

Prediksi Produksi Padi Di Pulau Sumatera Menggunakan Metode Ridge Regression

Noaga Swara Fahmi¹, Munjiat Setiani Asih², Rismayanti³

¹Universitas Harapan, Medan, Indonesia

^{1*}noaga03@gmail.com, ²munjiat.stth@gmail.com, ³risma.stth@gmail.com

^{*)}noaga03@gmail.com

Abstrak—Melakukan prediksi produksi padi di Sumatra menggunakan pemodelan ridge regression bertujuan untuk menyediakan platform yang diharapkan dapat membantu orang untuk dapat melakukan prediksi terhadap produksi padi yang akan terjadi beberapa tahun kedepan. Dengan memanfaatkan teknologi machine learning dengan menggunakan metode ridge regression, pemodelan ini akan memprediksikan serta memberikan visualisasi dari hasil produksi dari beberapa provinsi yang terdapat di Sumatra. Tahapan awal dimulai dengan melakukan install dan import library python, seperti numpy, pandas, seaborn, dll. Lalu, setelah itu akan kita lakukan pemuatan data dalam format file .CSV, setelah itu melakukan eksplorasi data. Dan yang terakhir adalah visualisasi data.

Kata Kunci: Pemodelan, Ridge regression, Machine Learning

Abstract—Predicting rice production in Sumatra using ridge regression modeling aims to provide a platform that is expected to help people predict the future rice production in the coming years. By utilizing machine learning technology with the ridge regression method, this model will predict and visualize the production results from several provinces in Sumatra. The initial steps begin with installing and importing Python libraries, such as numpy, pandas, seaborn, and others. Afterward, the data in .CSV format will be loaded, followed by data exploration. Finally, data visualization will be performed.

Keywords: Modeling, Ridge regression, Machine learning

1. PENDAHULUAN

Tanaman padi merupakan salah satu komoditas strategis yang memiliki peran vital dalam pemenuhan kebutuhan pangan dunia. Lebih dari separuh populasi dunia sangat bergantung pada padi sebagai sumber karbohidrat utama, menjadikannya tanaman budidaya yang tidak hanya penting bagi umat manusia karena lebih dari setengah penduduk di dunia tergantung pada tanaman ini sebagai sumber bahan pangan utama mereka [1], jika terjadi defisit produksi padi pada suatu negara nantinya akan menjadi masalah sangat berdampak di dalam masyarakat yang negaranya mengalami defisit produksi padi. Dalam rilis bulanan BPS pada 1 Maret mencatat bahwa Indonesia mengalami defisit beras pada Januari - Februari 2024. Minus Januari 2024 adalah 1,61 juta ton dan Februari 1,22 juta ton, sehingga total defisit beras 2,83 juta ton. Artinya, Indonesia masih punya utang 2,83 juta ton yang harus ditutup di luar jumlah total kenaikan yang dipenuhi untuk bulan Juni sampai akhir tahun. Produksi yang lebih rendah akan berdampak pada kenaikan harga beras. BPS mencatat kenaikan harga beras bisa mencapai 10 persen yang mendorong peningkatan inflasi sebesar 0,5-1 persen. Dampak dari kenaikan harga padi akan sangat berimbas bagi suatu negara yang mayoritas masyarakatnya mengkonsumsi nasi. Di satu sisi, petani dapat mendapatkan keuntungan dari naiknya harga jual beras. Namun di lain sisi, para petani juga akan menghadapi masalah dalam hal biaya produksi dan juga iklim yang semakin tidak stabil. Kondisi tersebut akan memperburuk posisi para petani kecil yang memiliki akses yang terbatas pada sumber daya pertanian yang lebih modern. Bagi petani akan sangat bervariasi meskipun mereka bisa mendapatkan keuntungan dari kenaikan harga, mereka juga dihadapkan pada tantangan produksi yang lebih mahal dan tidak stabil. Oleh karena itu pentinglah bagi kita untuk senantiasa dapat mengamati pertumbuhan pangan di pulau Sumatera karena dengan demikian diharapkan terciptanya ketahanan pangan yang akan berguna dimasa yang akan datang.

Solusi yang ada sebelumnya untuk mengatasi masalah dalam hal produksi padi ada banyak dan berikut ini adalah beberapa dari solusi dari masalah dalam hal produksi padi yang telah ada sebelumnya, hasil produksi pertanian dan khususnya produksi padi di Pulau Sumatera sangat amat rentan terhadap yang namanya perubahan iklim global terlebih lagi sering terjadinya kenaikan suhu suatu daerah dari tahun ke tahun yang nantinya akan berdampak ke hasil produksi. Prediksi hasil produksi pertanian dari suatu komoditas pertanian sangat amat perlu dilakukan agar nantinya diharapkan dapat memberikan suatu gambaran dari hasil produksi bahan pokok yang ada di Pulau Sumatera. Memprediksikan produksi dari suatu komoditas pertanian sangat lah penting, sama halnya dengan memprediksikan produksi padi menjadi sangat penting dilakukan agar nantinya didapatkan gambaran seperti apa angka produksi kedepannya dan hasil prediksi produksi padi tersebut juga nanti akan berguna untuk menunjang pembangunan nasional dalam sektor pertanian dalam upaya mewujudkan mandirinya negara untuk

memproduksi bahan makan utama rakyatnya pada suatu negara yang mayoritas masyarakatnya mengkonsumsi nasi sebagai bahan makanan utamanya. Maka dari itu kita harus mengupayakan untuk memprediksikan hasil produksi padinya, dan agar dapat memudahkan kita dalam hal melakukan upaya prediksi produksi tersebut kita harus menggunakan kemajuan teknologi sebagai alat bantu untuk melakukannya dan salah satu pendekatan yang bisa kita lakukan dan kita andalkan untuk mengatasi persoalan prediksi produksi pangan adalah dengan menerapkan metode berbasis kecerdasan buatan (*Artificial intelligence*) khususnya model *machine learning*. Teknologi ini akan memungkinkan melakukan suatu analisis data dalam skala yang sangat besar dan tentunya lebih efisien dibanding pendekatan konvensional. Memprediksikan hasil produksi pertanian dengan algoritma *machine learning* diharapkan dapat memberikan gambaran yang lebih awal terkait kinerja dalam sektor pertanian, sekaligus dapat menjadi dasar dalam hal perencanaan penguatan pangan dalam negeri. *Artificial Neural Network (ANN)* termasuk dalam metode yang baik dalam melakukan prediksi menggunakan algoritma *Back-propagation* yang mana algoritma ini telah banyak digunakan dalam melakukan suatu proses prediksi dan memberikan hasil yang akurat dengan error yang rendah [2]. Metode ini bekerja seperti meniru jaringan syaraf otak manusia, yang mana memungkinkan suatu model *machine learning* untuk belajar dari data-data historis yang diberikan padanya. Artikel ini bertujuan untuk merancang arsitektur *ANN* untuk melakukan prediksi terhadap produksi padi menggunakan *ANN* dengan algoritma *backpropagation*. Penelitian ini bertujuan untuk mengetahui algoritma terbaik dalam melakukan prediksi hasil pertanian komoditas tanaman pangan yaitu padi, jagung, kacang tanah, kedelai, ubi kayu dan ubi jalar di pulau Sumatera dengan pendekatan model regresi *machine learning* [3]. Dengan adanya suatu sistem prediksi yang berbasis *machine learning* ini, diharapkan pemerintah kita dan para pejabat khususnya di sektor pertanian ini dapat mengambil kebijakan yang berlandaskan pada data yang ada. Hal ini meliputi keputusan yang terkait waktu menanam, alokasi sumber daya alam, pendistribusian pupuk, dan kebijakan ekspor dan impor bahan pangan.

2. METODE PENELITIAN

2.1 Ridge Regression

Ridge regression memberikan perkiraan koefisien regresi dengan memodifikasi metode kuadrat terkecil, yang bertujuan untuk mencapai pengurangan varian dengan menambahkan konstanta k untuk menstabilkan koefisien. Karena ridge didasarkan pada penambahan konstanta deviasi k ke diagonal matriks XTX , koefisien estimasi ridge dipengaruhi oleh besarnya konstanta deviasi k , dimana k antara 0 dan 1. *Ridge regression* mengubah variabel independen X dan variabel dependen Y ke dalam bentuk standar. Regresi ridge digunakan untuk mengendalikan ketidakstabilan estimator kuadrat paling sedikit yang timbul karena *multicollinearity* [4]. Parameter regresi ridge diperkirakan dengan meminimalkan jumlah kuadrat kesalahan yang menambahkan batasan pada kotak yang menyusut koefisien mendekati nol-1000. Secara khusus, koefisien estimator ridge diperoleh dengan meminimalkan.

$$\beta^{\wedge}Ridge = (XTX + cI)^{-1}XTY \quad (1)$$

Keterangan:

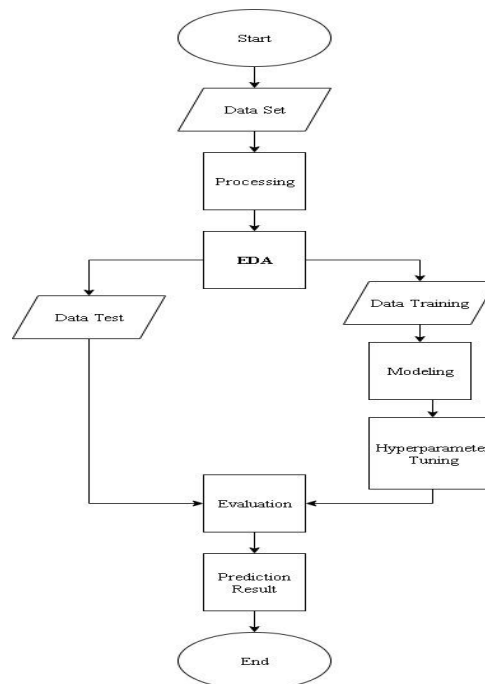
- $\beta^{\wedge}Ridge$: koefisien *Ridge Regression*
- X : matriks desain (matriks yang mengandung variabel *independent*).
- XT : transpos dari matriks X .
- Y : vektor *output* (target) yang sebenarnya.
- c : parameter regularisasi (sering dilambangkan dengan λ).
- I : identitas dari matriks yang memiliki ukuran yang sama dengan XTX .
- $(X^TX + cI)^{-1}$: invers dari $X^TX + cI$

2.2 Flowchart

Flowchart dapat diartikan sebagai langkah langkah penyelesaian masalah yang dituliskan dalam suatu simbol-simbol tertentu. Diagram alir ini akan menunjukkan alur di dalam program secara logika. Flowchart ini diperlukan tidak hanya sebagai alat komunikasi tetapi juga sebagai pedoman, dan sebelum komponen komponennya dapat lebih dipahami perludikomunikasikan aturan-aturan desain ord chart. Untuk menggambarkan langkah atau pemecahan masalah secara sederhana, dapat dimengerti, rapi dan tidak ambigu dengan menggunakan beberapa simbol-simbol yang bisa dibilang standart merupakan tujuan dari flowchart. Flowchart adalah langkah langkah pemecahan masalah yang ditulis dalam symbol-simbol tertentu. Flowchart ini akan menampilkan alur dalam program secara logika. Flowchart ini tidak hanya digunakan sebagai alat komunikasi tetapi juga diperlukan sebagai panduan dan sebelum pemahaman yang lebih baik dari komponen-komponen flowchart. Penulisan menggunakan flowchart adalah model penulisan dengan menggunakan bentuk penyusun bangun ruang flowchart atau bisa

disebut juga dengan bagian alir. Flowchart juga merupakan bagian yang menunjukkan aliran atau runtutan algoritma. Penulisan ini bagus secara visual akan tetapi bisa menjadi rumit jika algoritma yang dibuat terlalu Panjang. Tujuan dari flowchart adalah untuk menggambarkan langkah pemecahan masalah secara sederhana dengan dimengerti, rapi, dan tidak ambigu dengan menggunakan notasi standar. Simbol digunakan untuk menggambarkan algoritma dalam bentuk flowchart [5].

Penelitian ini diawali dengan tahap input dataset, yaitu memasukkan data historis produksi padi yang mencakup variabel seperti luas panen, curah hujan, suhu rata-rata, dan kelembapan. Selanjutnya dilakukan pre-processing untuk membersihkan dan menyiapkan data, termasuk normalisasi dan penanganan nilai yang hilang. Data kemudian dibagi pada tahap split data menjadi data latih (training) dan data uji (testing) dengan rasio 80:20. Proses modeling dilakukan dengan menerapkan metode Ridge Regression pada data training guna membangun model prediktif. Untuk meningkatkan performa, dilakukan hyperparameter tuning menggunakan teknik validasi silang (cross-validation) guna mencari nilai lambda (λ) yang optimal. Model yang telah dilatih kemudian dievaluasi menggunakan metrik R^2 , RMSE, dan MAE. Tahap akhir adalah prediction, yaitu menghasilkan output prediksi dari data uji yang telah diproses. Hasil ini menjadi dasar analisis untuk menilai efektivitas metode yang digunakan.



Gambar 1. Flowcart system

3. HASIL DAN PEMBAHASAN

3.1 Instal dan Import Library

Pertama sebelum kita mulai menuliskan *code* ada beberapa hal yang harus kita lakukan terlebih dahulu dan yang harus dilakukan adalah menginstall pustaka atau *library* dari Python yang bernama “Kaleido” dan “Phik” Setelah *library* “Kaleido” dan “Phik” telah terinstall maka setelah itu yang harus dilakukan adalah kita harus melakukan beberapa pemanggilan lagi dari *library* Python yang nantinya akan digunakan pada proyek ini, antara lain *library* tersebut adalah sebagai berikut:

1. *Pandas*
2. *Numpy*
3. *Matplotlib*
4. *Plotly*
5. *Scikit-Learn*
6. *Ridge Regression*

3.1.1 Pandas

Pandas adalah pustaka python yang dirancang untuk manipulasi dan analisis data. Diperkenalkan pada tahun 2008 oleh Wes McKinney, pandas menawarkan struktur data yang fleksibel dan alat analisis yang kuat, memungkinkan pengguna untuk bekerja dengan data terstruktur dengan mudah. Pandas sangat cocok untuk analisis data dari berbagai sumber, termasuk file CSV, Excel, dan database SQL.[6]

3.1.2 Numpy

Numpy (numerical python) sebuah pustaka yang sangat penting dalam dunia pemrograman python, khususnya dalam bidang manipulasi data dan komputasi numerik. Pustaka ini dirancang untuk memberikan dukungan yang kuat dalam melakukan operasi aljabar linear dan pengolahan matriks, yang merupakan elemen kunci dalam banyak aplikasi ilmiah dan teknik. Dengan Numpy, pengguna dapat dengan mudah melakukan operasi matematika yang kompleks pada array multidimensi, memungkinkan analisis data yang lebih efisien dan cepat. Salah satu keunggulan utama dari numpy adalah kemampuannya untuk memproses data dalam jumlah besar dengan waktu pemrosesan yang relatif singkat, berkat optimasi yang dilakukan pada algoritma dasar dan penggunaan struktur data yang efisien.[7]

3.1.3 Matplotlib

Matplotlib berfungsi untuk menciptakan visualisasi data dalam berbagai bentuk, termasuk scatter plot, line chart, bar chart, dan pie chart. Library ini juga memberikan kemampuan untuk melakukan modifikasi pada tampilan grafik, seperti penambahan judul dan lainnya. [8]

3.1.4 Plotly

Plotly adalah salah satu library visualisasi data interaktif yang digunakan dalam bahasa pemrograman python. Library ini memungkinkan pengguna untuk membuat berbagai jenis grafik seperti grafik garis, grafik batang, pie chart, scatter plot, box plot, dan lain-lain. Keunggulan utama plotly terletak pada kemampuannya menghasilkan grafik yang tidak hanya informatif tapi juga grafik lebihh interaktif, yang sangat berguna dalam analisis data dan presentasi hasil dari apa yang kita tengah kerjakan.

3.1.5 Scikit-Learn

Scikit learn adalah library python yang diperuntukkan untuk machine learning. Scikit learn memungkinkan untuk melakukan beragam pekerjaan dalam data science, seperti regresi (regression), klasifikasi (classification), pengelompokan / penggugusan (clustering), data preprocessing, dimensionality reduction, dan model selection (perbandingan, validasi, dan pemilihan parameter maupun model). Paket ini berfokus pada membawa pembelajaran mesin ke non-spesialis menggunakan bahasa tingkat tinggi tujuan umum. [9]

3.1.6 Ridge Regression

Metode ridge regression pertama kali dikemukakan oleh A.E Hoerl pada tahun 1962. Ridge regression (regresi ridge) merupakan metode yang dikembangkan untuk menstabilkan nilai koefisien regresi karena adanya multikolinearitas. Metode ini merupakan modifikasi dari metode kuadrat terkecil dengan cara menambahkan tetapan bias c yang kecil pada diagonal matriks XTX . Walaupun memiliki penduga yang bias, penduga ini memiliki variasi estimator koefisien minimum. Regresi ridge mengurangi dampak multikolinearitas dengan menentukan penduga yang bias tetapi mempunyai varians yang lebih kecil dari varians penduga regresi linear berganda.

Multikolinearitas yaitu kondisi dimana terjadi korelasi yang cukup tinggi antara variabel variabel bebas. Korelasi yang sangat tinggi akan menghasilkan penaksiran yang berbias, tidak stabil dan mungkin jauh dari nilai sasaran. Menurut montgomery dan hines (1990) dalam, multikolinearitas dapat menyebabkan koefisien regresi yang dihasilkan oleh analisis regresi berganda menjadi sangat lemah atau tidak dapat menghasilkan analisis yang mewakili sifat atau pengaruh dari variabel bebas yang bersangkutan[10]

3.2 Memuat Data

Setelah melakukan pemanggilan beberapa pustaka atau library yang diperlukan dalam proyek Python ini, langkah selanjutnya yang harus dilakukan adalah melakukan pemuatan data atau loading data yang mana artinya kita melakukan upload data berupa file yang berisi data historis dengan format excel yaitu (.CSV). pada gambar 2 kita dapat melihat pemuatan data.

Tabel 1. Pemuatan Data

No.	Provinsi	Tahun	Produksi	Luas Panen	Curah Hujan	Kelembapan	Suhu Rata-Rata
0	Aceh	1993	1329536.00	323589.00	1627.0	82.00	26.06
1	Aceh	1994	1299699.00	329041.00	1521.0	82.12	26.92

2	Aceh	1995	1382905.00	339253.00	1476.0	82.72	26.27
3	Aceh	1996	1419128.00	348223.00	1557.0	83.00	26.08
4	Aceh	1997	1368074.00	337561.00	1339.0	82.46	26.31
...	...	...	...	...	...	...	...
219	Lampung	2016	3831923.00	390799.0	2317.6	79.40	26.45
220	Lampung	2017	4090654.00	396559.0	1825.1	77.04	26.36
221	Lampung	2018	2488641.00	511940.0	1385.8	76.05	25.50
222	Lampung	2019	2164069.00	464103.0	1706.4	78.05	27.23
223	Lampung	2020	2604913.00	545149.0	2211.3	75.80	24.58

3.3 Eksplorasi Data

Setelah melakukan upload file, maka untuk tahap selanjutnya kita akan melakukan Exploratory data atau eksplorasi data dari data atau file .CSV yang telah kita upload tadi. Pada tahap ini kita harus melihat apakah ada missing values atau bagian yang tidak memiliki nilai didalam data kita atau tidak, dan jika terdapat bagian yang ber *value nill* atau tidak memiliki nilai kita harus membersihkan bagian tersebut agar nantinya tidak berdampak pada hasil model prediksi kita. Proses membersihkan data ini biasa disebut juga dengan proses *data cleaning*, yang merupakan tahap penting sebelum masuk ke proses pemodelan. Serta harus melihat untuk tipe tipe data dari kolom kolom yang ada dalam file .CSV yang sudah kita upload pada tahap sebelumnya. Pengecekan tipe data bertujuan agar tidak terjadi kesalahan saat proses analisis atau pengembangan model, seperti angka yang terbaca sebagai string atau masih ada data yang bernilai nill. Setelah mengetahui bahwa tidak ada data yang bernilai 0/nill maka selanjutnya kita bisa melihat untuk jumlah data, nilai minimal, nilai maksimal, nilai rata-rata, dll. Proses ini dilakukan agar dapat memahami distribusi data yang baik serta mengidentifikasi outlier atau nilai yang mungkin akan memengaruhi akurasi dari model itu sendiri.

3.4 Visualisasi Data

Untuk tahap selanjutnya adalah tahap dimana kita harus memvisualisasikan data kita. Visualisasi data adalah proses mewakili informasi kompleks dalam bentuk visual yang mudah dipahami. Visualisasi data memungkinkan analis dan pemangku kepentingan bisnis untuk menggali wawasan yang mendalam, mendeteksi pola, dan mengidentifikasi tren yang berharga dalam penjualan global. [11]. Visualisasi data harus dan wajib dilakukan agar kita mudah membaca data dan memahami isi dari data yang sedang kita coba kelola tanpa harus melihat baris demi baris yang ada dalam data kita, jika kita melihat kolom demi kolom yang ada dalam data kita akan merepotkan dan sangat memakan waktu yang sangat lama sekali dibandingkan dengan melihat visualisasi dari data yang sedang kita kerjakan saat ini. Untuk melihat data yang hanya berisi 10 baris mungkin akan mudah tapi tentunya data yang akan kita kelola tidak memiliki baris yang sedikit, seandainya kita mengelola data yang berisi sampai 1000 baris tentunya akan sangat memakan waktu jika harus melihatnya baris perbaris maka dari itu pentinglah untuk melakukan visualisasi data.

3.5 Transformasi Data dan Preprocessing

Preprocessing data adalah proses penting dalam analisis data mining yang bertujuan untuk membersihkan, mengubah format, dan mempersiapkan data agar lebih mudah dan akurat dalam proses analisis.[12] Maka dari itu tahap *preprocessing* sangat penting agar sesuai dengan kebutuhan dalam proses menganalisis baik dalam skala kecil maupun skala besar, tergantung pada seberapa kompleks suatu data dan apa tujuan dari menganalisa data tersebut. Transformasi data bertujuan agar data lebih terstruktur, konsisten dan sesuai dalam melakukan analisis hingga nantinya proses pengambilan keputusan dapat dilakukan dengan lebih akurat dan lebih efisien lagi. Data *Preprocessing* adalah langkah awal dalam analisis data, yang mana data mentah dibersihkan sebelum diolah lebih lanjut. Tahap ini bertujuan untuk menghilangkan data yang tidak diperlukan, menangani nilai yang hilang, dan memformat ulang data agar sesuai untuk kebutuhan analisis dengan tujuan akhir untuk memastikan bahwa data yang digunakan sudah bebas dari gangguan yang akan berdampak pada hasil kualitas dari analisis.

3.6 Data Splitting

Data splitting sesuai namanya jika diartikan ke bahasa Indonesia berarti membagi data yang mana data yang di bagi adalah berdasar dataset yang kita punya. Data dibagi dengan rasio 80% dan 20% yang mana 80% data pelatihan sementara 20% adalah data pengujian nya. Proses membagi data ini menjadi salah satu langkah yang penting dalam ranah *machine learning* karena memungkinkan model untuk belajar dari Sebagian besar data, namun tetap menyisakan Sebagian data yang belum pernah dilihat model untuk proses evaluasi. Dan alasan kenapa digunakan nya rasio 80% dan 20% selain karena ini adalah rasio yang cukup umum dalam praktik *machine learning* adalah karena rasio ini dinilai ideal dalam hal memberikan data yang banyak untuk melatih model dan juga menyediakan data yang cukup untuk mengukur performa model secara objektif. Sebanyak 80% dari total dataset akan dialokasikan sebagai data *training*, sedangkan 20% sisanya akan digunakan sebagai data *test*. Pembagian ini bertujuan untuk melatih model pada Sebagian besar dataset untuk memahami pola-pola yang ada, sementara Sebagian kecil akan digunakan untuk menguji kinerja model guna memastikan kemampuan generalisasinya dalam mengklarifikasikan data baru. [13]. Untuk kode nya bisa dilihat pada *code* berikut

3.7 Modelling

Modelling merupakan suatu konteks dalam *machine learning* adalah suatu proses membangun, melatih, dan mengevaluasi suatu model yang dapat membuat prediksi atau keputusan berdasarkan data input yang tersedia. Namun, adapun tujuan utama dari *modelling* adalah untuk menangkap suatu pola, hubungan, ataupun suatu informasi yang ada dalam data dan akan digunakan untuk membuat keputusan atau perkiraan terhadap data baru yang belum pernah dilihat sebelumnya.

3.8 Implementasi Pada Website

Pengimplementasian hasil prediksi ke dalam website merupakan langkah yang sangat penting dalam penelitian ini, karena dengan melakukan pengimplementasian ke dalam bentuk website akan mempermudah pengguna atau user untuk mengakses dan memanfaatkan model prediksi produksi padi yang telah dibuat. Dengan adanya website, pengguna tidak harus memahami atau menjalankan kode python secara langsung. Cukup hanya mengupload dataset milik mereka dan memilih tahun mulai prediksi dan tahun akhir prediksi, kemudian sistem akan memberikan hasil prediksinya secara otomatis dalam bentuk grafik dan tabel yang lebihh mudah dibaca bagi orang awam. Penggunaan website juga memungkinkan model ridge regression yang telah dilatih di google colab untuk diakses lebihh luas oleh siapa saja, baik itu petani, para pelaku usaha pertanian, para akademisi, dan masih banyak lagi yang bersangkutan dengan dunia pertanian. Hal ini tentu saja menjadikan hasil penelitian lebihh aplikatif dan tidak hanya terbatas dalam ruang lingkup akademik. Website ini juga dibuat sudah bersifat user-friendly, yang mana artinya website ini sudah mudah digunakan oleh para pengguna yang masih awam tanpa latar belakang di dunia IT. User hanya perlu mengunggah file dataset mereka dengan format .CSV dan menentukan tahun awal prediksi dan tahun akhir prediksi. Lalu hasil prediksi akan ditampilkan menggunakan visualisasi seperti grafik garis, pie chart, dan diagram batang.

Dengan implementasi model ke dalam bentuk website, penelitian ini tidak hanya berkontribusi dalam hal teknis pemodelan, tetapi juga dalam penyebaran teknologi prediksi secara praktis dan efisien. Langkah ini juga menjadi bentuk penerapan nyata dari integrasi antara machine learning dan pengembangan web dalam menyelesaikan masalah nyata di sektor pertanian.

Seperti halnya seperti saat awal melakukan build model ridge regression di google colab, di website juga harus melakukan import dan install library python yang akan kita pakai nantinya. Namun, ada yang berbeda disini yaitu untuk bisa publish pada web maka harus memakai framework yang bernama flask. Adapun fungsi flask antara lain, yaitu:

1. Membangun API dan aplikasi web
2. Routing URL
3. Mengelola request dan response
4. Templating HTML

Pada beberapa gambar dibawah dapat dilihat merupakan hasil prediksi produksi padi yang dimulai pada tahun 2021 dan berakhir pada tahun 2026, yang mana seperti yang terlihat produksi paling besar adalah Provinsi Sumatra Utara dan paling kecil adalah Bengkulu.



Gambar 2. Tampilan Halaman Prediksi

4. KESIMPULAN

Dataset yang dibutuhkan untuk memprediksi produksi padi di pulau Sumatra menggunakan metode ridge regression harus mencakup variable tahun, provinsi, luas panen, curah hujan, kelembapan dan suhu rata-rata. Harus di ingat bahwa hasil prediksi machine learning bersifat perkiraan dan mungkin berbeda dari kenyataan di kemudian hari karena potensi bias dan error pada model ataupun hal tidak terduga yang terjadi di kemudian hari. Akurasi prediksi sangat dipengaruhi oleh jumlah data yang dimiliki, semakin banyak data pada dataset maka akan semakin banyak data yang dapat dipelajari oleh model machine learning dengan begitu hasil yang didapat akan lebih akurat. Sama halnya dengan manusia yang menjadi lebih pintar dengan lebih banyak belajar. Namun, prediksi untuk tahun yang terlalu jauh cenderung kurang akurat karena keterbatasan jumlah data yang tersedia untuk model. Berdasarkan hasil evaluasi, model ridge regression mendapat nilai R2 sebesar 0.93 yang mana artinya model ini mampu menjelaskan 93% varian dalam data. Nilai RMSE sebesar 2,84 dan MAE sebesar 2,19 yang artinya model ini memiliki tingkat kesalahan yang relative rendah. Untuk hasil prediksi produksi dengan rentang waktu dari tahun 2021 sampai 2026 menunjukkan bahwa Provinsi Sumatra Utara menjadi wilayah dengan estimasi produksi paling tinggi, dengan angka rata-rata di atas 80 juta ton per tahun. Sementara itu, Bengkulu dan Riau diprediksikan memiliki tingkat produksi yang paling rendah, yaitu di bawah 20 juta ton. Implementasi model ke dalam bentuk website berbasis python dan flask menjadikan sistem ini mudah untuk diakses oleh berbagai kalangan. Diharapkan penelitian ini dapat menjadi kontribusi yang nyata dalam pengembangan teknologi prediktif untuk sektor pertanian yang berbasis data.

REFERENCES

- [1] Mergono Adi Ningrat, Carolina Diana Mual, and Yohanis Yan Makabori, "Pertumbuhan dan Hasil Tanaman Padi (*Oryza sativa* L.) pada Berbagai Sistem Tanam di Kampung Desay, Distrik Prati, Kabupaten Manokwari," *Pros. Semin. Nas. Pembang. dan Pendidik. Vokasi Pertan.*, vol. 2, no. 1, pp. 325–332, 2021, doi: 10.47687/snppvp.v2i1.191.
- [2] J. Pamungkas, R. Winanjaya, W. Robiansyah, and G. Artikel, "Metode Jaringan Saraf Tiruan dalam Memprediksi Jumlah Produksi Daging Itik Manila Berdasarkan Provinsi Di Indonesia," *JOMLAI J. Mach. Learn. Artif. Intell.*, vol. 1, no. 3, pp. 2828–9099, 2022, doi: 10.55123/jomlai.v1i3.918.
- [3] T. Nizami, M. A. Mustaqim, and W. Ariannor, "Analisis kinerja model machine learning dalam prediksi gagal panen gabah," *Progresif J. Ilm. Komput.*, vol. 21, no. 1, pp. 184–192, 2025, [Online]. Available: <https://ojs.stmik-banjarbaru.ac.id/index.php/progresif/article/view/2501>
- [4] I. Nurdin, Sugiman, and Sunarmi, "Penerapan Kombinasi Metode Ridge Regression (RR) dan Metode Generalized Least Square (GLS) untuk Mengatasi Masalah Multikolinearitas dan Autokorelasi," vol. 41, no. 1, pp. 58–68, 2018.
- [5] Khesya, Nurhaliza, "MENGENAL FLOWCHART DAN PSEUDOCODE DALAM ALGORITMA DAN PEMROGRAMAN," OSF Preprints, 29 Dec. 2021. Web.
- [6] Ariya Pannadhitthana Candra, "Analisis Data Menggunakan Python: Memperkenalkan Pandas dan NumPy," *JISED*, vol. 3, no. 1, pp. 11–16, Mar. 2025.
- [7] D. P. S. A. Diandra, "Monitoring Lahan Parkir Menggunakan Python, CV2, CVZONE dan NumPy," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 1, pp. 157, Feb. 2025.
- [8] N. M. Surbakti, Angelyca, A. Talia, C. B. Perangin-Angin, D. O. Nainggolan, N. D. Friskauly, and S. R.

-
- B. Tumorang, "Penggunaan Bahasa Pemrograman Python dalam Pembelajaran Kalkulus Fungsi Dua Variabel," *Algoritma: Jurnal Matematika, Ilmu Pengetahuan Alam, Kebumihan dan Angkasa*, vol. 2, no. 3, pp. 98–107, May 2024, doi: 10.62383/algoritma.v2i3.67.
- [9] A. Zuhairah, "Penerapan Algoritma Random Forest, Support Vector Machines (SVM) dan Gradient Boosted Tree (GBT) untuk Deteksi Penipuan (Fraud Detection) pada Transaksi Kartu Kredit," Skripsi, Jurusan Teknik Informatika, Universitas Islam Negeri Syarif Hidayatullah, Jakarta, 2022.
- [10] R. Arisandi, D. Ruhiat, and E. Marlina, "Implementasi Ridge Regression untuk Mengatasi Gejala Multikolinearitas pada Pemodelan Curah Hujan Berbasis Data Time Series Klimatologi," **Jurnal Riset Matematika dan Sains Terapan**, vol. 1, no. 1, pp. 1–11, Des. 2021.
- [11] S. N. Zahra and P. E. P. Utomo, "Visualisasi Data Penjualan Barang Retail di Seluruh Dunia Menggunakan Tableau," **Jurnal Nasional Ilmu Komputer**, vol. 4, no. 3, p. 12, Aug. 2023.
- [12] A. A. A. Daniswara and I. K. D. Nuryana, "Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru," **JINACS (Journal of Informatics and Computer Science)**, vol. 5, no. 1, p. 97, 2023.
- [13] C. V. Angkoso, M. A. N. Thrisna, B. D. Satoto, and A. Kusumaningsih, "Optimasi Klasifikasi Sentimen Menggunakan Random Forest dengan Preprocessing K-Means Clustering dan SMOTE," **EPIN (Jurnal Edukasi dan Penelitian Informatika)**, vol. 10, no. 3, p. 389, Dec. 2024.