

IMPLEMENTASI DATA MINING UNTUK PREDIKSI PRODUKSI BAWANG MERAH DI KABUPATEN BREBES

Suyono¹, Rini Astuti², Fadhil M. Basysyar³

^{1,3} Teknik Informatika, STMIK IKMI Cirebon

² Sistem Informasi, STMIK LIKMI Bandung

Jalan Perjuangan No. 10 B, Majasem, Cirebon Indonesia

soeyono199@gmail.com

ABSTRAK

Kabupaten Brebes merupakan lokasi utama di Indonesia yang menghasilkan bawang merah dalam jumlah besar. Bawang merah Brebes memiliki ciri bentuk lebih kering dan tahan busuk serta rasanya lebih pedas daripada bawang dari daerah lain menjadikan keunggulan bawang merah dari Brebes ini. Dinas Pertanian dan Ketahanan Pangan (DPKP) Kabupaten Brebes menyebutkan bahwa luas panen pada tahun 2022 mencapai 32,571 ha dengan jumlah total produksi mencapai 384,448 ton. Ketidakpastian hasil produksi bawang merah membuat harga bawang merah menjadi tidak stabil dan cenderung melonjak. Oleh karena itu, dengan memprediksi hasil produksi bawang merah akan membantu petani dalam melakukan manajemen pertanian bawang merah. Dalam penelitian ini, akan di terapkan metode regresi linier, dimana metode ini dapat menguji suatu hubungan dari variable yang satu dengan variable yang lain, dan memprediksi guna keperluan di bidang pertanian, sumber data yang digunakan berasal dari dataset Dinas Pertanian dan Ketahanan Pangan (DPKP) Kabupaten Brebes. Kemudian, membagi dua data menjadi data *training* sebesar 80% dan 20% untuk data *testing*, menghasilkan hasil prediksi jumlah produksi bawang merah pada tahun 2024 sebanyak 3.962.376 Kuintal, mengalami kenaikan daripada tahun 2022. Dengan nilai *Mean Absolute Error (MAE)* sebesar : 30116.24, *Mean Squared Error (MSE)* : 2989724823.94, *Root Mean Squared Error (RMSE)* : 54678.38 serta nilai *R2-Score* : 0.948.

Kata kunci : Kabupaten Brebes, Bawang Merah, Prediksi, *Data Mining*, *Regresi Linier Berganda*

1. PENDAHULUAN

Kabupaten Brebes terkenal menjadi daerah penghasil bawang merah yang unggul di Indonesia. Dengan memiliki 17 kecamatan dengan luas wilayah 1.770 km², dengan mata pencaharian penduduknya sebagian besar menjadi petani bawang merah, tidak heran jika Kabupaten Brebes dijuluki kota bawang. Dinas Pertanian dan Ketahanan Pangan (DPKP) Kabupaten Brebes menyebutkan bahwa luas panen pada tahun 2022 mencapai 32,571 ha dengan jumlah total produksi mencapai 384,448 ton, ini menjadikan Kabupaten Brebes menjadi penghasil bawang merah terbesar di Indonesia. Seiring kemajuan teknologi yang semakin masif, memaksa sektor pertanian khususnya petani bawang merah untuk lebih mendalami teknologi untuk bisnis.

Namun, pada praktiknya teknologi belum benar-benar dimanfaatkan dengan baik di bidang pertanian khususnya untuk pertanian bawang merah yang berada di Kabupaten Brebes, ini menjadi masalah sangat serius dikarenakan kebutuhan bawang merah dalam dan luar negeri akan terus meningkat seiring berjalannya waktu, serta ketidakstabilan hasil produksi membuat harga bawang merah tidak stabil dan cenderung naik. Dengan rata-rata luas wilayah untuk penanaman bawang merah mengalami penurunan yang disebabkan oleh pembangunan berbagai infrastruktur yang ada di Kabupaten Brebes, ini menyebabkan masalah keberlangsungan bawang merah. Curah hujan juga sangat mempengaruhi kualitas bawang merah, ini sangat penting untuk menjaga kualitas bawang merah itu sendiri agar tetap

disukai oleh pasar dan tetap menjadi primadona. Dengan berbagai masalah tersebut, penerapan prediksi untuk produksi bawang merah diharapkan bisa memberikan dampak yang positif untuk keberlangsungan bawang merah dari Brebes ini dengan memperhatikan berbagai aspek yang ada.

Dengan melakukan kajian yang terdahulu, penulis melakukan kajian penelitian terdahulu [1] dengan judul memprediksi jumlah penduduk Jakarta Selatan dengan memakai Regresi Linier Berganda dengan menggunakan data dari tahun 2010 hingga 2018 dengan 2 variable yang kemudian menghasilkan peningkatan jumlah penduduk pada tahun 2019 sebesar 4.195 jiwa. Kemudian penelitian [2] dengan judul penelitian memprediksi produksi jagung dengan menggunakan 2 metode yakni algoritma apriori dan regresi linier berganda yang kemudian menghasilkan beberapa angka yang menghasilkan data dari masing-masing metode. Selanjutnya penelitian [3] dengan judul implementasi regresi linier berganda guna memprediksi produksi padi pada Kabupaten Bantul dengan menggunakan dataset penelitian tahun 2009 hingga 2017 menghasilkan mean absolute deviation (MAD) sebesar 0,101 dengan persamaan $Y = 8307,561443282 + 5,9294543706657x_1 + 118,28063200866x_2 + 175,71009241484x_3$.

Penelitian ini bertujuan melakukan prediksi jumlah produksi di bidang pertanian khususnya petani bawang merah dengan menggunakan metode regresi linier dengan menggunakan dataset yang sudah didapat dari Dinas Pertanian dan Ketahanan Pangan (DPKP) Kabupaten Brebes. Kemudian diharapkan

dengan adanya penelitian ini bisa memberikan dampak positif untuk masyarakat terkhusus pada bidang informatika dan memberikan hasil yang bermanfaat untuk pertanian bawang merah di Kabupaten Brebes.

Di dalam penelitian ini, digunakan metode estimasi regresi linear untuk menjelaskan keterkaitan antara dua variabel menggunakan garis lurus. Regresi linear secara ringkas menggambarkan hubungan sebab-akibat. Penggunaan metode ini memiliki keunggulan dalam memprediksi ketidakpastian, menyelidiki pola dalam data serta meningkatkan kecepatan perhitungan dengan menerapkan metode perhitungan yang berjalan secara simultan atau paralel. Penelitian ini akan menggunakan 3 variabel independent dan 1 data dependent[4].

Hasil dari penelitian ini ialah sebuah prediksi atau estimasi hasil jumlah produksi bawang merah yang berada di Kabupaten Brebes, dimana diharapkan bisa memberikan hasil yang positif terutama soal bagaimana pengambilan keputusan di dalam bidang pertanian dan memaksa petani bawang merah berinovasi terutama di soal teknologi, ini berguna untuk terus meningkatkan hasil produksi bawang merah di atas permasalahan yang begitu kompleks terutama soal cuaca, hama, lahan pertanian yang terus tergerus dan petani bawang merah tetap harus memperhatikan kualitas mutu dari bawang merah tersebut agar tetap jadi yang terbaik.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Paper [5] Artikel ini membicarakan Prediksi Produksi Kelapa Sawit menggunakan Metode Regresi Linier Berganda dengan variabel predictor termasuk bulan, curah hujan, usia lahan, jumlah pokok, jumlah tandan, rata-rata berat, dan hasil produksi kelapa sawit sebagai variabel terikat. Dataset bermula dari PT. Perkebunan Nusantara I serta data eksternal, seperti curah hujan dari Badan Meteorologi, Klimatologi, dan Geofisika. Terdapat 180 data untuk pelatihan dan 20% dari data pelatihan digunakan untuk pengujian. Hasil penelitian mengungkap persamaan Regresi Linier Berganda: $Y = -415337,95 + 1073,82208X_1 + 3736,68741X_2 + -15306,629X_3 + -621,89932X_4 + 11,7449262X_5 + 7,47948459X_6 + 33441,5621X_7$ dengan nilai MAPE sekitar 14.28%.

Paper [6] artikel membahas estimasi produksi daging sapi menggunakan metode regresi linier serta regresi polinomial. Penelitian bertujuan untuk meramalkan produksi daging sapi dalam skala nasional guna memahami hasil produksi di masa depan. Hasil prediksi dari kedua metode dibandingkan untuk menentukan model terbaik berdasarkan dataset yang mencakup tahun 1984 hingga 2019. Hasil prediksi menunjukkan produksi tahun 2020 sebesar 531.367 ton, tahun 2021 sebanyak 538.567 ton, kemudian tahun 2022 sebesar 545.343 ton. Hasil menunjukkan bahwa model terbaik dalam penelitian ini ialah menggunakan metode regresi polinomial orde 4, dibuktikan dengan perbandingan hasil prediksi

menggunakan nilai $R^2 = 0.9217$, $MSE = 80.549.907$, dan $RMSE = 8.975$

2.2. Data Mining

Data mining adalah proses berjenjang untuk mengungkap informasi baru dari volume data besar yang sebelumnya tidak diketahui. Proses ini melibatkan pengidentifikasian dan pemisahan informasi dari data menggunakan metode yang menghasilkan informasi bernilai. Dalam penelitiannya, data mining memanfaatkan teknik-teknik matematika, statistika, machine learning, dan kecerdasan buatan untuk memperoleh informasi yang tepat dari beragam jenis data yang besar[7].

Data besar lalu diolah menggunakan teknik pemisahan dan identifikasi menghasilkan informasi yang bernilai. Dalam proses penelitian, data mining mengadopsi teknik-teknik matematika, statistika, machine learning, dan kecerdasan buatan untuk mendapatkan informasi yang akurat dari berbagai jenis data yang berskala besar.

2.3. Prediksi

Memprediksi melibatkan keahlian dan pengetahuan untuk menaksir kejadian di masa mendatang. Ini bisa dilakukan dengan menganalisis data masa lalu dan mengekstrapolasikannya ke masa yang akan datang menggunakan model matematika (dalam hal kuantitatif), atau dengan menggunakan naluri yang lebih subjektif (dalam hal kualitatif)[6].

2.4. Regresi Linier Berganda

Regresi linier memiliki dua bentuk, yakni regresi linier sederhana dan regresi linier berganda. Regresi linier berganda digunakan untuk mengeksplorasi pola keterkaitan antara variabel terikat dan dua atau lebih variabel bebas. Menurut suatu penelitian, regresi linier berganda terbukti lebih efektif daripada metode fuzzy dan jaringan syaraf tiruan[8].

Metode ini adalah pendekatan analisis yang melibatkan variabel bebas lebih dari satu dan bertujuan untuk menentukan apakah ada pengaruh yang signifikan dari dua atau lebih variabel bebas ($x_1, x_2, \dots, x_n$) terhadap variabel terikat (Y). Metode ini sangat cocok digunakan untuk memprediksi dengan variabel bebas lebih dari satu. Penjelasan mengenai model regresi linear berganda disampaikan melalui rumus atau persamaan yang sesuai. Secara umum, persamaan untuk regresi linier berganda dinyatakan sebagai berikut[9]:

$$Y = a + b_1 + X_1 + b_2 + X_2 + \dots + b_n + X_n \quad (1)$$

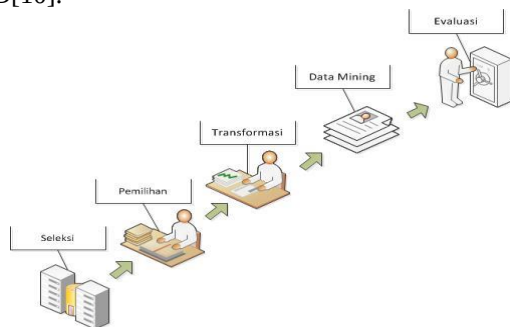
Artinya:

Y : Variable tak bebas
 X_1, X_2 : Variable bebas
 a : Konstanta
 b : Koefisien regresi

3. METODE PENELITIAN

3.1. Metode Penelitian

Dalam penelitian ini, pendekatan yang dilakukan ialah metode kuantitatif, suatu metode yang menggunakan perhitungan, pengukuran, observasi dan eksperimen, lalu diolah dengan menggunakan teknik statistik guna menghasilkan hasil yang diinginkan dengan angka. Kemudian, penulis menggunakan Knowledge Discovery in Database (KDD). KDD ialah proses penggalian dan menganalisa sejumlah data besar yang berguna memberikan informasi serta pengetahuan yang bermanfaat. Berikut tahapan dalam KDD[10].



Gambar 1. Tahapan Knowledge Data Discovery (KDD)

Penjelasan tahapan yang ada di Knowledge Discovery in Database (KDD). Berikut penjelasannya:

1. Selection

Tahap seleksi ialah tahap penyeleksian Data yang akan dimanfaatkan dalam studi ini akan disesuaikan dengan kebutuhan yang ada. Kemudian data diubah ke format yang dibutuhkan.

2. Preprocessing

Tahap pemilihan data, setelah penyeleksian data kemudian dilakukan seleksi data kembali dengan tujuan membuang data atau membersihkan data yang tidak diperlukan agar tidak ada data yang bermasalah.

3. Transformation

Tahap transformasi ialah proses transformasi data atau perubahan format ke dalam format tertentu. Proses ini menjadi hal terpenting dalam tahap penelitian ini karena sumber data yang digunakan berasal dari multiple sources.

4. Data Mining

Setelah data diseleksi dan di ubah ke dalam format tertentu, kemudian tahap selanjutnya adalah data mining, tahap ini ialah bentuk interpretasi dari berbagai tahap yang bertujuan mencari hubungan yang ada dalam berbagai variable dalam data tersebut.

5. Interpretation/Evaluation

Tahap terakhir yakni evaluasi, tahap ini ialah tahap menyimpulkan serta mengevaluasi hasil dari data mining, yang dimana tahap ini merupakan bagian integral dari proses Knowledge Data Discovery (KDD), di mana dilakukan pengecekan untuk memeriksa sejauh mana kesesuaian pola informasi tersebut dengan fakta dan hipotesis yang telah ada sebelumnya.

3.2. Sumber Data

Data yang dipakai dalam penelitian ini ialah, data berasal dari sebuah dataset Dinas Pertanian dan Ketahanan Pangan (DPKP) Kabupaten Brebes, dari tahun 2014 sampai tahun 2023 serta dataset dari Badan Pusat Statistik (BPS) Kabupaten Brebes, yang bisa diakses melalui link <https://brebeskab.bps.go.id>.

3.3. Training dan Testing

Sebelum melakukan penelitian, penulis melakukan pemisahan dataset menjadi dua bagian, di mana salah satunya digunakan untuk melatih model dan yang lainnya untuk melakukan pengujian[5]. Lalu penulis memutuskan membagi dataset untuk training sebesar 80% dan 20% digunakan untuk testing.

3.4. Teknik Pengumpulan Data

3.4.1. Observasi

Teknik pengumpulan data dalam studi ini adalah dengan menggunakan teknik observasi ke Dinas Pertanian dan Ketahanan Pangan (DPKP) Kabupaten Brebes. Kemudian di dapat data bawang merah dari tahun 2014 sampai 2023 yang berupa jumlah tanam, jumlah panen dan jumlah produksi. Kemudian data seperti curah hujan dan hari hujan didapatkan dari website resmi Badan Pusat Statistik (BPS) Kabupaten Brebes dari publikasi dengan judul 'Kabupaten Brebes dalam Angka' dari tahun 2014 sampai 2023.

3.4.2. Studi literature

Studi Literature adalah upaya untuk menelusuri teori yang relevan dengan penelitian yang sedang dikerjakan, semacam teori data mining, berbagai jurnal tentang prediksi jumlah produksi, dan metode penelitian yang sesuai dengan penelitian yang sedang dilakukan ini[11].

3.5. Teknik Analisis Data

Dalam analisa ini, penulis menggunakan regresi linier yang umumnya digunakan dalam prediksi, karena tingkat keakuratannya yang terbukti. Regresi linier ialah suatu pendekatan statistik yang digunakan guna menciptakan sebuah metode yang menggambarkan keterkaitan diantara variabel *dependent* berhubungan lebih atau hanya satu variabel bebas. Kalo hanya terdapat satu variabel *independent*, dinyatakan sebagai regresi linier sederhana, tetapi jika terdapat beberapa variabel *independent*, disebut sebagai regresi linier berganda. Kedua jenis regresi ini memiliki perbedaan dalam jumlah variabel bebas yang terlibat. Koefisien regresi dapat diklasifikasikan ke dalam dua kategori.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

4.1.1. Selection

Dalam tahap *selection*, penulis melakukan seleksi data yang akan digunakan dalam penelitian prediksi produksi bawang merah ini. Berikut data yang sudah diseleksi secara manual:

No	Kecamatan	Tahun	Luastanam	Luaspapan	Curahhujan	Produksi
1	Bantarkawung	2014	130	127	2552	12927
2	Larangan	2014	8281	8335	1646	1004865
3	Ketanggungan	2014	1338	1272	1849	190800
4	Banjarharjo	2014	211	223	2322	26490
5	Losari	2014	901	938	1838	86660
6	Tanjung	2014	1966	1873	1624	184886
7	Kersana	2014	1093	947	1275	90399
8	Bulakamba	2014	3106	3817	1469	363035
9	Wanasari	2014	7068	7075	1592	1025680
10	Songgom	2014	1316	1336	1945	157288
...	...	...	...	...	...	...
127	Brebes	2023	3743	4737	0	431454

Gambar 2. Hasil seleksi data

4.1.2. Preprocessing

Tahap *preprocessing* ini penulis melakukan seleksi ulang di *Google Colaboratory*, tahap ini mencari apakah terdapat missing value yang akan mempengaruhi penelitian ini. Setelah diperiksa ulang tidak ditemukan missing values yang bisa dilihat di gambar bawah ini:

```
bawangdf = pd.read_csv('bawang1.csv')
print(bawangdf.isnull().sum())
```

Kecamatan 0
Tahun 0
Luastanam 0
Luaspapan 0
Curahhujan 0
Produksi 0
dtype: int64

Gambar 3. Hasil pengecekan missing value

4.1.3. Transformation

Dalam tahap transformasi, penulis mengubah format yang sebelumnya dari .xlsx diubah menjadi .csv, sekaligus meng-upload file bawang.csv ke dalam *Google Colaboratory*, kemudian penulis memanggil dataset tersebut dengan bawangdf sebagai nama data frame, yang tersaji pada gambar bawah ini:

```
bawangdf = pd.read_csv('bawang1.csv')
print(bawangdf)
```

	Kecamatan	Tahun	Luastanam	Luaspapan	Curahhujan	Produksi
0	BANTARKAWUNG	2014	130	127	2552	12927
1	LARANGAN	2014	8281	8335	1646	1004865
2	KETANGGUNGAN	2014	1338	1272	1849	190800
3	BANJARHARJO	2014	211	223	2322	26490
4	LOSARI	2014	901	938	1838	86660
...	...	...	...	...	...	...
122	BULAKAMBA	2023	1324	1638	0	170449
123	WANASARI	2023	5184	3718	0	492335
124	SONGGOM	2023	416	375	0	53252
125	JATIBARANG	2023	924	866	0	84951
126	BREBES	2023	3743	4737	0	431454

[127 rows x 6 columns]

Gambar 4. Tampilan dataset di dalam dataframe bawangdf

Lalu penulis melakukan pengecekan data standar deviasi guna mengecek banyaknya nilai, nilai rata-rata, nilai minimal hingga nilai maksimal, yang tersaji pada gambar berikut ini:

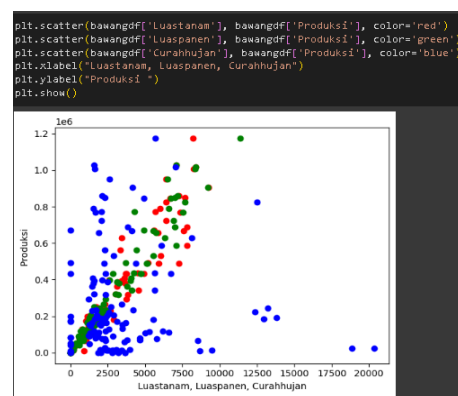
```
bawangdf.describe()
```

	Tahun	Luastanam	Luaspapan	Curahhujan	Produksi
count	127.000000	127.000000	127.000000	127.000000	1.270000e+02
mean	2018.614173	2371.748031	2333.740157	3435.464567	2.619039e+05
std	2.862032	2466.935542	2476.991093	3426.036212	2.826012e+05
min	2014.000000	0.000000	0.000000	0.000000	0.000000e+00
25%	2016.000000	633.000000	611.000000	1662.000000	6.598600e+04
50%	2019.000000	1328.000000	1375.000000	2403.000000	1.639390e+05
75%	2021.000000	3527.500000	3493.500000	3818.500000	3.914615e+05
max	2023.000000	9261.000000	11385.000000	20332.000000	1.174452e+06

Gambar 5. Tampilan tabel standar deviasi

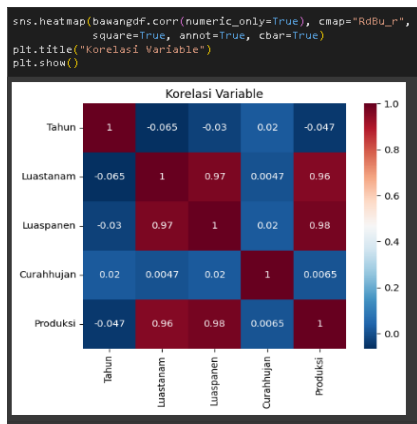
4.1.4. Data Mining

Dataset yang sudah di upload ke dalam *Goggle Colaboratory*, Setelah itu, dilakukan penelusuran terkait keterkaitan antar variabel x dan variabel y. Variabel x di kasus ini merujuk pada data luas tanam yang direpresentasikan dengan warna merah, kemudian luas panen direpresentasikan dengan warna hijau, lalu variabel x lainnya yakni Curah hujan direpresentasikan dengan warna biru sedangkan variabel y adalah data produksi, yang tersaji di gambar bawah ini:



Gambar 6. Visualisasi hubungan variable x dan y

Kemudian penulis melihat korelasi antar variabel x dan variabel y, yang dimana jika nilai korelasinya mendekati angka 1 (satu), itu bisa diartikan bahwa hubungan variabel tersebut sangat berpengaruh ke dalam hasil produksi. Berikut tampilan hasil korelasi yang bisa dilihat dibawah ini:



Gambar 7. Tampilan hasil korelasi antar variable

Setelah itu penulis membuat *dataframe* baru dengan nama *bawangdf2*, berisi dengan data luas tanam, luas panen dan produksi, dapat dilihat dibawah ini:

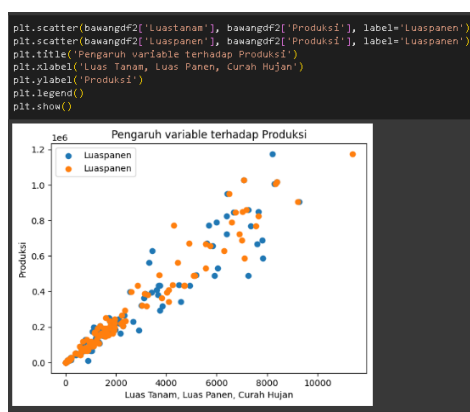
```
bawangdf2=bawangdf[['Luastanam','Luaspanen','Produksi']]
bawangdf2
```

	Luastanam	Luaspanen	Produksi
0	130	127	12927
1	8281	8335	1004865
2	1338	1272	190800
3	211	223	26490
4	901	938	86660
...	...	...	...
122	1324	1638	170449
123	5184	3718	492335
124	416	375	53252
125	924	866	84951
126	3743	4737	431454

127 rows x 3 columns

Gambar 8. Tampilan isi *dataframe* baru

Lalu penulis membuat grafik dari *dataframe* *bawangdf2*, dengan membuat grafik ini agar mempermudah pengamatan variable x dan y melalui visualisasi grafik, berikut tampilannya:

Gambar 9. Tampilan grafik di *dataframe* baru

4.1.5. Interpretation/Evaluation

Tahap *Interpretation* penulis melakukan penyimpulan hasil dari data mining. Kemudian penulis melakukan *split training* dan *testing*, 80% dari data digunakan sebagai bagian pelatihan (*training*),

sementara 20% dipakai sebagai bagian pengujian (*testing*). Yang tersaji pada gambar dibawah ini:

```
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.23,random_state=0)
```

Gambar 10. Kode perintah *split training* dan *testing*

Karena penelitian ini hasil akhirnya ialah prediksi, kemudian penulis membuat objek regresi linier dengan nama *reg2*, yang bisa dilihat di bawah ini:

```
reg2=LinearRegression()
reg2.fit(x_train, y_train)
```

► LinearRegression

Gambar11. Membuat objek prediksi linier regresi berganda

Setelah itu, penulis melakukan pengecekan nilai *Coefficient* untuk Luastanam dan Luaspanen dan nilai *Intercept*, yang tersaji di gambar bawah ini:

```
print('Coefficient: ',reg2.coef_)
print('Intercept: ',reg2.intercept_)
```

Coefficient: [18.6341758 92.48510643]
Intercept: 1344.9603055058396

Gambar 12. Tampilan Nilai *coefficient* dan *intercept*

Kemudian, penulis melakukan pencarian data prediksi yang akan digunakan perhitungan prediksi jumlah produksi bawang merah pada tahun 2024. Berikut data prediksinya:

```
y_pred = reg2.predict(x_test)
```

```
dataframe = pd.DataFrame({'Data sebenarnya':y_test,'Data Prediksi':y_pred})
jumlah_atas = 5
jumlah_bawah = 5
dataframe_atas = dataframe.head(jumlah_atas)
dataframe_bawah = dataframe.tail(jumlah_bawah)
print(dataframe_atas)
print(dataframe_bawah)
```

	Data sebenarnya	Data Prediksi
0	1025690	787383.442865
1	11739	27825.968930
2	431454	509194.629491
3	16031	15345.989867
4	119728	128026.664863
...	...	...
25	91805	112296.771588
26	53252	43778.692380
27	880720	733087.221346
28	115676	80639.197653
29	195	1622.415625

Gambar 13. Tampilan hasil data prediksi

Setelah data prediksi di dapat seperti yang diatas dengan jumlah 30 data prediksi, penulis melakukan penghitungan secara *random* guna menghasilkan prediksi jumlah produksi bawang merah pada tahun 2024 dengan hasil sebanyak 3.962.376 Kuintal.

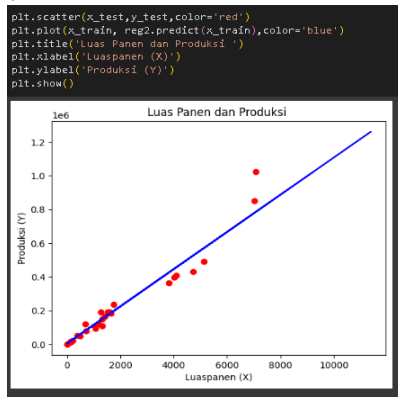
```
df = pd.DataFrame({
    "y_pred": [787383, 509194, 128026, 585865, 446917,
              585865, 43778, 793087, 80639, 1622]
})

Jumlah = df["y_pred"].sum()
print("Prediksi Jumlah produksi Bawang Merah tahun 2024 sebanyak: ", Jumlah)

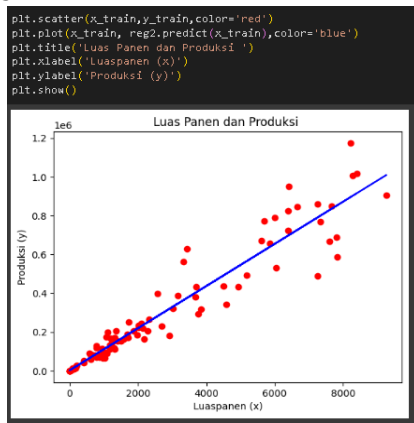
Prediksi Jumlah produksi Bawang Merah tahun 2024 sebanyak: 3962376
```

Gambar 14. Hasil penghitungan data prediksi

Penulis mencari visualisasi hasil dari data *testing* terhadap hasil produksi, yang dimana data *testing* ini linier atau tegak lurus terhadap hasil produksi.

Gambar 15. Tampilan visual data *testing*

Kemudian penulis memvisualisasikan hasil dari data *training* terhadap produksi, yang dimana hasil data *training* ini linier atau tegak lurus terhadap produksi

Gambar 16. Tampilan visual data *training*

Setelah itu penulis mencari nilai *MSE*, *MAE*, *RMSE*, dan *R2-Score*, tersaji pada gambar dibawah ini:

```
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
import math

print("Mean absolute error (MAE): %.2f" % mean_absolute_error(y_test, y_pred))
print("Mean squared error (MSE): %.2f" % mean_squared_error(y_test, y_pred))
print("Root Mean squared error (RMSE): %.2f" % math.sqrt(mean_squared_error(y_test, y_pred)))
r2 = r2_score(y_test, y_pred)
print("R2 Score: ", r2)

Mean absolute error (MAE): 30116.24
Mean squared error (MSE): 2989724823.94
Root Mean squared error (RMSE): 54678.38
R2 Score: 0.9488071314468131
```

Gambar 17. Hasil nilai *MSE*, *RMSE*, *MAE* dan *R2-Score*

4.2. Pembahasan

Dalam tahap pembahasan hasil ini telah didapatkan beberapa hasil dari prediksi produksi bawang merah dengan menggunakan metode regresi linier berganda menggunakan bahasa pemrograman python di *Google Colaboratory*. Berikut penjelasan pembahasan berikut ini:

1. Hasil prediksi jumlah produksi bawang merah pada tahun 2024 adalah sebanyak 3.962.376 Kuintal, ini artinya hasil prediksi di tahun 2024 mengalami kenaikan dari pada tahun sebelumnya terutama di tahun 2022 sebanyak 3.844.480 Kuintal.
2. Penerapan prediksi dengan menggunakan metode regresi linier berganda dengan menggunakan dataset yang penulis dapatkan dari Dinas Pertanian dan Ketahanan Pangan (DPKP) Kabupaten Brebes dan Badan Pusat Statistik (BPS), menunjukan nilai *MAE* : 30116.24, *MSE* : 2989724823.94, *RMSE* : 54678.38 dan nilai *R2-Score* : 0.948.
3. Karakteristik yang berpengaruh dalam prediksi jumlah produksi bawang merah untuk tahun 2024 yaitu Luas tanam, Luas panen dan Curah hujan. Luas tanam sangat berpengaruh di susul dengan luas panen kemudian curah hujan yang dimana curah hujan berpengaruh dalam jumlah produksi serta kualitas bawang merah itu sendiri.
4. Rekomendasi yang bisa diberikan penulis yaitu meningkatkan atau memperluas luas tanam bawang merah pada musim kemarau dikarenakan pada musim kemarau, akan meningkatkan jumlah produksi bawang merah dan meningkatkan kualitas bawang merah.
5. Dengan diperoleh nilai *R2-Score* sebesar: 0.948, artinya mendekati angka 1 yang artinya penelitian dengan menggunakan metode regresi linier berganda cukup akurat.

5. KESIMPULAN DAN SARAN

Dari analisa yang sudah dilaksanakan, bisa di simpulkan dengan hasil prediksi produksi bawang merah pada tahun 2024 sebanyak 3.962.376 kuintal artinya mengalami kenaikan dari pada tahun 2022. Kemudian Faktor yang mempengaruhi produksi bawang merah ialah luas tanam, luas panen dan curah hujan. Curah hujan sangat mempengaruhi kualitas bawang merah. Hasil akurasi metode yang digunakan untuk melakukan prediksi produksi bawang merah memiliki nilai *Mean Absolute Error (MAE)* sebesar: 30116.24, *Mean Squared Error (MSE)* : 2989724823.94, *Root Mean Squared Error (RMSE)* : 54678.38 dan nilai *R2-Score* : 0.948.

Saran untuk penelitian berikutnya, bisa menggunakan lebih banyak dataset dan bisa menambahkan variable-variable yang relevan terhadap penelitian prediksi bawang merah berikutnya supaya mendapatkan angka prediksi dengan nilai akurat dan tinggi.

DAFTAR PUSTAKA

- [1] P. Kurniawan *et al.*, “Prediksi Jumlah Penduduk Jakarta Selatan Menggunakan Metode Regresi Linear Berganda,” *J. Sist. dan Teknol. Inf.*, vol. 10, no. 4, p. 518, 2022, doi: 10.26418/justin.v10i4.48331.
- [2] M. Adha, E. Utami, and H. Hanafi, “Prediksi Produksi Jagung Menggunakan Algoritma Apriori Dan Regresi Linear Berganda (Studi Kasus: Dinas Pertanian Kabupaten Dompu),” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 7, no. 3, pp. 803–820, 2022, doi: 10.29100/jupi.v7i3.3139.
- [3] E. Triyanto, H. Sismoro, and A. D. Laksito, “Implementasi Algoritma Regresi Linear Berganda Untuk Memprediksi Produksi Padi Di Kabupaten Bantul,” *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 4, no. 2, pp. 66–75, 2019, doi: 10.36341/rabit.v4i2.666.
- [4] A. S. Lubis, T. Tugiono, and H. Hafizah, “Data Mining Estimasi Biaya Produksi Ikan Kembung Rebus Dengan Regresi Linier Berganda,” *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 6, p. 888, 2022, doi: 10.53513/jursi.v1i6.5732.
- [5] A. Prasetyo, S. Salahuddin, and A. Amirullah, “Prediksi Produksi Kelapa Sawit Menggunakan Metode Regresi Linier Berganda,” *J. Infomedia*, vol. 6, no. 2, p. 76, 2021, doi: 10.30811/jim.v6i2.2343.
- [6] A. Eka, A. Juarna, T. Informatika, F. T. Industri, and U. Gunadarma, “Prediksi Pro duksi Daging Sapi Nasional dengan Meto de Regresi Linier dan Regresi Polinomial,” *J. Ilm. Komputasi*, vol. 20, no. 2, pp. 209–215, 2021, doi: 10.32409/jikstik.20.2.2722.
- [7] D. N. Aini, B. Oktavianti, M. J. Husain, D. A. Sabillah, S. T. Rizaldi, and M. Mustakim, “Seleksi Fitur untuk Prediksi Hasil Produksi Agrikultur pada Algoritma K-Nearest Neighbor (KNN),” *J. Sist. Komput. dan Inform.*, vol. 4, no. 1, p. 140, 2022, doi: 10.30865/json.v4i1.4813.
- [8] V. V. Utari, A. Wanto, I. Gunawan, and Z. M. Nasution, “Prediksi Hasil Produksi Kelapa Sawit PTPN IV Bahjambi Menggunakan Algoritma Backpropagation,” *J. Comput. Syst. Informatics (JoSYC)*, vol. 2, no. 3, pp. 271–279, 2021.
- [9] T. N. Padilah and R. I. Adam, “Analisis Regresi Linier Berganda Dalam Estimasi Produktivitas Tanaman Padi Di Kabupaten Karawang,” *FIBONACCI J. Pendidik. Mat. dan Mat.*, vol. 5, no. 2, p. 117, 2019, doi: 10.24853/fbc.5.2.117-128.
- [10] A. Diyanti; Martanto; Bahtiar, “PREDIKSI HASIL PANEN PADI TAHUN 2023 MENGGUNAKAN METODE REGRESI LINIER DI KABUPATEN INDRAMAYU,” *J. Inform. Terpadu*, vol. 9, no. 1, pp. 18–23, 2023.
- [11] D. P. L. Mardi Hardjianto, “Data Mining untuk Memprediksi Jumlah Penjualan Hasil Pertanian Menggunakan Algoritma Forecasting (Studi Kasus : Dinas Pertanian Kabupaten Banggai),” *Respati*, vol. 16, no. 2, p. 94, 2021, doi: 10.35842/jtir.v16i2.405.