



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com

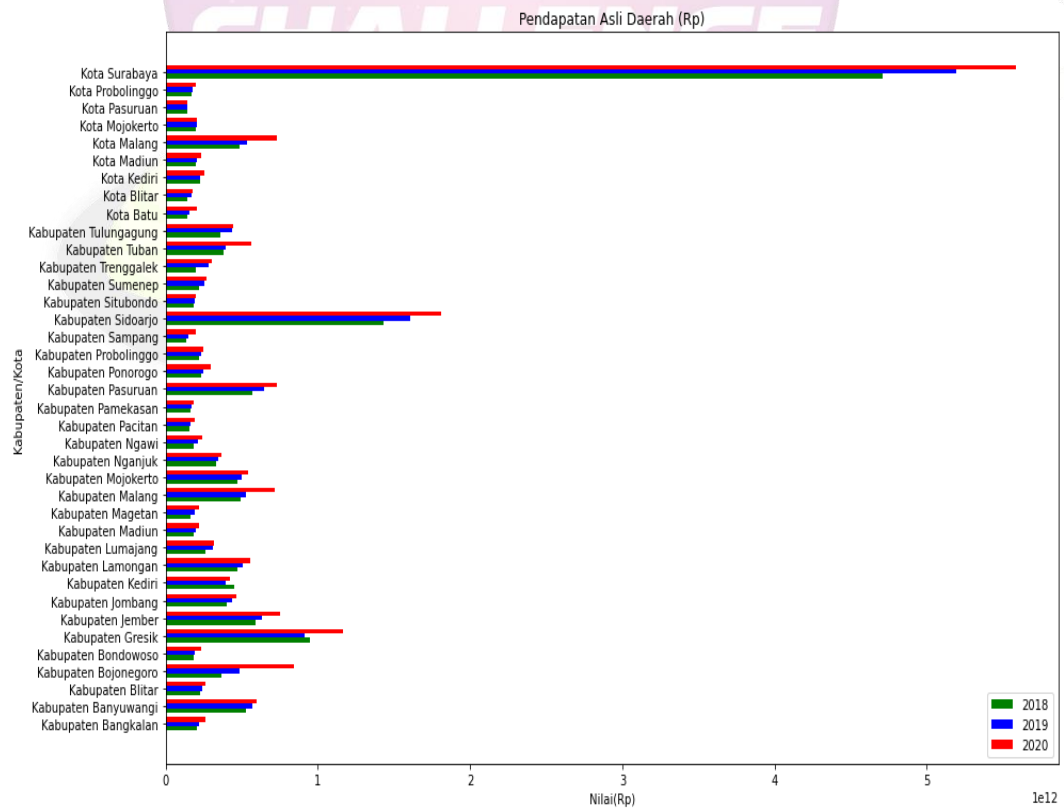


BIODATA UMUM NSC 2021

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

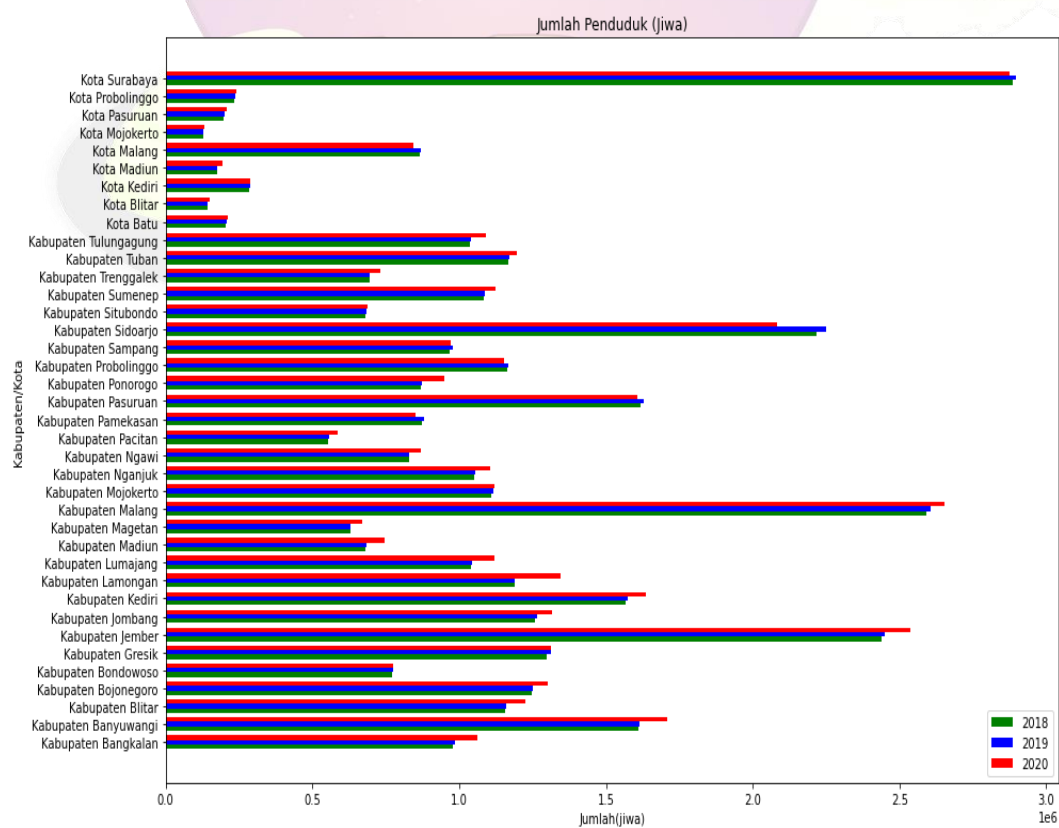
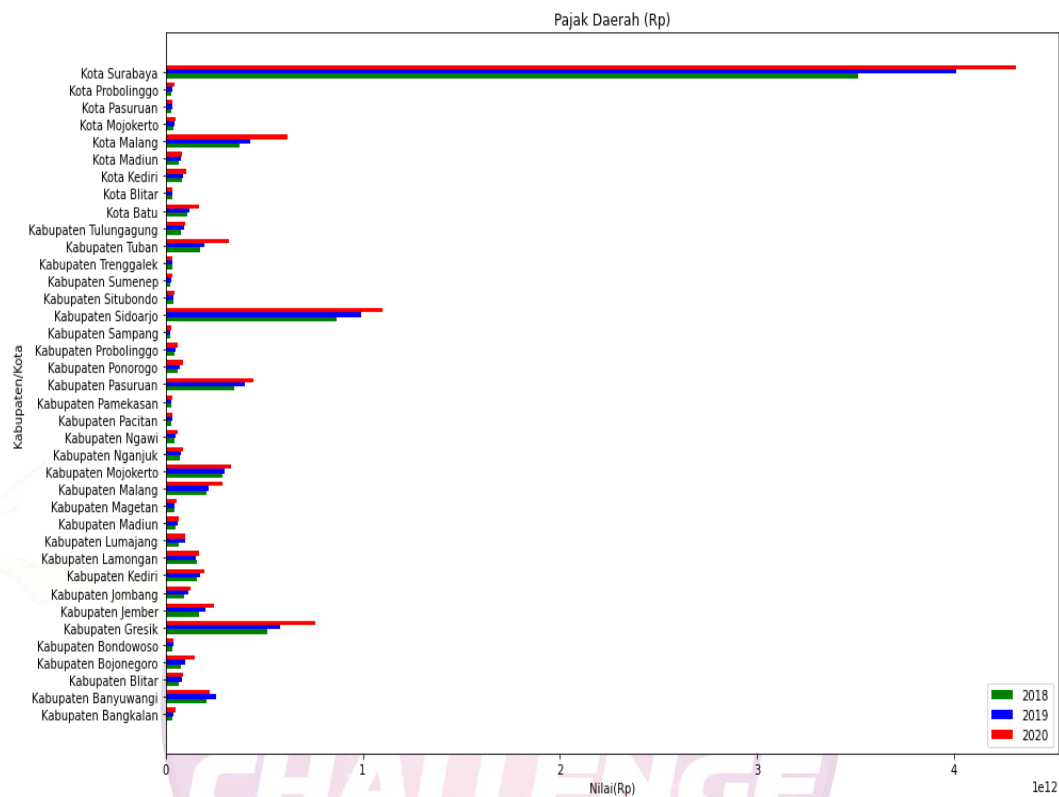
Berikut adalah diagram batang berdasarkan data dari file “Data Penyisihan NSC 2021.xlsx” yang dibagi menjadi 4 variabel yaitu “Pendapatan Asli Daerah”, “Pajak Daerah”, “Jumlah Penduduk” dan “Upah Minimum Kota/Kabupaten”.





KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

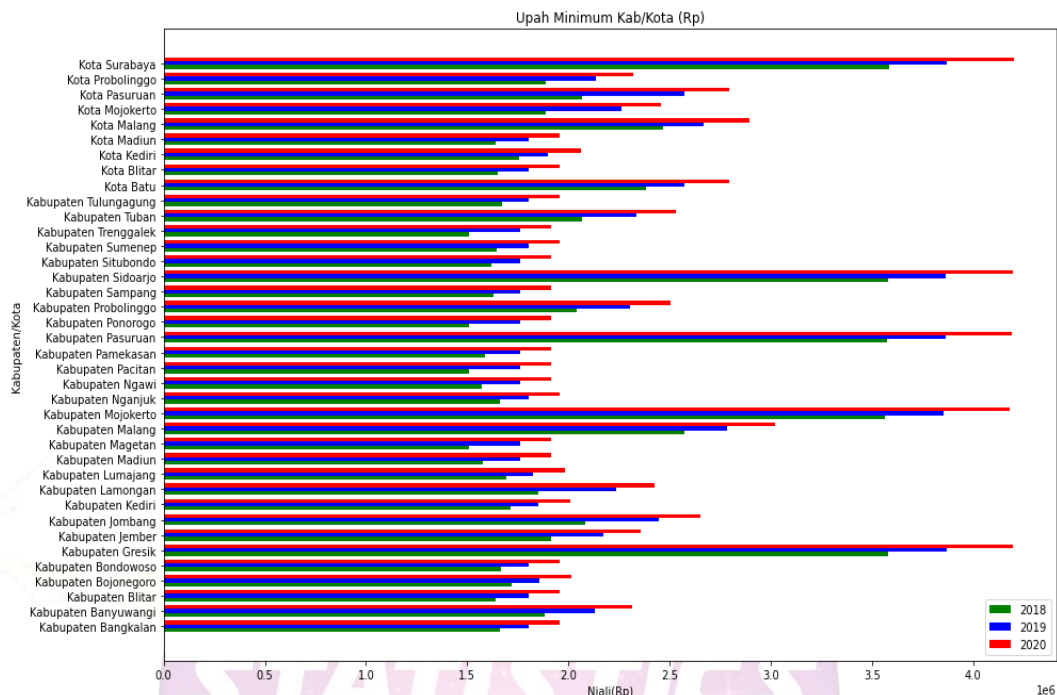
Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com





KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com



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 mengalokasikan 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

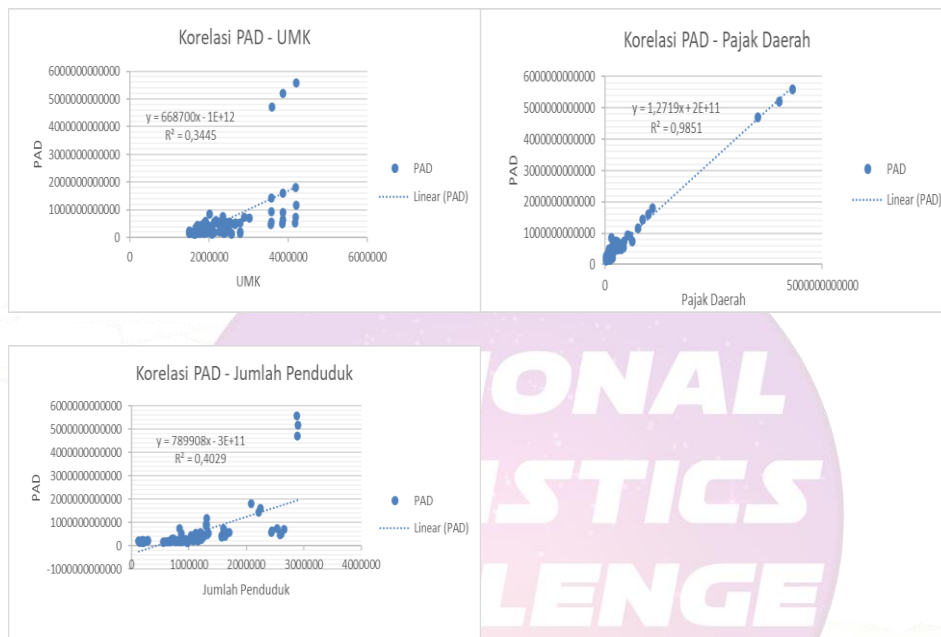


KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com



menggunakan scatter plot (tabel sebaran) dan mencari koefisien determinasi (R^2) untuk mengetahui nilai kekuatan hubungan antara variabel dependen dengan variabel-variabel independennya. Berikut adalah tabel sebaran antara variabel dependen dengan masing-masing variabel independennya.



Berdasarkan tabel di atas, didapat nilai R^2 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 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.



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com



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

$$\check{y}_1 = 130078180857.9956 + 1.23x_1 + 1.387x_2 - 4.43x_3$$

- Keterangan Variabel:

\check{y}_1 = Pendapatan Asli Daerah model ke-1

x_1 = Pajak Daerah (Rp)

x_2 = Jumlah Penduduk (Jiwa)

x_3 = UMK (Rp)

- Keterangan Angka:

Intersep: 130,078,180,857.9956

Koefisien: {1.23 , 1.387 , -4.43}

2. Tahun 2019

$$\check{y}_2 = 133169491380.92224 + 1.18x_1 + 1.55x_2 - 3.79x_3$$

- Keterangan Variabel:



\check{y}_2 = Pendapatan Asli Daerah model ke-2

x_1 = Pajak Daerah (Rp)

x_2 = Jumlah Penduduk (Jiwa)

x_3 = UMK (Rp)

- Keterangan Angka:

Intersep: 133169491380.92224

Koefisien: {1.18, 1.55, -3.79}

3. Tahun 2020

$$\check{y}_3 = 124289698935.15778 + 1.16x_1 + 2.11x_2 - 4.11x_3$$

- Keterangan Variabel:

\check{y}_3 = Pendapatan Asli Daerah model ke-3

x_1 = Pajak Daerah (Rp)

x_2 = Jumlah Penduduk (Jiwa)

x_3 = UMK (Rp)

- Keterangan Angka:

Intersep: 124289698935.15778

Koefisien: {1.16, 2.11, -4.11}

4. Tahun 2018 – 2020

$$\check{y}_4 = 115210250977.48486 + 1.21x_1 + 1.20x_2 - 1.87x_3$$

- Keterangan Variabel:

\check{y}_4 = Pendapatan Asli Daerah model ke-4

x_1 = Pajak Daerah (Rp)



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com



x_2 = Jumlah Penduduk (Jiwa)

x_3 = UMK (Rp)

- Keterangan Angka:

Intersep: 115210250977.48486

Koefisien: {1.21, 1.20, -1.87}

Uraian arti dari masing-masing angka tersebut adalah sebagai berikut (Pada \check{y}_1) :

- 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: Rp130,078,180,857.
- 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 (R^2) 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**

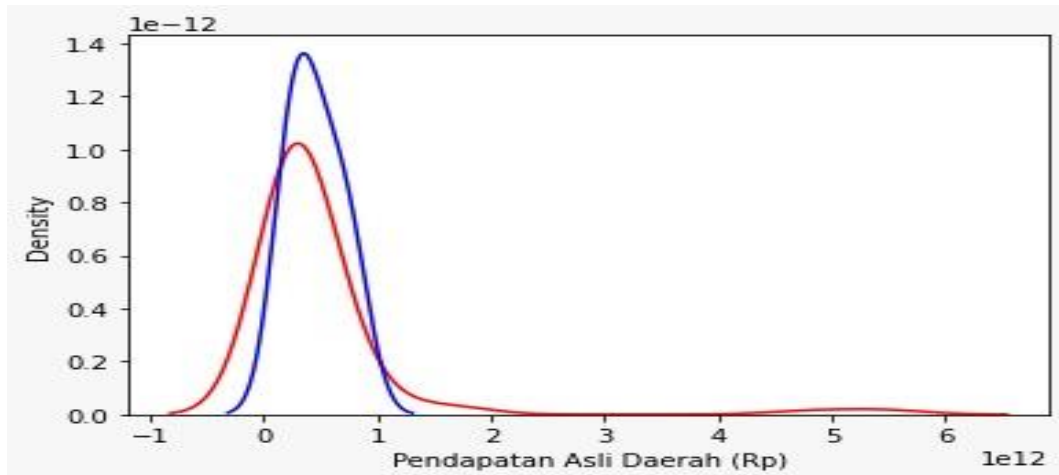
$R^2 = 0.9089256118570053$

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



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com



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**

$$R^2 = 0.8915587161116589$$

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**

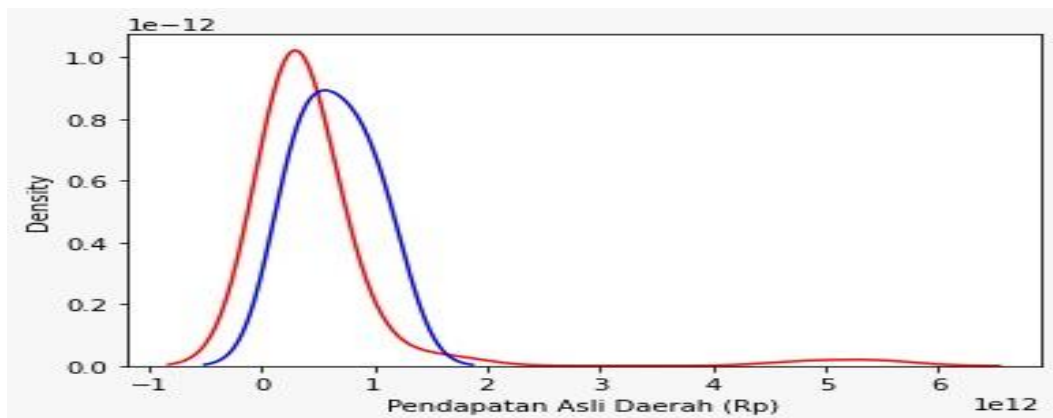
$$R^2 = 0.8821992739114471$$

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



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com



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**

$$R^2 = 0.9926471092133367$$

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 membandingkan 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:



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com



- OLS 2018

```
OLS Regression Results: 2018
X = data_2018[['Pajak Daerah (Rp)', 'Jumlah Penduduk (Jiwa)', 'Upah Minimum Kab/Kota (Rp)']]
X = sm.add_constant(X) # adding a constant
olsmod = sm.OLS(data_2018[['Pendapatan Asli Daerah (Rp)']], X).fit()
print(olsmod.summary())
✓ Bk.
```

OLS Regression Results						
Dep. Variable:	Pendapatan Asli Daerah (Rp)	R-squared:	0.992			
Model:	OLS	Adj. R-squared:	0.994			
Method:	Least Squares	F-statistic:	2044.			
Date:	Mon, 25 Aug 2021	Prob (F-statistic):	1.61e-38			
Time:	20:17:04	Log Likelihood:	-993.63			
No. Observations:	38	AIC:	1999.			
DF Residuals:	34	BIC:	2002.			
DF Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	1.07e+11	3.6e+10	2.999	0.005	3.4e+10	1.81e+11
Pajak Daerah (Rp)	1.3878	0.823	55.168	0.000	1.782	1.794
Jumlah Penduduk (Jiwa)	1.07e+05	1.03e+04	5.876	0.000	7.85e+04	1.45e+05
Upah Minimum Kab/Kota (Rp)	-2.071e+04	1.8e+04	-1.181	0.279	-5.84e+04	1.70e+04
Omnibus:	0.074	Durbin-Watson:	2.100			
Prob(Omnibus):	0.964	Jarque-Bera (JB):	0.826			
Skew:	0.004	Prob(JB):	0.987			
Kurtosis:	2.875	Cond. No.	2.33e+12			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.33e+12. This might indicate that there are strong multicollinearity or other numerical problems.

- OLS 2019

```
OLS Regression Results: 2019
X = data_2019[['Pajak Daerah (Rp)', 'Jumlah Penduduk (Jiwa)', 'Upah Minimum Kab/Kota (Rp)']]
X = sm.add_constant(X) # adding a constant
olsmod = sm.OLS(data_2019[['Pendapatan Asli Daerah (Rp)']], X).fit()
print(olsmod.summary())
✓ Bk.
```

OLS Regression Results						
Dep. Variable:	Pendapatan Asli Daerah (Rp)	R-squared:	0.994			
Model:	OLS	Adj. R-squared:	0.993			
Method:	Least Squares	F-statistic:	1845.			
Date:	Mon, 25 Aug 2021	Prob (F-statistic):	1.00e-37			
Time:	20:17:06	Log Likelihood:	-999.39			
No. Observations:	38	AIC:	2007.			
DF Residuals:	34	BIC:	2013.			
DF Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	1.27e+11	4.35e+10	2.929	0.006	3.9e+10	2.16e+11
Pajak Daerah (Rp)	1.2075	0.823	52.632	0.000	1.161	1.254
Jumlah Penduduk (Jiwa)	1.141e+05	2.11e+04	5.405	0.000	7.12e+04	1.57e+05
Upah Minimum Kab/Kota (Rp)	-2.502e+04	2.02e+04	-1.241	0.223	-6.6e+04	1.6e+04
Omnibus:	0.943	Durbin-Watson:	2.766			
Prob(Omnibus):	0.624	Jarque-Bera (JB):	0.730			
Skew:	0.332	Prob(JB):	0.694			
Kurtosis:	2.857	Cond. No.	2.74e+12			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.74e+12. This might indicate that there are strong multicollinearity or other numerical problems.

- OLS 2020

```
OLS Regression Results: 2020
X = data_2020[['Pajak Daerah (Rp)', 'Jumlah Penduduk (Jiwa)', 'Upah Minimum Kab/Kota (Rp)']]
X = sm.add_constant(X) # adding a constant
olsmod = sm.OLS(data_2020[['Pendapatan Asli Daerah (Rp)']], X).fit()
print(olsmod.summary())
✓ Bk.
```

OLS Regression Results						
Dep. Variable:	Pendapatan Asli Daerah (Rp)	R-squared:	0.999			
Model:	OLS	Adj. R-squared:	0.998			
Method:	Least Squares	F-statistic:	1035.			
Date:	Mon, 25 Aug 2021	Prob (F-statistic):	2.17e-33			
Time:	20:17:04	Log Likelihood:	-1015.5			
No. Observations:	38	AIC:	2035.			
DF Residuals:	34	BIC:	2042.			
DF Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	1.18e+11	6.44e+10	1.840	0.074	-1.24e+10	2.49e+11
Pajak Daerah (Rp)	1.1908	0.811	38.364	0.000	1.129	1.255
Jumlah Penduduk (Jiwa)	1.529e+05	2.97e+04	5.147	0.000	9.26e+04	2.13e+05
Upah Minimum Kab/Kota (Rp)	-2.446e+04	2.71e+04	-0.904	0.372	-7.95e+04	3.05e+04
Omnibus:	32.312	Durbin-Watson:	2.000			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	382.379			
Skew:	1.956	Prob(JB):	6.17e-23			
Kurtosis:	10.401	Cond. No.	3.05e+12			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 3.05e+12. This might indicate that there are strong multicollinearity or other numerical problems.

- OLS 2018 – 2020

```
OLS Regression Results: 2018-2020
X = data_ghd[['Pajak Daerah (Rp)', 'Jumlah Penduduk (Jiwa)', 'Upah Minimum Kab/Kota (Rp)']]
X = sm.add_constant(X) # adding a constant
olsmod = sm.OLS(data_ghd[['Pendapatan Asli Daerah (Rp)']], X).fit()
print(olsmod.summary())
✓ Bk.
```

OLS Regression Results						
Dep. Variable:	Pendapatan Asli Daerah (Rp)	R-squared:	0.992			
Model:	OLS	Adj. R-squared:	0.991			
Method:	Least Squares	F-statistic:	4295.			
Date:	Mon, 25 Aug 2021	Prob (F-statistic):	8.81e-114			
Time:	20:17:04	Log Likelihood:	-3017.0			
No. Observations:	114	AIC:	6042.			
DF Residuals:	110	BIC:	6053.			
DF Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	1.042e+11	2.75e+10	3.783	0.000	4.36e+10	1.59e+11
Pajak Daerah (Rp)	1.2092	0.815	79.511	0.000	1.179	1.239
Jumlah Penduduk (Jiwa)	1.246e+05	1.37e+04	9.112	0.000	9.75e+04	1.52e+05
Upah Minimum Kab/Kota (Rp)	-1.666e+04	1.22e+04	-1.317	0.191	-4.17e+04	8415.044
Omnibus:	59.483	Durbin-Watson:	2.435			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	369.917			
Skew:	1.612	Prob(JB):	4.72e-81			
Kurtosis:	11.213	Cond. No.	2.63e+12			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.63e+12. This might indicate that there are strong multicollinearity or other numerical problems.

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.



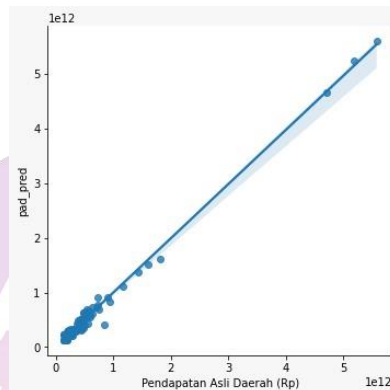
KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com



- **Uji Linearitas**

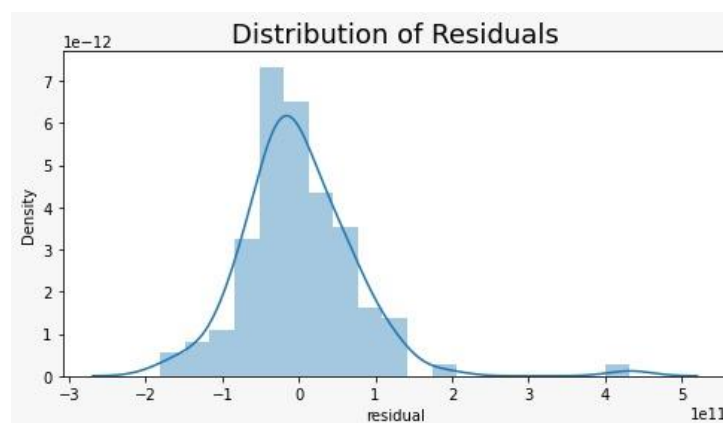
Ini mengasumsikan bahwa terdapat hubungan linier antara variabel bebas dan variabel terikat. Dalam kasus kami karena kami memiliki beberapa variabel independen, kami dapat melakukan ini dengan menggunakan plot pencar untuk melihat nilai prediksi kami versus nilai sebenarnya.



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.





KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

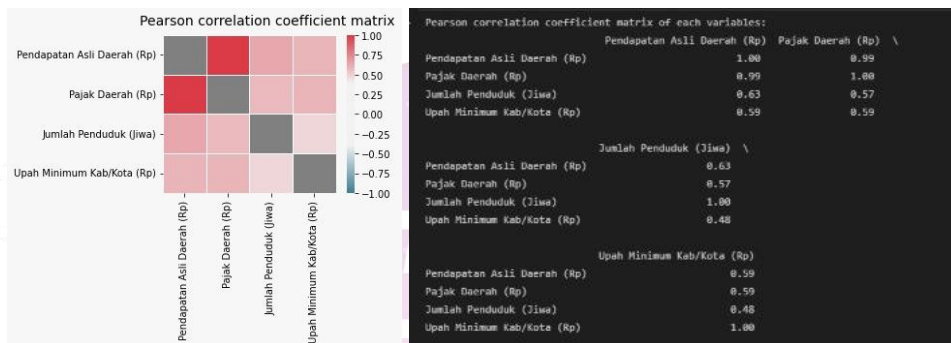
Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com



Uji Normalitas menunjukkan 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.



Berdasarkan data peta panas tersebut, **kita mendapatkan hasil bahwa hubungan antara variabel dependen PAD dengan variabel-variabel independennya memiliki koefisien 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.

```
Durbin-Watson: 2.4350374619758997
Little to no autocorrelation

Assumption satisfied
```




KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

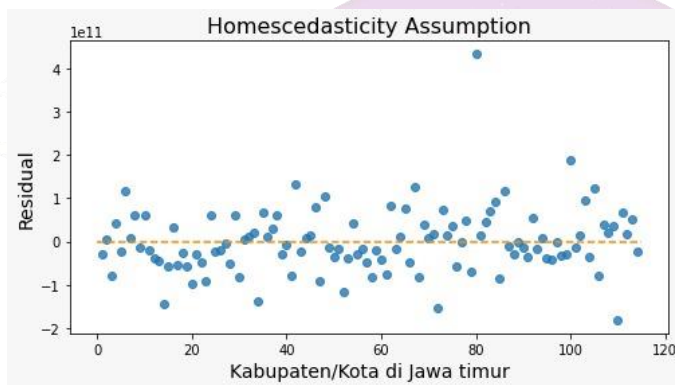
Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com



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.



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 R^2 sebesar 99%, yang berarti ketiga variabel independen di atas menjadi faktor yang sangat memengaruhi Pendapatan Asli Daerah di Provinsi Jawa Timur dari Tahun 2018 – 2020.

$$\hat{y}_4 = 115210250977.48486 + 1.21x_1 + 1.20x_2 - 1.87x_3$$

Maka dapat disimpulkan bahwa model persamaan linear berganda tersebut sudah menjelaskan hubungan dari variabel dependen dan variabel independennya. 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.



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com



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['P
ajak Daerah (Rp)']), np.array(df_2018['Jumlah Penduduk (Jiwa)']), np.array(df_20
18['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['P
ajak Daerah (Rp)']), np.array(df_2019['Jumlah Penduduk (Jiwa)']), np.array(df_20
19['Upah Minimum Kab/Kota (Rp)'])
```



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com



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['Pa
jak Daerah (Rp)']), np.array(df_2020['Jumlah Penduduk (Jiwa)']), np.array(df_202
0['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 Banyuw
angi', 'Kabupaten Blitar', 'Kabupaten Bojonegoro', 'Kabupaten Bondowoso', 'Kabupat
en Gresik', 'Kabupaten Jember', 'Kabupaten Jombang', 'Kabupaten Kediri', 'Kabupate
n Lamongan', 'Kabupaten Lumajang', 'Kabupaten Madiun', 'Kabupaten Magetan', 'Ka
bupaten Malang', 'Kabupaten Mojokerto', 'Kabupaten Nganjuk', 'Kabupaten Ngawi', '
Kabupaten Pacitan', 'Kabupaten Pamekasan', 'Kabupaten Pasuruan', 'Kabupaten Pon
orogo', 'Kabupaten Probolinggo', 'Kabupaten Sampang', 'Kabupaten Sidoarjo', 'Kabu
paten Situbondo', 'Kabupaten Sumenep', 'Kabupaten Trenggalek', 'Kabupaten Tuban
', 'Kabupaten Tulungagung', 'Kota Batu', 'Kota Blitar', 'Kota Kediri', 'Kota Madiun', 'K
ota 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 Banyuw
angi', 'Kabupaten Blitar', 'Kabupaten Bojonegoro', 'Kabupaten Bondowoso', 'Kabupat
en Gresik', 'Kabupaten Jember', 'Kabupaten Jombang', 'Kabupaten Kediri', 'Kabupate
n Lamongan', 'Kabupaten Lumajang', 'Kabupaten Madiun', 'Kabupaten Magetan', 'Ka
bupaten Malang', 'Kabupaten Mojokerto', 'Kabupaten Nganjuk', 'Kabupaten Ngawi', '
Kabupaten Pacitan', 'Kabupaten Pamekasan', 'Kabupaten Pasuruan', 'Kabupaten Pon
orogo', 'Kabupaten Probolinggo', 'Kabupaten Sampang', 'Kabupaten Sidoarjo', 'Kabu
paten Situbondo', 'Kabupaten Sumenep', 'Kabupaten Trenggalek', 'Kabupaten Tuban
```



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com



```
','Kabupaten Tulungagung','Kota Batu','Kota Blitar','Kota Kediri','Kota Madiun','K  
ota 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 Banyuw  
angi','Kabupaten Blitar','Kabupaten Bojonegoro','Kabupaten Bondowoso','Kabupat  
en Gresik','Kabupaten Jember','Kabupaten Jombang','Kabupaten Kediri','Kabupate  
n Lamongan','Kabupaten Lumajang','Kabupaten Madiun','Kabupaten Magetan','Ka  
bupaten Malang','Kabupaten Mojokerto','Kabupaten Nganjuk','Kabupaten Ngawi','  
Kabupaten Pacitan','Kabupaten Pamekasan','Kabupaten Pasuruan','Kabupaten Pon  
orogo','Kabupaten Probolinggo','Kabupaten Sampang','Kabupaten Sidoarjo','Kabu  
paten Situbondo','Kabupaten Sumenep','Kabupaten Trenggalek','Kabupaten Tuban  
,','Kabupaten Tulungagung','Kota Batu','Kota Blitar','Kota Kediri','Kota Madiun','K  
ota 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 Banyuw  
angi','Kabupaten Blitar','Kabupaten Bojonegoro','Kabupaten Bondowoso','Kabupat  
en Gresik','Kabupaten Jember','Kabupaten Jombang','Kabupaten Kediri','Kabupate  
n Lamongan','Kabupaten Lumajang','Kabupaten Madiun','Kabupaten Magetan','Ka  
bupaten Malang','Kabupaten Mojokerto','Kabupaten Nganjuk','Kabupaten Ngawi',
```



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com



```
Kabupaten Pacitan','Kabupaten Pamekasan','Kabupaten Pasuruan','Kabupaten Pon  
orogo','Kabupaten Probolinggo','Kabupaten Sampang','Kabupaten Sidoarjo','Kabu  
upaten Situbondo','Kabupaten Sumenep','Kabupaten Trenggalek','Kabupaten Tuban  
,','Kabupaten Tulungagung','Kota Batu','Kota Blitar','Kota Kediri','Kota Madiun','K  
ota 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)', 'U  
pah Minimum Kab/Kota (Rp)'], y_vars=['Pendapatan Asli Daerah (Rp)'], kind='re  
g')
```

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 Ka  
b/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_
```



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com



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)
```




KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com



```
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 Pe
nduduk (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')
```



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
STUDIO STATISTIKA

Sekretariat: Gedung Graha Santa Lantai 1 Universitas Brawijaya - Jl. Veteran, Malang
E-mail: studiostatistika@gmail.com



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('Homoscedasticity Assumption', fontsize=16)
plt.show()
```

