

# Penerapan Estimasi: Linear Regression Ganda Untuk Memprediksi Hasil Nilai Kuesioner Mahasiswa

Rizka Irijiba

Fakultas Rekayasa Sistem, Universitas Teknologi Sumbawa

Email: [rizkairjibakyungsoo@gmail.com](mailto:rizkairjibakyungsoo@gmail.com)

**Abstrak:** Salah satu aktivitas yang rutin diselenggarakan pada proses akademik adalah pengisian kuesioner yang dilakukan mahasiswa. Salah satu item dalam kuesioner adalah umpan balik untuk dosen dalam melaksanakan proses pembelajaran. Mahasiswa mengisi form yang sudah disediakan. Dan hasil kuesioner ini akan menghasilkan nilai kinerja dosen dari sisi kuesioner mahasiswa dengan rentangangka 0-4. Penelitian yang ini melakukan pemrosesan analisis prediksi dengan menggunakan regresi linear. Datasheet diolah sesuai dengan data-data yang diperlukan. Data yang digunakan merupakan data hasil kuesioner dari data tahun 2016-2020. Datasheet terdiri dari 7 kolom dan 800 baris. Dari datasheet akan dilakukan pemisahan data. Dari data yang ada, sebanyak 80% akan digunakan untuk data training dan data test sebanyak 20%. Penelitian diawali dengan analisis sistem dengan melakukan pencarian dan pengumpulan datasheet, desain sistem dilakukan dengan pemisahan data-data yang tidak diperlukan serta koding dan implementasi. Implementasi dalam prediksi hasil kuesioner akan menggunakan Python dengan menggunakan berbagai library, diantaranya Pandas yang digunakan analisis data, Matplotlib digunakan visualisasi data dan Scikit-Learn digunakan dalam proses machine learning. Penggunaan Python dengan dukungan library memberikan kemudahan dalam implementasi machine learning khusus regresi linear.

**Kata Kunci:**

Regresi linear, Python, Data nilai

*Abstract: One activity that is routinely held in the academic process is filling out questionnaires by students. One of the items in the questionnaire is feedback for lecturers in carrying out the learning process. Students fill out the form provided. And the results of this questionnaire will produce lecturer performance scores from the student questionnaire side with a range of 0-4. This research performs predictive analysis processing using linear regression. The datasheet is processed according to the required data. The data used is data from a questionnaire from the 2016-2020 data. The datasheet consists of 7 columns and 800 rows. From the datasheet data processing will be carried out. From the existing data, as much as 80% will be used for training data and 20% for test data. The research begins with system analysis by searching and collecting data sheets, system design is done by purifying unnecessary data as well as coding and implementation. Implementation in predicting the results of the questionnaire will use Python using various libraries, including Pandas which is used for data analysis, Matplotlib is used for data visualization and Scikit-Learn is used in the machine learning process. Using Python with library support makes it easy to implement machine learning, especially linear regression.*

**Keywords:**

Linear regression, python, value data

## 1 PENDAHULUAN

Data science merupakan salah satu ilmu yang kini semakin dibutuhkan di perusahaan. Posisi ini membantu perusahaan untuk mengembangkan bisnis berdasarkan data yang dimiliki. Salah satu kelebihan dari data science adalah mampu memprediksi masa depan secara akurat, efektif, dan efisien. Tentunya hal ini dapat dilakukan dengan bantuan beragam algoritma yang ada. Algoritma ini akan membaca data yang dimiliki dan memodelkannya sesuai tujuan yang diinginkan. Algoritma data science ada beragam. Terbagi menjadi supervised dan unsupervised learning. Jenis ini bergantung dari kondisi data dan tujuan yang ingin dicapai. Salah satu algoritma yang cukup populer dan mudah diaplikasikan adalah linear regression dalam estimasi.

Linear regression adalah salah satu algoritma yang digunakan data science dan tergolong pada algoritma supervised learning. Algoritma ini menggunakan prinsip regresi. Regresi membuat model prediksi untuk target variabel berdasarkan dari variabel bebasnya. Jenis algoritma ini sering digunakan untuk mencari hubungan antara variabel-variabel yang ada dan prediksinya.

Perguruan tinggi sebagai lembaga pendidikan yang mndidik mahasiswa menajdi lulusan bermutu harus selalu memperhatikan sistem serta proses pendidikan. Sistem dan proses yang berjalan dengan baik tentunya akan menghasilkan produk yang berkualitas. Salah satu faktor utama agar sistem dan proses berjalan dengan baik adalah dosen. Dalam proses melaksanakan fungsi dan tugas pendidikan dengan baik, dosen harus memahami teori-teori pendidikan serta metode yang digunakan dalam proses perkuliahan, Dosen dalam proses perkuliahan harus menjadi sumber belajar, fasilitator dan mediator serta evaluator dalam menentukan output perguruan tinggi. Di samping unsur pendidikan, kinerja dosen juga dinilai dari peran aktif dalam menghasilkan penelitian dan pelaksanaan pengabdian pada masyarakat.

Di setiap perguruan tinggi parameter penilaian yang tidak sama tetapi mempunyai maksud dan tujuan yang sama yaitu menilai kinerja dosen dari sisi proses perkuliahan, melakukan evaluasi terhadap penilaian kuesioner yang diisi mahasiswa. Kuesioner diberikan dengan sifat tertutup dan berdasar hasil kuesioner tersebut akan dinilai kepuasan mahasiswa terhadap dosen, melakukan analisis hasil kuesioner yang diisi mahasiswa, khususnya proses pembelajaran yang berlangsung. Salah satu hasil penelitian adalah adanya korelasi antara nilai kuesioner yang diisi mahasiswa dengan prestasi mahasiswa. Variabel yang diteliti

diantaranya adalah kedisiplinan dosen selama proses perkuliahan. Prestasi mahasiswa meningkat dapat dipengaruhi dengan nilai yang tinggi yang dicapai dosen . Hal ini menunjukkan jika proses perkuliahan berjalan dengan baik, maka prestasi mahasiswa dalam perkuliahan juga baik.

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik daripada kearah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai variabel dari target sebagai nilai prediksi, selanjutnya pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Dengan menggunakan metode estimasi yang ada pada data mining yang sesuai dengan keperluan perusahaan biasanya yaitu untuk mengetahui tingkat kekuatan beton yang nantinya akan dihasilkan. Menggunakan linear regression yang dimana nanti dengan memanfaatkan data yang sudah ada pada suatu perusahaan, maka nanti untuk menghasilkan persamaan yang digunakan untuk menghitung tingkat kekuatan beton yang nantinya akan dihasilkan.

Regresi linier adalah satu satu jenis analisis peramalan yang paling sering digunakan pada data kuantitatif yang lebih tepatnya jenis tipe data interval atau rasio. Adapun tujuan dilakukannya analisis regresi linier adalah apabila seperangkat variabel prediktor signifikan dalam memprediksi variabel respon, dan variabel prediktor manakah yang memiliki signifikan dalam menjelaskan variabel respon. Hal ini ditunjukkan dengan koefisien estimasi regresi. Dengan analisis regresi kamu dapat menunjukkan hasil rata-rata dan nilai hubungan variabel yang dihitung berdasarkan variabel bebasnya, serta menguji hipotesis yang ingin diketahui hasilnya oleh penguji.

## 2 PENELITIAN YANG BERHUBUNGAN

Pada tabel dibawah ini mengenai beberapa penelitian yang serupa diengkapi dengan metode dan hasil penelitiannya, yaitu sebagai berikut:

Tabel 1 Penelitian yang berhubungan

No	Nama Peneliti	Judul Penelitian dan Tahun Penelitian	Metode	Hasil Penelitian
1	Robi Yanto	Implementasi Data Mining Estimasi Ketersediaan Lahan Pembuangan Sampah menggunakan Algoritma Regresi Linear, Vol. 2 No. 1 (2018) 361 – 366.	Metode yang digunakan dalam penelitian ini dilakukan analisa data terhadap estimasi ketersediaan lahan pembuangan sampah dalam jangka panjang dengan menggunakan teknik data mining dengan algoritma regresi linear.	Penelitian ini memperoleh hasil prediksi yang sama yaitu prediksi sampah pada tahun 2025 adalah 4.853,967 ton dengan jumlah penduduk tahun 2025 sebesar 201484 Jiwa.

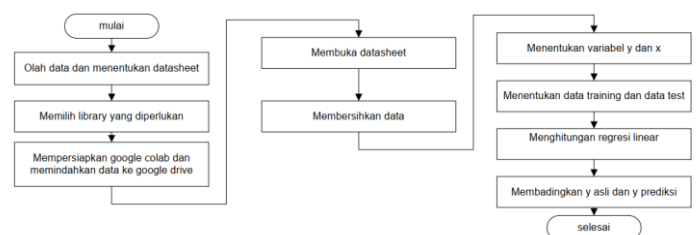
2	Tyas Setiyor ini, Rizky Tri Asmono	Komparasi Metode Neural Network, Support Vector Machine Dan Linear Regression Pada Estimasi Kuat Tekan Beton, Vol. 15, No. 1 Maret 2018.	Penelitian ini melakukan komparasi antara metode Neural Network (NN), Support Vector Machine (SVM) dan Linear Regression (LR) dengan menggunakan dataset concrete compressive strength dan slump.	Hasil penelitian pada dataset concrete compressive strength dengan menggunakan metode NN didapatkan RMSE 5,667, dengan menggunakan metode SVM didapatkan RMSE 5,165 dan dengan metode LR didapatkan RMSE 10,501.
---	------------------------------------	--	---	--

## 3 METODE PENELITIAN

Dalam menganalisis data pada penelitian ini, yaitu menggunakan CRISP-DM (Cross-Industry Standard Process for Data Mining). Dalam data mining, ada beberapa model teknik analisis data yang bisa diterapkan oleh praktisi data. Salah satunya adalah model CRISP-DM yang merupakan standardisasi data mining. Berdasarkan jajak pendapat yang dilakukan oleh situs datascience-pm, CRISP-DM masuk di urutan pertama dengan model data science yang seringkali digunakan oleh praktisi data dengan perolehan nilai sebanyak 49 persen.

Hasil survei “Penggunaan Metodologi dalam Proyek Data Mining”, memperlihatkan pengguna CRISP-DM di tahun 2002 mencapai 51%, kemudian menurun menuju 41% di tahun 2004. Meskipun persentasi penggunaan CRISP-DM menurun 10%, jumlah pengguna metodologi ini masih terbilang lebih banyak daripada pengguna metodologi lain. CRISP-DM memiliki 6 tahapan dalam melakukan analisis data mining.

Langkah penelitian diawali dengan pengumpulan data hasil kuesioner dandisimpan dataset dalam bentuk file csv. Sebelum dilakukan pengolahan pada dataset, library yang diperlukan dilakukan proses instalasi. Dari library yang diperlukan dan sudah dipasang, proses dilakukan proses pengecekan datasheet dari data-data yang mengandung kesalahan seperti data yang kosong. Datasheet yang sudah tidak ada kesalahan dan untuk memvari regresi linear, datasheet akan dibedakan menjadi data x dan data y serta menentukan data yang menjadi data training dan data yang menjadi data test. Hasil dari data training akan menghasilkan rumus regresi linear dan untuk menguji prediksi menggunakan data test. Hasil data tes ini akan ditemukan persentasi keakuratan rumus regresi linear. Alur penelitian lebih detail ada pada Gambar 1 di bawah ini.



Gambar 1 Alur Penelitian

## 4 HASIL DAN PEMBAHASAN

### 4.1 Business Understanding

Kita masuk ke dalam tahap pertama dari model CRISP-DM. Tahapan yang pertama ini bisa dibilang sangat vital karena praktisi data butuh pijakan dasar untuk melakukan analisis data. Salah satunya adalah business knowledge ataupun pengetahuan dari segi objek bisnis. Dalam tahap ini, praktisi data akan mulai belajar bagaimana membangun atau mendapatkan data, bagaimana untuk mencocokkan tujuan permodelan untuk tujuan bisnis sehingga model terbaik dapat dibangun.

Adapun kegiatan dalam business understanding ini mulai dari menentukan tujuan, menentukan fenomena dan persyaratan dengan jelas secara keseluruhan, menerjemahkan tujuan tersebut serta menentukan pembatasan dalam perumusan masalah yang diaplikasikan dalam data mining dan selanjutnya mempersiapkan strategi awal untuk mencapai tujuan tersebut.

### 4.2 Data Understanding

Tahap kedua adalah data understanding. Secara garis besar, data understanding dipakai untuk memeriksa data sehingga dapat mengidentifikasi masalah pada data yang kita dapatkan. Tahapan ini memberikan pondasi analitik untuk sebuah penelitian dengan membuat ringkasan (summary) dan mengidentifikasi potensi masalah dalam data. Tahap ini juga harus dilakukan secara cermat dan tidak terburu-buru oleh praktisi data. Contoh visualisasi data yang dilakukan oleh praktisi data. Biasanya kalau tidak berhati-hati maka insight atau kesimpulan juga tidak bisa langsung ditemukan.

Apalagi kalau tidak dihubungkan dengan summary datanya. Apabila praktisi data sedang mengalami masalah ini maka ada gangguan pada tahap modelling. Maka dari itu, praktisi data perlu jeli juga untuk mengamati tahapan dalam membuat visualisasi data dan segera menemukan insight dari data yang akan atau sudah kita visualisasikan. Ringkasan dari data dapat berguna untuk mengonfirmasi apakah data terdistribusi seperti yang diharapkan. Biasanya permasalahan dalam data yang sering kali ditemui oleh praktisi data adalah nilai-nilai yang hilang, outlier, berdistribusi spike, berdistribusi bimodal, missing value dan lain-lain.

### 4.3 Data Preparation

Lanjut ada data preparation. Secara garis besar, data preparation di kalangan data mining dipakai untuk memperbaiki masalah dalam data, kemudian membuat variabel turunan. Tahap data preparation sangat jelas untuk membutuhkan pemikiran yang cukup matang dan usaha yang cukup tinggi untuk memastikan data tepat sesuai dengan algoritma yang dipakai.

Proses data preparasi merupakan proses data treatment menuju model berkualitas yang berguna. Tahapan ini adalah yang paling menguras resources dari tim analisis. Model yang baik dan akurat berawal dari data preparasi yang baik. Beberapa hal yang umum dilakukan pada tahapan di bawah ini adalah:

- Melakukan pengecekan kembali pada kebenaran data  
Pengecekan pada data perlu di desain bertingkat sehingga akuntabilitas terjaga. Pengecekan juga diperlukan terhadap konsistensi inputing data. System yang baik dalam pengumpulan data antara lain

menggunakan default akan dapat menjaga konsistensi data.

- Mengelola data outlier  
Data Outlier perlu dikelola dengan baik. Data Outlier dapat berupa Univariate Outlier, dan Multivariate Outlier serta dapat berada pada variable dependent maupun variable independent. Data Mining untuk tujuan generalisir akan terpengaruh dengan dengan Data Outlier sehingga perlu dinetralisir. Sebelum melakukan treatment atas data Outlier alangkah baiknya terlebih dahulu dilakukan pengecekan terhadap pengambilan dan pengisian data.
- Memberlakukan data missing dan data inkonsistensi  
Perlakuan terhadap data missing harus match dengan tujuan dari data mining itu sendiri. Misalkan data missing yang di isi dengan average mungkin masih dapat diterima untuk tujuan prediction dan forecasting, namun untuk klastering bisa jadi akan mengarahkan pada kelompok yang kurang tepat. Sebaliknya menggunakan data yang sering muncul untuk mengisi missing data untuk multi variable data mining akan berpengaruh pada hasil untuk tujuan prediction dan forecasting

### 4.4 Modelling

Tahap keempat yaitu Modelling. Secara garis besar untuk membuat model prediktif atau deskriptif. Pada tahap ini dilakukan metode statistika dan Machine Learning untuk penentuan terhadap teknik data mining, alat bantu data mining, dan algoritma data mining yang akan diterapkan. Lalu selanjutnya adalah melakukan penerapan teknik dan algoritma data mining tersebut kepada data dengan bantuan software.

Jika diperlukan penyesuaian data terhadap teknik data mining tertentu, dapat kembali ke tahap data preparation. Beberapa teknik modeling yang biasa dilakukan dan dipakai oleh praktisi data antara lain classification, scoring, ranking, clustering, finding relation, dan characterization.

### 4.5 Evaluation

Tahap kelima yaitu Evaluation. Setelah didapatkan sebuah atau beberapa model sehingga dilakukan penilaian terkait kualitas dan efektifitas-nya. Kemudian ditentukan model seperti apa yang digunakan agar sesuai dengan objective pada fase 1 hingga diambil sebuah keputusan penggunaan dari hasil data mining.

### 4.5 Deployment

Tahap terakhir dalam model CRISP-DM adalah Deployment. Perencanaan untuk Deployment dimulai selama Business Understanding dan harus menggabungkan tidak hanya bagaimana untuk menghasilkan nilai model, tetapi juga bagaimana mengkonversi skor keputusan, dan bagaimana untuk menggabungkan keputusan dalam sistem operasional.

Pada akhirnya, rencana sistem Deployment mengakui bahwa tidak ada model yang statis. Model tersebut dibangun dari data yang diwakili data pada waktu tertentu, sehingga perubahan waktu dapat menyebabkan berubahnya karakteristik data. Modelpun harus dipantau dan mungkin diganti dengan model yang sudah diperbaiki.

## Persiapan data

Hasil proses pengolahan data yang digunakan disimpan dalam bentuk file CSV. Dari data CSV ini akan dilakukan proses pembersihan data yang tidak digunakan. Gambar 2, file CSV yang disimpan dalam Microsoft Excel. Data yang diolah sebanyak 800 baris dan 7 kolom.

Gambar 2 Data CSV

## Menyiapkan library dan data

Proses awal dalam melakukan perhitungan analisis regresi berganda dengan menggunakan bahasa pemrograman python adalah menentukan *library* yang digunakan dalam proses analisis dan memindahkan data ke *drive*. Hasil proses running skrip di atas ditampilkan pada gambar 3 di bawah ini:

```
import pandas as pd
from google.colab import files
uploaded = files.upload()
with open('data2.csv', 'r') as data:
    df3 = pd.read_csv(data, encoding = ('ansi'))
    print(df3)
```

Pilih File data2.csv

- data2.csv(application/vnd.ms-excel) - 29690 bytes, last modified: 17/1/2022 - 100% done

Saving data2.csv to data2 (2).csv

	Aspek1	Aspek2	Aspek3	Aspek4	IPK_responden	jumlah_responden	Rata-rata
0	2.88	2.93	2.87	2.93	3.76	99.0	2.90
1	2.87	3.02	2.97	3.00	3.70	43.0	2.96
2	2.72	2.91	2.94	2.95	3.37	112.0	2.88
3	2.98	2.97	2.97	2.97	2.90	40.0	2.98
4	3.21	3.12	3.12	3.12	2.80	26.0	3.14
...	...	...	...	...	...	...	...
795	3.17	3.00	3.00	3.00	2.97	4.0	3.04
796	2.89	3.00	3.00	3.00	2.41	6.0	2.97
797	3.00	3.00	3.00	3.00	3.94	1.0	3.00
798	3.00	3.00	3.00	3.00	3.10	4.0	3.00
799	3.00	3.00	3.00	3.00	2.67	8.0	3.00

[800 rows x 7 columns]

Gambar 3 Hasil proses running data yang digunakan

## Membersihkan data

Pada tahapan ini dilakukan pengecekan terhadap isi datasheet yang digunakan, Proses pengecekan diantaranya pengecekan data kosong, tipe data, data yang sama, outliers yaitu data yang nilainya terlalu jauh dengan data lainnya dan proses membersihkan data lainnya.

Berdasarkan Gambar 4 di bawah ini menjelaskan contoh proses dan hasil proses untuk melakukan data kosong.

```
#Mencari dan menangani missing values.
df3.isnull().sum()
```

Aspek1	2
Aspek2	2
Aspek3	2
Aspek4	2
IPK_responden	2
jumlah_responden	2
Rata-rata	2

dtype: int64

Gambar 4, Hasil Pencarian data kosong

Hasil pada gambar 4 terdapat 2 data yang kosong dan data kosong ini akan dihapus. Pada Gambar 5 proses penghapusan data dan proses pengecekan kembali apakah ada data kosong.

```
[ ] df3=df3.dropna()

df3.isnull().sum()
```

Aspek1	0
Aspek2	0
Aspek3	0
Aspek4	0
IPK_responden	0
jumlah_responden	0
Rata-rata	0

dtype: int64

Gambar 5 Perintah dan hasil penghapusan data

## Korelasi Antar Data

Analisis korelasi digunakan untuk melihat keterkaitan tentang derajat hubungan variabel sehingga dapat mengetahui hubungan variabel yang ada. Pada Gambar 6 Perintah dan hasil analisis korelasi data seperti di bawah ini:

```
df3.corr().style.set_precision(3)
```

	Aspek1	Aspek2	Aspek3	Aspek4	IPK_responden	jumlah_responden	Rata-rata
Aspek1	1.000	0.481	0.467	0.456	0.035	-0.016	0.856
Aspek2	0.481	1.000	0.955	0.935	0.035	-0.023	0.852
Aspek3	0.467	0.955	1.000	0.947	0.037	-0.009	0.847
Aspek4	0.456	0.935	0.947	1.000	0.035	-0.005	0.836
IPK_responden	0.035	0.035	0.037	0.035	1.000	0.452	0.041
jumlah_responden	-0.016	-0.023	-0.009	-0.005	0.452	1.000	-0.016
Rata-rata	0.856	0.852	0.847	0.836	0.041	-0.016	1.000

Gambar 6 Perintah dan hasil analisis korelasi data

## Menghitung hasil regresi linear

Langkah selanjutnya adalah menghitung regresi linear. Library yang digunakan adalah library sklearn. Gambar 7, berikut merupakan proses penghitungan nilai koefisien.

```
[31] #mencari cari tau nilai intercept (b).
print(lin_reg.intercept_)

-0.012897918703604017

[33] #mencari nilai koefisien dan menampilkan hasilnya
feature_cols = ['Aspek1', 'Aspek2', 'Aspek3', 'Aspek4', 'IPK_responden', 'jumlah_responden']
X = df3[feature_cols]
y = df3['Rata-rata']
lin_reg = LinearRegression()
lin_reg.fit(X, y)

[('Aspek1', 0.24743301702419157),
 ('Aspek2', 0.257319654680142),
 ('Aspek3', 0.2662633268960668),
 ('Aspek4', 0.2332290731664389),
 ('IPK_responden', 9.697801443559612e-05),
 ('jumlah_responden', -1.4522339706890541e-06)]
```

Gambar 7 Proses penghitungan nilai koefisien.

Berdasar hasil pada perhitungan di atas, nilai regresi linear adalah :  $y = 0.25 X_1 + 0.26 X_2 + 0.27 X_3 + 0.23 X_4 + 0.0001 X_5 - 1e-06 X_6$

## Evaluasi untuk model linear regresi

Evaluasi untuk model linear regresi dapat dilakukan dengan beberapa cara evaluasi seperti menggunakan *mean absolute error (MAE)*, *mean squared error (MSE)* atau *root mean squared error (RMSE)* . Gambar 8 merupakan hasil perhitungan MSE dan RMSE

```
[ ] ypredict=lin_reg.predict(x_test)
print ("Coefficient of determination :",r2_score(y_test,ypredict))
print ("MSE: ",mean_squared_error(y_test,ypredict))
print ("RMSE: ",np.sqrt(mean_squared_error(y_test,ypredict)))

Coefficient of determination : 0.9979745042302917
MSE: 8.352756861727575e-06
RMSE: 0.002890113641663174
```

Gambar 8 Hasil perhitungan MSE dan RMSE

## Data Asli dan data prediksi

Pada tahap ini akan dilakukan perbandingan antara nilai y awal dengan nilai y prediksi. Nilai y prediksi dihitung pada rumus regresi linear berganda yang telah dihitung pada



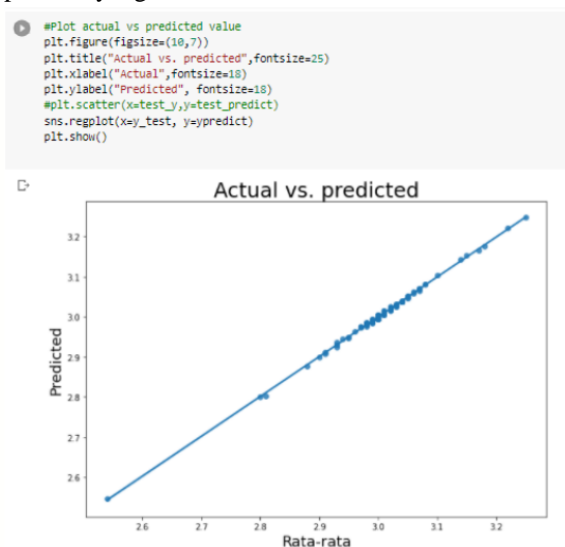
langkah di atas. Gambar 9 Hasil perhitungan membandingkan data asli dengan data prediksi dibuat dalam bentuk grafik.

```
df_best_predict = pd.DataFrame({'Actual': y_test, 'Predicted': ypredict})
df_best_predict.head(10)
```

	Actual	Predicted
267	3.15	3.152753
637	3.04	3.040133
464	2.96	2.962937
721	3.00	3.000125
796	2.97	2.972845
669	3.02	3.017842
791	3.00	3.000167
504	2.99	2.990139
539	3.01	3.010014
147	3.01	3.015032

Gambar 9 Proses dan hasil perhitungan

Gambar 10 Hasil y asli dengan y prediksi dalam bentuk grafik. Hasil grafik memperlihatkan antara data asli dan data prediksi menghasilkan selisih angka antara data aktual dan data prediksi yang sedikit.



Gambar 10. Grafik y asli dengan y prediksi

### Perhitungan dengan library statsmodels

Library lain yang dapat digunakan untuk melakukan analisis regresi linear adalah dengan menggunakan library statsmodel. Hasil perhitungan dengan library statsmodel ada pada gambar 11 di bawah ini:

```
[ ] import statsmodels.api as sm
x = x_train.to_numpy()
y = y_train.to_numpy()
x = sm.add_constant(x)
model = sm.OLS(y, x).fit()
print(model.summary())
```

```
=====
OLS Regression Results
=====
Dep. Variable: y                R-squared: 0.997
Model: OLS                    Adj. R-squared: 0.997
Method: Least Squares         F-statistic: 3.794e+04
Date: Tue, 18 Jan 2022         Prob (F-statistic): 0.00
Time: 06:43:40                Log-likelihood: 2028.6
No. Observations: 638         AIC: -5643.
Df Residuals: 631             BIC: -5612.
Df Model: 6
Covariance Type: nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
const	-0.0129	0.008	-1.687	0.092	-0.028	0.002
x1	0.2474	0.001	184.485	0.000	0.245	0.250
x2	0.2573	0.009	29.622	0.000	0.248	0.274
x3	0.2563	0.010	25.449	0.000	0.245	0.256
x4	0.2332	0.008	28.395	0.000	0.217	0.249
x5	9.690e-05	0.000	0.760	0.447	-0.000	0.000
x6	-1.452e-06	2.01e-06	-0.723	0.470	-5.4e-06	2.49e-06

```
=====
Omnibus: 3.270 Durbin-Watson: 1.964
Prob(Omnibus): 0.194 Jarque-Bera (JB): 2.910
Skew: 0.088 Prob(JB): 0.233
Kurtosis: 2.720 Cond. No. 9.97e+03
=====
```

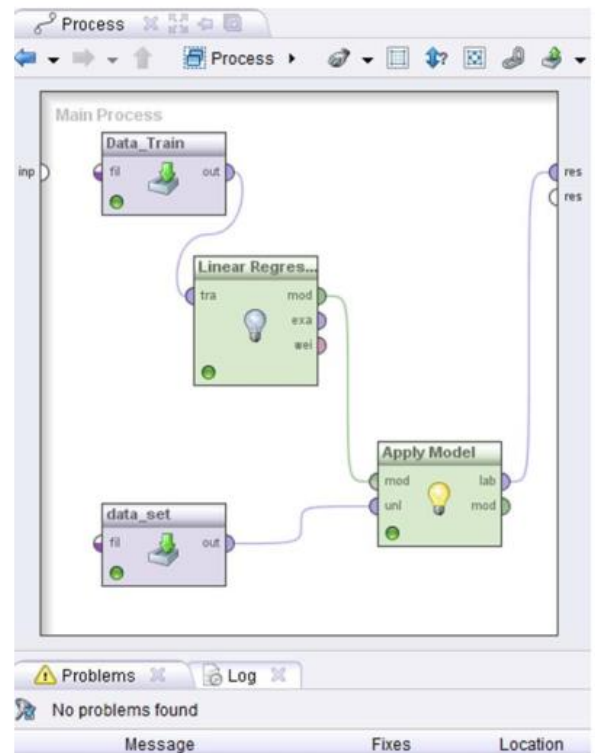
Warnings:

```
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 9.97e+03. This might indicate that there are strong multicollinearity or other numerical problems.
```

Gambar 11. Hasil perhitungan dengan library Statsmodel

### Pengujian Hasil Prediksi

Pengujian data dilakukan dengan menggunakan aplikasi pengujian *rapidminer* 5 terhadap hasil perhitungan prediksi sampah sampai dengan tahun 2025 menggunakan algoritma regresi linier seperti ditunjukkan pada Gambar 12 di bawah ini.



Gambar 12 Proses Pengujian Data

## 5 KESIMPULAN

Linear regression memiliki beberapa keunggulan jika digunakan untuk memprediksi suatu model, seperti: mudah diimplementasikan dan diinterpretasikan, menghasilkan model prediksi yang paling akurat untuk data yang bersifat linear serta mudah untuk dilakukan evaluasi dan memiliki beragam metode atau metrik yang mudah diterapkan.

Sehingga dapat ditarik kesimpulan, bahwa analisis regresi bertujuan untuk mengetahui pengaruh dari suatu variabel terhadap variabel lainnya. Pada analisis regresi suatu variabel yang mempengaruhi disebut variabel bebas atau independent variable, sedangkan variabel yang dipengaruhi disebut variabel terkait atau dependent variable. Analisis regresi juga dapat digunakan untuk memahami variabel – variabel bebas mana saja yang dapat berhubungan dengan variabel terikat, serta untuk mengetahui bentuk hubungan tersebut.

Dalam penelitian ini data yang digunakan disimpan dalam drive dan proses harus dilakukan secara online. Penelitian menghasilkan rumus regresi linear dengan nilai dengan nilai koefisien adalah 0.997. Hasil analisis korelasi antara variabel jumlah korespon dengan variabel yang lain menunjukkan nilai negatif. Hal menunjukkan variabel jumlah respondn tidak ada kaitannya dengan variabel yang lain.

## DAFTAR PUSTAKA

- A. S. T. Nishadi, "Predicting Heart Diseases In Logistic Regression Of Machine Learning Algorithms By Python Jupyterlab," International Journal of Advanced Research and Publications, vol. 3, no. 8, pp. 69–74, 2019.

- A. R. Shidiq, I. Ermawati, M. M. Santoni, J. Barat, O. Keluar, and R. Linier, "Penerapan Algoritma Regresi Linier Untuk Prediksi Pada," pp. 102–111, 2020.
- A. Wibisono, M. Rofik, and E. Purwanto, "Penerapan Analisis Regresi Linier Berganda dalam Penyelesaian Skripsi Mahasiswa," *Jurnal ABDINUS: Jurnal Pengabdian Nusantara*, vol. 3, no. 1, p. 30, 2019, doi:10.29407/ja.v3i1.13512.
- B. D. Purba, "Analisa Korelasi Kinerja Dosen Terhadap Prestasi Mahasiswa Dengan Menggunakan Logika Fuzzy," *Jurnal Teknik Informatika Unika St.Thomas (JTIUST)*, vol. 02, no. 479, pp. 64–71, 2017.
- D. S. SUSANTI, Y. SUKMAWATY, and N. SALAM, *ANALISIS REGRESI DAN KORELASI*. IRDH, 2019.
- E. Karamazova, T. Jusufi Zenku, and Z. Trifunov, "Analysing and Comparing the Final Grade in Mathematics by Linear Regression Using Excel and SPSS," *International Journal of Mathematics Trends and Technology*, vol. 52, no. 5, pp. 334–344, 2017, doi: 10.14445/22315373/ijmtt-v52p549.
- E. Triyanto, H. Sismoro, and A. D. Laksito, "IMPLEMENTASI ALGORITMA REGRESI LINEAR BERGANDA UNTUK MEMPREDIKSI PRODUKSI PADI DI KABUPATEN BANTUL," *RABIT: Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 4, no. 2, pp. 73–86, 2019.
- I. Sonata, Nuraida, and E. Sofian, "Pengaruh Kepribadian Dan Continuance Commitment Terhadap Kinerja Pegawai Di Kantor Dinas Kelautan Dan Perikanan Provinsi Sumatera Utara," *Tijarah: Jurnal Ekonomi dan Bisnis*, vol. 1, no. 21, pp. 12–22, 2021.
- J. Enterprise, *Mengolah Data dengan Python dan Pandas*. Jakarta: ElexJudul Penelitian (First Author, Second Author, Third Author) Media Komputindo, 2021.
- K. Sa'adah and A. Febrianto, "ANALISA KINERJA DOSEN BERDASARKAN TINGKAT KEPUASAN MAHASISWA DI POLITEKNIK MANUFAKTUR BANDUNG," *JIMEA | Jurnal Ilmiah MEA (Manajemen, Ekonomi, dan Akuntansi)*, vol. 5, no. 1, pp. 1339–1351, 2021, doi: 10.31955/mea.vol3.iss2.p.
- K. Puteri and A. Silvanie, "MACHINE LEARNING UNTUK MODEL PREDIKSI HARGA SEMBAKO," *Jurnal Nasional Informatika*, vol. 1, no.2, pp. 82–94, 2020.
- M. S. Sholehuddin, A. Akhwanudin, and U. Khasanah, *PENGLOLAAN KINERJA DOSEN DAN BUDAYA AKADEMIK*. Penerbit NEM, 2018.
- N. N. Dewi, *KIAT-KIAT MERANGSANG KINERJA DOSEN PTS*. MEDIA SAHABAT CENDEKIA, 2019.
- S. C. Agrawal, "Deep learning based non-linear regression for Stock Prediction," *IOP Conference Series: Materials Science and Engineering*, vol. 1116, no. 1, p. 012189, 2021, doi: 10.1088/1757-899x/1116/1/012189.
- S. Muhammad, Rr. Yuliana Rachmawati, Eko Nur Cahyo, "Penerapan Regresi Linear Ganda Untuk Memprediksi Hasil Nilai Kuesioner Mahasiswa Dengan Menggunakan Python," *Jurnal Dinamika Informatika*. Volume 11, No 1, Februari 2022. ISSN 1978-1660 : 13-24
- S. N. Waghmare and C. N. Sakhale, "Formulation of Experimental Data Based model using SPSS ( Linear Regression ) for Stirrup Making Operation by Human Powered Flywheel Motor," *International Research Journal of Engineering and Technology (IRJET)*, vol. 02, no. 04, pp. 461–468, 2015.
- Sudaryatno et al., "Multiple linear regression analysis of remote sensing data for determining vulnerability factors of landslide in PURWOREJO," *IOP Conference Series: Earth and Environmental Science*, vol. 500, no. 1, 2020, doi: 10.1088/1755-1315/500/1/012046.
- U. Haryaka, *Faktor-faktor yang Mempengaruhi Kinerja Dosen*. CV. Amerta Media, 2021.
- T. N. Padilah and R. I. Adam, "Analisis regresi linier berganda dalam estimasi produktivitas tanaman padi di kabupaten karawang 1,2)," *Fibonacci Jurnal Pendidikan Matematika dan Matematika*, vol. 5, no. 2, pp. 117–128, 2019.
- T. Sulastri, "Analisis Kepuasan Mahasiswa Terhadap Kinerja Dosen," *Optimal: Jurnal Fakultas Ekonomi Universitas Islam "45" Bekasi*, vol. 10, no. 2, pp. 167–184, 2017.
- Y. Robi, "Implementasi Data Mining Estimasi Ketersediaan Lahan Pembuangan Sampah menggunakan Algoritma Regresi Linear," Vol. 2 No. 1 (2018) 361 – 366. ISSN : 2580-0760.

<https://www.dqlab.id/teknik-analisis-data-crisp-dm-dalam-data-mining>

diakses pada pukul `12.30 Wita tanggal 12 Januari 2023.

<https://mmsi.binus.ac.id/2020/09/18/cross-industry-standard-process-for-data-mining-crisp-dm/>

diakses pada pukul `10.00 Wita tanggal 13 Januari 2023.

<https://www.dqlab.id/linear-regression-algoritma-data-science-terpopuler>

diakses pada pukul `10.13 Wita tanggal 13 Januari 2023.

<https://www.dqlab.id/mengenal-analisis-regresi-dalam-metode-statistik-data-science>

diakses pada pukul `12.00 Wita tanggal 13 Januari 2023.

<https://bbs.binus.ac.id/management/2019/12/analisis-regresi-sederhana/>

diakses pada pukul 02.30 Wita tanggal 14 Januari 2023.