

Peramalan Pendapatan dari Penjualan Bawang Merah Menggunakan Metode Regresi Linier Berganda

Agusta Aldhi Saputra¹, Misbakhul Munir², Zamima Daffa Rizki A.P.³

^{1,2,3}Teknik Informatika, Fakultas Teknik, Universitas Nusantara PGRI Kediri

E-mail: ¹aldhiagst@gmail.com, ²misbakhulm187@gmail.com, ³Rizkidaffa890@gmail.com

Abstrak – Peramalan (*forecasting*) adalah suatu seni dan ilmu pengetahuan dalam memprediksi peristiwa pada masa mendatang. Peramalan akan melibatkan pengambilan data historis (penjualan tahun lalu) dan memproyeksi mereka ke masa yang akan datang dengan model matematika. Peramalan penjualan merupakan bagian penting dari manajemen rantai pasokan baik pada pengecer akhir dan distributor, manufaktur dan pemasok. Perlunya mengetahui peralaman jumlah pendapatan dari penjualan bawang merah adalah untuk menentukan strategi pemasaran yang tepat, mengoptimalkan kapasitas produksi, dan meningkatkan keuntungan. Kami mengeksplorasi metode regresi linier berganda untuk memprediksi jumlah pendapatan penjualan bawang merah. Kami menggunakan data historis tentang harga bawang merah di pasar, jumlah penjualan, dan faktor-faktor lain yang mungkin mempengaruhi pendapatan penjualan bawang merah. Setelah melakukan pemodelan dan validasi, kami menemukan bahwa metode ini memberikan tingkat akurasi sebesar 60 persen dan nilai mape sebesar 13 persen. Hasil ini menunjukkan bahwa metode regresi linier berganda dapat digunakan dengan efektif untuk memprediksi pendapatan penjualan bawang merah. Selanjutnya, kami menyarankan agar penelitian lebih lanjut dilakukan untuk meningkatkan tingkat akurasi dan mengevaluasi keandalan metode ini pada data yang lebih luas.

Kata Kunci — Mean Absolute Percentage Error (MAPE) , Regresi Linier Berganda.

1. PENDAHULUAN

Bawang merah adalah salah satu jenis bawang yang banyak digunakan sebagai bahan dasar masakan di seluruh dunia. Bawang merah memiliki rasa yang khas dan kandungan gizi yang tinggi, sehingga menjadi bahan yang sangat diandalkan dalam industri makanan. Di Jawa Timur, produksi bawang merah tahun 2019 sebesar 407,8 ribu ton, meningkat sebanyak 40,8 ribu ton atau 11.13 % dari tahun 2018[1].

Penting bagi petani bawang merah untuk meramalkan jumlah panen bawang merah yang akan dihasilkan di masa yang akan datang. Dengan demikian, petani dapat mengambil keputusan yang tepat dalam mengelola tanaman bawang merah, seperti memilih teknik pertanian yang tepat atau menentukan jumlah bibit yang akan ditanam[2].

Penelitian ini bertujuan untuk meramalkan jumlah panen bawang merah menggunakan metode regresi linier berganda. Regresi linier berganda adalah salah satu metode statistik yang digunakan untuk memprediksi variabel terikat (variabel yang ingin diprediksi) berdasarkan beberapa variabel bebas (variabel yang mungkin mempengaruhi variabel terikat)[4][5]. Regresi linier berganda merupakan versi yang lebih kompleks dari regresi linier sederhana, dimana regresi linier sederhana hanya menggunakan satu variabel bebas untuk memprediksi variabel terikat[7].

Dengan memprediksi jumlah panen bawang merah, petani dapat memperkirakan berapa banyak bawang merah yang akan tersedia untuk dipasarkan dan mengambil keputusan tentang strategi pemasaran yang tepat. Selain itu, penelitian ini juga dapat membantu pemerintah dalam mengelola sumber daya alam dan mengatur kebijakan pertanian yang efektif.

Hasil dari penelitian ini diharapkan dapat memberikan rekomendasi yang bermanfaat bagi petani bawang merah dalam mengelola tanaman bawang merah dan memprediksi jumlah panen yang akan dihasilkan di masa yang akan datang. Dengan demikian, diharapkan dapat membantu meningkatkan produksi bawang merah dan meningkatkan keuntungan di masa yang akan datang.

2. METODE PENELITIAN

2.1 Regresi Linier Berganda

Regresi Linear Berganda adalah model regresi linear dengan melibatkan lebih dari satu variable bebas atau predictor. Dalam bahasa inggris, istilah ini disebut dengan multiple linear regression.

Model regresi linear berganda dilukiskan dengan persamaan sebagai berikut:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_n X_n + \varepsilon \dots\dots\dots(1)$$

Keterangan:

- Y = Variabel terikat atau variabel response.
- X = Variabel bebas atau variabel predictor.
- α = Konstanta.
- β = Slope atau Koefisien estimate.

2.2 Mean Absolute Percentage Error (MAPE)

MAPE adalah kesalahan persentase rata-rata absolut. Definisi Mean Absolute Percentage Error adalah ukuran statistik dari akurasi prediksi dari metode peramalan. Pengukuran Mean Absolute Percentage Error (MAPE) dapat digunakan oleh masyarakat luas karena MAPE mudah dipahami dan digunakan untuk memprediksi keakuratan ramalan. Metode MAPE (Mean Absolute Percentage Error) memberikan informasi seberapa besar kesalahan peramalan dibandingkan dengan nilai deret yang sebenarnya. Semakin kecil nilai persentase error pada MAPE, maka semakin akurat hasil prediksinya.

Mean Absolute Percentage Error (MAPE) dapat dirumuskan sebagai berikut

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| 100}{n} \dots\dots\dots(2)$$

Keterangan

- At = Aktual permintaan ke t
- Ft = hasil peramalan ke t
- N = besarnya data peramalan

Terdapat simbol absolut pada rumus MAPE, yang menunjukkan bahwa hasil nilai negatif dari perhitungan akan tetap bernilai positif.

2.3 Dataset

Dataset adalah kumpulan objek dan fitur atau karakteristik (atribut) dari objek itu sendiri. Dataset yang digunakan pada penelitian ini yaitu data panen bawang merah selama 8 tahun terakhir. Data tersebut merupakan hasil wawancara dari salah satu petani di kota Kediri, yang kemudian kita olah menjadi dataset yang bisa digunakan untuk melakukan peramalan (*forecasting*). Untuk dataset bisa dilihat pada tabel berikut.

Tabel 1. Dataset

<i>panen_ke</i>	<i>Tanggal</i>	<i>Jumlah Panen</i>	<i>Harga Per Kg</i>	<i>Total Pendapatan</i>
1	Februari 2014	1093	8000	8744000
2	April 2014	1430	11000	15730000
3	Juni 2014	1470	10000	14700000
4	Agustus 2014	1268	13000	16484000
5	Januari 2015	1447	14000	20258000
6	Maret 2015	1049	10000	10490000
7	Juli 2015	1024	15000	15360000
8	September 2015	1481	7000	10367000
9	Januari 2016	1527	24000	36648000
10	April 2016	1514	32000	48448000
11	Juli 2016	1588	23000	36524000
12	September 2016	1316	25000	32900000

13	Januari 2017	1000	25000	25000000
14	Maret 2017	1309	30000	39270000
15	Juni 2017	1064	20000	21280000
16	Agustus 2017	1216	20000	24320000
17	Januari 2018	1581	12500	19762500
18	Maret 2018	1014	17500	17745000
19	Juni 2018	1314	15000	19710000
20	Agustus 2018	1373	15000	20595000
21	Januari 2019	1040	25000	26000000
22	Maret 2019	1203	20000	24060000
23	Juli 2019	1564	17000	26588000
24	September 2019	1215	17000	20655000
25	Januari 2020	1096	30000	32880000
26	Maret 2020	1397	30000	41910000
27	Juni 2020	1296	15000	19440000
28	Agustus 2020	1316	20000	26320000
29	Frebuari 2021	1537	17000	26129000
30	April 2021	1122	10000	11220000
31	Juli 2021	1220	7000	8540000
32	September 2021	1533	12000	18396000
33	Januari 2022	1398	18000	25164000
34	Maret 2022	1590	18000	28620000
35	Juni 2022	1216	14000	17024000
36	September 2022	1183	14000	16562000
<i>Total</i>		<i>47004</i>		<i>823843500</i>

Pada tabel diatas, variable yang digunakan yaitu, panen ke, tanggal, jumlah panen, harga jual per kg, lalu total pendapatan dari penjualan bawang merah tersebut.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Data Set Penelitian

3.1.1 Import Dataset

Dataset penelitian dengan nama “data_brambang.csv” terlebih dahulu dilakukan proses input di google colab. Dataset tersebut di import, lalu dipilih kolom mana saja yang akan digunakan dan ditampilkan di sistem.

Mengimport dataset yang digunakan. Disini nama dataset yang digunakan adalah "data_brambang.csv" dan memilih kolom yang akan digunakan. Setelah itu menampilkan 5 baris pertama dari dataset.

```
[2] df = pd.read_csv('data_brambang.csv', usecols=['jumlah_panen(kg)', 'harga_per_kg(rp)', 'total_pendapatan'])
df.head()
```

	jumlah_panen(kg)	harga_per_kg(rp)	total_pendapatan
0	1000	8000	8744000
1	1400	11000	15730000
2	1400	10000	14700000
3	1200	13000	16484000
4	1400	14000	20258000

Gambar 1. Import Dataset

Pada penelitian ini, kolom yang digunakan untuk melakukan proses peramalan yaitu kolom 'jumlah_panen(kg)' yaitu kolom yang menunjukkan total produksi bawang merah saat panen dalam kilogram, lalu kolom 'harga_per_kg(rp)' merupakan harga bawang merah per kilogram pada masa panen saat itu, kolom 'total_pendapatan' merupakan kolom yang berisi total pendapatan dari hasil penjualan bawang merah dalam rupiah.

3.1.2 Cek Data

Melihat informasi data dan mengecek apakah data kita terdapat missing value.

```
Melihat informasi data

[3] df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36 entries, 0 to 35
Data columns (total 3 columns):
#   Column          Non-Null Count  Dtype
---  ---            -
0   jumlah_panen(kg) 36 non-null     int64
1   harga_per_kg(rp) 36 non-null     int64
2   total_pendapatan 36 non-null     int64
dtypes: int64(3)
memory usage: 992.0 bytes

Mengecek jika di data kita ada missing value

[4] df.isnull().sum()

jumlah_panen(kg)    0
harga_per_kg(rp)    0
total_pendapatan    0
dtype: int64
```

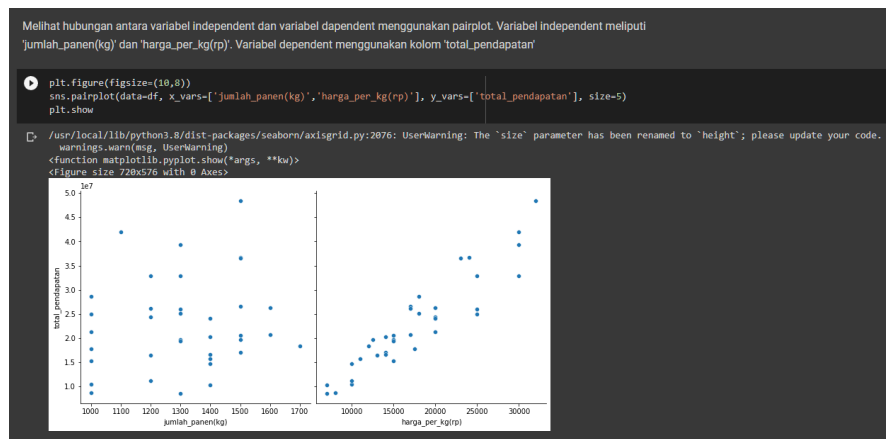
Gambar 2. Cek Data

Dari hasil pengecekan data diatas, data kita berjumlah 36 baris dengan tipe data adalah integer, serta tidak terdapat missing value dan dapat digunakan untuk melakukan peramalan pada proses selanjutnya.

3.1.3 Bivariate analysis

Bivariate analysis adalah analisis yang melibatkan satu variabel variabel dependen dan independen. *Bivariate analysis* dapat digunakan untuk melihat apakah ada perbedaan atau signifikansi statistik dari hasil tabulasi silang 2x2. Tingkat *confidence* yang digunakan adalah 95% dan $P < 0,05$ yang berarti hipotesis diterima jika $P < 0,05$ dan atau *Confidence Interval (CI)* tidak mengandung angka satu.

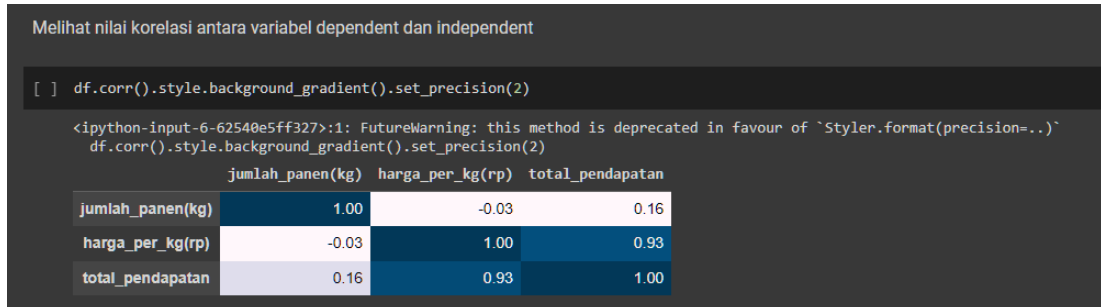
Dengan *bivariate analysis* kita bisa mengetahui hubungan antara *independent* dan *dependent variable* menggunakan *pairplot*.



Gambar 3. Bivariate Analysis

Berdasarkan *bivariate analysis* diatas yang mana variabel independent atau 'x_vars' menggunakan kolom 'jumlah_panen(kg)' dan 'harga_per_kg(rp)', sedangkan variabel dependen atau 'y_vars' menggunakan kolom 'total_pendapatan', menghasilkan bahwa korelasi antar kolom tersebut bernilai positif.

Korelasi linear paling kuat terdapat pada kolom 'harga_per_kg(rp)' dengan kolom 'total_pendapatan', seperti yang tertera pada gambar 3.4 berikut.

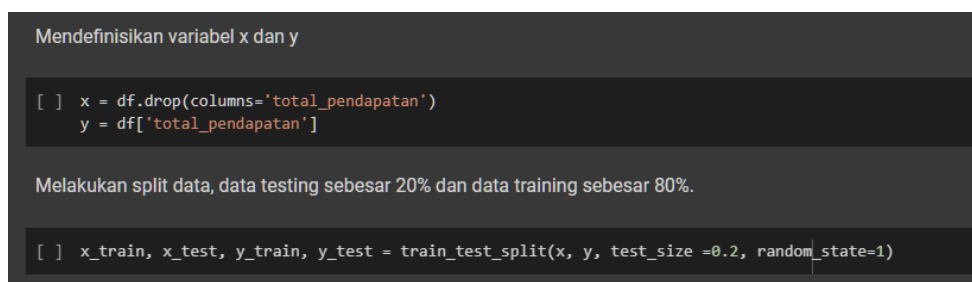


Gambar 4. Korelasi Paling Kuat

Dari tabel korelasi diatas, kolom 'harga_per_kg(rp)' memiliki korelasi yang sangat kuat dengan kolom 'total_pendapatan' dengan nilai 0.93.

3.2 Split Data

Dataset dibagi menjadi data training (data latih) sebesar 80% dan data testing (data uji) sebesar 20% dari 36 data yang digunakan.



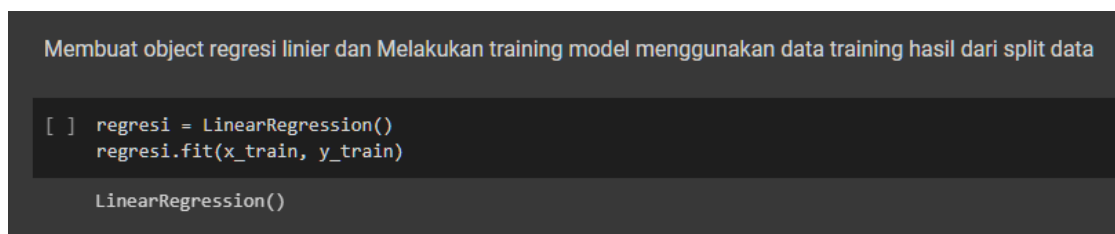
Gambar 5. Split Data

Pada proses split data diatas, kita definisikan dulu variabel x dan y. Untuk variabel x kita mendrop kolom yang tidak digunakan dalam variabel x yaitu kolom 'total_pendapatan'. Sedangkan untuk variabel y kita menggunakan kolom 'total_pendapatan'.

3.3 Peramalan dengan Algoritma yang digunakan

3.3.1 Membuat Model Algoritma Regresi

Membuat model regresi linier dan melakukan training model menggunakan data training hasil dari split data sebelumnya.



Gambar 6. Model Regresi Linier

Gambar diatas merupakan pembuatan model regresi linier dan pengujian model menggunakan data training.

3.3.2 Prediksi dengan data testing

Sebelum melakukan prediksi yang sesungguhnya, kita melakukan prediksi dengan menggunakan data testing, hasil dari split data.

```
[ ] y_pred = regresi.predict(x_test)
print(y_pred)

[ 9384191.89745551 20709113.36954853 21361586.29568085 16134962.55219289
22015769.30542052 19836584.34262774 27459884.02428592 25718246.13765905]
```

Gambar 7. Prediksi dengan Data Testing

Hasil dari prediksi menggunakan data testing, dapat dilihat pada gambar 3.7. Terdapat hasil prediksi total pendapatan dari penjualan bawang merah dari data testing yang digunakan.

3.3.3 Skor Akurasi Model

Untuk menampilkan skor akurasi dari model yang telah dibuat, menggunakan modul ‘.score’.

```
Menampilkan skor akurasi dari model yang telah dibuat

[ ] regresi.score(x_test, y_test)

0.6020262725691172
```

Gambar 8. Skor Akurasi

Pada gambar diatas, skor akurasi yang didapat sebesar 60%. Jadi bisa disimpulkan bahwa model yang telah dibuat bisa digunakan untuk melakukan peramalan.

3.3.3 Peramalan pada Masa Panen Selanjutnya

Variabel input yang digunakan untuk melakukan peramalan selanjutnya menggunakan 2 buah variabel yaitu variabel total hasil panen dan harga bawang merah per kilogram.

```
Meramalkan pendapatan pada panen berikutnya

[ ] regresi.predict([[1500, 10000]])

/usr/local/lib/python3.8/dist-packages/sklearn/base.py:450: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
warnings.warn(
array([15482489.6260658])

Jadi pada panen berikutnya total pendapatan yang diperoleh dari penjualan bawang merah adalah Rp.15.482.489
```

Gambar 9. Hasil Prediksi

Hasil dari peramalan pada masa panen selanjutnya menunjukkan total pendapatan yang didapat sebesar Rp. 15.482.489.

3.4 Pengetesan Model dengan Mean Absolute Percentage Error (MAPE)

Pengetesan menggunakan MAPE bertujuan untuk mengetahui rata-rata persentase kesalahan pada model yang telah dibuat.

```
Pengtesan Model Peramalan dengan Mean Absolute Percentage Error (MAPE)

[ ] def MAPE(y_actual,y_predicted):
    mape = np.mean(np.abs((y_actual - y_predicted)/y_actual))*100
    return mape

[ ] regresi_mape= MAPE(y_test,y_pred)
print("Nilai MAPE: ",regresi_mape)

Nilai MAPE: 13.65307745956596

Nilai MAPE yang dihasilkan sebesar 13.65. Jadi berdasarkan hasil MAPE diatas, bisa diblilng tingkat error pada model yang dibuat sebesar 13% dan nilai tersebut terblilng rendah. Lalu bisa disimpulkan bahwa kemampuan model peramalan tersebut baik.
```

Gambar 10. MAPE

Hasil dari pengetesan model menggunakan MAPE memperoleh nilai sebesar 13.65, atau 13%. Angka tersebut menunjukkan bahwa model yang telah dibuat memiliki kesalahan yang kecil dan dapat disimpulkan bahwa kemampuan dari model peramalan yang telah dibuat tersebut baik.

4. SIMPULAN

Dari hasil peramalan yang sudah dilakukan, dapat disimpulkan bahwa metode regresi linier berganda cukup efektif dalam memprediksi jumlah pendapatan penjualan bawang merah, namun masih terdapat sekitar 40 persen kesalahan dalam peramalannya. Nilai mape sebesar 13 persen menunjukkan bahwa terdapat sekitar 13 persen perbedaan antara nilai prediksi dengan nilai sebenarnya..

5. SARAN

Untuk meningkatkan akurasi peramalan, diusulkan agar dilakukan analisis terhadap faktor-faktor yang mempengaruhi jumlah pendapatan penjualan bawang merah, seperti harga pasar, kondisi cuaca, dan faktor lainnya. Selain itu, diusulkan juga untuk melakukan penelitian lebih lanjut untuk meningkatkan tingkat akurasi dan mengevaluasi keandalan metode ini pada data yang lebih luas.

DAFTAR PUSTAKA

- [1] Produksi Bawang Merah Menurut Provinsi, Tahun 2015-2019. <https://www.pertanian.go.id/home/index.php?show=repo&fileNum=286>. Diakses pada tanggal 20 Desember 2022.
- [2] Amankan Produksi Bawang Merah, Petani Pati Terapkan Budidaya Ramah Lingkungan <https://www.pertanian.go.id/home/?show=news&act=view&id=4393>. Diakses pada tanggal 20 Desember 2022.
- [3] Prasetya, Hery dan Lukiasuti Fitri. (2009). Manajemen Operasi. Yogyakarta:Media Presindo.
- [4] Vining, G.G. 1998. Statistical Methods for Engineers. Duxbury Press. Pacific Grove, CA
- [5] Walpole, R.E.; Myers, R.H.; Myers, S.L.; and Ye, K. 2007. Probability and Statistics for Engineers & Scientists. Pearson Prentice Hall. Upper Saddle River, NJ
- [6] Y. Moon and T. Yao, "A robust mean absolute deviation model for portfolio optimization," Comput. Oper. Res., vol. 38, no. 9, pp. 1251–1258, 2011.
- [7] Yuliara, I. M. 2016. Regresi Linier Berganda. Denpasar: Universitas Udayana.