

The Relation Between Personal Outcome Index (W875RX1) and CPI Index to Manufacturing Index in USA

A correlation study using linear regression & Decision Tree

Your Nme Here

Your Major & Study Here

Your Campus Here

Datasets

Dataset yang digunakan dalam tugas ini berasal dari Federal Reserve Economic Data USA. Dimana data yang digunakan adalah indeks pengeluaran personal (W875RX1), CPI Index (CPILFESL) dan Indeks Produksi Manufaktur (IPMAN) dari tahun 2000 sampai 2007.

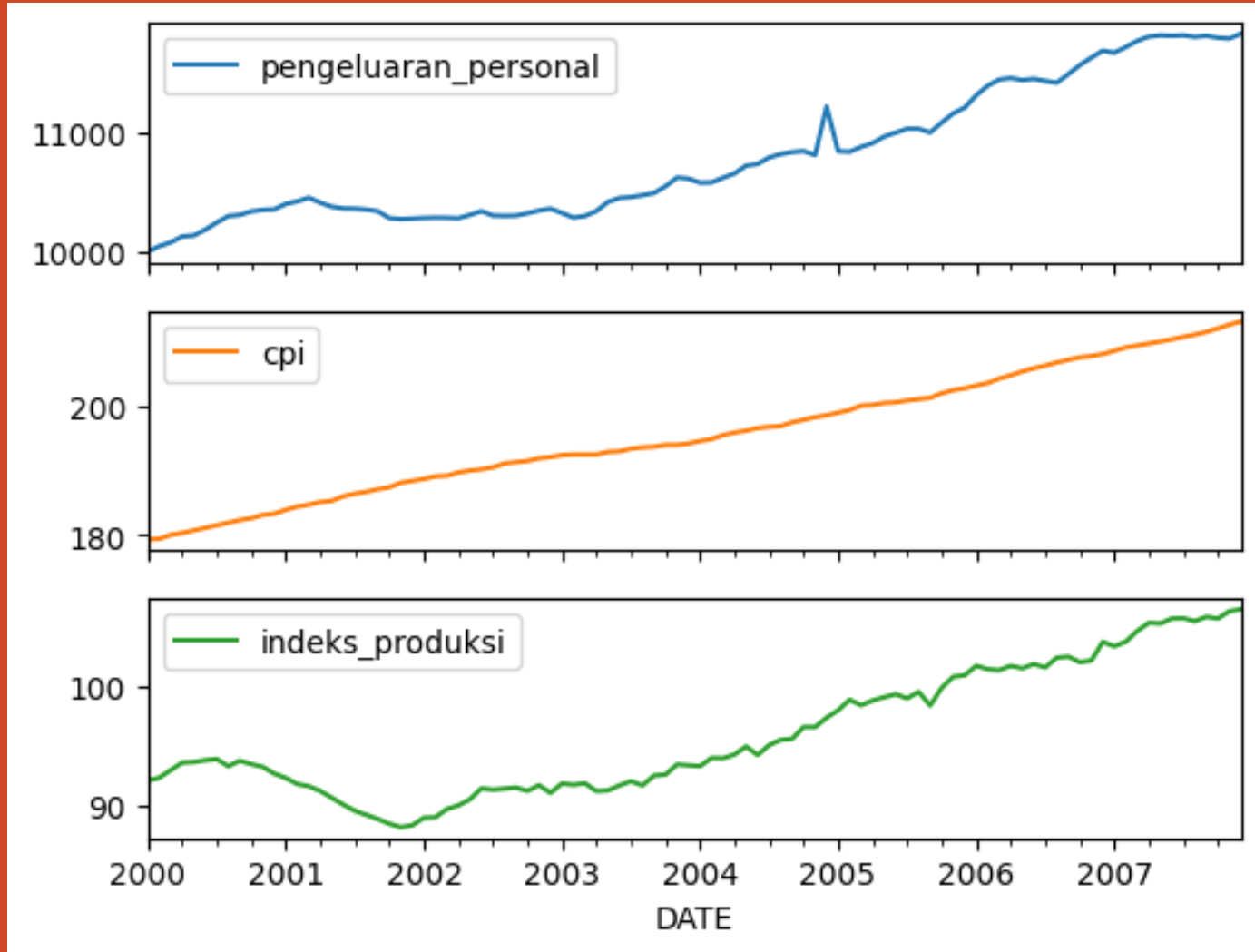
	pengeluaran_personal	cpi	indeks_produksi
DATE			
2000-01-01	10000.3	179.300	92.0231
2000-02-01	10045.9	179.400	92.2694
2000-03-01	10078.6	180.000	92.9096
2000-04-01	10127.6	180.300	93.5374
2000-05-01	10135.7	180.700	93.5876

Deskripsi Statistik

Adapun hasil dari deskripsi statistic dari setiap data yang kita miliki adalah sebagai berikut:

	pengeluaran_personal	cpi	indeks_produksi
count	96.000000	96.000000	96.000000
mean	10791.695833	195.654177	95.902233
std	556.843391	9.425305	5.368942
min	10000.300000	179.300000	88.119000
25%	10334.350000	188.625000	91.624400
50%	10595.950000	194.400000	93.789850
75%	11240.000000	202.900000	100.981500
max	11825.300000	213.168000	106.420200

Eksplanatory



Uji Korelasi

Dari hasil visualisasi grafik sebelumnya, dapat dilihat secara sekilas bahwa baik pengeluaran personal maupun indeks CPI memiliki korelasi yang positif terhadap indeks produksi manufaktur.

Namun untuk memastikannya, kita perlu melakukan uji korelasi. Uji yang digunakan dalam tugas ini adalah korelasi pearson.

	pengeluaran_personal	cpi	indeks_produksi
pengeluaran_personal	1.000000	0.950695	0.957423
cpi	0.950695	1.000000	0.887779
indeks_produksi	0.957423	0.887779	1.000000

Regresi Linear

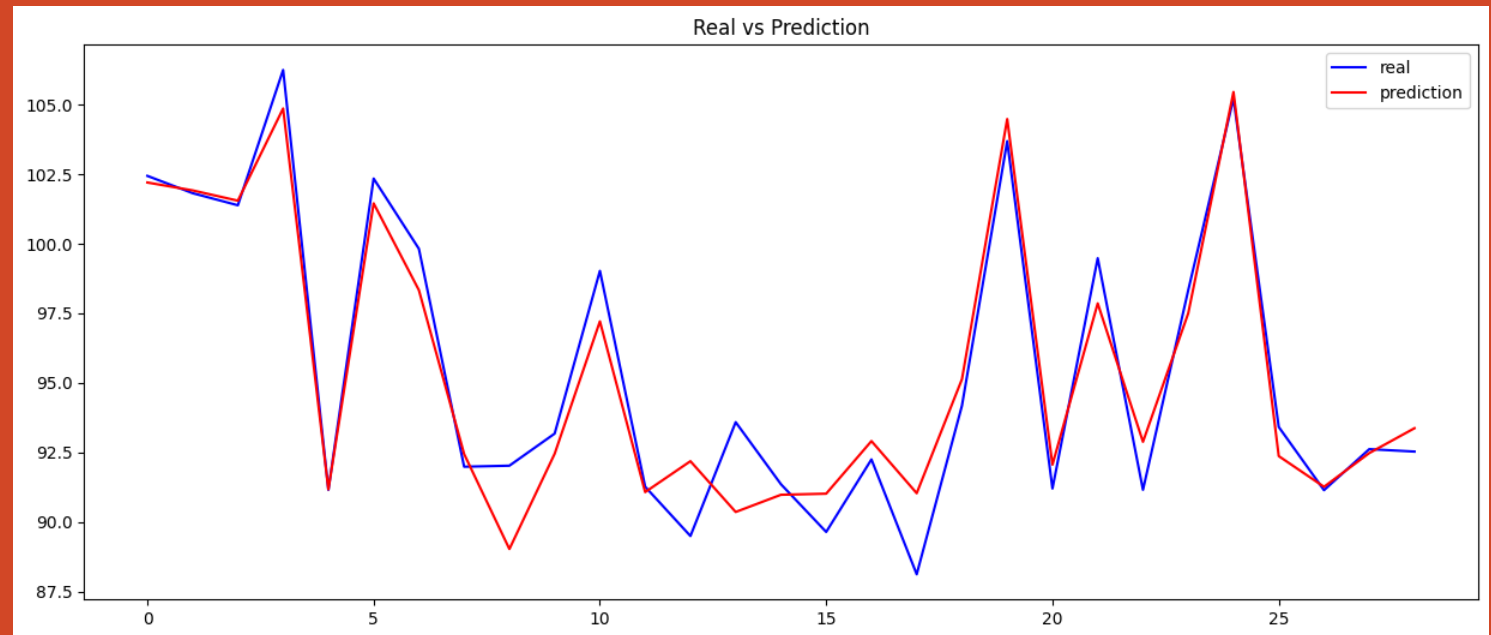
Langkah berikutnya adalah melakukan uji regresi linear untuk mendapatkan fungsi regresi linear berganda. Dimana variable bebas yang digunakan adalah indeks pengeluaran personal dan CPI index. Sementara variable terikatnya adalah Indeks Produksi Manufaktur.

Dari pengujian regresi linear, didapati rumus regresi linier sebagai berikut:

$$y = -0.590792 + 0.01102665 * x_1 - 0.11517985 * x_2$$

Pengujian

Hasil pengujian regresi linear adalah sebagai berikut:



MSE = 1.9705631509522659

MAE = 1.0647484806352407

$R^2 = 0.9289301803065451$

Datasets

Klasifikasi

Kemudian membuat dummy class dimana data dengan indeks produksi manufaktur berada diatas rata-rata akan diberi label '1', sementara sisanya akan diberi label '0'.

DATE	pengeluaran_personal	cpi	indeks_produksi	label
2000-01-01	10000.3	179.300	92.0231	0
2000-02-01	10045.9	179.400	92.2694	0
2000-03-01	10078.6	180.000	92.9096	0
2000-04-01	10127.6	180.300	93.5374	0
2000-05-01	10135.7	180.700	93.5876	0

Datasets

Klasifikasi

Kemudian membuat dummy class dimana data dengan indeks produksi manufaktur berada diatas rata-rata akan diberi label '1', sementara sisanya akan diberi label '0'.

DATE	pengeluaran_personal	cpi	indeks_produksi	label
2000-01-01	10000.3	179.300	92.0231	0
2000-02-01	10045.9	179.400	92.2694	0
2000-03-01	10078.6	180.000	92.9096	0
2000-04-01	10127.6	180.300	93.5374	0
2000-05-01	10135.7	180.700	93.5876	0

Decision Tree Classifier

Model klasifikasi yang digunakan adalah decision tree classifier dengan akurasi mencapai 100%. Adapun metric pengujian lainnya dapat dilihat pada table berikut:

Label	Precision	Recall	F1-Score
1	100%	100%	100%
0	100%	100%	100%

Hasil pengujian k-fold cross validation dengan $k = 10$ juga menghasilkan akurasi 100% pada setiap fold nya.

[illegible]