

Analisis Faktor-Faktor yang Mempengaruhi Produksi Padi di Sumatera Menggunakan Metode Regresi Linier

Mohammad Yusuf Nugroho^{1*}, Nurmalitasari²

^{1,2}Sistem Informasi

Universitas Duta Bangsa Surakarta

^{1*}202040239@mhs.udb.ac.id, ²nurmalitasari.udb.ac.id

Abstrak— Dalam penelitian data harus melalui proses pengolahan agar dapat digunakan dalam penelitian tersebut. Data yang di gunakan haruslah valid untuk dapat menghasilkan solusi yang tepat guna. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi produksi padi di sawah. Pulau Sumatera mempunyai lebih dari 50 persen lahan pertanian setiap provinsinya dengan komoditas pangan utama paling dominan adalah padi, sedangkan sisanya adalah jagung, kacang tanah, dan ubi. Hasil pertanian di Sumatera sangat rentan terhadap perubahan iklim yang dapat memengaruhi pola tanam, waktu tanam, produksi dan kualitas hasil. Perubahan iklim dapat memberikan dampak negatif terhadap produksi bahan pokok tersebut. Apalagi bertambahnya suhu bumi akibat dampak dari pemanasan global yang akan mempengaruhi pola presipitasi, evaporasi, water-run off, kelembaban tanah, dan variasi iklim yang sangat fluktuatif secara keseluruhan dapat mengancam keberhasilan hasil produksi pertanian. Prediksi hasil pertanian komoditas bahan pangan banyak dipengaruhi oleh perubahan cuaca (climate change). Metode yang di gunakan untuk analisis yaitu Regresi Linier dan juga menggunakan library python.

Kata kunci — Regresi Linier, Python, Pulau Sumatera

Abstract— In research data must go through a processing process so that it can be used in the research. The data used must be valid to be able to produce an appropriate solution. This study aims to analyze the factors that influence rice production in paddy fields. The island of Sumatra has more than 50 percent of agricultural land in each province with the most dominant main food commodity being rice, while the remainder is corn, peanuts and sweet potatoes. Agricultural products in Sumatra are very vulnerable to climate change which can affect cropping patterns, planting time, production and yield quality. Climate change can have a negative impact on the production of these basic commodities. Moreover, an increase in the earth's temperature due to the impact of global warming which will affect the pattern of precipitation, evaporation, water runoff, soil moisture, and climate variations which are very fluctuating as a whole can threaten the success of agricultural production. Predictions of agricultural yields for food commodities are heavily influenced by climate change. The method used for analysis is Linear Regression and also uses the python library.

Keywords— Linear Regression, Python, Sumatra Island

I. PENDAHULUAN

Pulau Sumatera mempunyai lebih dari 50 persen lahan pertanian setiap provinsinya dengan komoditas pangan utama paling dominan adalah padi, sedangkan sisanya adalah jagung, kacang tanah, dan ubi. Pertanian memegang peran penting bagi perekonomian negara, karenanya banyak studi yang dilakukan untuk menyokong kemajuan pertanian di Indonesia. Studi yang dilakukan membutuhkan data valid sebagai bahan kajian agar penelitian tersebut dapat menghasilkan sebuah solusi yang tepat guna.

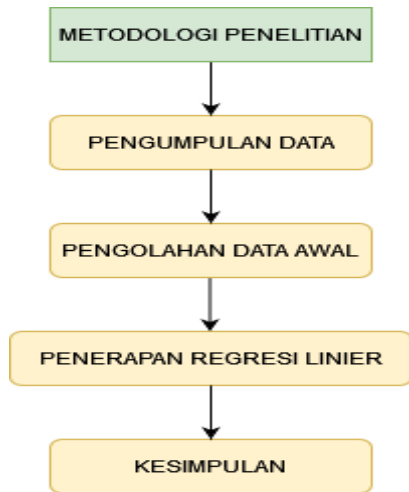
Pengambilan data diperoleh melalui website BPS pada kategori tanaman pangan utama dari 8 provinsi di pulau Sumatera yaitu Nanggroe Aceh Darussalam (NAD), Sumatera Utara, Riau, Jambi, Sumatera Selatan, Bengkulu dan Lampung. Data yang digunakan adalah data dari tahun 1993 hingga tahun 2020 untuk dataset padi. Data memuat hasil produksi tahunan dan luas panen atau luas lahan. Kemudian data perubahan cuaca diperoleh melalui

website BMKG untuk data harian curah hujan, kelembapan, dan temperatur rata-rata atau suhu rata-rata dari tahun 1993 hingga tahun 2020

Kehadiran BPS (Badan Pusat Statistik) sebagai Lembaga Pemerintahan non Kementrian yang menyediakan kebutuhan data bagi pemerintah maupun masyarakat menjadi salah satu keuntungan bagi akademisi dalam menjalankan penelitiannya. Data-data yang disediakan oleh BPS telah melalui proses pengolahan data menggunakan teknologi komputer, sehingga data ditampilkan berupa stastistik yang mudah dipahami oleh masyarakat dan dapat diakses secara mudah oleh siapa saja melalui website BPS.

II. METODOLOGI PENELITIAN

Bagian metodologi penelitian menjelaskan secara rinci tentang tahapan penelitian yang dilakukan. Untuk mengetahui analisis faktor-faktor yang mempengaruhi produksi padi di sawah Sumatera



Gambar 1. Tahapan Penelitian

1. Pengumpulan Data

Peneliti menggunakan pengumpulan data faktor faktor yang mempengaruhi produksi padi di sawah Sumatera. Data tersebut telah di kumpulkan menjadi 6 Variabel yaitu tahun, produksi, luas panen, curah hujan, kelembapan, suhu rata rata

2. Pengelolahan Data Awal

Tahap ini digunakan untuk membantu metode yang digunakan agar mendapatkan hasil yang lebih baik dan tepat. Dengan melakukan pembersihan (*Cleaning*) pada data untuk memperbaiki atau menghapus data yang rusak. Untuk Langkah selanjutnya dilakukan uji standarisasi dan outlier untuk mengetahui kualitas data yang telah dilakukan pembersihan (*cleaning*).

3. Penerapan Regresi Linier

Regresi linier adalah metode statistik yang digunakan untuk memodelkan hubungan linier antara satu atau lebih variabel independen (variabel prediktor) dengan satu variabel dependen (variabel respons). Tujuan dari regresi linier adalah untuk memahami dan memprediksi bagaimana variabel dependen berubah sebagai respons terhadap perubahan variabel independen.

III. HASIL DAN PEMBAHASAN

a. Data yang digunakan

Berdasarkan penelitian ini, dapat di simpulkan dengan hasil observasi yang udah di simpan dalam

bentuk MS.Excel. File MS.Excel tersebut di ubah dalam bentuk CSV agar mudah untuk pembacaan dalam program python dan dataset kurang lebih ada 225 data.

Gambar 2. Data dalam bentuk MS.Excel

b. Prossesing Data

Table yang udah di import dalam model python

	Provinsi	Tahun	Produksi	Luas Panen	Curah hujan	Kelembapan	Suhu rata-rata
0	Aceh	1993	1329536.00	323589.00	1627.00	82.00	26.06
1	Aceh	1994	1299699.00	329041.00	1521.00	82.12	26.92
2	Aceh	1995	1382905.00	339253.00	1476.00	82.72	26.27
3	Aceh	1996	1419128.00	348223.00	1557.00	83.00	26.08
4	Aceh	1997	1368074.00	337561.00	1339.00	82.46	26.31
...
219	Lampung	2016	3831923.00	390799.00	2317.60	79.40	26.45
220	Lampung	2017	4090654.00	396559.00	1825.10	77.04	26.36
221	Lampung	2018	2488641.91	511940.93	1385.80	76.05	25.50
222	Lampung	2019	2184089.33	464103.42	1706.40	78.03	27.23
223	Lampung	2020	2604913.29	545149.05	2211.30	75.80	24.58

Gambar 3. Table di import dalam python

Dan gambar selanjutnya adalah faktor-faktor yang mempengaruhi produksi padi di Sumatera

```

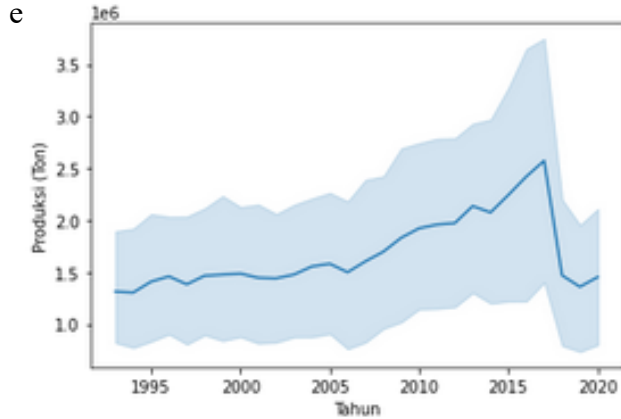
3]:
Provinsi
Tahun
Produksi
Luas Panen
Curah hujan
Kelembapan
Suhu rata-rata
    
```

Gambar 4. variabel faktor-faktor yang mempengaruhi produksi padi

c. Visualisasi Data

Visualisasi data adalah proses mewakili informasi atau data menggunakan elemen visual seperti grafik, diagram, peta, dan visual lainnya. Tujuan utama

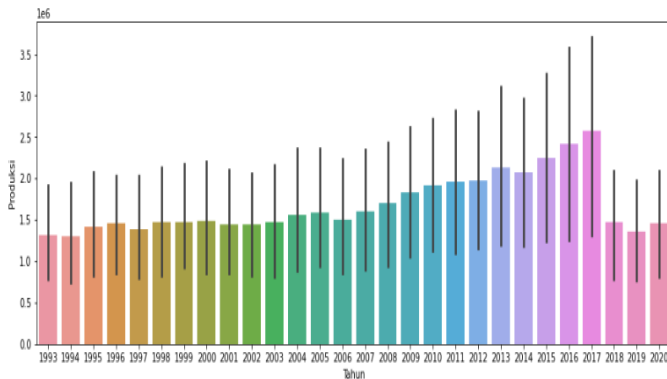
visualisasi data adalah untuk memudahkan pemahaman dan interpretasi data yang kompleks atau besar dengan cara yang lebih intuitif dan



Gambar 5. lineplot produksi padi pada setiap tahunnya

Line plot, juga dikenal sebagai line chart atau grafik garis, adalah jenis grafik yang menampilkan data dalam bentuk garis yang menghubungkan titik-titik data yang terurut. Grafik ini sering digunakan untuk memvisualisasikan tren atau pola dalam data sepanjang waktu atau urutan lainnya.

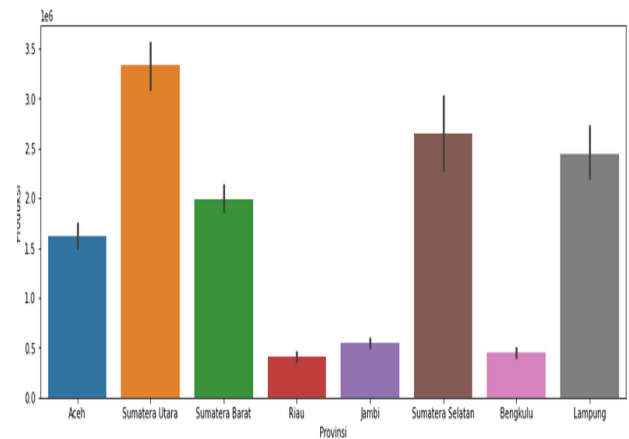
Seperti yang kita lihat gambar di atas Lineplot dari hasil Produksi padi pada tiap tahunnya



Gambar 6. barplot produksi padi tiap tahun

Seperti yang kita liat gambar diatas yaitu barplot produksi padi pada setiap tahun ketahunnya, jenis grafik yang menggunakan batang vertikal atau horizontal untuk memvisualisasikan data kategorikal. Grafik ini memperlihatkan perbandingan atau distribusi data antara beberapa kategori yang

berbeda. Setiap batang mewakili nilai atau frekuensi yang terkait dengan kategori yang spesifik.

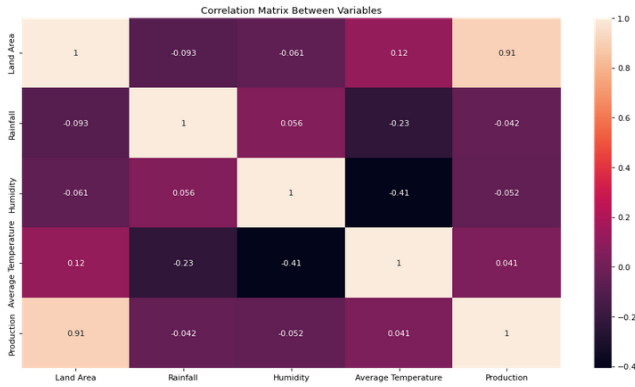


Gambar 7 Subplot produksi padi tiap provinsi

Subplot di atas menggambarkan produksi padi tiap provinsi dan Subplot bisa dikenal dengan visualisasi data yang mengacu pada pembuatan beberapa grafik atau subplot dalam satu gambar atau ruang tampilan yang sama. Dalam subplot, beberapa grafik disusun secara teratur dalam bentuk matriks atau grid, memungkinkan perbandingan dan pemahaman yang lebih baik antara berbagai aspek data.

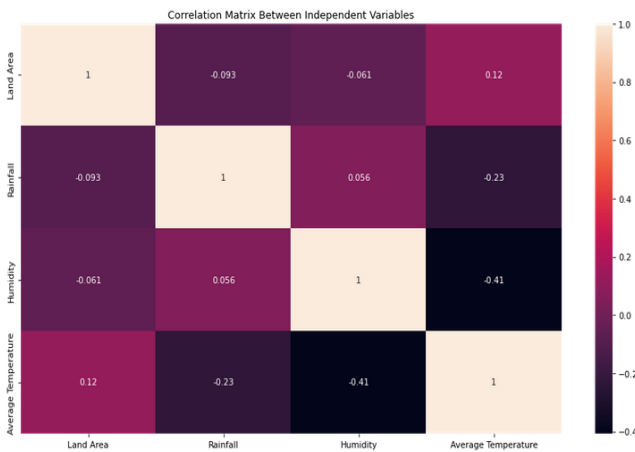
Subplot memungkinkan kita untuk menggabungkan beberapa grafik yang saling terkait atau membandingkan beberapa variabel dalam satu tampilan, sehingga mempermudah analisis dan pemahaman data secara komprehensif.

Untuk Selanjutnya pengujian Normalitas telah dilakukan dengan baik yang menghasilkan data normal. Dan untuk Langkah selanjutnya yaitu pengujian analisis kolerasi memiliki dua macam diantaranya Kolerasi Matrix Variabel dan Kolerasi Matrix Independent Variabel



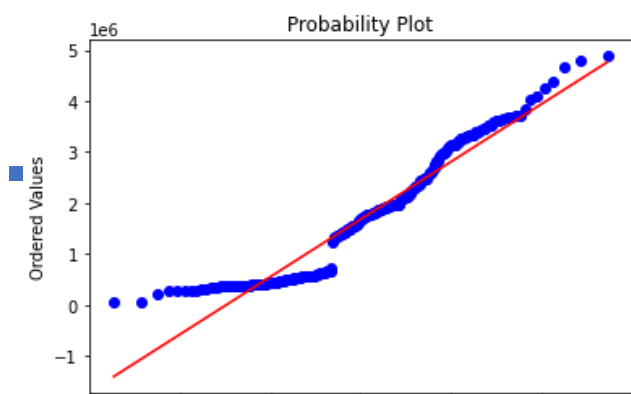
Gambar 8. korelasi Matrix variable

Gambar diatas menunjukkan korelasi matrix variabel tertentu. Variabel yang berpengaruh besar terhadap produksi padi yaitu di angka 0.41, Korelasi matriks mengacu pada perhitungan koefisien korelasi antara dua atau lebih variabel dalam bentuk matriks. Matriks korelasi digunakan untuk menunjukkan hubungan antara variabel-variabel tersebut dan memberikan informasi tentang sejauh mana variabel-variabel tersebut bergerak bersama-sama.



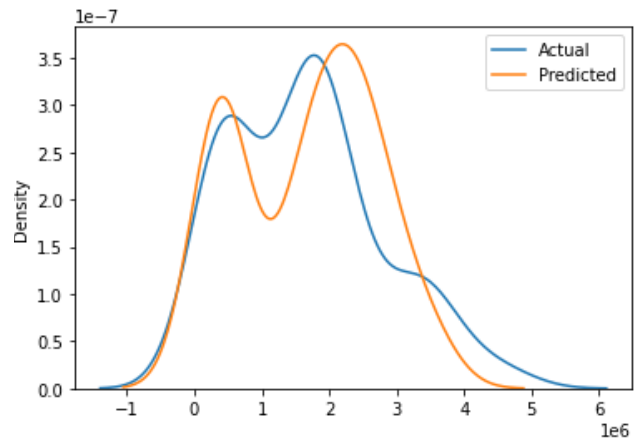
Gambar 9. kolerasi matrix independent variabel

Dan sedangkan gambar di atas yaitu kolerasi Matrix Independent Variabel bisa disebut juga Korelasi matriks untuk variabel independen mengacu pada perhitungan koefisien korelasi antara dua atau lebih variabel independen dalam bentuk matriks. Ini berguna untuk memahami hubungan dan ketergantungan antara variabel independen sebelum melakukan analisis lebih lanjut, seperti regresi atau analisis multivariat. Yang menghasilkan



Gambar 10. Probability Plot

Gambar diatas bisa disebut juga QQ-plot di atas menunjukkan bagaimana distribusi data dari variabel terpilih yaitu variabel Produksi. Plot di atas merupakan plot nilai kuantil teoritis dengan nilai kuantil dari variabel yang dipilih. Jadi, plot akan menggambarkan berapa banyak nilai dalam suatu distribusi berada di atas atau di bawah batas tertentu. Dari hasil plot di atas diperoleh garis-garis yang cenderung lurus yang berarti bahwa secara keseluruhan sebaran data variabel-variabel tersebut cenderung normal.



Gambar 11. grafik prediksi vs aktual

Grafik prediksi vs aktual dikenal dengan sebutan "grafik scatterplot" atau "grafik scatterplot prediksi vs aktual". Grafik ini digunakan untuk memvisualisasikan perbandingan antara nilai aktual (data yang sebenarnya) dan nilai prediksi dari suatu model atau metode analisis

Dan yang kita bisa liat Dari grafik diatas dapat ditafsirkan bahwa nilai prediksi (biru) tidak terlalu jauh dengan nilai aktual (orange) sehingga

keakuratan dari hasil prediksi cukup tinggi dengan nilai 3.2

IV. KESIMPULAN

Berdasarkan analisis yang telah dilakukan dengan metode regresi linier dapat memprediksi faktor-faktor yang mempengaruhi produksi padi di Sumatera dengan baik dan benar. Yang di bantu dengan implementasikan pemrograman Python dan Google Collab secara Online. Dan hasil kolerasi dari table 2 matriks tersebut diketahui bahwa mempunyai perbedaan pendapat yaitu yang kolerasi variable independent matrix bisa mengidentifikasi adanya multikolinearitas

Dan yang kolerasi variable matrix semakin kecil kedua variabel tersebut maka produksi akan meningkat (pergerakan grafik bergerak terbalik), meskipun hubungannya tidak terlalu kuat.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak yang sudah membantu dalam penulisan jurnal, terutama pada pihak dosen yang senantiasa membantu dalam konsultasi penulisan.

REFERENSI

- [1] Arum, G. (2017). Perbandingan Kepekaan Uji Anderson Darling dan Uji Geary pada Masalah Kenormalan Data menggunakan Simulasi Monte Carlo. Jakarta: Universitas Islam Negeri Syarif Hidayatullah.
- [2] Kotler, P., & Armstrong, G. (2010). Prinsip-Prinsip Pemasaran, Jilid 1 dan 2 Edisi Kedua Belas. Jakarta: Erlangga
- [3] Yayan Adrianova Eka Tuah(2020),Implementasi Linear Sederhana Untuk Prediksi Gaji Berdasarkan Pengalaman Lama Bekerja,Jutech
- [4] Gujarati, D.N. and Porter, D.C. (2009) Basic Econometrics. 5th Edition, McGraw Hill Inc., New York.
- [5] Supranto, J. 2005. Ekonometri (1st ed). Bogor: Ghalia Indonesia.
- [6] Walpole, R.E. & Myers, R.H. 1995. Ilmu peluang dan statistika untuk insinyur dan ilmuwan. Edisi ke-4. ITB. Bandung.
- [7] Cramer, D. & Howitt, D. 2006. The Sage Dictionary of Statistics. London: Sage Publication.
- [8] Drapper. N.R. & Smith. 1992. Analisis Regresi Terapan. Edisi Kedua. Jakarta: PT Gramedia Pustaka
- [9] Setiawan & Kusriani, D.E. 2010. Ekonometrika. Yogyakarta: ANDI.
- [10] Sembiring, R.K. 2003. Analisis Regresi. Edisi Kedua. Bandung: ITB Bandung
- [11] T. Khotimah and R. Nindiyasari, "Forecasting Dengan Metode Regresi Linier Pada Sistem Penunjang Keputusan Untuk Memprediksi Jumlah Penjualan Batik (Studi Kasus Kub Sarwo Endah Batik Tulis Lasem)," J. Mantik Penusa, vol. 1, no. 1, pp. 71–92, 2017.
- [12] M. Sadikin and F. Alfiandi, "Comparative Study of Classification Method on Customer Candidate Data to Predict its Potential Risk," Int. J. Electr. Comput. Eng., vol. 8, no. 6, p. 4763, 2018.