja formal dan informal. Contohnya, Jawa Barat memiliki porsi tenaga kerja formal sebesar 47.90%, sementara Jawa Timur cenderung didominasi tenaga kerja informal dengan 62.41%. Analisis ini menunjukkan bahwa ketimpangan distribusi tenaga kerja formal dan informal masih menjadi tantangan besar di Indonesia. Provinsi dengan urbanisasi dan industrialisasi tinggi, seperti DKI Jakarta, memiliki proporsi tenaga kerja formal yang lebih baik, sementara daerah tertinggal seperti Papua dan NTT masih bergantung pada sektor informal. Upaya strategis perlu difokuskan pada pengembangan ekonomi daerah untuk mendorong transformasi tenaga kerja informal ke sektor formal yang lebih stabil dan berkelanjutan.

2. Perbandingan dengan Solusi yang sudah ada

Solusi yang diusulkan dalam capstone project ini, yaitu model prediksi tingkat pengangguran berbasis machine learning dan dashboard interaktif, memiliki pendekatan yang berbeda namun saling melengkapi dengan solusi yang sudah ada untuk mengatasi masalah pengangguran di Indonesia. Solusi yang ada, seperti program penciptaan lapangan kerja melalui pembangunan infrastruktur dan kemudahan berusaha, berfokus pada penciptaan peluang kerja secara langsung melalui kebijakan ekonomi dan investasi. Program-program ini bertujuan untuk mengurangi pengangguran dengan menciptakan lebih banyak lapangan kerja, namun tidak memberikan analisis prediktif yang membantu dalam merumuskan kebijakan yang lebih tepat sasaran.

Berbeda dengan itu, model prediksi pengangguran yang dikembangkan dalam proyek ini memberikan alat analisis berbasis data yang dapat membantu memprediksi tren pengangguran di masa depan dan mengidentifikasi faktor-faktor ekonomi yang paling berpengaruh terhadap tingkat pengangguran. Hal ini memberi pemangku kepentingan, seperti pemerintah, sektor swasta, dan lembaga pendidikan, wawasan berbasis data untuk merancang kebijakan yang lebih tepat sasaran dan proaktif. Dashboard interaktif yang disediakan memungkinkan pengguna untuk mengevaluasi pengangguran berdasarkan provinsi dan melihat prediksi pengangguran secara visual,

memberikan pemahaman yang lebih mendalam tentang dampak dari kebijakan atau perubahan ekonomi yang sedang diterapkan.

Namun, solusi yang ada, seperti program pelatihan vokasi, Kartu Prakerja, dan beasiswa, lebih terfokus pada peningkatan keterampilan dan kualitas sumber daya manusia. Ini adalah langkah penting untuk mengurangi pengangguran, tetapi kurang memberikan alat prediktif untuk mengantisipasi kebutuhan keterampilan di masa depan. Di sisi lain, model prediksi pengangguran dapat mendukung program pelatihan vokasi dengan memberikan wawasan tentang keterampilan yang paling dibutuhkan dalam pasar kerja, sehingga lebih relevan dan tepat sasaran.

Sementara itu, program jaring pengaman sosial seperti PKH dan BLT yang bertujuan untuk mendukung mereka yang terdampak oleh pengangguran atau krisis ekonomi, solusi berbasis data dalam capstone project ini akan lebih memfokuskan pada pencegahan dan perencanaan jangka panjang melalui analisis prediktif. Dengan mengintegrasikan teknologi digital dan pemanfaatan e-commerce, yang juga menjadi bagian dari kebijakan pemerintah, model ini dapat memberikan pemahaman yang lebih dalam mengenai bagaimana perubahan dalam perekonomian Indonesia dapat mempengaruhi pengangguran dan sektor pekerjaan.

Pada akhirnya, kedua solusi ini yang sudah ada maupun yang diusulkan dalam proyek ini saling melengkapi satu sama lain, dengan solusi berbasis machine learning menawarkan analisis data yang lebih mendalam dan berbasis prediksi, sementara program-program yang ada bertujuan untuk memberikan intervensi langsung dalam menciptakan lapangan kerja dan meningkatkan keterampilan masyarakat.

3. Keterbatasan dari solusi yang dikembangkan

Keterbatasan utama dari solusi yang dikembangkan dalam proyek ini adalah ketergantungan pada ketersediaan dan kualitas data. Untuk model prediksi tingkat pengangguran berbasis machine learning yang akurat, diperlukan variabel-variabel yang lengkap dan relevan, seperti data ekonomi, demografi, serta faktor-faktor lain yang memengaruhi tingkat pengangguran. Jika variabel-variabel ini tidak tersedia secara lengkap atau memiliki kualitas yang rendah, maka hasil prediksi yang dihasilkan oleh model bisa kurang akurat. Beberapa variabel yang mungkin diperlukan, seperti data mengenai sektor pekerjaan, penyebaran lapangan kerja, dan tingkat pendidikan, harus dikumpulkan

dengan konsisten dan dalam periode waktu yang cukup panjang agar model dapat memberikan hasil yang lebih valid.

Selain itu, model ini juga memiliki keterbatasan dalam hal generalitas. Meskipun prediksi yang dihasilkan dapat memberikan wawasan berguna untuk kebijakan pengangguran di tingkat provinsi, model ini mungkin tidak dapat secara akurat memprediksi perubahan yang terjadi dalam jangka panjang, terutama ketika ada perubahan besar dalam kebijakan ekonomi, krisis ekonomi, atau faktor eksternal lainnya. Hal ini karena model prediksi ini bergantung pada data historis yang ada, dan perubahan mendadak dalam kondisi makroekonomi dapat mempengaruhi prediksi yang dihasilkan.

Keterbatasan lain adalah kemampuan model untuk menangani faktor non-ekonomi yang memengaruhi pengangguran. Meskipun model ini menggunakan data ekonomi dan demografi yang relevan, beberapa faktor sosial dan politik yang dapat memengaruhi tingkat pengangguran, seperti perubahan dalam kebijakan pemerintah, disparitas regional, atau krisis kesehatan global (seperti pandemi), tidak sepenuhnya dapat dimodelkan dengan baik oleh machine learning. Faktor-faktor ini sering kali bersifat lebih kompleks dan tidak selalu tercermin dalam data historis.

Selain itu, model ini mungkin memerlukan penyesuaian dan pemeliharaan yang berkelanjutan. Data yang digunakan untuk melatih model harus terus diperbarui agar tetap relevan dengan kondisi ekonomi terkini. Jika data terbaru tidak tersedia atau tidak diperbarui dengan cukup cepat, akurasi model dapat menurun, yang akan mempengaruhi hasil prediksi.

Terakhir, keberhasilan dashboard interaktif sangat bergantung pada desain antarmuka yang intuitif dan kemudahan penggunaan. Meskipun dashboard memungkinkan pengguna untuk memilih provinsi dan melihat data pengangguran, pemangku kepentingan dengan latar belakang non-teknis mungkin menghadapi kesulitan dalam menginterpretasi data yang lebih kompleks atau memilih variabel yang tepat tanpa panduan yang memadai. Ini bisa membatasi efektivitas dashboard dalam menyampaikan informasi secara menyeluruh kepada semua pengguna.

F. Kesimpulan

Proyek ini bertujuan untuk memberikan solusi berbasis data dalam mengatasi masalah pengangguran di Indonesia melalui model prediksi tingkat pengangguran yang dikembangkan menggunakan teknik machine learning, khususnya Random Forest Regression. Dengan memanfaatkan data ekonomi dan demografi, model ini mampu memberikan prediksi yang lebih akurat dibandingkan dengan metode regresi linier, karena mampu menangani hubungan non-linier antar variabel. Optimisasi model menggunakan GridSearchCV menunjukkan peningkatan yang signifikan dalam akurasi prediksi.

Selain itu, dashboard interaktif berbasis Power BI dikembangkan untuk menyajikan hasil prediksi pengangguran secara visual dan memberikan alat analisis bagi pemangku kepentingan, memungkinkan mereka untuk menganalisis pengangguran berdasarkan provinsi dan faktor ekonomi lainnya. Model ini juga mengidentifikasi faktor-faktor signifikan, seperti peningkatan tenaga kerja informal dan penurunan tenaga kerja formal, yang berkontribusi pada tingginya tingkat pengangguran terbuka.

Namun, terdapat beberapa keterbatasan, seperti ketergantungan pada kualitas data yang digunakan dan keterbatasan dalam menangani faktor non-ekonomi yang mempengaruhi pengangguran. Selain itu, perubahan mendadak dalam kebijakan ekonomi atau krisis global dapat mempengaruhi hasil prediksi.

Meskipun demikian, solusi ini memberikan wawasan yang sangat berharga untuk membantu merumuskan kebijakan ketenagakerjaan yang lebih tepat sasaran. Dengan pengembangan lebih lanjut dan pembaruan data secara berkala, model prediksi ini dapat terus memberikan manfaat dalam merencanakan kebijakan ketenagakerjaan yang lebih efektif di masa depan.

G. Lampiran

Link Github:

<https://github.com/evanfulka/CapstoneProjectBangsawan>

Link Kumpulan Data Raw:

<https://docs.google.com/document/d/1XDosxFQSZdaH5sgVrMmFClg7OzkW39XzFZ4cKi5jpJw/edit?usp=sharing>

Link PPT Pitching Video:

https://www.canva.com/design/DAGZVjeZ-E8/-Qu45S-RybdQgrwAnAondw/edit?utm_content=DAGZVjeZ-E8&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton

Link Video Pitching:

[Video Pitching Kelompok 14 Bangsawan](#)

Link Dashboard:

[Dashboard Kelompok 14 Bangsawan](#)

Link Video Demonstrasi:

[Video Demonstrasi Kelompok 14 Bangsawan](#)