**BIODATA UMUM NSC 2021**

**Nama Tim : Tim SinCosTan**

**Nomor Peserta : NSC21A0028**

**Nama Peserta 1 : Bhayu Phermana Sachty Muktar**

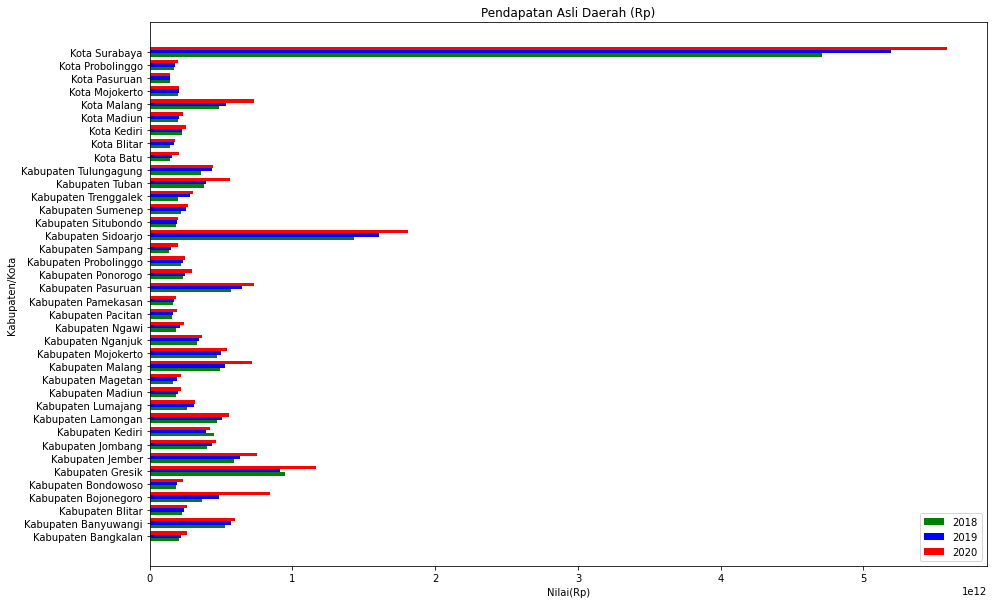
**Nama Peserta 2 : Muhammad Jamaalullail**

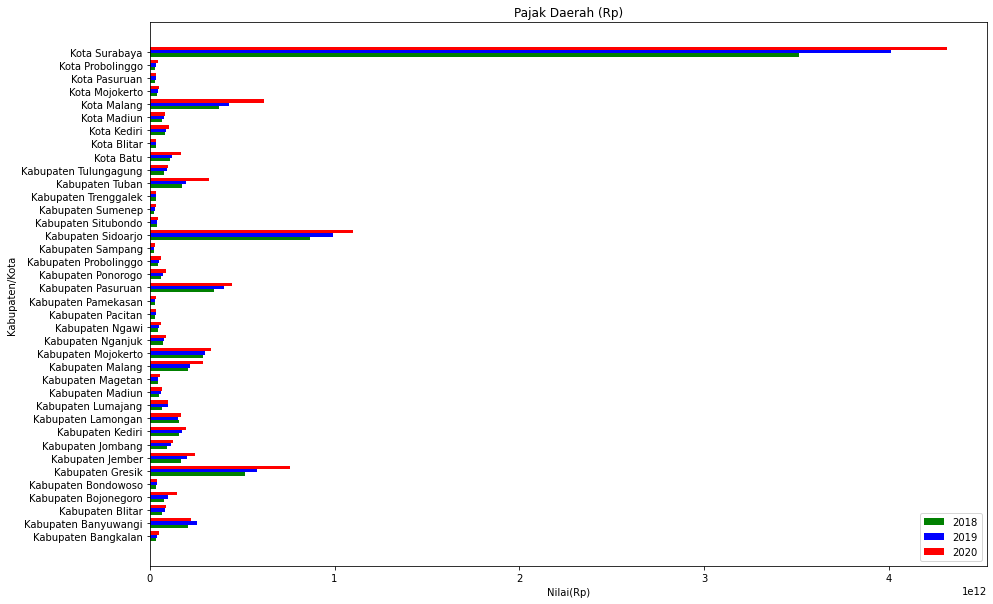
**Nomor Telepon / ID Line : 085778071720/bhayu\_psm**

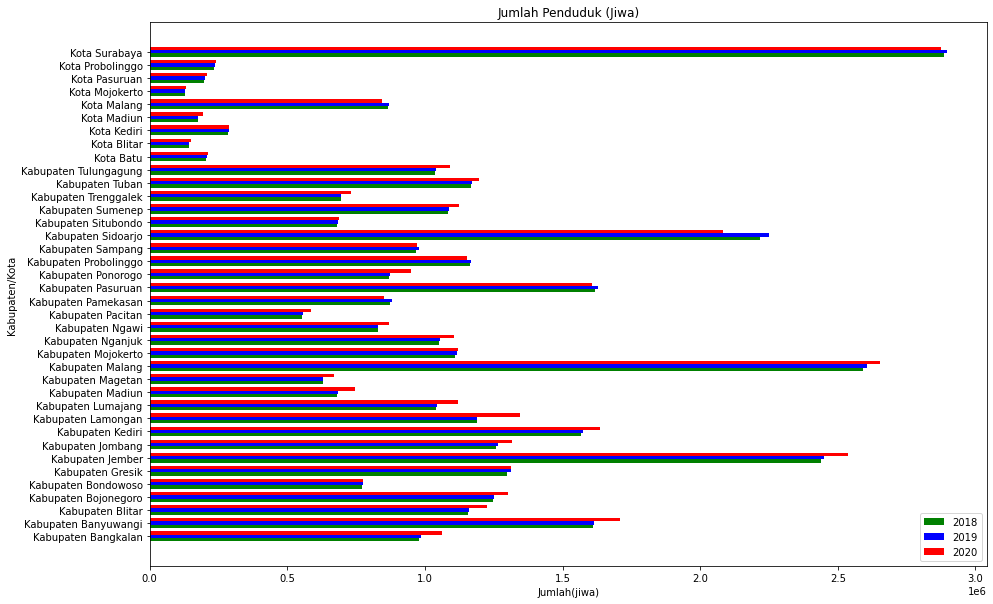
**088291235054/jamaalullail22**

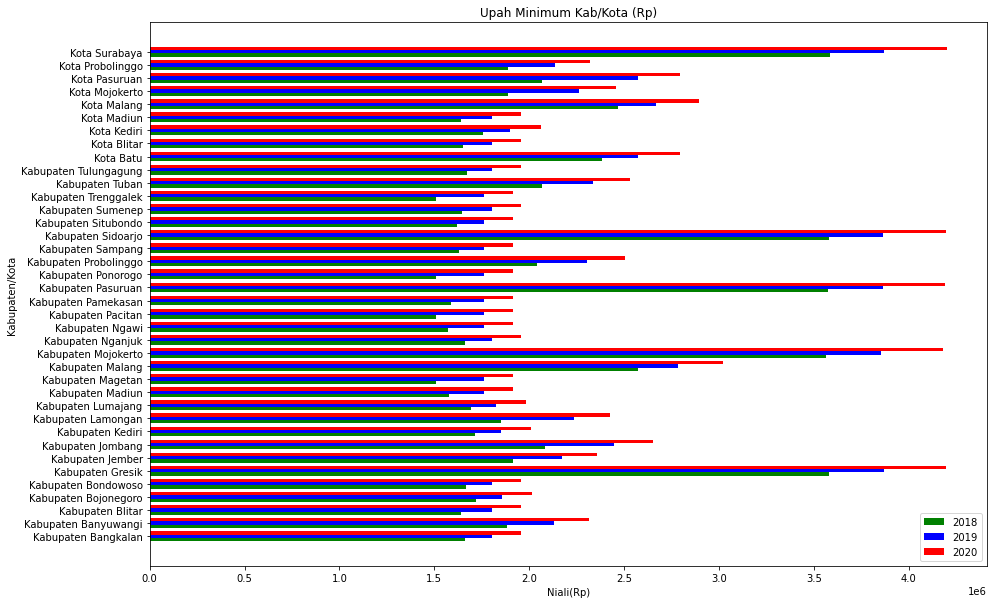
**LEMBAR JAWABAN NSC 2021**

Berikut adalah diagram batang berdasarkan data dari file “Data Penyisihan NSC 2021.xlsx” yang dibagi menjadi 4 variabel yaitu “Pendapatan Asli Daerah”, “Pajak Daerah”, “Jumlah Penduduk” dan “Upah Minimum Kota/Kabupaten”.



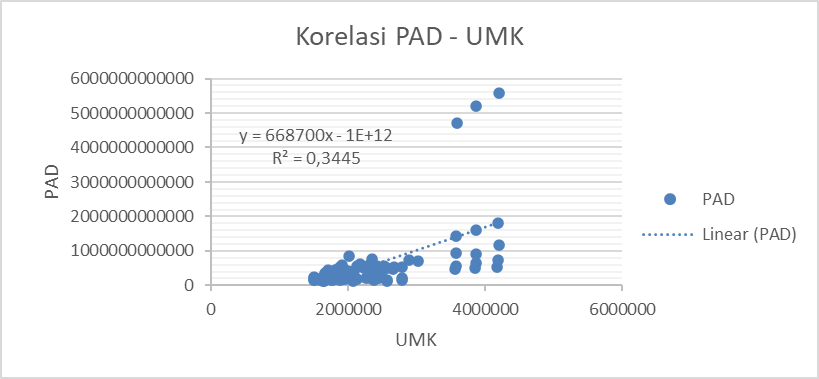




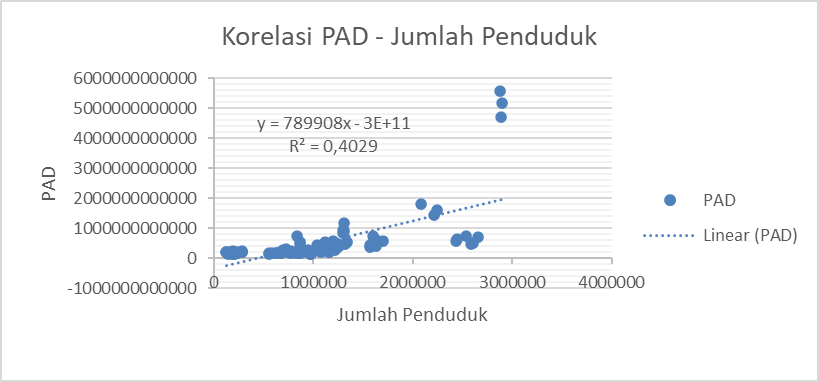


Berdasarkan data tersebut, tanggapan kami terkait Pendapatan Asli Daerah (PAD), Pajak Daerah, Jumlah Penduduk dan Upah Minimum Kota/Kabupaten dari kabupaten/kota di provinsi Jawa Timur setiap tahunnya dari tahun 2018 – 2020 secara umum terjadi peningkatan yang signifikan terutama pada tahun 2020. Menurut kami, terdapat pengaruh dari adanya pandemi Covid-19 yang terjadi pada awal tahun 2020 dan kami berasumsi bahwa kenaikan tersebut dapat terjadi karena pemerintah daerah provinsi Jawa Timur menganggarkan dana kesehatan untuk setiap kota/kabupaten. Namun di sisi lain, hal ini mengakibatkan beberapa kabupaten/kota dengan dampak pandemi yang tinggi, membuat jumlah penduduk berkurang. Dan dengan adanya pandemi Covid-19 upah minimum kabupaten/kota mengalami kenaikan yang signifikan dikarenakan kebutuhan yang meningkat.

Untuk mengetahui hubungan penerimaan pajak daerah, jumlah penduduk, penghasilan asli daerah dan upah minimum kota/kabupaten di provinsi Jawa Timur, kami menggunakan metode analisis regresi. Untuk mengetahui regresi apa yang digunakan kami membagi variabel menjadi beberapa jenis, yaitu variabel independen dan variabel dependen. Kami menentukan variabel yang merupakan variabel dependen yaitu “Pendapatan Asli Daerah (PAD)” dan variabel independen yaitu “Pajak Daerah”, “Jumlah Penduduk” dan “Upah Minimum Kabupaten/Kota”. Dikarenakan terdapat dua atau lebih variabel independen, maka akan digunakan metode analisis regresi linear berganda. Selanjutnya kami mencari korelasi antara variabel dependen dan variabel-variabel independennya menggunakan scatter plot (tabel sebaran) dan mencari koefisien determinasi (R2) untuk mengetahui nilai kekuatan hubungan antara variabel dependen dengan variabel-variabel independennya. Berikut adalah tabel sebaran antara variabel dependen dengan masing-masing variabel independennya.

Chart, scatter chart

Description automatically generated



Berdasarkan tabel di atas, didapat nilai R2 terbesar yaitu 0.9851 pada korelasi antara variabel dependen “Pendapatan Asli Daerah (PAD)” dengan variabel independen “Pajak Daerah”. Maka dapat disimpulkan korelasi terkuat adalah antara variabel “Pendapatan Asli Daerah (PAD)” dengan variabel “Pajak Daerah”. Untuk melihat hubungan yang ada pada variabel tersebut, kami tidak hanya melihat pada pada satu variabel saja, namun kami harus melihat semua variabel pada data tersebut. Sehingga untuk melihatnya kami akan menggunakan metode analisis regresi berganda.

Berikut adalah tahapan yang akan kami lakukan:

1. Melakukan persiapan dari data yang akan dianalisis berserta dengan *software* yang akan digunakan.

2. *Software* yang kami gunakan dalam proses analisis adalah Python dan Microsoft Excel.

3. Akan dilakukan analisis deskriptif untuk melihat hubungan setiap variabelnya. Hasilnya akan digunakan untuk menjawab model satu dan menyimpulkan analisis apa yang digunakan.

4. Membentuk model persamaan regresi linear berganda berdasarkan tahun. Maka akan terbentuk tiga model persamaan regresi.

5. Melakukan validasi model

6. Melakukan uji asumsi yang terdiri dari: Linearitas, Normalitas, Multikolinieritas, Autokorelasi, Homoskedastisitas.

7. Menentukan model yang tepat

8. Membuat kesimpulan disertai interpretasi dari model yang telah ditentukan.

Berdasarkan data yang kami dapat, akan dibentuk model regresi linear berganda untuk masing-masing tahun dan juga model regresi secara keseluruhan. Setiap data kami olah dengan menggunakan program python, kami menggunakan data uji sebanyak 30% untuk setiap model. Setelah dilakukan pengolahan pada setiap data, didapat model persamaan linear berganda yaitu sebagai berikut:

**1. Tahun 2018**

* Keterangan Variabel:

= Pendapatan Asli Daerah model ke-1

= Pajak Daerah (Rp)

= Jumlah Penduduk (Jiwa)

= UMK (Rp)

* Keterangan Angka:

Intersep:

Koefisen:

**2. Tahun 2019**

* Keterangan Variabel:

= Pendapatan Asli Daerah model ke-2

= Pajak Daerah (Rp)

= Jumlah Penduduk (Jiwa)

= UMK (Rp)

* Keterangan Angka:

Intersep:

Koefisen:

**3. Tahun 2020**

* Keterangan Variabel:

= Pendapatan Asli Daerah model ke-3

= Pajak Daerah (Rp)

= Jumlah Penduduk (Jiwa)

= UMK (Rp)

* Keterangan Angka:

Intersep:

Koefisen:

**4. Tahun 2018 – 2020**

* Keterangan Variabel:

= Pendapatan Asli Daerah model ke-4

= Pajak Daerah (Rp)

= Jumlah Penduduk (Jiwa)

= UMK (Rp)

* Keterangan Angka:

Intersep:

Koefisen:

Uraian arti dari masing-masing angka tersebut adalah sebagai berikut (Pada ) :

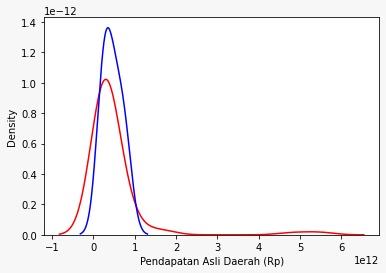
* Nilai Intersep adalah perkiraan nilai rata-rata variabel dependen kami ketika semua nilai variabel independen kami adalah 0. Dalam kasus kami pada model pertama, ini berarti jika tidak diberlakukan pajak daerah atau pajak daerahnya Rp0, jumlah penduduknya 0, dan Upah Minimumnya Rp0 maka Pendapatan Asli Daerah tersebut adalah sekitar: Rp.
* Untuk koefisien, kami memiliki 3 nilai untuk variabel Pajak Daerah, Jumlah Penduduk dan Upah Minimum Kota/Kabupaten. Nilai ini merepresentasikan hubungan ketiga variabel independen tersebut dengan variabel dependen yaitu Pendapatan Asli Daerah, di mana perubahan tepat 1 pada variabel independen akan mengubah nilai variabel dependen sebesar koefisien.

Setelah mendapatkan model persamaan regresi tersebut, akan dilakukan prediksi pada subset pengujian dan mencari koefisien determinasi (R2) menggunakan program python. Selanjutnya akan diuji dengan melakukan perbandingan plot distribusi antara nilai prediksi dengan nilai sebenarnya pada masing-masing tahun dan juga pada tahun secara keseluruhan, sebagai berikut:

* **Plot distribusi 2018**

R2 = 0.9089256118570053

90% variasi dalam Pendapatan Asli Daerah (Rp) dijelaskan oleh Variabel bebas, Nilai R kuadrat mendekati 1, sehingga dapat dikatakan model sudah sesuai.

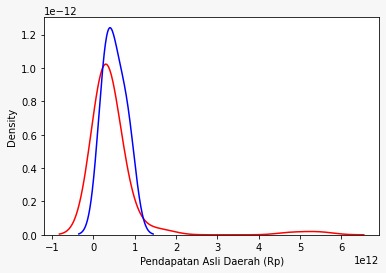


Terlihat bahwa nilai prediksi (biru) tidak jauh berbeda dengan nilai aktual (merah), sehingga dapat diambil model Regresi Linear Berganda pada tahun 2018.

* **Plot distribusi 2019**

R2 = 0.8915587161116589

89% variasi dalam Pendapatan Asli Daerah (Rp) dijelaskan oleh Variabel bebas, Nilai R kuadrat mendekati 1, sehingga dapat dikatakan model sudah sesuai.

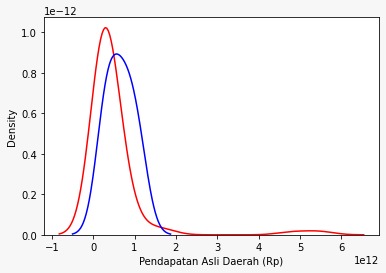


Terlihat bahwa nilai prediksi (biru) tidak jauh berbeda dengan nilai aktual (merah), sehingga dapat diambil model Regresi Linear Berganda pada tahun 2019.

* **Plot distribusi 2020**

R2 = 0.8821992739114471

88% variasi dalam Pendapatan Asli Daerah (Rp) dijelaskan oleh Variabel bebas, Nilai R kuadrat mendekati 1, sehingga dapat dikatakan model sudah sesuai.

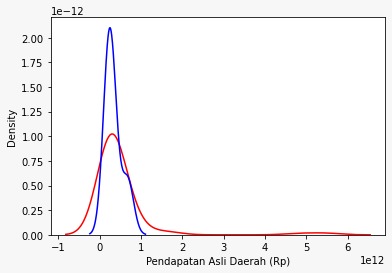


Terlihat bahwa nilai prediksi (biru) tidak jauh berbeda dengan nilai aktual (merah), sehingga dapat diambil model Regresi Linear Berganda pada tahun 2020.

* **Plot distribusi 2018 – 2020**

R2 = 0.9926471092133367

99% variasi dalam Pendapatan Asli Daerah (Rp) dijelaskan oleh Variabel bebas, Nilai R kuadrat mendekati 1, sehingga dapat dikatakan model sudah sesuai.



Terlihat bahwa nilai prediksi (biru) tidak jauh berbeda dengan nilai aktual (merah), sehingga dapat diambil model Regresi Linear Berganda pada tahun 2018 - 2020.

Setelah melakukan uji pada persamaan regresi berganda dengan membandingan plot distribusi nilai prediksi dan nilai sebenarnya, selanjutnya akan dilakukan uji validasi model Regresi Linear Berganda dengan OLS (Ordinary Least Square) dengan mencetak ringkasan model yang berisi banyak nilai penting yang dapat kita gunakan untuk mengevaluasi model kami, sebagai berikut:

|  |  |
| --- | --- |
| * OLS 2018   Text  Description automatically generated | * OLS 2019   Text  Description automatically generated |
| * OLS 2020   Text  Description automatically generated | * OLS 2018 – 2020   Text  Description automatically generated |

Selanjutnya kami akan memvalidasi model dengan melakukan analisis residual, berikut adalah daftar pengujian atau asumsi yang akan kita lakukan untuk mengecek validitas model kita: Linearitas, Normalitas, Multikolinieritas, Autokorelasi dan Homoskedastisitas.

* **Uji Linearitas**

Ini mengasumsikan bahwa terdapat hubungan linier antara variabel bebas dan variabel terikat. Dalam kasus kami karena kami memiliki beberapa variabel independen, kami dapat melakukan ini dengan menggunakan plot pencar untuk melihat nilai prediksi kami versus nilai sebenarnya.

Chart, scatter chart

Description automatically generated

Plot sebar menunjukkan titik sisa yang tersebar merata di sekitar garis diagonal, sehingga **kita dapat mengasumsikan bahwa ada hubungan linier antara variabel independen dan dependen kita.**

* **Uji Normalitas**

Ini mengasumsikan bahwa istilah kesalahan model terdistribusi normal. Kami akan memeriksa normalitas residual dengan memplotnya ke dalam histogram dan melihat nilai p dari uji normalitas Anderson-Darling. Residual adalah perbedaan antara nilai observasi dengan nilai prediksi yang diperoleh dengan menggunakan model regresi estimasi. Kemudian dibandingkan dengan threshold 0,05, jika p-value yang kita peroleh lebih tinggi dari threshold maka kita dapat mengasumsikan bahwa residual kita terdistribusi normal.

Chart, histogram

Description automatically generated

**Uji Normalitas menunjukkkan Residual data tidak terdistribusi normal.**

* **Uji Multikolinieritas**

Ini mengasumsikan bahwa prediktor yang digunakan dalam regresi tidak berkorelasi satu sama lain. Untuk mengidentifikasi apakah ada korelasi antara prediktor kami, kami dapat menghitung koefisien korelasi Pearson antara setiap kolom dalam data kami menggunakan Peta Panas.

Chart, treemap chart

Description automatically generated A picture containing text, monitor, screenshot

Description automatically generated

Berdasarkan data peta panas tersebut, **kita mendapatkan hasil bahwa hubungan antara variabel dependen PAD dengan variabel-variabel independennya memiliki koefisen korelasi di atas 50% yang berarti bahwa ada hubungan kuat antara variabel dependen dan independen.**

* **Uji Autokorelasi**

Kita dapat mendeteksi autokorelasi dengan melakukan uji Durbin-Watson untuk menentukan apakah ada korelasi positif atau negatif. Pada langkah ini kami akan menghitung skor Durbin-Watson dan kemudian menilai nilainya dengan kondisi sebagai berikut:

* Jika skor Durbin-Watson kurang dari 1,5 maka terdapat autokorelasi positif dan asumsi tidak terpenuhi.
* Jika skor Durbin-Watson antara 1,5 dan 2,5 maka tidak ada autokorelasi dan asumsi puas.
* Jika skor Durbin-Watson lebih dari 2.5 maka terdapat autokorelasi negatif dan asumsi tidak puas.

Graphical user interface, text, application

Description automatically generated

**Dari Pengujian diatas didapat bahwa tidak ada autokorelasi dan asumsi puas.**

* **Uji Homoskedastisitas**

Ini mengasumsikan homoskedastisitas, yang merupakan varian yang sama dalam istilah kesalahan kami. Heteroskedastisitas, pelanggaran homoskedastisitas, terjadi ketika kita tidak memiliki varian genap di seluruh istilah kesalahan. Untuk mendeteksi homoskedastisitas, kita dapat memplot residual kita dan melihat apakah variansnya tampak seragam.

Chart, scatter chart

Description automatically generated

Dari titik-titik yang terlihat, residual kami tampaknya memiliki varian yang konstan dan seragam, sehingga **kami dapat mengasumsikan bahwa itu memenuhi asumsi homoskedastisitas.**

Berdasarkan uji asumsi yang diperoleh maka untuk model regresi linear keseluruhan tahun didapatkan menjadi model yang paling tepat dikarenakan model kami berhasil melewati semua pengujian dalam langkah validasi model, sehingga kami dapat menyimpulkan bahwa model kami dapat berperforma baik. Model kami juga memiliki skor R2 sebesar 99%, yang berarti ketiga variabel independen di atas menjadi faktor yang sangat memengaruhi Pendapatan Asli Daerah di Provisi Jawa Timur dari Tahun 2018 – 2020.

Maka dapat disimpulkan bahwa model persamaan linear berganda tersebut sudah menjelaskan hubungan dari variabel dependen dan variabel indepennya. Dapat ditarik kesimpulan bahwa meningkatnya variabel Pajak Daerah dan Jumlah penduduk membuat besaran Pendapatan Asli Daerah juga meningkat. Sebaliknya, apabila Upah Minimum Kabupaten/Kota (UMK) terjadi peningkatan, maka hal itu dapat menyebabkan turunan nilai Pendapatan Asli Daerah (PAD) di Kabupaten/Kota tersebut.

**LAMPIRAN CODE PYTHON**

1. Import Library Python

# Basic libs

import pandas as pd

import numpy as np

## Building Model

import statsmodels

import statsmodels.api as sm

## Data Visualization

import seaborn as sns

import matplotlib.pyplot as plt

2. Mempersiapkan Data

data\_apbd = pd.read\_excel('Data-Penyisihan-NSC-2021.xlsx', header=[3], index\_col=0)

data\_apbd

data\_2018 = data\_apbd.iloc[0:38]

data\_2018

data\_2019 = data\_apbd.iloc[38:76]

data\_2019

data\_2020= data\_apbd.iloc[76:114]

data\_2020

3. Mendefinisikan data-data yang diperlukan pada tahun 2018-2020

pd.set\_option('display.float\_format', '{:.2f}'.format)

df\_2018 = data\_2018.drop(['Tahun'], axis=1)

df\_2018

x = np.array(df\_2018['Kab/Kota di Jawa Timur'])

X = np.arange(len(x))

a, b, c, d = np.array(df\_2018['Pendapatan Asli Daerah (Rp)']), np.array(df\_2018['Pajak Daerah (Rp)']), np.array(df\_2018['Jumlah Penduduk (Jiwa)']), np.array(df\_2018['Upah Minimum Kab/Kota (Rp)'])

df\_2019 = data\_2019.drop(['Tahun'], axis=1)

df\_2019

x = np.array(df\_2019['Kab/Kota di Jawa Timur'])

e, f, g, h = np.array(df\_2019['Pendapatan Asli Daerah (Rp)']), np.array(df\_2019['Pajak Daerah (Rp)']), np.array(df\_2019['Jumlah Penduduk (Jiwa)']), np.array(df\_2019['Upah Minimum Kab/Kota (Rp)'])

4. Membuat Diagram Batang Pendapatan Asli Daerah

df\_2020 = data\_2020.drop(['Tahun'], axis=1)

df\_2020

x = np.array(df\_2020['Kab/Kota di Jawa Timur'])

i, j, k, l = np.array(df\_2020['Pendapatan Asli Daerah (Rp)']), np.array(df\_2020['Pajak Daerah (Rp)']), np.array(df\_2020['Jumlah Penduduk (Jiwa)']), np.array(df\_2020['Upah Minimum Kab/Kota (Rp)'])

plt.figure(figsize=[15, 10])

plt.barh(X, a, color='green', height=0.25)

plt.barh(X+0.25, e, color='blue', height=0.25)

plt.barh(X+0.5, i, color='red', height=0.25)

plt.legend(['2018', '2019', '2020'])

plt.yticks([i+0.25 for i in range(38)], ['Kabupaten Bangkalan','Kabupaten Banyuwangi','Kabupaten Blitar','Kabupaten Bojonegoro','Kabupaten Bondowoso','Kabupaten Gresik','Kabupaten Jember','Kabupaten Jombang','Kabupaten Kediri','Kabupaten Lamongan','Kabupaten Lumajang','Kabupaten Madiun','Kabupaten Magetan','Kabupaten Malang','Kabupaten Mojokerto','Kabupaten Nganjuk','Kabupaten Ngawi','Kabupaten Pacitan','Kabupaten Pamekasan','Kabupaten Pasuruan','Kabupaten Ponorogo','Kabupaten Probolinggo','Kabupaten Sampang','Kabupaten Sidoarjo','Kabupaten Situbondo','Kabupaten Sumenep','Kabupaten Trenggalek','Kabupaten Tuban','Kabupaten Tulungagung','Kota Batu','Kota Blitar','Kota Kediri','Kota Madiun','Kota Malang','Kota Mojokerto','Kota Pasuruan','Kota Probolinggo','Kota Surabaya'])

plt.xlabel('Nilai(Rp)')

plt.ylabel('Kabupaten/Kota')

plt.title('Pendapatan Asli Daerah (Rp)')

plt.show()

5. Membuat Diagram Batang Pajak Daerah

plt.figure(figsize=[15, 10])

plt.barh(X, b, color='green', height=0.25)

plt.barh(X+0.25, f, color='blue', height=0.25)

plt.barh(X+0.5, j, color='red', height=0.25)

plt.legend(['2018', '2019', '2020'])

plt.yticks([i+0.25 for i in range(38)], ['Kabupaten Bangkalan','Kabupaten Banyuwangi','Kabupaten Blitar','Kabupaten Bojonegoro','Kabupaten Bondowoso','Kabupaten Gresik','Kabupaten Jember','Kabupaten Jombang','Kabupaten Kediri','Kabupaten Lamongan','Kabupaten Lumajang','Kabupaten Madiun','Kabupaten Magetan','Kabupaten Malang','Kabupaten Mojokerto','Kabupaten Nganjuk','Kabupaten Ngawi','Kabupaten Pacitan','Kabupaten Pamekasan','Kabupaten Pasuruan','Kabupaten Ponorogo','Kabupaten Probolinggo','Kabupaten Sampang','Kabupaten Sidoarjo','Kabupaten Situbondo','Kabupaten Sumenep','Kabupaten Trenggalek','Kabupaten Tuban','Kabupaten Tulungagung','Kota Batu','Kota Blitar','Kota Kediri','Kota Madiun','Kota Malang','Kota Mojokerto','Kota Pasuruan','Kota Probolinggo','Kota Surabaya'])

plt.xlabel('Nilai(Rp)')

plt.ylabel('Kabupaten/Kota')

plt.title('Pajak Daerah (Rp)')

plt.show()

6. Membuat Diagram Batang Jumlah Penduduk

plt.figure(figsize=[15, 10])

plt.barh(X, c, color='green', height=0.25)

plt.barh(X+0.25, g, color='blue', height=0.25)

plt.barh(X+0.5, k, color='red', height=0.25)

plt.legend(['2018', '2019', '2020'])

plt.yticks([i+0.25 for i in range(38)], ['Kabupaten Bangkalan','Kabupaten Banyuwangi','Kabupaten Blitar','Kabupaten Bojonegoro','Kabupaten Bondowoso','Kabupaten Gresik','Kabupaten Jember','Kabupaten Jombang','Kabupaten Kediri','Kabupaten Lamongan','Kabupaten Lumajang','Kabupaten Madiun','Kabupaten Magetan','Kabupaten Malang','Kabupaten Mojokerto','Kabupaten Nganjuk','Kabupaten Ngawi','Kabupaten Pacitan','Kabupaten Pamekasan','Kabupaten Pasuruan','Kabupaten Ponorogo','Kabupaten Probolinggo','Kabupaten Sampang','Kabupaten Sidoarjo','Kabupaten Situbondo','Kabupaten Sumenep','Kabupaten Trenggalek','Kabupaten Tuban','Kabupaten Tulungagung','Kota Batu','Kota Blitar','Kota Kediri','Kota Madiun','Kota Malang','Kota Mojokerto','Kota Pasuruan','Kota Probolinggo','Kota Surabaya'])

plt.xlabel('Jumlah(jiwa)')

plt.ylabel('Kabupaten/Kota')

plt.title('Jumlah Penduduk (Jiwa)')

plt.show()

7. Membuat Diagram Batang Upah Minimum Kabupaten/Kota

lt.figure(figsize=[15, 10])

plt.barh(X, d, color='green', height=0.25)

plt.barh(X+0.25, h, color='blue', height=0.25)

plt.barh(X+0.5, l, color='red', height=0.25)

plt.legend(['2018', '2019', '2020'])

plt.yticks([i+0.25 for i in range(38)], ['Kabupaten Bangkalan','Kabupaten Banyuwangi','Kabupaten Blitar','Kabupaten Bojonegoro','Kabupaten Bondowoso','Kabupaten Gresik','Kabupaten Jember','Kabupaten Jombang','Kabupaten Kediri','Kabupaten Lamongan','Kabupaten Lumajang','Kabupaten Madiun','Kabupaten Magetan','Kabupaten Malang','Kabupaten Mojokerto','Kabupaten Nganjuk','Kabupaten Ngawi','Kabupaten Pacitan','Kabupaten Pamekasan','Kabupaten Pasuruan','Kabupaten Ponorogo','Kabupaten Probolinggo','Kabupaten Sampang','Kabupaten Sidoarjo','Kabupaten Situbondo','Kabupaten Sumenep','Kabupaten Trenggalek','Kabupaten Tuban','Kabupaten Tulungagung','Kota Batu','Kota Blitar','Kota Kediri','Kota Madiun','Kota Malang','Kota Mojokerto','Kota Pasuruan','Kota Probolinggo','Kota Surabaya'])

plt.xlabel('Niali(Rp)')

plt.ylabel('Kabupaten/Kota')

plt.title('Upah Minimum Kab/Kota (Rp)')

plt.show()

8. Regresi Linear Sederhana (Membandingkan hubungan per variabel independen)

#Pairplot per Variabel independen dengan Variabel dependen

sns.pairplot(data\_apbd, x\_vars=['Pajak Daerah (Rp)', 'Jumlah Penduduk (Jiwa)','Upah Minimum Kab/Kota (Rp)'], y\_vars=['Pendapatan Asli Daerah (Rp)'], kind='reg')

9. Regresi Linear Berganda (Membandingkan dengan seluruh variabel independen yang ada) pada tahun 2018-2020

#Regresi Linear Berganda

from sklearn.linear\_model import LinearRegression

from sklearn.model\_selection import train\_test\_split

z = data\_apbd[['Pajak Daerah (Rp)','Jumlah Penduduk (Jiwa)','Upah Minimum Kab/Kota (Rp)']]

y = data\_apbd['Pendapatan Asli Daerah (Rp)']

10. Membuat model persamaan linear berganda

z\_train, z\_test, b\_train, b\_test = train\_test\_split(z, y, test\_size=0.3, random\_state=123)

print(z\_train.shape, b\_train.shape)

print(z\_test.shape, b\_test.shape)

lm2 = LinearRegression()

model2= lm2.fit(z\_train, b\_train)

lm2.intercept\_

lm2.coef\_

11. Melakukan Prediksi pada subset pengujian

#Prediksi

w = lm2.predict(z\_test)[0:5]

w

lm2.score(z\_test, b\_test)

12. Menampilkan Plot distribusi nilai prediksi vs nilai sebenarnya

ax1= sns.distplot(data\_apbd['Pendapatan Asli Daerah (Rp)'], hist=False, color='red', label='Actual Value')

sns.distplot(w, hist=False, color='blue', label='Fitted Values', ax=ax1)

13. Melakukan Validasi Model tahun 2018-2020 menggunakan OLS Regression

#OLS Regression Results 2018-2020

X = data\_apbd[['Pajak Daerah (Rp)', 'Jumlah Penduduk (Jiwa)','Upah Minimum Kab/Kota (Rp)']]

X = sm.add\_constant(X) # adding a constant

olsmod = sm.OLS(data\_apbd['Pendapatan Asli Daerah (Rp)'], X).fit()

print(olsmod.summary())

14. Akan dilakukan Pengujian Asumsi dengan Analisis Residual

data\_apbd['pad\_pred'] = olsmod.predict(X)

data\_apbd['residual'] = olsmod.resid

data\_apbd

15. Linearitas

sns.pairplot(data\_apbd, x\_vars=['Pendapatan Asli Daerah (Rp)'], y\_vars=['pad\_pred'], kind='reg', size=5)

16. Normalitas

from statsmodels.stats.diagnostic import normal\_ad

# Performing the test on the residuals

p\_value = normal\_ad(data\_apbd['residual'])[1]

print('p-value from the test Anderson-Darling test below 0.05 generally means non-normal:', p\_value)

# Plotting the residuals distribution

plt.subplots(figsize=(8, 4))

plt.title('Distribution of Residuals', fontsize=18)

sns.distplot(data\_apbd['residual'])

plt.show()

# Reporting the normality of the residuals

if p\_value < 0.05:

    print('Residuals are not normally distributed')

else:

    print('Residuals are normally distributed')

17. Multikolinearitas

corr = data\_apbd[['Pendapatan Asli Daerah (Rp)', 'Pajak Daerah (Rp)', 'Jumlah Penduduk (Jiwa)','Upah Minimum Kab/Kota (Rp)']].corr()

print('Pearson correlation coefficient matrix of each variables:\n', corr)

# Generate a mask for the diagonal cell

mask = np.zeros\_like(corr, dtype=np.bool)

np.fill\_diagonal(mask, val=True)

# Initialize matplotlib figure

fig, ax = plt.subplots(figsize=(4, 3))

# Generate a custom diverging colormap

cmap = sns.diverging\_palette(220, 10, as\_cmap=True, sep=100)

cmap.set\_bad('grey')

# Draw the heatmap with the mask and correct aspect ratio

sns.heatmap(corr, mask=mask, cmap=cmap, vmin=-1, vmax=1, center=0, linewidths=.5)

fig.suptitle('Pearson correlation coefficient matrix', fontsize=14)

ax.tick\_params(axis='both', which='major', labelsize=10)

18. Autokorelasi

from statsmodels.stats.stattools import durbin\_watson

durbinWatson = durbin\_watson(data\_apbd['residual'])

print('Durbin-Watson:', durbinWatson)

if durbinWatson < 1.5:

    print('Signs of positive autocorrelation', '\n')

    print('Assumption not satisfied')

elif durbinWatson > 2.5:

    print('Signs of negative autocorrelation', '\n')

    print('Assumption not satisfied')

else:

    print('Little to no autocorrelation', '\n')

    print('Assumption satisfied')

19. Homoskedastisitas

# Plotting the residuals

plt.subplots(figsize=(8, 4))

plt.scatter(x=data\_apbd.index, y=data\_apbd.residual, alpha=0.8)

plt.plot(np.repeat(0, len(data\_apbd.index)+2), color='darkorange', linestyle='--')

plt.ylabel('Residual', fontsize=14)

plt.xlabel('Kabupaten/Kota di Jawa timur', fontsize=14)

plt.title('Homescedasticity Assumption', fontsize=16)

plt.show()