Impelementasi Algoritma Linear Regression Untuk Prediksi Produksi Tanaman Padi Di Kabupaten Grobogan

M Adin Musababa[1)*]
[1)] Universitas Teknologi Yogyakarta

Received: 5 November 2023 Accepted: 12 November 2023 Published: 19 January 2024

***adhienmuzhba@gmail.com**

Kata Kunci: Impelementasi, Algoritma, Regeresi Linier, Prediksi, Produksi, Tanaman Padi

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : Produksi padi di Indonesia memiliki nilai yang cukup besar. Pada 2020 sendiri Indonesia menempati posisi ke 3 dalam produksi padi setelah Cina dan India. Namun produksi padi di Indonesia jika melihat data Badan Pusat Indonesia pada 2019 mengalami penurunan sebesar 6,16% jika dibandingkan dengan tahun 2020. Pada objek penelitian ini yaitu Kabupaten Groban merupakan penghasil beras terbanyak di Provinsi Jawa Tenggah. Namun pada 2022 mengalami penurunan sebesar 4,62% jika dibandingkan dengan tahun 2021. Penurunan jumlah produksi itu diiringi dengan semakin bertambahnya jumlah penduduk menurut DISPENDUKCAPIL Kabupaten Grobogan pada tahun 2022 mengalami kenaikan sebesar 3,27% jika dibandingkan dengan tahun 2021. Selain itu peralihan lahan pertanian menjadi nonpertanian juga menjadi faktor pendukung penurunan. Untuk mengatasi masalah tersebut salah satunya adalah dengan membangun sebuah model prediksi untuk memproyeksikan hasil panen padi di masa depan. Model ini akan dibangun dengan *Algoritma Linier Regresion*. Data yang akan digunakan untuk melatih serta menguji model ini digunakan data yang bersumber dari Badan Pusat Statistika untuk variable data luas penan, jumlah produksi dan Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) untuk variabel suhu rata-rata (*cencius*), kelembapan rata-rata , curah hujan (mm). data yang diambil mencakup 10 Tahun terakhir dari tahun 2011-2022 dengan jumlah kecamatan di Kabupaten Grobogan yaitu 19 kecamatan. Setelah data diperoleh data akan dilakukan preprocessing untuk memastikan data sehat *(fit),* agar model yang di bangun memiliki nilai MSE dan MAE yang rendah. Setelah data dipastikan sehat (fit) data akan dilakukan proses pembangunan model Linier Regresion. Setelah model dibangun dapat dilihat bahwa model memiliki nilai MSE 6550,241810 dan MSE 121247657,9756. Kemudian model dilakukan proses pengujian dan menghasilkan nilai prediksi yang cukup bagus. Dari model tersebut dapat disipulkan bahwa model yang telah dibangun dapat meberikan solusi untuk masalah produksi padi di Kabupaten Grobogan.

PENDAHULUAN

Indonesia adalah negara agraris diamana sektor pertanian memegang kedudukan penting, hal ini karenakan Indosensia kaya akan sumber energi alam, seperti hortikultura serta tanaman panggan[1]. Pada 2020 Indonesia menepati posisi ketiga dalam produksi beras setalelah China kurang lebih 30% serta India sekitar 21%. Tetapi Indonesia mengalami penurangan dalam hasil produktivitas tanaman padi, pada tahun 2019 mengalai penurunan besar 700,05 ribu hektar ataupun sebesar 6,15% dibanding dengan tahun 2018 (Badan Statistik Indonesia). Pada tahun 2019 jumlah hasil panen padi sebanyak 54,60 juta ton serta mengalai pengurangan sebesar 4,60 juta ton ataupun 7,76%[2].

Pada objek kali ini yaitu Kabupaten Grobogan sendiri menempati posisi pertama penghasil beras di Provinsi Jawa Tenggah. pada tahun 2021 jumlah produksitivitas mencapai 60,37 (Ku/Ha) dengan jumlah produksi

pertonnya 800.945,00. Tetapi bila melihat data pada tahun 2022, Kabupaten Grobogan menggalami penunuran sebesar 4,62% dengan jumlah produktivitasnya 57,58 (Ku/Ha) serta jumlah produksi pertonnya 787.275,00. Mengingat Kabuapten Grobogan merupakan penghasil Beras terbanyak pertama di Provinsi Jawa tenggah. (Badan Pusat Statistika 2023).

Kenaikan jumlah penduduk masing-masing tahunnya yang menyebabkan kebutuhan produksi melonjak [3], di Kabupaten Grobogan memiliki jumlah penduduk pada tahun 2020 adalah 1.453.526 jiwa, jika melihat pada tahun 2022 Kabupaten Grobogan mengalai kenaikan jumlah penduduk sebanyak 3,27% dengan jumlah penduduknya adalah 1.501.145 jiwa. (DISPENDUKCAPIL Kabupaten Grobogan)

Peralihan lahan pertania menjadi lahan non-pertanian membuat jumlah produksi yang diperlukan semakin melonjak tetapi lahan yang merupakan aspek penting untuk penunjang produktifitas padi terus menjadi berkurang[4].

Ketersiaan panggan di sesuatu negara sangatlah penting, sebab beras adalah bahan makanan utama penduduk indonesia [5]. Tetapi dengan melihat fenomena pengurangan jumlah produksi dengan berkuranya lahan pertanian menjadi nonpertanian namun jumlah penduduknya yang semakin banyak, memberikan gambaran jika ketersediaan bahan panggan khususnya beras tidak menentu serta mengakibtkan pemerintah mesti menangulangi ketahan pangan di indonesia.

Untuk mengatasi masalah tersebut, penelitian ini akan menggunakan model *Machine Learning* dengan algorita *Linier Regresion*. *Machine Learning* adalah cabang dari kecerdasan buatan (Artificial Intelligence) yang memungkinkan komputer untuk belajar dari data dan pengalaman tanpa harus secara eksplisit diprogram. Ini memungkinkan sistem untuk mengidentifikasi pola, membuat prediksi, dan mengambil keputusan dengan minimnya intervensi manusia [6][7]. Tujuan dibangunnya model *machine learning* ini adalah untuk membangun model prediksi yang dapat memproyeksikan hasil panen padi di masa depan. Dengan model ini, diharapkan dapat mengestimasi kebutuhan produksi padi yang lebih akurat. Ini akan membantu dalam mengatasi ketidakpastian dan fluktuasi yang terjadi dalam pertanian, termasuk perubahan cuaca dan fluktuasi harga pasar yang mempengaruhi hasil produksi.

Prediksi jumlah hasil panen merupakan salah satu kasus sangat kritis yang ditemui dalam bidang pertanian, hal ini disebabkan kurangnya pengetahuan para petani tentang panen yang baik dan melimpah, ketidakpastian keadaan cuaca hingga naik-turunnya harga pasar yang mempengaruhi terhadap hasil produksi tanaman padi[2].

Hasil penelitian ini diharapkan dapat digunakan untuk perencanaan produksi padi yang lebih efisien dan untuk menjaga ketahanan pangan di Indonesia. Selain itu, model ini juga dapat digunakan untuk mengantisipasi berbagai perubahan dan tantangan, seperti lonjakan permintaan, inflasi, atau bencana alam, yang dapat memengaruhi pasokan beras di negara ini.

TINJAUAN LITERATUR

Penelitian terkait dengan prediksi hasil produksi tanaman padi telah dilakukan pada penelitian sebelumnya, pada penelitian sebelumnya mengunakan berbagai macam jenis Algoritma Machine Learning, peneltia yang mengunakan jenis Algoritma Regresi Linier telah dilakukan oleh Trianto dkk. Pada penelitian yang telah dilakukan mendapatkan hasil sebuah aplikasi website untuk memprediksi hasil produksi padi di Kabupaten Bantul menggunakan algoritma Regresi Linier berganda. Hasil yang diperoleh dapat melakukan prediksi dengan akrurat, dibuktikan dengan proses pengujian validasi dengan metode MAD, yang mendapatkan hasil uji untuk prediksi produksi tanaman padi sebesar 0,101 sehingga dapat disipulkan bahwa hasil prediksi pada model yang dibuat dalam kategori sangat baik [8].

Pengunaan Algoritma Support Vector Machine (SVM) untuk mempredisi hasil produksi tanaman padi juga telah dilakukan. Penelitian ini membanagun model Machine Learning SVM dengan menghasilkan sebuah model regresi yang memliki akurasi tinggi. Ini dibuktikan dengan hasil perhitunggan MAPE (Mean Absolute Percentage Error) sebesar 6635,53, dan RMSE (Root Mean Square Error) sebesar 1094810,74. Dari model tersebut dapat disimpulkan bahwa model memiliki hasil predisi yang cukup baik [9].

Jaringan Syaraf Tiruan (JST) yang juga merupaka salah satu algoritma pembelajaran mesin yang populer juga telah dilakukan penelitian dan pembuatan model. Model JST yang digunakan yaitu Backpropragation, model ini disusun dengan struktur dan fungsi otak manusia sebagai model untuk ditiru. Penelitian ini menghasilkan

sebuah model machine learning yang memiliki hasil evalusi kinerja algoritma yaitu Root Mean Square Error (RMSE). Berdasarkan hasil pengujian yang telah dilakukan nilai RMSE yang paling minimum yakni 8.6918 dengan nilai parameter learning rate = 0.8, hidden layer = 3, hidden neuron = 4 dengan jumlah epoch 10000 terhadap 135 data latih dan 13 data uji. Hasil RMSE 8.6918 menandakan ukuran besar kesalahan yang dihasilkan oleh sistem pada prediksi hasil produktivitas padi [10].

METODE PENELITIAN

1. Pengumpulan Dataset

   Pengumulan dataset dilakukan melalui rekap data dari sumber website Badan Pusat Statistika (BPS) Kabupaten Grobogan untuk variabel luas panen(Ha), Hasil Produksi (Ton) menurut kecamatan di Kabupaten Grobogan, dan

   Dari proses pengumpulan Dataset didapatkan jumlah 190 data untuk 19 kecamatan dalam 10 tahun dan 7 variable yaitu kecamatan, tahun, produksi, luas area, kelembapan, curah hujan dan suhu. Berikut gambaran singkat terkait data yang akan digunakan :

	Province	Year	Production	Land Area	Rainfall	Humidity	Average Temperature
0	Brati	2011	24462	4587	162	76.794	27.964
1	Brati	2012	27236	4782	128	65.576	24.199
2	Brati	2013	27304	4856	219	76.029	27.996
3	Brati	2014	24433	4416	176	77.748	26.632
4	Brati	2015	30280	4691	121	84.543	27.055
...	...	...	...	...	...	...	...
185	Wirosari	2016	65369	10566	152	75.793	30.982
186	Wirosari	2017	41295	7590	111	71.839	31.981
187	Wirosari	2018	62203	9956	144	84.672	25.681
188	Wirosari	2019	10014	10014	144	71.966	24.809
189	Wirosari	2020	59641	10014	119	81.782	26.276

190 rows × 7 columns

Gambar 1 Dataset yang Digunakan

Dari table tersebut berisikan gambaran singkat tentang data yang akan digunakan untuk pelatihan model

2. *Pre-processing*

   Pre-processing adalah langkah awal yang dilakukan dalam pengolahan data sebelum dilakukan proses pemodelan menggunakan algoritma untuk membantu algoritma yang digunakan agar dapat berjalan dengan baik dan nantinya menghasilkan faktor kesalahan Root Mean Error (RME) dan Mean Square Error (MSE) yang rendah sehingga model dapat memiliki akurasi yang baik [11]. Pada tahapan ini data yang telah terkumpul akan dilakukan akan dilakukan proses-proses sebagai berikut :

   2.1 Deskripsi Data

   Deskripsi data berisikan gambaran singkat setiap variable data yang akan digunakan, ini bertujuan untuk melihat apakah jumlah rentang variable berbeda jauh atau tidak

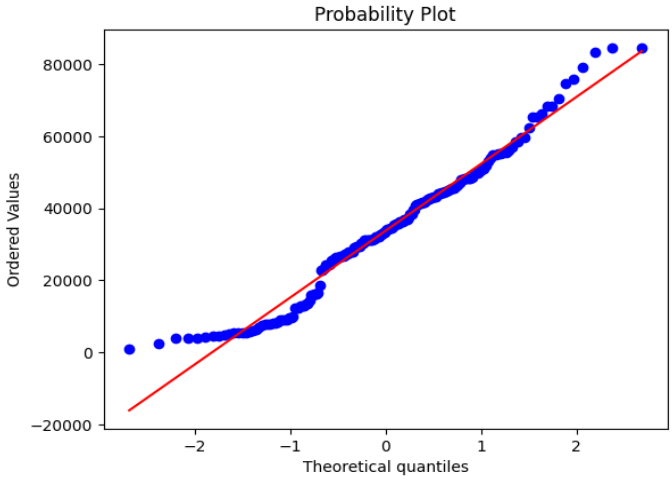
	Year	Production	Land Area	Rainfall	Humidity	Average Temperature
count	190.00000	190.000000	190.000000	190.000000	190.000000	190.000000
mean	2015.50000	33737.747368	6371.847368	159.342105	75.052021	27.940300
std	2.87987	18666.943229	2759.032785	43.444592	5.747407	2.447078
min	2011.00000	1065.000000	738.000000	47.000000	62.788000	22.008000
25%	2013.00000	22848.000000	4640.000000	135.000000	70.337000	26.188000
50%	2015.50000	33527.500000	6460.000000	159.000000	75.272500	28.039000
75%	2018.00000	45271.000000	7997.750000	178.500000	79.517500	29.853250
max	2020.00000	84542.000000	12795.000000	540.000000	88.028000	32.543000

Gambar 2 Deskripsi Dataset

Daari data yang didapat Rata-rata hasil panen di 19 kecamatan selama 10 tahun adalah 33737.747368 ton dengan hasil terendah 1065.000000 ton dan tertinggi 4881089 ton. Rata-rata luas lahan pertanian adalah 6371.847368 hektar. Dari uraian data di atas terlihat bahwa nilai mean dan median setiap atribut tidak jauh berbeda

2.2 Sebaran data pada variable target *Production*

Untuk memeriksa variable target berdistribusi normal atau tidak meengunakan grafik QQ-plot, Tujuannya adalah untuk memeriksa sejauh mana distribusi data pada variable class yang akan dijadikan sebagai target prediksi sesuai dengan distribusi normal. Jika titik-titik dalam variable produksi berada dekat dengan garis diagonal, maka data memiliki distribusi yang mirip dengan distribusi normal. Jika titik-titik tersebut terletak jauh dari garis diagonal, maka data mungkin tidak mengikuti distribusi normal. Dengan melihat sebaran data pada variable target, akan membantu untuk mengevaluasi apakah asumsi bahwa data Anda terdistribusi normal adalah wajar. Jika data tidak mengikuti distribusi normal, itu mungkin mempengaruhi pemilihan model *Machine Learning* yang akan diuat pada proses selanjutnya.

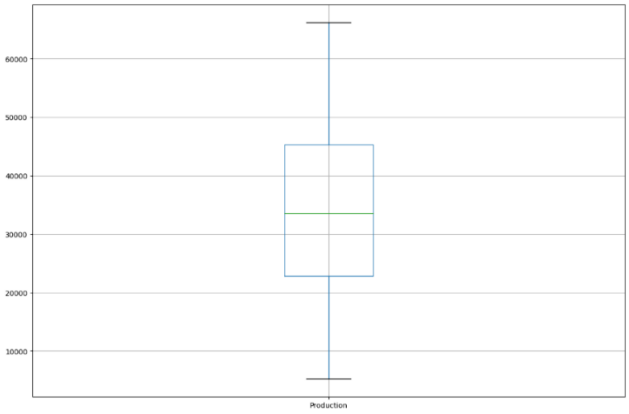


Gambar 3 QQ-Plots Sebaran Dataset Variable Target

Dari data yang akan digunakan sebagai class target, terlihat bahwa data berada dekat pada garis diagonal, maka data yang akan digunakan sebagai class target adalah berdistribusi normal.

2.3 Memeriksa Outliner

Outlier adalah data yang berbeda secara signifikan dari sebagian besar data lainnya dalam kumpulan data, nilai yang jauh lebih besar atau lebih kecil dari nilai-nilai lain dalam data dan dapat mempengaruhi hasil model [12] . Outlier dapat memengaruhi hasil analisis, seperti rerata (mean) dan deviasi standar yang menyebabkan data tidak berdistribusi noral

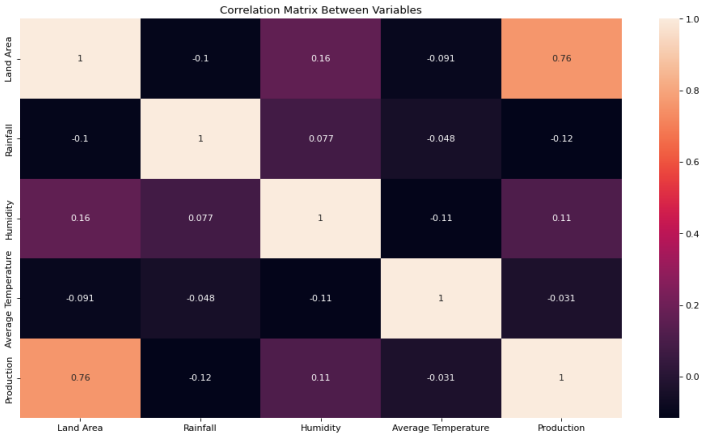


Gambar 4 Bloxplot Outlier Pada Variable Target

Dari hasil visualisasi bloxplot pada gabar 2.3 diatas, terlihat bahwa data variable output tidak memiliki outlier yang membuktikan bahwa variabel tersebut berdistribusi normal.

2.4 Correlation Matrix Between Variables

Untuk melihat jenis korelasi anatara variable indepenent dan variable output, tujuannya adalah akan memeberikan gambaran seberapa pengaruh antara variable independent dan output dalam menentukan hasil predisi, jika nilai korelasi >= 0,5 maka terdapat hubungan yang erat antar variable dan jika nilai korelasi <0,5 maka tidak terdapat hubungan yang erat antar variable :



Gambar 5 Correlation Matrix

Dari matriks korelasi dapat diketahui jenis korelasi antara variabel independen dengan variabel output. Luas lahan pertanian dan suhu rata-rata mempunyai nilai korelasi negatif, artinya semakin besar nilai variabel bebas tersebut maka produksi juga akan menurun walaupun tidak signifikan (pengaruh suhu). Sedangkan variabel curah hujan dan kelembaban mempunyai nilai korelasi positif yang berarti semakin kecil kedua variabel tersebut maka produksi akan meningkat (pergerakan grafiknya bergerak terbalik), meskipun hubungannya tidak terlalu kuat.

2.5 Feature Selection

Feature Selection adalah pemilihan fitur yang akan digunakan sebagai variable indepenteng dan variable target [13] :

```
x = df[['Land Area', 'Rainfall', 'Humidity', 'Average Temperature']]
y = df[['Production']]
```

Figure 6 Variable untuk Menapung Nilai Input dan Target

Variable x adalah variabel independent yang berisi subset fitur yang dipilih. Subset ini terdiri dari kolom-kolom 'Land Area', 'Rainfall', 'Humidity', dan 'Average Temperature' dari dataset yang akan digunakan sebagai model. fitur-fitur ini akan dijadikan sebagai input untuk model.
Variable y adalah variabel yang berisi target atau variabel yang akan dijadikan output prediksi, yaitu 'Production'.

2.6 Train and Test Split

Train and Test Split adalah proses pembagian dataset enjadi 2 subset yaitu dataset pelatihan *(train set)* dan dataset pengujian *(test set)*.Tujuannya adalah untuk melatih model pada sebagian data dan menguji kinerja model pada bagian data yang tidak digunakan dalam pelatihan. Data akan dibagi 70% untuk data pelatihan dan 30% akan dijadikan pengujian.

```
x_train : (133, 4)
x_test : (57, 4)
y_train : (133, 1)
y_test : (57, 1)
```

Figure 7 Jumlah Pembagian Dataset Untuk Training dan Testting

Setelah dilakukan proses pembagian data, data yang akan digunakan untuk pelatihan adalah 133 data dengan 4 variabel untuk variable independent untuk digunakan sebagai input dan 57 data dengan 1 variabel dependent yang akan digunakan sebagai variable target.

2.7 Feature Scaling

Feature Scaling adalah proses normalisasi atau penskalaan fitur-fitur dalam dataset sehingga mereka memiliki besaran yang sama atau setidaknya dalam rentang yang serupa. Tujuannya adalah untuk memastikan bahwa perbedaan dalam skala atau besaran fitur tidak memengaruhi kinerja algoritma machine learning yang digunakan [14] .

Data Train Test:

```
np.set_printoptions(suppress=True)
print(x_train[:5])
print(y_train[:5])
```

```
[[ 0.52755707 -0.74522166  1.00211073  0.14276532]
 [ 0.28976549 -0.22690073  1.681204   -0.78892522]
 [-0.68806233 -0.25569633 -0.30732769  1.56071504]
 [ 0.35786035 -0.16930951 -0.17318142  0.0202819 ]
 [-0.42144752 -0.25569633  0.65927271  0.43590897]]
[[ 0.08705141]
 [ 0.4666291 ]
 [-0.57661071]
 [ 0.39381151]
 [-0.31301953]]
```

Gambar 8 Scalling Data Train

Data test set

```
print(x_test[:5])
print(y_test[:5])
```

```
[[ 1.20054568  0.06668367 -0.98447198 -1.95311237]
 [-1.98322003 -0.87077516 -1.22204807  0.6859436 ]
 [-0.61758239 -0.17194221  0.27047301  0.86393825]
 [ 0.36200927  0.73142721 -0.45918906  0.18366893]
 [ 0.31375353  0.10077309 -0.43152494 -0.63147043]]
[[-1.32789818]
 [-1.5547762 ]
 [-0.3964522 ]
 [ 0.58703355]
 [ 0.16906857]]
```

Gambar 9 Scalling Data Test

Setelah penskalaan fitur, rentang nilai setiap fitur telah disesuaikan sehingga semua fitur memiliki rata-rata nol dan deviasi standar satu. Ini membantu model machine learning berfungsi secara efektif dengan memastikan bahwa setiap fitur memiliki kontribusi yang setara dalam analisis.

3. Pembuatan model *Linier Regresion*

Model yang akan dibuat untuk melakukan proses prediksi adalah dengan *Algoritma Linier Regresion,* Regresi merupakan teknik membangun model yang digunakan untuk prediksi nilai dari data masukan yang diberikan. Regresi adalah ukuran statistik yang digunakan untuk menentukan kekuatan hubungan antara variabel dependen (tak bebas) dengan variabel independen (bebas) [4].

Metode utama untuk melakukan prediksi yakni membangun model regresi dengan mencari hubungan antara satu atau lebih variabel independen atau prediktor (X) dengan variabel dependen atau respons (Y) yang telah dilakukan pada proses pre-processing. Linear regression memodelkan hubungan antara variabel skalar dan satu atau lebih variabel penjelas.

Metode linear regression tersusun atas dasar pola hubungan data yang relevan di masa lalu [10]. Secara umum, algoritme linear regression dibagi menjadi dua jenis, yaitu simple linear regression dan multiple linear regression. Simple linear regression merupakan hubungan antara satu variabel dependen dengan satu variabel independen, sedangkan multiple linear regression merupakan hubungan antara satu variabel dependen dengan dua atau lebih variabel independen.

Pada penelitian ini digunakan multiple linear regression, karena terdapat lebih dari satu buah variable independent. Dataset penelitian terdiri atas satu variabel dependen (Y) dan empat variabel independen (X). Variabel dependen tersebut adalah atribut Production, sedangkan variabel independen adalah atribut Luas Lahan, kelembapan, rata-rata hujan, dan suhu rata-rata.

Pembuatan model akan dilakukan dengan mengunakan bahasa pemograan python dengan librari sklearn. Langkah awal yang dilakukan adalah mengimport library sklearn.

```
from sklearn.metrics import r2_score
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import RandomizedSearchCV
```

Gambar 10 Library Python untuk Pembuatan Model

Modul dari sklearn yang digunakan adalah sebagai berikut :

Setalah library sklearn di import, selanjutnya akan dibuat model *Regresi Linier :*

```
from sklearn.linear_model import LinearRegression
LinReg = LinearRegression()
LinReg.fit(x_train, y_train)
```

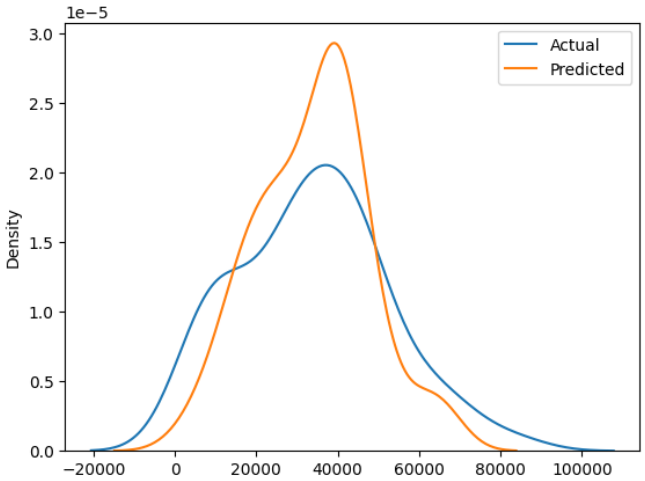
```
▾ LinearRegression
LinearRegression()
```

Gambar 11 Pebuatan Model Linier Regresion

Pertama lakukan import modul pada library sklearn yaitu *LinearRegression,* kemudian buat variable untuk menampung model yang telah dibuat untuk meudahkan pemanggilan selanjutnya, setelah model jadi dilakukan proses pelatihan model menggunakan data pelatian *(train)* yaang telah dibuat pada pre-processing (x_train adalah fitur, dan y_train adalah label).
Karena data telah dilakukan proses standarisasi yang menyebabkan Interpretasi maka data asli akan dikembalikan ke skala aslinya

```
ypred_LinReg = sc.inverse_transform(LinReg.predict(x_test).reshape(-1,1))
y_pred_LinReg = pd.DataFrame(ypred_LinReg)
y_pred_LinReg.tail()
```

Ini akan mengmbalikan hasil akhir predisi sesuai dengan satuan yang telah dilakukan yaitu Ton.
Setelah model jadi dan siap dilakukan proses uji model, akan dilihat nilai akurasi dan nilai prediksi melalui grafik *plots,* tujuannya adalah untuk memberikan gambran seberapa akurat model yang telah dibuat untuk melakukan prediksi :



Gambar 12 Nilai Aktual dan Prediksi Dari Model yang dibuat

Dari grafik tersebut dapat dilihat bahwa nilai prediction > actual. Ini memberikan gambaran bahwa nilai yang sebenarnya yang inggin model ini diprediksi lebih besar dari nilai yang dihasilkan oleh model mesin yang telah dibuat. Perbandingan itu dapat dijadikan tolak ukur sejauhmana model benar-benar dapat memprediksi nilai-nilai yang lebih rendah dari data yang akan digunakan untuk inputan prediksi.

4. Evaluasi Model

Evaluasi kinerja dilakukan untuk menguji atau mengevaluasi kebenaran dan mengukur kinerja yang diinginkan pada sistem yang dibuat [07]. Teknik pengujian yang digunakan pada penelitian ini adalah Mean Squared Error (MSE), Mean Absolute Error (MAE) dan koefisien determinasi (R2), untuk mengukur ketepatan ramalan untuk mengetahui hasil Algoritma Linier Regression dalam memprediksi hasil panen padi di kabupaten grobogan.
MSE adalah matriks yang mengukur seberapa baik model regresi memprediksi nilai yang sebenarnya dalam data MSE Ini menghitung rata-rata dari kuadrat selisih antara nilai sebenarnya dan nilai yang diprediksi oleh model, MAE metrik evaluasi alternatif yang mengukur kesalahan prediksi dengan cara yang lebih sederhana dari pada MSE. MAE menghitung rata-rata dari selisih absolut antara nilai sebenarnya dan nilai yang diprediksi oleh model, Semakin kecil nilai MAE dan MSE maka semakin baik model yang dihasilkan karena tingkat akurasinya akan tinggi, R2 adalah metrik yang mengukur sejauh mana model regresi cocok dengan data, R2 mengukur proporsi varians dalam variabel dependen yang dapat dijelaskan oleh variabel independen dalam model semakin besar nilai R2 maka model akan semakin bagus.

	R2-score	Mean Absolute Error	Mean Squared Error
Linear Regression	63.463462	6550.241810	121247657.975668

Gambar 13 Nilai Evaluasi Model

Dari hasil model Regresi Linier yang telah dibuat, didapatkan hasil nilai R2, MAE dan MSQ seperti tabel 2.2 tersebut. Dapat disimpulkan bahwa skor yang didapat nilai R2 meeiliki nilai 63.463462, nilai tersebut cukup besar untuk memberikan gabaran dari model yang dibuat bahwa model regresi cocok dengan data. Nilai MAE dan MSE memiliki nilai yang rendah yang berarti model meiliki akurasi yang tinggi untuk melakukan proses prediksi pada tahapan uji model .

5. Uji Model

Uji model adalah proses untuk melakukan pengujian terhadapan odel esin yang telah dibuat. Sebelum melakukan proses uji model, model akan disimpan menjadi sebuah file. Proses penyimpanan ini mengunakan python dengan library pickle.

```
import pickle

with open ('LinReg', 'wb') as r:
  pickle.dump(LinReg, r)
```

Gambar 14 Menyimpan Model

Setelah model tersimpan dengan nama file LinReg.pkl, kemudian model akan dilakukan proses pengujian mengunakan data yang sebelunya belu pernah diuji

```
with open ('LinReg', 'rb') as r:
  perc = pickle.load(r)

prediction = perc.predict([[12566, 162, 65.793, 32.982 ]])
prediction

array([[9367.08582878]])
```

Gambar 15 Proses Uji Model

Dari uji odel yang dihasilkan dengan memasskukan nilai variable independet seperti digambar menghasilan nilai prediksi yaitu 9376 Ton. Jika elihat data yang dilatih, bisa dilihat bahwa odel telah mendekati nilai predisi yang baik.

HASIL PENELITIAN

Pada penelitian ini, telah menjalani beberapa tahapan penting dalam metodologi penelitian yang telah dilakukan pada metode penelitian, khususnya dalam membangun model prediksi produksi padi berdasarkan data yang telah kumpulkan yang mempertimbangkan faktor-faktor penting dalam proses produksi padi. Namun pada penelitian ini difokuskan hanya pada beberapa faktor yaitu suhu, kelembapa, curah hujan, luas lahan. Didapatkan hasil model prediksi yang memiliki akurasi cukup akurat. Ini bisa dilihat pada gambar 13 yang memiliki hasil nilai MSE dan MAE pada evaluasi model yang cukup rendah dan hasil nilai predisi yang lebih besal dari aktual. Hal itu juga dibuktikan dengan melakukan uji model yang menginputkan nilai faktor-faktor yang mempengaruhi hasil produksi pada variable variable independent yang nilainnya sebelumnya belum pernah diuji yang berhasil mempredisi hasil yang akurat jika dilihat dengan nilai pada dataset yang telah direkap dan akan dilakukan prosses pembuatan model.

PEMBAHASAN

Penelitian ini membangun sebuah model untuk melakukan prediksi produksi tanaman padi. Pada proses sebelumnya telah dilakukan proses rangkaian untuk pembuatan model Regresi Linier, Linier Regresion adalah metode statistik yang digunakan untuk mengetahui hubungan antara satu atau lebih variable dengan satu buah variable[4], [15].

Dengan mengunakan faktor-faktor yang terjasi di lapangan untuk mendukung proses prediksi jumlah produksi tanaman padi. Banyak faktor yang mempengaruhi hasil produksi, Namun penelitian ini hanya terbatas pada

luas lahan, suhu, temperatur, curah hujan yang terjadi pada daerah penelitian ini yaitu Kabupaten Grobogan. Data ini didapatkan dari Badan Pusat Statistika(BPS) dan Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) dalam rentang waktu 2011 sampai 2021. Daerah penelitian ini memiliki jumlah produksi padi terbanyak pertama di Provinsi Jawa Tenggah sehingga Kabupaten Grobogan menjadi studi penelitian ini.

Dari rumusan latar belakang dan masalah yang telah diketahui, maka didapatkan solusi untuk mengatasi masalah tersebut, salah satunya dengan menggunaka model prediksi produksi tanaman padi menggunakan Linier Regresion.

Hal tersebut bisa dilihat dari hasil akhir model yang telah dilakukan tahapan penting yang telah dilakukan pada metode penelitian. Yang memberikan gambaran akhir bahwa model dapat memprediksi dengan akurat.

Dari model itu akan bisa membantu untuk memeperkirakan hasil produksi di Kabupaten Grobogan kedepannya.

Untuk penelitian selanjutnya dapat lebih di kembangkan lagi untuk data yang akan digunakan agar model yang dibuat lebih bagus

KESIMPULAN

Berdasarkan proses-proses yang dilakaukan pada seluruh tahapan sebelumnya, penelitian ini dapat disipulkan:

1. Algoritme linear regression dapat diterapkan untuk prediksi hasil panen tanaman padi dengan beberapa variabel yang memengaruhi, yaitu Luas Lahan, suhu rata-rata, kelebapan rata-rata dan curah hujan.
2. Model yang telah dibuat mampu untuk melakukan prediksi cukup baik, dengan nilai MSE 6542.458597 dan MAE 117823406.915991 yang cukup rendah memeberikan informasi bahwa model mendekati akurat dan dapat memprediksi cukup baik.
3. Berdasarkan penelitian yang telah dilakukan beberapa saran yang diberikan untuk penelitian selanjutnya antara lain, mengunakan data bulanan atau harian dengan menambahkan parameter faktor-faktor yang mempengauhi hasil produksi padi, dapat memodifikasi dengan menambahkan metode lain yang memiliki nilai MSE dan MAE yang lebih rendah agar model dapat lebih bagus dan maksimal untuk melakukan prediksi.

REFERENCES

[1] F. Joanda Kaunang, R. Rotikan, and G. Stella Tulung, "Pemodelan Sistem Prediksi Tanaman Pangan Menggunakan Algoritma Decision Tree Crop Prediction System Using Decision Tree Algorithm," *Cogito Smart Journal*, vol. 4, no. 1, 2018.

[2] "583-1273-1-PB".

[3] N. Nur and F. Wajidi, "Jurnal Politeknik Caltex Riau Implementasi Algoritma Random Forest Regression untuk Memprediksi Hasil Panen Padi di Desa Minanga," *Jurnal Komputer Terapan*, vol. 9, no. 1, p. 58, doi: 10.35143/jkt.v9i1.5917.

[4] H. W. Herwanto, T. Widiyaningtyas, and P. Indriana, "Penerapan Algoritme Linear Regression untuk Prediksi Hasil Panen Tanaman Padi," 2019.

[5] A. H. Krisdianto, Rais, N. Fiskia, and H. Sain, "Prediksi Tingkat Produksi Padi di Sulawesi Tengah Menggunakan Analisis Algoritma FBprophet," *JURNAL ILMIAH MATEMATIKA DAN TERAPAN*, vol. 19, no. 2, pp. 2014–214, Dec. 2022, doi: 10.22487/2540766x.2022.v19.i2.16062.

[6] A. S. Diantika, Y. Firmanto, A. J. Akuntansi, F. Ekonomi, and D. Bisnis, "IMPLEMENTASI MACHINE LEARNING PADA APLIKASI PENJUALAN PRODUK DIGITAL (STUDI PADA GRABKIOS)."

[7] R. G. Wardhana, G. Wang, and F. Sibuea, "PENERAPAN MACHINE LEARNING DALAM PREDIKSI TINGKAT KASUS PENYAKIT DI INDONESIA," 2023.

[8] 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. 66–75, Jul. 2019, doi: 10.36341/rabit.v4i2.666.

[9] "admin,+Journal+manager,+06-553-1121-1-SM+-+UNISLA+-+OK".

[10] G. Ramadhona, B. Darma Setiawan, and F. A. Bachtiar, "Prediksi Produktivitas Padi Menggunakan

Jaringan Syaraf Tiruan Backpropagation," 2018. [Online]. Available: http://j-ptiik.ub.ac.id
[11] P. Meilina, "PENERAPAN DATA MINING DENGAN METODE KALSIFIKASI MENGGUNAKAN DECISION TREE DAN REGRESI," 2015.
[12] "13104-Article Text-26629-1-10-20170227".
[13] A. Rahmansyah, O. Dewi, P. Andini, T. Hastuti, P. Ningrum, and M. E. Suryana, "Membandingkan Pengaruh Feature Selection Terhadap Algoritma Naïve Bayes dan Support Vector Machine," 2018.
[14] O. Somantri and M. Khambali, "Feature Selection Klasifikasi Kategori Cerita Pendek Menggunakan Naïve Bayes dan Algoritme Genetika," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, vol. 6, no. 3, Sep. 2017, doi: 10.22146/jnteti.v6i3.332.
[15] I. M. Kamal, T. Hendro, and R. Ilyas, "PREDIKSI PENJUALAN BUKU MENGGUNAKAN DATA MINING DI PT. NIAGA SWADAYA," 2017.