

Penerapan Algoritma Regresi Linear Sederhana Untuk Prediksi Pengaruh Nilai Omset Terhadap Besaran Pendapatan Bersih Usaha Mikro Kecil Dan Menengah

Naufal Musthofa¹, Ifan Aulia Wahyudi², Ridwan AripNurazazi³, Nico Candilo Gumay⁴

¹Teknik Informatika, Teknologi Informasi, Universitas Serang Raya

²Teknik Informatika, Teknologi Informasi, Universitas Serang Raya

³Teknik Informatika, Teknologi Informasi, Universitas Serang Raya

⁴Teknik Informatika, Teknologi Informasi, Universitas Serang Raya

¹naufal.msth@gmail.com, ²Ifanaulia100@gmail.com, ³ridwanarip2@gmail.com, ⁴Kandilonico@gmail.com

Abstract

Micro, small and medium enterprises (MSMEs) are one of the business fields that can develop and be consistent in the national economy. Micro, small and medium enterprises (MSMEs) are good forums for job creation, which are planned by the government, private sector and individual business actors. The purpose of this study is to predict the increase in net income of micro, small and medium enterprises that affect the economic growth of Cilegon City. In this study using quantitative methods, namely analyzing and interpreting the relationship between variables through accurate and up-to-date data according to the problems in this study. The research method is carried out by examining data from data that is wrong or can interfere with the analysis process, performs analysis on the dataset and divides the data into training data and test data. The process of data sharing is 80% used for training data and 20% for test data. All processes are processed using Python programming. Python libraries used are numpy, pandas, matplotlib, seaborn, sklearn. The analysis process is done with Kaggle Notebook. The results of the study resulted in a simple linear regression model $y = -374359.94312876 + (0.20259073 (X))$ with a Correlation value: 0.9999546662293942 Coefficient of determination 0.7244150380582826, Mean Absolute Error: 243269.23966642626, Mean Squared Error: 71893932500.72606 and 934012073. The final result of the analysis is a comparison between the actual y and the predicted y in the form of tables and graphs.

Keywords: SMEs, Linear Regression, Python

Abstrak

Kegiatan usaha mikro, kecil dan menengah (UMKM) merupakan salah satu bidang usaha yang dapat berkembang dan konsisten dalam perekonomian nasional. Usaha mikro, kecil dan menengah (UMKM) menjadi wadah yang baik bagi penciptaan lapangan pekerjaan yang direncanakan baik oleh pemerintah, swasta dan pelakunya usaha perorangan. Adapun yang menjadi tujuan penelitian ini yaitu untuk memprediksi peningkatan pendapatan bersih usaha mikro, kecil dan menengah berpengaruh terhadap pertumbuhan ekonomi Kota cilegon. Dalam penelitian ini menggunakan metode kuantitatif, yaitu menganalisis dan menginterpretasikan hubungan antar variabel melalui data yang akurat serta terbaru sesuai dengan permasalahan dalam penelitian ini. Metode penelitian dilakukan dengan memeriksa data dari data yang salah atau dapat mengganggu proses analisis, melakukan analisis pada dataset serta membagi data menjadi data training dan data test. Proses pembagian data adalah 80% digunakan untuk data training dan 20% untuk data test. Semua proses diolah dengan menggunakan pemrograman Python. Library Python yang digunakan numpy, pandas, matplotlib, seaborn, sklearn. Proses analisis dikerjakan dengan kaggle Notebook. Hasil penelitian menghasilkan model regresi linear sederhana $y = -374359.94312876 + (0.20259073 (X))$ dengan nilai Correlation: 0.9999546662293942 Coefficient of determination 0.7244150380582826, Mean Absolute Error:

243269.23966642626, Mean Squared Error: 71893932500.72606 dan Root Mean Squared Error: 268130.43934012053. Hasil akhir analisis dilakukan perbandingan antara y aktual dengan y prediksi baik dalam bentuk tabel maupun grafik.

Kata kunci: UMKM, Regresi Linear, Python.

1. Pendahuluan

Usaha Kecil Mikro dan Menengah (UMKM) merupakan usaha yang saat ini berpengaruh sangat besar akan pertumbuhan ekonomi masyarakat, hal itu juga UMKM sebagai penunjang untuk memenuhi kebutuhan hidup manusia, dengan adanya UMKM ini masyarakat dapat mendirikan usahanya dengan modal yang memadai [1]. Usaha Miko Kecil dan Menengah ini sangat digalakkan oleh pemerintah khususnya di Kota Cilegon. Semakin banyak masyarakat berwirausaha maka semakin baik dan kokohnya perekonomian suatu daerah karena sumber daya lokal, pekerja lokal serta pembiayaan lokal dapat terserap dan bermanfaat secara optimal [2]. Populasi sektor UMKM pada dasarnya merupakan suatu potensi bagi perusahaan skala besar, baik dalam sektor industri pengolahan maupun jasa [3]. Disisi lain, UMKM juga dapat dianggap sebagai penyedia barang dan jasa input bagi perusahaan besar melalui berbagai macam mekanisme diantaranya melalui skema kemitraan atau *subcontracting* [4].

Cilegon merupakan kota baja yang perekonomiannya dikatakan cukup berkembang dan kini memasuki kota dengan Pendapatan Perkapita Produk Domestik Regional Bruto (PDRB) tertinggi di Provinsi Banten pada tahun 2019. Peneliti memiliki beberapa alasan mengapa memilih untuk meneliti perkembangan UMKM yang ada di Kota Cilegon. Permasalahan perkembangan UMKM yang ada di Cilegon ini menarik dibahas untuk mengetahui kebenaran data yang memang mempengaruhi perkembangan UMKM, selain itu UMKM di Kota Cilegon ditahun 2020 dan memasuki tahun 2021 terlihat begitu berkembang pesat dibanding tahun-tahun sebelumnya. Menurut peneliti hal itu dipengaruhi karena adanya motivasi dalam bentuk kontribusi positif dari tangan pemerintah melalui berbagai bentuk binaan. Selain itu adanya faktor modal atau omset dan psikologis, karenanya dapat menunjang keberhasilan usaha itu sendiri. Jika pendapat bersih pada umkm mengalami kenaikan dan penurunan setiap tahun, kejadian ini dapat menyebabkan kekhawatiran jika suatu saat pendapatan bersihnya terus menurun. Seperti pada tabel berikut :

Produk	Omset	Pendapatan Bersih
Jual Aqua Galon	144.000.000	28.800.000
Bensin Eceran	90.000.000	18.000.000
Warung Sembako	126.000.000	25.200.000
Rumah Makan	126.000.000	25.200.000
Aneka Buah Segar	126.000.000	25.200.000
Jajanan Anak	36.000.000	7.200.000
Es Kelapa	108.000.000	21.600.000
Warung Sembako	72.000.000	14.400.000
Bubur Kacang ljo,Soto,Sop,Bubur Ayam Madura	288.000.000	57.600.000
Sembako,Isi Ulang Galon	126.000.000	25.200.000
Gorengan,Bubursum Sum,Urab	54.000.000	10.800.000
Konter Pulsa	108.000.000	21.600.000
Sosis,Baso Bakar,Dll	144.000.000	28.800.000
Konter Hp,Pulsa	90.000.000	18.000.000
Konter Hp,Pulsa,Kuota	90.000.000	18.000.000
Peralatan Memancin	180.000.000	36.000.000
Es Batu,Pop Ice	18.000.000	3.600.000
Baju Muslim,Kerudung, Daster Dll, Jasa jahit	180.000.000	36.000.000
Warung Sembako	180.000.000	36.000.000
Jajanan,Es,Pulsa	54.000.000	10.800.000
Warung Sembako	72.000.000	14.400.000
Risol, Onde-Onde, Dadar Gulung	36.000.000	7.200.000
Petis Buah,Es,Jajanan	54.000.000	10.800.000
Makanan Ringan	72.000.000	14.400.000
Nasi Uduk & Gorengan	72.000.000	14.400.000
Sr12	72.000.000	14.400.000
Warung Sembako	108.000.000	21.600.000
Nasi Uduk & Gorengan	72.000.000	14.400.000
Jual Makanan Ringan	72.000.000	14.400.000
Jual Tahu Tempe	72.000.000	14.400.000
Jual Pulsa,Bensin,Gas	108.000.000	21.600.000
Minuman,Makanan	72.000.000	14.400.000
Bengkel	108.000.000	21.600.000
Pampers	72.000.000	14.400.000
Nasi Uduk	108.000.000	21.600.000

Rokok,Es,Kopi	108.000.000	21.600.000
Sembako	108.000.000	21.600.000
Buah-Buahan	180.000.000	36.000.000
Emping	180.000.000	36.000.000
Sembako	54.000.000	10.800.000
Sembako	72.000.000	14.400.000
Jual Baju	36.000.000	7.200.000
Kue Basah	54.000.000	10.800.000
Kue Basah	36.000.000	7.200.000
Kue Basah	72.000.000	14.400.000
Nasi Uduk	54.000.000	10.800.000
Sembako	72.000.000	14.400.000
Sembako	72.000.000	14.400.000
Butik Kebutuhan Bayi dan Balita	144.000.000	28.800.000
Kue Basah Dan Es	72.000.000	14.400.000
Jual Baju	108.000.000	21.600.000
Jajanan Anak	18.000.000	3.600.000
Jawara Ayam Potong	288.000.000	57.600.000
Jual Makanan Dan Minuman	72.000.000	14.400.000
Sembako	72.000.000	14.400.000
Sandal Dan Sosis Bakar	54.000.000	10.800.000
Jual Kue Basah	54.000.000	10.800.000
Dawet, Kurma, Susu Kacang Ijo	72.000.000	14.400.000
Pejual, Kue, Pisang, Sayur	36.000.000	7.200.000
Kosmetik	72.000.000	14.400.000
Penyewaan Tenda & Rias	120.000.000	24.000.000
Warung Jajanan	108.000.000	21.600.000
Baju	72.000.000	14.400.000
Lontong Sayur	72.000.000	14.400.000
Menjahit Permak	25.200.000	420.000
Sembako	108.000.000	21.600.000
Gorengan	54.000.000	10.800.000
Bubur Ayam	72.000.000	14.400.000
Sembako & Sayuran	108.000.000	21.600.000
Baju	108.000.000	21.600.000
Sembako	72.000.000	14.400.000
Dhum-Dhum Thai Tea,Ramen Meledak	108.000.000	21.600.000
Pempek	54.000.000	10.800.000
Jahit Baju	18.000.000	3.600.000
Sembako	36.000.000	7.200.000
Sembako	36.000.000	7.200.000
Mie Ayam	36.000.000	7.200.000
Sayur Mateng	72.000.000	14.400.000
Nugget,Sosis,Alat- Alat Listrik	180.000.000	36.000.000
Sembako	108.000.000	21.600.000
Warung Sembako	108.000.000	21.600.000

Jualan Lampu	36.000.000	7.200.000
Jajanan Anak	36.000.000	7.200.000
Warung Sembako	108.000.000	21.600.000
Sembako	72.000.000	14.400.000
Sayuran	288.000.000	57.600.000
Somay	72.000.000	14.400.000
Warung Kiki	54.000.000	10.800.000
Makanan Ringan	36.000.000	7.200.000
Warung Sembako	108.000.000	21.600.000
Bensin	54.000.000	10.800.000
Kue Basah	36.000.000	7.200.000
Kue Basah	36.000.000	7.200.000
Sembako	72.000.000	14.400.000
Warung Sembako	72.000.000	14.400.000
Warung Makanan Ringan	108.000.000	21.600.000
Token Listrik & Pulsa	144.000.000	28.800.000
Kue Basah	36.000.000	7.200.000
Makanan Ringan	54.000.000	10.800.000
Warung Jajanan	18.000.000	3.600.000
Pakaian	54.000.000	10.800.000
Warung Sembako	54.000.000	10.800.000
Emping	90.000.000	18.000.000
Parut Kelapa, Kacang/Bumbu	18.000.000	3.600.000
Jual Bakso	36.000.000	7.200.000
Bensin/Pom Mini	36.000.000	7.200.000
Sayuran	90.000.000	18.000.000
Bubur Ayam	72.000.000	14.400.000
Warung Jajanan	36.000.000	7.200.000
Air Mineral,Beras	54.000.000	10.800.000
Es Kelapa Dan Makanan Ringan	36.000.000	7.200.000
Tahu Bandung	288.000.000	57.600.000
Warung Jajanan	36.000.000	7.200.000
Anora	36.000.000	7.200.000
Warung Sembako	36.000.000	7.200.000
Sosis Bakar	72.000.000	14.400.000
Sembako	36.000.000	7.200.000
Lontong Sayur	25.200.000	4.200.000
Sembako	72.000.000	14.400.000
Sembako	36.000.000	7.200.000
Jual Baju	96.000.000	19.200.000
Jual Sate	360.000.000	72.000.000
Kacang Sangrai/Kerupuk	14.400.000	2.880.000
Bakso, Es	36.000.000	7.200.000
Warung Sembako	36.000.000	720.000
Bakso , Bakwan	54.000.000	10.800.000
Jual Sayuran	288.000.000	57.600.000
Warung Sembako	36.000.000	7.200.000

Sembako	180.000.000	36.000.000
Warung Sembako	36.000.000	720.000
Warung Minuman	72.000.000	14.400.000
Jual Bensin	54.000.000	10.800.000
Minuman Kopi & Es	108.000.000	21.600.000
Minuman Kopi, Es & Makanan	108.000.000	21.600.000
Sembako	360.000.000	72.000.000
Sayur Mateng	180.000.000	36.000.000
Sayuran	288.000.000	57.600.000
Sayuran	180.000.000	36.000.000
Penjahit	72.000.000	14.400.000
Jual Donat	72.000.000	14.400.000
Jualan Baju	48.000.000	9.600.000
Bakso	144.000.000	28.800.000
Ikan Bakar	108.000.000	21.600.000
Nasi Bakar,Jus	19.200.000	3.840.000
Plastik Dan Bahan Kue	108.000.000	21.600.000
Air Mineral	72.000.000	14.400.000
Ketoprak	72.000.000	14.400.000
Penjahit	18.000.000	3.600.000
Air Mineral dan Pakaian	72.000.000	14.400.000
Kue Basah	36.000.000	7.200.000

Tabel 1 Data Umkm Kota Cilegon 2021

Dari data yang ada pada tabel 1 usaha umkm pada tahun 2021 pengaruh omset terhadap pendapatan bersih mengalami penurunan berturut-turut pada tahun 2020 dan 2019 karna pandemic covid - 19. pada tahun 2020 pendapatan bersih mengalami kenaikan dari tahun sebelumnya. Akan tetapi pada tahun 2021 mengalami penurunan kembali. Jika permasalahan ini tetap dibiarkan maka akan berdampak negatif bagi kelangsungan umkm di kota cilegon seperti bangkrutnya sebuah usaha, harga usaha pada produk umkm tertentu akan mengalami kenaikan, serta pekerja akan kehilangan sumber mata pencaharianya. Pemerintah telah berupaya meningkatkan pendapatan umkm dengan berbagai cara, namun kurangnya teknologi pendukung menjadi salah satu penghambat dalam meningkatkan pendapatan umkm di Kota cilegon. Oleh karena itu, perlunya prediksi untuk memperkirakan seberapa besar pendapatan di tahun yang akan datang.

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil [5]. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat

mungkin yang akan terjadi. Prediksinya mempunyai makna sebagai sebuah informasi, pemberitahuan, peringatan, bahkan tentang segala sesuatu yang akan terjadi berdasarkan fakta maupun tidak [6]. Dalam memprediksi untuk pembelian suatu produk, ada juga perusahaan yang membutuhkan prediksi untuk kelangsungan dari produk yang dibeli. Perusahaan harus memiliki data omset sebelumnya untuk dapat memprediksinya.

Omset merupakan akumulasi dari kegiatan penjualan suatu produk barang dan jasa yang dihitung secara keseluruhan selama kurun waktu tertentu secara terus menerus [2]. Omset penjualan disebut untung jika hasil dari penjualan lebih besar daripada modal dan sebaliknya, disebut rugi jika hasilnya lebih kecil daripada modal [7].

Untuk memprediksi pengaruh nilai omset terhadap besaran pendapatan bersih pada UMKM dengan memanfaatkan data mining dan Algoritma Regresi Linear Sederhana. Data Mining merupakan serangkaian proses yang dilakukan untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [8]. Data mining menjadi alat yang semakin penting digunakan untuk mengubah data menjadi sebuah informasi [9]. Sedangkan, regresi linear sederhana adalah analisis regresi yang melibatkan hubungan antara satu variabel tak bebas dihubungkan dengan satu variabel bebas. Yaitu merupakan satu cara prediksi yang menggunakan garis lurus untuk menggambarkan hubungan di antara dua variabel (atau lebih). Dalam kasus di atas variabelnya adalah jarak dan waktu tempuh. Jarak di gambarkan sebagai sumbu X (dalam kilometer) dan waktu tempuh sebagai sumbu Y (dalam menit), setiap pasang jarak dan waktu tempuh di gambar sebagai titik.

Regresi linier juga merupakan metode statistik yang berfungsi untuk menguji sejauh mana hubungan sebab-akibat antara variabel faktor penyebab (x) terhadap variabel akibatnya. Faktor penyebab pada umumnya dilambangkan dengan X sedangkan variabel akibat dilambangkan dengan Y [10]. Regresi linear sederhana atau sering disingkat dengan SLR (Simple Linier Regression) juga merupakan salah satu metode statistik yang dipergunakan dalam produksi untuk melakukan peramalan atau pun prediksi tentang karakteristik kualitas maupun kuantitas [8].

Berdasarkan latar belakang yang dipaparkan, sehingga dapat diambil keputusan yang diharapkan dapat menyelesaikan permasalahan di atas untuk melakukan penelitian dengan membangun perangkat lunak dengan membuat machine learning dengan pemodelan regresi linear sederhana dengan menggunakan pemrograman Python, yang mempunyai kelebihan yaitu metode ini cukup simpel dan mudah dipahami untuk memprediksi produksi tanaman pangan di Indonesia. Oleh karena itu dapat dipetik sebuah kesimpulan untuk melakukan penelitian yang berjudul "Penerapan Algoritma Regresi Linear

Sederhana Untuk Prediksi Pengaruh Nilai Omset Terhadap besaran pendapatan bersih” dengan harapan pemerintah memberikan perhatian lebih pada pelaku UMK.

Dalam melakukan penelitian dibutuhkan adanya batasan masalah yang bertujuan untuk membatasi proses penelitian agar tujuan yang sebenarnya tercapai. Adapun batasan-batasan dalam penelitian ini yaitu : Masalah yang diteliti adalah bagaimana proses penerapan metode simple regressi linear, penelitian ini menggunakan beberapa variabel yaitu: omset (X) sebagai variabel independennya dan pendapatan bersih tahun sesudahnya (Y) sebagai variabel dependennya, data penelitian yang digunakan dari tahun 2020 sampai 2021.

2. Metode Penelitian

2.1 Tipe Penelitian

Penelitian ini merupakan penelitian terapan (applied research) yakni metodologi yang digunakan untuk memecahkan masalah praktis dunia modern, baik masalah praktis dari individu atau kelompok. Atau dengan kata lain penelitian terapan merupakan dirancang untuk menjawab pertanyaan spesifik yang bertujuan memecahkan masalah praktis bukan untuk memperoleh pengetahuan baru atau teori baru.

2.2 Dataset

Dataset yang digunakan dalam proses penelitian ini diolah dari dinas koperasi dan UMKM Kota Cilegon, Dataset insurance.csv berisi 150 data dan terdiri dari 3 kolom. Ke tiga kolom tersebut terdiri dari 1 kolom kategori (Produk, dan 2 kolom numerik ((X) yaitu omset dan (Y) pendapatan bersih.

2.3 Pre-processing Data

Pre-processing dilakukan dalam melakukan pengolahan data agar data yang digunakan dapat diolah dengan baik dan terhindarkan dari data-data yang salah.

Dalam proses pengolahan yang dilakukan, data awal yang digunakan masih berupa data mentah. Dalam proses yang dilakukan data-data yang diperlukan akan diformat dengan cara tertentu dan sesuai kebutuhan.

2.3. Analisis Data

Dalam proses penelitian yang dilakukan proses analisis data yang dilakukan adalah regresi linier sederhana dan diolah dengan menggunakan python dan diimplementasikan menggunakan Kaggel.

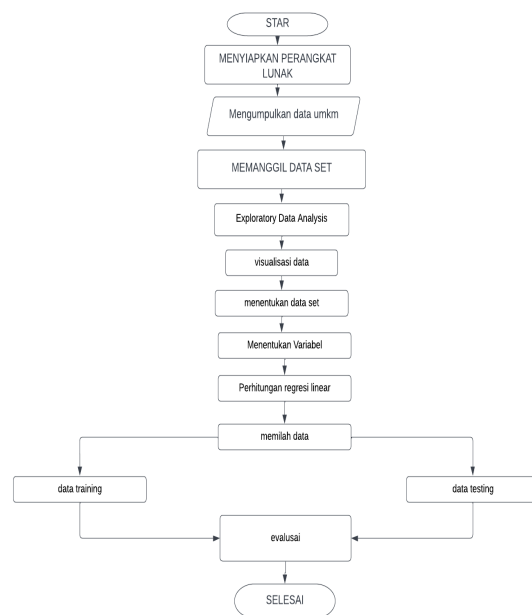
Analisis data yang dilakukan pertama adalah melakukan pembersihan dataset dari data yang tidak diinginkan, seperti data kosong, data yang diluar ambang batas dan tipe data yang tidak sesuai. Hasil proses pembersihan data akan dipilah menjadi 2, yaitu data yang digunakan sebagai data training dan data yang digunakan sebagai data test. Data training digunakan untuk melatih algoritma dan data testing digunakan untuk mengetahui performa algoritma yang

sudah dilatih sebelumnya. Komposisi pembagian ini adalah 80% dari data yang ada untuk data training dan 20% untuk data test. Awal penelitian dilakukan dengan menentukan kebutuhan perangkat keras dan perangkat lunak yang diperlukan.

Kebutuhan perangkat lunak yang diperlukan diantaranya adalah datasheet yang digunakan dalam proses simulasi data dan *library/packages* yang digunakan dalam proses olah data.

Dari datasheet yang ditentukan, datasheet dilakukan proses pengecekan data dari data yang tidak diperlukan atau menghapus data kosong. Langkah lain adalah menentukan persentase untuk data training dan data test.

Hasil data training dan data test dilakukan proses pengujian keakuratan data dan dilakukan pengujian dengan data diluar datasheet. Alur penelitian yang dilakukan ada pada gambar 1.



Gambar 1.

3. Hasil dan Pembahasan

3.1 Menyiapkan Perangkat Lunak

Proses analisis dengan menggunakan python diawali dengan menyiapkan library yang digunakan. Library digunakan adalah numpy, pandas, matplotlib, seaborn, sklearn dan math. Library numpy digunakan untuk data analysis tools, library matplotlib dan seaborn untuk visualisasi data serta library scikit-Learn untuk machine learning dan library math untuk perhitungan. Perintah untuk memanggil library ada pada gambar 2

```
[90]:
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_absolute_error
import math
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
```

Gambar 2 Library yang digunakan

3.2 Memanggil Dataset dan Analisis

Dataset yang digunakan merupakan dataset yang diambil dari Dinas Koperasi dan Umkm Kota Cilegon Proses penggunaan dataset menggunakan perintah read yang ada pada library pandas. Proses pemanggilan dengan menggunakan perintah print (dataset). Perintah ini menampilkan datasheet dari awal sampai akhir. Hasil Perintah tersebut pada gambar 3.

```
dataset = pd.read_csv('../input/kecwarnasari/kecwarnasari.csv')
print(dataset)
```

	PRODUK	omset	pendapatan bersih
0	JUAL AQUA GALON	144000000	28800000
1	BENSIN ECEHAN	90000000	18000000
2	WARUNG SEMBAKO	126000000	25200000
3	RUMAH MAKAN	126000000	25200000
4	ANEKA BUAH SEGAR	126000000	25200000
...
145	AIR MINERAL	72000000	14400000
146	KETOPRAK	72000000	14400000
147	PENJAHIT	18000000	3600000
148	AIR MINERAL DAN PAKAIAN	72000000	14400000
149	KUE BASAH	36000000	7200000

[150 rows x 3 columns]

Gambar 3 Contoh datasheet kecwarnasari.csv

Pada gambar 3, dataset kec, terdiri dari 3kolom. Berdasar pada dataset tersebut, akan dilakukan analisis prediksi (Y) pendapatan bersih dan yang dipengaruhi hasil produksi (X) omset. Prediksi dalam bentuk regresi linear sederhanayang dinotasikan dengan

$$Y = a + bX$$

Keterangan:

Y = variabel tak bebas

X = variabel bebas

a = Konstanta (intercept)

b = Koefisien regresi (kemiringan) besaran response yang ditimbulkan oleh variabel

Menentukan koefisien persamaan a dan b dapat dengan menggunakan metode kuadrat terkecil, yaitu cara yang dipakai untuk menentukan koefisien persamaan dan dari jumlah pangkat dua (kuadrat) antara titik-titik dengan garis regresi yang dicari yang terkecil Dengan demikian, dapat ditentukan

Menghitung konstanta :

$$a = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n(\sum X^2) - (\sum X)^2}$$

Menghitung koefisien :

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2}$$

dalam melakukan perhitungan menggunakan linear regresi sederhana :

1. Menentukan tujuan dari melakukan analisis Regresi Linear yaitu mempelajari hubunganyang diperoleh dan dinyatakan dalam perasamaan matematika yang menyatakan hubungan antar variabel.
2. Mencari variabel dependen (Y) dan independen (X).
3. mengumpulkan data yang diperlukan.
4. menghitung nilai-nilai X^2 , Y^2 dan XY .
5. melakukan perhitungan prediksi

3.3 Exploratory Data Analysis

Salah satu cara dalam melakukan analisis data adalah dengan melakukan exploratory data analysis (EDA). Pada EDA, dilakukan eksplorasi data sehingga akan mendapatkan data yang sesuai dengan proses yang dilakukan. EDA merupakan suatu kegiatan untuk mempelajari data yang dimiliki serta menentukan bagaimana proses pengolahannya terhadap data tersebut.

Pada tahap ini dilakukan pemeriksaan pada data seperti data kosong, menghapus datayang sama. Hasil pengecekan informasi dari dataset yang digunakan ditampikan padagambar 4 dan gambar 5.

```
dataset.describe()
```

[92]:

	omset	pendapatan bersih
count	1.500000e+02	1.500000e+02
mean	8.884000e+07	1.764520e+07
std	6.609086e+07	1.334870e+07
min	1.440000e+07	4.200000e+05
25%	3.900000e+07	7.800000e+06
50%	7.200000e+07	1.440000e+07
75%	1.080000e+08	2.160000e+07
max	3.600000e+08	7.200000e+07

Gambar 4 Data Statistik

Gambar 4 merupakan analisis yang dapat digunakan untuk melihat data statistika, Dari data tersebut dapat digunakan untuk melihat apakah ada data yang tidak wajar.

Dari data tersebut, nilai dari data yang dapat dilakukan analisis, misal omset tertinggi pada kolom(X) adalah 3.600000e+08. Nilai (X) maksimal 3.600000e+08, tentunya masih wajar, demikian banyak Y maksimal 7.200000e+07 juga masih wajar dan disimpulkan data sudah benar.

```
dataset.isnull().sum()
#untuk melihat ada atau tidak missing values.
```

```
PRODUK      0
omset       0
pendapatan bersih  0
dtype: int64
```

```
+ Code + Markdown
```

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   PRODUK      150 non-null    object
1   omset       150 non-null    int64
2   pendapatan bersih  150 non-null    int64
dtypes: int64(2), object(1)
memory usage: 3.6+ KB
```

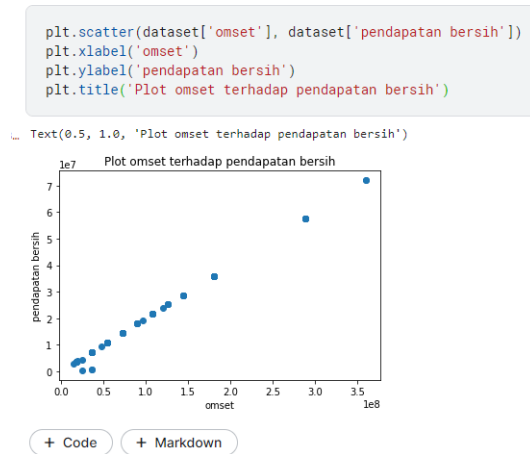
Gambar 5 Informasi datasheet

Pada gambar 5, menampilkan informasi dataset dan dapat disimpulkan banyak dataset sebanyak 150 data dan 3 kolom. Informasi pada gambar 5 menunjukkan semua data tidak ada data kosong dan tipe data dari masing-masing kolom. Kolom Provinsi mempunyai tipe object, kolom (omset) dan kolom (pendapatan bersih) mempunyai tipe integer.

3.4 Visualisasi Data

Visualisasi data merupakan proses yang mengubah data mentah menjadi informasi yang ditampilkan secara grafik. Visualisasi data dapat ditampilkan dalam bentuk scatter plot, box plot, histogram dan bentuk lainnya.

Scatter plot adalah Scatter Plot (scatterplot, scatter graph, scatter chart, scattergram, atau scatter diagram) merupakan sebuah tipe grafik yang digunakan untuk menggambarkan suatu data dengan menggunakan koordinat cartesian. Data-data yang di tampilkan menggunakan scatter plot direpresntasikan dengan titik yang terletak di antara 2 axis. Scatter Plot baik digunakan ketika kita ingin melihat relasi antara 2 variabel. Scatter Plot juga dapat digunakan untuk melihat trend jumlah suatu data lebih besar di variabel mana, (Gambar 6). Box plot adalah jenis visualisasi data yang secara statistik merepresentasikan distribusi data melalui lima dimensi utama, yaitu nilai minimum, kuartil 1, kuartil 2 (median), kuartil 3, dan nilai maksimum. Box plot digunakan untuk memeriksa keberadaan outlier dalam dataset. Histogram adalah jenis visualisasi data untuk merepresentasikan distribusi frekuensi dari dataset numerik menampilkan visualisasi data dalam bentuk scatter plot.



Gambar 6 Visualisasi data set

3.5 Analisis Regresi Linear

Perhitungan regresi linear dilakukan dengan penggunaan library sklearn dan library math. langkah - langkah yang dilakukan adalah menentukan kolom yang menjadi variabel dependen yaitu kolom (Y) dan kolom yang menjadi variabel independen yaitu (X).

Proses analisis regresi linear sederhana disajikan pada gambar 7, 8, 9 atau pada tabel 1

```
x = dataset['omset'].values.reshape(-1,1)
y = dataset['pendapatan bersih'].values.reshape(-1,1)
```

```
regressor = LinearRegression()
regressor.fit(x, y)
```

```
] LinearRegression()
```

```
+ Code + Markdown
```

Gambar 7 Menentukan kolom yang akan menjadi variabel dependen dan menjadi variabel independent

```
#Cara auto
#bisa pakai ini print(regressor.coef_) print(regressor.intercept_) print(regressor.coef_)
print(a)
```

```
[99]: b = regressor.coef_
print(b)
```

```
[[0.20158609]]
```

```
[100]: r = regressor.score(x, y)
print(r)
```

```
0.9961533377954608
```

Gambar 8 Menentukan nilai a, b dan r


```
[101]: regressor.predict([[7200000]])

[101]: array([[14250490.18508364]])

[102]: y_pred = regressor.predict

+ Code + Markdown
```

Gambar 9 Menghitung nilai y

Tabel 2. Perintah Python Dalam Proses Analisis Regresi Linear Sederhana

1	x = dataset['omset'].values.reshape(-1,1) y = dataset['pendapatan bersih'].values.reshape(-1,1)	Menentukan kolom yang akan menjadi variabel dependen dan menjadi variabel independen
2	regressor = LinearRegression() regressor.fit(x, y)	
3	a = regressor.intercept_ print(a)	Mencari dan menampilkan nilai intercept atau nilai a. Dari proses ini menghasilkan nilai a = [-263708.5485255]
4	b = regressor.coef_ print(b)	Mencari dan menampilkan nilai coef atau nilai b. Dari proses ini menghasilkan nilai b = [[0.20158609]]
6	r = regressor.score(x, y)	Mencari dan menampilkan nilai r. Dari proses ini menghasilkan nilai r = [[0.9961532337984608]]
7	print(r)	
8	regressor.predict([[masukan nilai x]])	Mencari dan menampilkan nilai prediksi atau y

Langkah pada tabel 1, melakukan proses perhitungan nilai inception dan nilai coef Hasil perhitungan ditampilkan pada tabel 2

Tabel 3. Nilai Interception & Nilai Coef

variabel	nilai
interception	-102.39422984
coef	0.99643956

Dari data pada tabel 2, model simple linear regression adalah

$$y = -102.39422984 + 0.99643956.X$$

Contoh perhitungan

$$y = -263708.5485255 + 0.20158609 * 7200000 \text{ (omset)}$$

$$= 14250490.18508364$$

Maka nilai pendapatan yang akan di dapat lebih besar atau sama dengan nilai y (14250490.18508364).

3.6 Perhitungan regresi linear menggunakan data training dan data test.

Proses pembagian data adalah 80% digunakan untuk data training dan 20% untuk data test. Proses analisis regresi linear sederhana disajikan pada gambar 10,11,12, atau pada tabel 3.

```
#MENGGUNAKAN data testing dan data train

[112]: print(reg.coef_)
print(reg.intercept_)

[[0.20259873]]
[-374359.94312876]

[113]: print('Correlation: ', math.sqrt(regressor.score(x_test,y_test)))

Correlation: 0.999546662293042

+ Code + Markdown
```

Gambar 10 Menentukan Data training sebanyak 80% dari jumlah data dan data testsebanyak 20%

```
[112]: print(reg.coef_)
print(reg.intercept_)

[[0.20259873]]
[-374359.94312876]

[113]: print('Correlation: ', math.sqrt(regressor.score(x_test,y_test)))

Correlation: 0.999546662293042

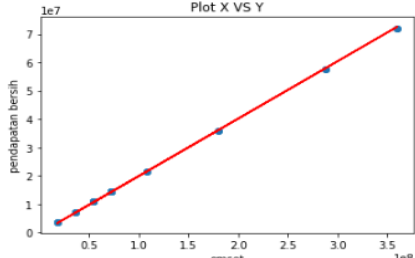
+ Code + Markdown
```

Gambar 11 Menentukan nilai a, b dan rdari data training dan data test

```
[114]: y_prediksi = reg.predict(x_test)

plt.scatter(x_test, y_test)
plt.plot(x_test, y_prediksi, c='r')
plt.xlabel('omset')
plt.ylabel('pendapatan bersih')
plt.title('Plot X VS Y')

[114]: Text(0.5, 1.0, 'Plot X VS Y')
```



+ Code + Markdown

Gambar 12 Menentukan scatter plot dari data training dan data test

```
[116]: from sklearn import metrics
print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_prediksi))
print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_prediksi))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_prediksi)))

Mean Absolute Error: 243269.2396642626
Mean Squared Error: 71893932506.72686
Root Mean Squared Error: 268138.49394812893

+ Code + Markdown
```

Gambar 13 menentukan niali mae, mse,danrmse, dari data training dan data test.

Tabel 4.

1	x = dataset['X'].values.reshape(-1,1)	Menentukan kolom yang akan menjadi variabel dependen dan menjadi variabel independen
2	y = dataset['Y'].values.reshape(-1,1)	
3	x_train, x_test, y_train, y_test = train_test_split(x, y, train_size = 0.8, test_size = 0.2, random_state = 0)	Menentukan dan membagi data-data menjadi data training dan data test. Data training sebanyak 80% dari jumlah data dan data test sebanyak 20%
4	reg = LinearRegression()	Mencari dan menampilkan nilai variabel dependen. Dari proses ini menghasilkan nilai coef
5	reg.fit(x_train, y_train)	
6	print(reg.coef_) print(reg.intercept_)	[[0.20259073]] dan nilai interception atau nilai a = [-374359.94312876]
7	print('Correlation: ', math.sqrt(regressor.score(x_test, y_test)))	Mencari dan menampilkan nilai korelasi, maka diperoleh nilai korelasi = 0.9999546662293942
8	y_prediksi = regressor.predict(x_test)	untuk menampilkan scatter plot X vs Y
9	plt.scatter(x_test, y_test) plt.plot(x_test, y_prediksi, c='r') plt.xlabel('omset') plt.ylabel('pendapatan bersih') plt.title('Plot X VS Y')	
#	np_table = np.concatenate((x_test, y_test, y_prediksi), axis=1)	untuk membuat dan menampilkan data baru dari hasil x test dan y test
#	new_dataframe = pd.DataFrame(data=np_table, columns=['x_test', 'y_test', 'y_prediksi'])	
#	new_dataframe	

#	from sklearn import metrics print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_prediksi)) print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_prediksi)) print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_prediksi))) Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_prediksi)))	Mean Absolute Error: 243269.23966642626
#		Mean Squared Error: 71893932500.72606
#		Root Mean Squared Error: 268130.43934012053
#	plt.title('Perbandingan nilai Y dalam pengujian dan nilai Prediksi')	menampilkan Perbandingan nilai Y dalam pengujian dan nilai Prediksi

3.7 Uji Korelasi dan Uji Koefisien Determinasi

Nilai Korelasi Untuk mengetahui seberapa besar keterkaitan masing-masing variabel bebas terhadap variabel tidak bebas maka perlu dihitung korelasi parsial. Uji Koefisien determinasi digunakan untuk mengetahui seberapa besar pengaruh variabel independen terhadap variabel dependen sehingga dapat diketahui kesamaan dan kecocokan model regresi linier. Berdasarkan tabel 5, proses perhitungan uji koefisien determinasi dan hasil perhitungan disajikan pada tabel 5.

Tabel 5.

Nilai Korelasi	
Coef	0.20259073
Correlation	0.9999546662293942
Mean Absolute Error:	243269.23966642626
Mean Squared Error:	71893932500.72606
Root Mean Squared Error:	268130.43934012053

Dari hasil perhitungan yang disajikan pada tabel 4 dan 5, dapat diketahui bahwa keterkaitan antara variabel dependen dengan variabel independen sangat kuat.

3.8 Pengujian Pada Data Prediksi

Pengujian dilakukan untuk melakukan perhitungan yang digunakan memprediksi hasil regresi linier dengan nilai y yang asli. Hasil pengujian dilakukan sebanyak 30 data. Hasil pengujian pada data prediksi

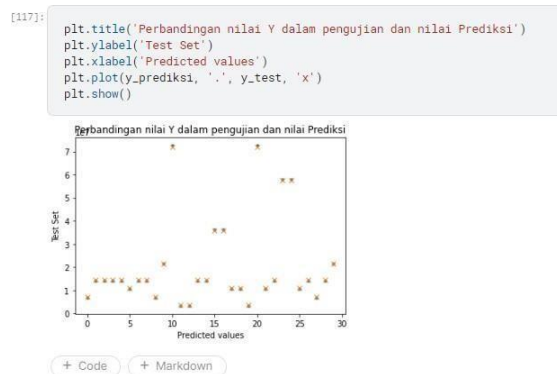
disajikan pada gambar 14. Data yang ditampilkan sebanyak 30 dari 150 data.

```
np_table = np.concatenate((x_test,y_test,y_prediksi), axis=1)
new_dataframe = pd.DataFrame(data=np_table, columns=['x_test','y_test','y_predict'])
new_dataframe
```

	x_test	y_test	y_predict
0	36000000.0	72000000.0	6.918906e+06
1	72000000.0	144000000.0	1.421217e+07
2	72000000.0	144000000.0	1.421217e+07
3	72000000.0	144000000.0	1.421217e+07
4	72000000.0	144000000.0	1.421217e+07
5	54000000.0	108000000.0	1.056554e+07
6	72000000.0	144000000.0	1.421217e+07
7	72000000.0	144000000.0	1.421217e+07
8	36000000.0	72000000.0	6.918906e+06
9	108000000.0	216000000.0	2.150544e+07
10	360000000.0	720000000.0	7.255830e+07
11	180000000.0	360000000.0	3.272273e+06
12	180000000.0	360000000.0	3.272273e+06
13	72000000.0	144000000.0	1.421217e+07
14	72000000.0	144000000.0	1.421217e+07
15	180000000.0	360000000.0	3.609197e+07
16	180000000.0	360000000.0	3.609197e+07
17	54000000.0	108000000.0	1.056554e+07
18	54000000.0	108000000.0	1.056554e+07
19	180000000.0	360000000.0	3.272273e+06
20	360000000.0	720000000.0	7.255830e+07
21	54000000.0	108000000.0	1.056554e+07
22	72000000.0	144000000.0	1.421217e+07
23	288000000.0	576000000.0	5.797177e+07
24	288000000.0	576000000.0	5.797177e+07
25	54000000.0	108000000.0	1.056554e+07
26	72000000.0	144000000.0	1.421217e+07
27	360000000.0	720000000.0	6.918906e+06
28	72000000.0	144000000.0	1.421217e+07
29	108000000.0	216000000.0	2.150544e+07

Gambar 14

Uji Hasil pada gambar 14 antara data Y test dan Y predicted. ada perbedaan hasil dan hasil tampilan dalam bentuk scatter plot ditampilkan pada gambar 15.



Gambar 15

Salah satu bagian dari machine learning adalah proses mencari prediksi dengan menggunakan regresi linier ganda. Penelitian yang dilakukan adalah membuat simulasi penerapan machine learning terutama regresi linier sederhana dengan menggunakan python dan menggunakan Kaggel notebook. Implementasi machine learning regresi linier diterapkan pada Pengaruh Nilai Omset Terhadap Besaran Pendapatan Bersih Usaha Mikro Kecil Dan Menengah dipengaruhi oleh omset dan pendapat bersih. Berdasarkan hasil uji korelasi antara omset dan pendapatan bersih 0.99. Hasil ini menunjukkan omset mempunyai korelasi yang tinggi dengan pendapatan bersih dan dapat diprediksi pelaku umkm yang mempunyai nilai omset yang tinggi akan mendapat pendapatan bersih yang tinggi, semakin pendapatan bersih maka dapat diprediksi biaya omset juga bertambah.

Penggunaan bahasa Python dalam implementasi machine learning, khususnya analisis regresi linier sederhana dapat diimplementasikan dengan mudah dan tidak memerlukan koding yang rumit. Hal ini karena dukungan library di Python yang banyak dan pengguna tinggal menyesuaikan library-library yang digunakan.

DAFTAR PUSTAKA

- [1] E. Nanda Utami and H. Dwi Mulyaningsih, “Pengaruh Kompetensi Kewirausahaan Terhadap Kinerja Umkm,” *Bisnis dan Iptek*, vol. 9, no. 2, pp. 98–109, 2016.
- [2] D. Sartika, “Pengaruh Modal Terhadap Omset Pada Pelaku Umkm Di Seluruh Kecamatan Pekanbaru,” *J. Daya Saing*, vol. 4, no. 2, pp. 234–239, 2018.
- [3] D. M. Pangesti and A. N. Yushita, “Pengaruh Kesadaran Membayar Pajak, Persepsi Atas Efektivitas Sistem Perpajakan, Dan Pemahaman Peraturan Pemerintah Nomor 23 Tahun 2018 Terhadap Kemauan Membayar Pajak (Pada Umkm Sektor Perdagangan Di Kabupaten Klaten),” *Nominal Barom. Ris. Akunt. dan Manaj.*, vol. 8, no. 2, pp. 166–178, 2019.
- [4] N. A. Sari, “Pengaruh Perkembangan Ekonomi Digital Terhadap Pendapatan Pelaku Usaha Umkm Di Kota Makassar,” *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2019.
- [5] F. Ginting, E. Buulolo, and E. R. Siagian, “Implementasi Algoritma Regresi Linear Sederhana Dalam Memprediksi Besaran Pendapatan Daerah (Studi Kasus: Dinas Pendapatan Kab. Deli Serdang),” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 274–279, 2019.
- [6] E. T. Novalyn, G. Ginting, and H. K. Siburian, “Pemanfaatan Metode Cart dalam Memprediksi Omset Pakaian Pria Remaja (Studi Kasus : PT. Matahari Departement Store Thamrin Plaza Medan),” *J. Pelita Inform.*, vol. 7, no. 2, pp. 199–206, 2018.
- [7] S. Adiguno, Y. Syahra, and M. Yetri, “Prediksi Peningkatan Omset Penjualan Menggunakan Metode Regresi Linier Berganda,” *J. Sist. Inf. TGD*, vol. 1, no. 4, pp. 275–281, 2022.
- [8] M. Sholeh, S. Suraya, and D. Andayati, “Machine Linear untuk Analisis Regresi Linier Biaya Asuransi Kesehatan dengan Menggunakan Python Jupyter Notebook,” *JEPIN (Jurnal Edukasi dan Penelit. Inform.)*, vol. 8, no. 1, pp. 20–27, 2022.
- [9] T. Syahputra, J. Halim, and K. Perangin-angin, “Penerapan Data Mining Dalam Memprediksi Tingkat Kelulusan Uji Kompetensi (UKOM) Bidan Pada STIKes Senior Medan Dengan Menggunakan Metode Regresi Linier Berganda,” *Sains dan Komput.*, vol. 17, no. 1, pp. 1–7, 2018.
- [10] Supriyanta and N. Hasan, “Data Mining dengan Metode Regresi Linier untuk Melakukan Prediksi,” *Semin. Nas. Inov. dan Teknol.*, pp. 215–220, 2013.