

PREDIKSI HASIL PRODUKSI JAGUNG DI JAWA BARAT DENGAN METODE ALGORITMA REGRESI LINEAR MENGGUNAKAN GOOGLE COLLAB

Guruh Maulana¹, Raditya Danar Dana²

¹ Komputerisasi Akuntansi, STMIK IKMI Cirebon

² Manajemen Informatika, STMIK IKMI Cirebon

Jalan Perjuangan No. 10B Karyamulya Kec. Kesambi Kota Cirebon Jawa Barat 45135, Indonesia

guruhmaulana555@gmail.com

ABSTRAK

Jagung merupakan makanan pokok penting di Indonesia karena mempunyai sumber karbohidrat terbesar kedua setelah nasi. Jagung memiliki sumber karbohidrat yang penting bagi Sumber Daya manusia karena jagung memiliki kandungan tinggi lemak dan kalori dari pada nasi serta dapat berperan penting untuk pencegahan penyakit. Sebagai permasalahan penelitian ini, terdapat tantangan terkait dengan permintaan produksi Jagung di Jawa Barat yang cukup signifikan. Akibat dari pertumbuhan penduduk yang terus meningkat namun produksi jagung sendiri masih rendah, menyebabkan ketimpangan dalam pemenuhan kebutuhan produksi jagung di Jawa Barat. Seiring dengan berjalannya waktu, pertumbuhan penduduk di Jawa Barat semakin meningkat dan produksi jagung mengalami perubahan setiap tahunnya, maka diperlukan suatu metode untuk mengetahui apakah produksi jagung akan menurun atau meningkat. Selain itu, beberapa faktor yang mempengaruhi peran penting terhadap peningkatan atau penurunan produksi jagung yaitu produktivitas jagung, luas panen, dan luas areal tanaman yang di mana beberapa faktor tersebut tidak dapat diprediksi. Dalam mengatasi permasalahan ini, peneliti akan menggunakan metode Algoritma Regresi Linear untuk memprediksi hasil produksi jagung. Penelitian ini bertujuan untuk mengatasi permasalahan permintaan produksi jagung ke depannya. Berdasarkan hasil yang peneliti lakukan dengan metode algoritma linear regresi hasil produksi jagung sebanyak 2.056.890 ton dengan nilai *MAE*: 5741.21, *MSE*: 124802556, *RMSE*: 124802556.11, dan *Performance*: 0.991046 (99.10%) dan hasil persamaan regresi linear yaitu $Y = -2753.018316202004 * x + [7.72294428]$.

Kata kunci : *Prediksi, Produksi Jagung, Jawa Barat, Regresi Linear, Perekonomian*

1. PENDAHULUAN

Menurut Badan Pusat Statistik Republik Indonesia (BPS), Jawa Barat merupakan provinsi dengan jumlah penduduk terpadat dan menempati urutan pertama dari tujuh provinsi terbesar di Indonesia. Jumlah penduduk provinsi Jawa Barat pada tahun 2019 sebanyak 48.274.162 jiwa, dan pada tahun 2022 sebanyak 49.405.810 jiwa. Jumlah penduduk tersebut mengalami peningkatan sebesar 0.8% atau setara 1.131.648 jiwa dari jumlah penduduk pada tahun 2019. Pertumbuhan penduduk di Jawa Barat yang terus mengalami peningkatan setiap tahunnya, menyebabkan ketimpangan bahan pangan pokok akan terus bertambah salah satunya yaitu produksi jagung.

Jagung merupakan makanan pokok penting di Indonesia karena mempunyai sumber karbohidrat terbesar kedua setelah nasi. Jagung memiliki sumber karbohidrat yang penting bagi Sumber Daya Manusia karena memiliki kandungan tinggi lemak dan kalori dari pada nasi serta dapat berperan penting untuk pencegahan [1]. Sebagai permasalahan penelitian ini, terdapat tantangan terkait dengan permintaan produksi jagung di Jawa Barat yang cukup signifikan. Akibat dari pertumbuhan penduduk yang terus meningkat namun produksi jagung sendiri masih rendah, menyebabkan ketimpangan dalam pemenuhan kebutuhan produksi jagung di Jawa Barat.

Seiring dengan berjalannya waktu, pertumbuhan penduduk di Jawa barat semakin meningkat dan produksi jagung mengalami perubahan setiap

tahunnya. Selain itu, beberapa faktor yang mempengaruhi peran penting terhadap peningkatan atau penurunan produksi jagung yaitu produktivitas jagung, luas panen, dan luas areal tanaman yang di mana beberapa faktor tersebut tidak dapat di prediksi.

Selanjutnya menurut data yang peneliti per-oleh mengenai produksi jagung tahun 2018 – 2022 di Jawa Barat, total produksi jagung tahun 2018 sebanyak 1.429.148 ton, tahun 2019 sebanyak 1.306.548 ton, tahun 2020 sebanyak 1.343.181 ton, tahun 2021 sebanyak 1.176.476 ton, dan tahun 2022 sebanyak 1.165.047 ton. Adapun penurunan jumlah produksi jagung di Jawa Barat tahun 2018 hingga 2022 disebabkan karena variabel *x* yang mempengaruhi variabel *y*. Kemudian variabel yang digunakan dalam penelitian ini yaitu untuk variabel *x* produktivitas jagung, luas panen, dan luas areal tanaman dan variabel *y* produksi jagung.

Berdasarkan latar belakang di atas, maka perumusan masalah yang dihasilkan adalah Bagaimana prediksi hasil produksi jagung di Jawa Barat menggunakan algoritma linear regresi dan bagaimana pengukuran *performance* hasil produksi jagung di Jawa Barat menggunakan algoritma linear regresi.

Berdasarkan permasalahan yang ada, maka peneliti akan melakukan penelitian mengenai prediksi hasil produksi jagung di Jawa Barat dan mencari hasil *performance* berdasarkan hasil produksi jagung di Jawa Barat menggunakan algoritma regresi linear.

Oleh karena itu, peneliti mengambil judul “Prediksi Hasil Produksi Jagung di Jawa Barat Dengan Metode Algoritma Linear Regresi Menggunakan *Google Colab*”

2. TINJAUAN PUSTAKA

2.1. Literatur Review

Penelitian yang dilakukan oleh Muh Adha, Ema Utami, dan Hanafi yang berjudul “Prediksi Produksi Jagung Menggunakan Algoritma A-priori dan Regresi Linear Berganda”. Jurnal ini membahas tentang permasalahan permintaan jagung padahal produksi dalam negeri masih rendah sehingga menimbulkan ketimpangan dalam pemenuhan permintaan jagung. Metode yang digunakan dalam penelitian ini adalah algoritma regresi linear berganda. Hasil penelitian diperoleh dengan menggunakan uji MAD 54, MSE 231372, RMSE 481, dan hasil persamaan regresi berganda adalah $Y = -70.860 + -0.505x_1 + 7.068x_2 + -4.349x_3$. Metode asosiasi digunakan dalam perhitungan yang menghasilkan dua aturan dengan nilai dukungan 10% dan tingkat kepercayaan 70%. Diperoleh *IF BUY Gandasil THEN BUY Ricestar with confidence 100%* [1].

Penelitian yang dilakukan oleh Surya Fajri, Heru Gunawan, Lokot Ridwan Barubara, Zunaida Sitorus yang berjudul “Prediksi Hasil Produksi Tanaman Tomat di Indonesia Menurut Provinsi Menggunakan Algoritma Fletcher - Reeves” Jurnal ini membahas tentang isu meningkatnya permintaan sayuran tomat akibat pertumbuhan penduduk di Indonesia. Metode yang digunakan dalam penelitian ini menggunakan algoritma gradien konjugasi Fletcher-Reeves. Hasil Penelitian Berdasarkan perbandingan jaringan ketiga model arsitektur diperoleh hasil terbaik pada model 5-17-1. Model ini memiliki nilai MSE terendah dibandingkan dua model lainnya, yaitu 0,0009915 dibandingkan dengan 0,0010851 dan 0,0049764, serta akurasi yang lebih tinggi, yaitu 94% 91% dan 88%. Oleh karena itu, model 5-17-1 digunakan untuk memprediksi hasil produksi tanaman tomat di masa depan (2022 dan 2023). Berdasarkan hasil perkiraan untuk akhir tahun 2022 dan 2023, produksi tomat diperkirakan sedikit meningkat di 18 negara bagian. Kesimpulannya adalah prediksi produksi tomat menggunakan algoritma Fletcher-Reeves sangat baik karena tingkat kesalahan yang rendah dan akurasi yang tinggi [2].

Penelitian yang dilakukan oleh Ardika Satria, Rizty Mualida badri, dan Ira Safitri yang berjudul “Prediksi Hasil Panen Tanaman Pangan Sumatera dengan Metode Machine Learning”. Jurnal ini membahas isu peningkatan produksi tanaman pangan di Sumatera meningkat sejak tahun 1993. Perluasan lahan pertanian juga meningkat secara signifikan. Di sisi lain, produk pertanian di Sumatera sangat rentan terhadap perubahan iklim global terutama kenaikan suhu dari tahun ke tahun. Untuk mencapai hasil yang maksimal dalam produksi tanaman pangan utama Pulau Sumatera, maka perlu dilakukan peramalan hasil

panen. Tujuan dari penelitian ini adalah untuk menemukan algoritma terbaik dalam memprediksi hasil tanaman pangan di Pulau Sumatera yaitu padi, jagung, kacang tanah, kedelai, singkong, dan ubi jalar. Metode yang digunakan dalam penelitian ini menggunakan Metode yang digunakan adalah Machine Learning, yaitu *Random Forest (RF)*, *Decision Tree (DT)*, *Gradient Boosting (GB)*, *Extra Tree (ET)*, *Support Vector Machine (SVM)*, dan *Artificial Neural Network (ANN)*. Hasil penelitian yang dilakukan menunjukkan nilai koefisien R2 produksi padi sebesar 0,897. 0,893; 0,957; 0,968; 0,928; dan 0,909. Sedangkan produksi barang kebutuhan pokok lainnya sebesar 0,754. 0,786; 0,721; 0,913; 0,509; dan 0,90. Model *Extra Tree* memperoleh nilai koefisien R2 tertinggi dan lebih akurat dibandingkan model lainnya [3].

Penelitian yang dilakukan oleh Nurhayati, Mhd.Buhari Sibuea, Dedi Kusbiantoto, Martina Silaban, dan Anjar Wanto yang berjudul “Implementasi Algoritma Resilient untuk Prediksi Potensi Produksi Bawang Merah di Indonesia”. Jurnal ini meramalkan potensi produksi bawang merah di Indonesia sehingga pemerintah mempunyai acuan dan informasi untuk menetapkan kebijakan ekonomi yang tepat dan dapat lebih meningkatkan produksi bawang merah atau setidaknya menstabilkannya setiap tahunnya. Metode yang digunakan dalam penelitian ini mengadopsi metode algoritma Resilient. Hasil penelitian ini dianalisis menggunakan empat model arsitektur jaringan: 6-5-1, 6-10-1, 6-17-1, dan 6-29-1. Berdasarkan analisis terhadap keempat model yang digunakan diperoleh hasil bahwa model 6-17-1 merupakan model yang sangat baik karena memiliki nilai uji mean squared error (MSE) yang lebih rendah dibandingkan ketiga model lainnya yaitu sebesar 0,0337792. Tingkat akurasi yang bagus. 79, dengan tingkat kesalahan 0,04. Model arsitektur ini digunakan untuk memprediksi potensi produksi bawang merah di Indonesia. Berdasarkan hasil ramalan keseluruhan tiap provinsi, potensi produksi rakkyo Indonesia cenderung menurun pada akhir tahun 2022 dibandingkan tahun 2021. Kesimpulan yang dapat diambil adalah penerapan algoritma Resilient pada permasalahan data produksi bawang merah Indonesia sudah sangat baik, namun tingkat akurasi belum terlalu tinggi sehingga diperlukan kajian yang lebih detail [4].

Penelitian yang dilakukan oleh Nazarullah Moha Lalapa, Warid Yunus, dan Serwin yang berjudul “Implementasi Metode Regresi Linear Sederhana Untuk Prediksi Harga Cabai”. ini membahas mengenai permasalahan produksi cabai rawit di Provinsi Gorontalo yang merupakan daerah penghasil cabai rawit terbesar kedua setelah Lahir Bolango. akibat produksi cabai rawit di dataran tinggi Tanjung Gorontalo, harga cabai rawit di pasaran kerap turun hingga Rp 20.000/kg, namun menjelang hari raya keagamaan, harga cabai rawit naik signifikan. Tujuan dari penelitian ini adalah membangun sistem prediksi

harga cabai rawit. Variabel yang digunakan adalah kuantitas produksi sebagai nilai X dan harga sebagai nilai Y. Metode yang digunakan dalam penelitian ini adalah metode prediksi dengan menggunakan regresi linier sederhana dan metode MAPE. Berdasarkan hasil prediksi, diperoleh tingkat kesalahan masing-masing sebesar 24,00% dan tingkat kebenaran sebesar 76,00%. Oleh karena itu, sistem yang dibuat cocok untuk memprediksi harga cabai rawit [5].

2.2. Jagung

Jagung (*Zea Mays*) adalah tanaman biji-bijian yang berasal dari Amerika Tengah dan Selatan. Tanaman yang mempunyai batang yang tinggi dan daun yang lebar. Buah jagung, disebut juga tongkol jagung atau sekam jagung atau sekam jagung, menumbuhkan biji-bijian yang biasa disebut biji jagung. Biji jagung dapat dimakan dan banyak digunakan sebagai bahan makanan pokok manusia, peternakan, dan industri. Jagung merupakan tanaman pangan penting di seluruh dunia. Biji jagung mengandung karbohidrat, protein, serat, serta beberapa vitamin dan mineral. Selain dikonsumsi langsung, jagung juga dimanfaatkan sebagai bahan produk olahan, antara lain tepung jagung, minyak jagung, *sirup* jagung, dan pakan jagung [6].

Ada beberapa jenis jagung yang dapat dibedakan berdasarkan berbagai cirinya seperti warna, tekstur, dan kegunaannya [7]. Berikut beberapa jenis-jenis jagung yang sering digunakan :

- 1) Jagung Manis
Jagung ini mempunyai rasa yang manis dan biasanya di konsumsi secara mentah atau dimasak. Jagung manis ini memiliki butiran yang lebih lembut dan kandungan gula yang lebih tinggi dibandingkan jagung keras.
- 2) Jagung Ladang
Jagung jenis ini mempunyai butiran yang lebih keras dan rasa manisnya kurang dibandingkan jagung manis.
- 3) Jagung Keras
Umumnya digunakan untuk pakan ternak, keperluan industri, dan produk olahan seperti tepung jagung.
- 4) Jagung Berwarna
Jagung yang warna bijinya berbeda-beda seperti merah, biru, kuning, dan ungu. Jagung berwarna sering digunakan untuk tujuan dekoratif dan sebagai bahan makanan khusus.
- 5) Jagung Pipilan (*Popcorn*)
Jagung jenis ini mempunyai cangkang yang kuat dan dapat meledak bila dipanaskan. *Popcorn* sering disajikan sebagai camilan populer.
- 6) Jagung Hortikultura (jagung Hias)
Jagung Hortikultura digunakan terutama untuk keperluan hias. Mereka hadir dalam berbagai warna dan bentuk yang menarik dan sering digunakan dalam industri dalam penataan dekoratif.

7) Jagung Pulut (Flint Corn)

Jagung varietas ini berbiji keras dan banyak digunakan dalam industri sebagai bahan baku berbagai produk olahan jagung.

8) Jagung Lilin

Memiliki biji seperti “lilin” atau lengket saat dimasak. Jagung jenis ini sering digunakan dalam industri pangan sebagai bahan pengental atau pembuatan produk tepung jagung *waxy*.

2.3. Data Mining

Data Mining merupakan proses penggalian informasi dan pola yang bermanfaat dari data yang sangat besar. Data Mining mencakup pengumpulan data, ekstraksi data, analisis data, dan statistik data. Data Mining juga dikenal sebagai *Knowledge discovery*, *Knowledge extraction*, *data/pattern analysis*, *information harvesting*, dan lain-lain [8]. Proses dan tujuan KDD akan sangat bergantung pada penentuan metode atau algoritma yang tepat. Berikut tahapan mengenai *Knowledge Discovery Database* (KDD) :

1) Data Selection

Sebelum memulai proses di KDD, pilih data dari dataset yang ingin digunakan. Data yang sudah dipilih kemudian diproses melalui data mining dan disimpan dalam file yang terpisah.

2) Pre-Processing

Sebelum diproses dengan data *mining*, data yang terpilih selanjutnya dibersihkan terlebih dahulu dengan data *cleaning* yang menjadi fokus KDD. Proses pembersihan berfungsi untuk menghapus *missing values*, duplikat data, dan memperbaiki kesalahan pada data.

3) Data Transformation

Proses transformasi adalah proses data yang telah dibersihkan, sehingga sesuai dengan ketentuan Data Mining. Proses KDD merupakan proses kreatif dan sangat bergantung pada jenis atau pola informasi yang akan dicari dalam *database*.

4) Data Mining

Data *mining* adalah proses menemukan pola atau informasi menarik dari data terpilih dengan menggunakan teknik atau metode tertentu. Pemilihan metode algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5) Interpretasi/Evaluation

Pola informasi yang diperoleh dari proses data *mining* harus ditampilkan dalam bentuk yang mudah dipahami oleh pihak berkepentingan. Tahap ini merupakan bagian dari KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

Menurut penelitian yang dilakukan oleh [9], ada beberapa metode data mining, antara lain :

- 1) *Classification*
Klasifikasi adalah proses pencarian pola atau fungsi yang menggambarkan dan membedakan kelas data atau konsep data.
- 2) *Clustering*
Digunakan untuk menemukan mencari atribut – atribut grup kedalam segmentasi berdasarkan kesamaan.
- 3) *Association*
Berfungsi untuk menemukan hubungan antara sekumpulan atribut atau sekumpulan elemen, berdasarkan elemen yang muncul dan aturan asosiasi yang ada.
- 4) *Regression*
Bertujuan untuk menemukan prediksi dari pola yang ada.
- 5) *Forecasting*
Bertujuan untuk menemukan prediksi dari pola yang ada.
- 6) *Sequence Analysis*
Berfungsi untuk menemukan pola urutan dari urutan kejadian.
- 7) *Deviation Analysis*
Digunakan untuk mencari kejadian langka yang sangat berbeda dengan kondisi normal (kejadian abnormal).

2.4. Algoritma Regresi Linear

Regresi linear adalah analisis statistik yang digunakan untuk menganalisis dan memahami hubungan antara dua variabel atau lebih. Regresi linier merupakan metode yang umum digunakan untuk melakukan prediksi dengan menggunakan garis lurus sehingga bisa dilihat dan bisa menggambarkan hubungan di antara dua variabel atau lebih [10]. Regresi linier termasuk dalam kategori *supervised learning*. Rumus regresi linear secara umum adalah :

$$Y = f(x) \quad (1)$$

Keterangan :

Y = Variabel Dependen

X = Variabel Independen

Persamaan Regresi Linear adalah sebagai berikut :

$$Y = a + bX \quad (2)$$

Di mana :

Y : Variabel Terikat

a : nilai intercept / konstanta

b : koefisien

X : variabel bebas

Rumus untuk mengetahui nilai a dan b :

$$a = \frac{\sum y(\sum x^2) - \sum x \sum xy}{n \sum x^2 - (\sum x)^2} \quad (3)$$

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} \quad (4)$$

Langkah – langkah perhitungan regresi linear adalah :

1. Identifikasi sebuah regresi variabel y dan x
2. Menyiapkan data yang akan dijadikan variabel x dan y
3. Hitung nilai x^2 , xy , dan masing-masing jumlah nilai tersebut
4. Kemudian lakukan perhitungan nilai a & b dengan rumus [2] dan [3]

2.5. Rancangan Dataset

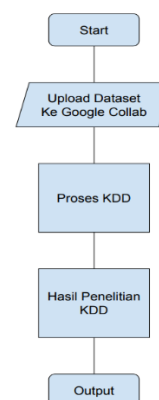
Dalam menentukan prediksi hasil produksi jagung di Jawa Barat terdapat beberapa variabel yang akan digunakan untuk memprediksi hasil produksi jagung ke depannya. Berdasarkan *dataset* tersebut terdapat 140 baris data dengan beberapa variabel yang akan digunakan dalam memprediksi data jagung di Jawa Barat yakni variabel x (produktivitas tanaman, luas panen, luas areal tanaman) dan variabel y (produksi jagung) [10]. Berikut ini adalah tabel penjelasan dari variabel tersebut yaitu :

Tabel 1. Keterangan variabel *dataset* produksi jagung

Variabel	Keterangan
Nama kabupaten kota	Menjelaskan kabupaten / kota yang berada di Jawa Barat
Tahun	Menjelaskan Tahun dari data yang diambil
Produktivitas Jagung	Menjelaskan jumlah hasil panen jagung yang diperoleh per-unit
Luas Panen	Menjelaskan total lahan yang digunakan untuk menanam jagung dan kemudian dipanen dalam suatu periode
Luas Areal Tanaman	Memberikan gambaran tentang seberapa besar wilayah yang dialokasikan untuk budidaya jagung.
Produksi Jagung	Menjelaskan hasil panen jagung dalam periode tertentu dalam satuan ton

2.6. Flowchart Algoritma Regresi Linear

Berikut adalah uraian dari metode algoritma regresi linear ditunjukkan pada gambar 1 :



Gambar 1. Flowchart algoritma regresi linear

Pada gambar 1 di atas merupakan *flowchart* dari proses penelitian *dataset* menggunakan metode regresi linear. Di mana dimulai dari penguploadan data produksi jagung ke dalam *google collab*. Selanjutnya di proses ke dalam tahapan KDD hingga menemukan hasil nilai *performance* dan hasil prediksi produksi jagung.

2.7. Google Colaboratory

Google Colaboratory atau *Google Colab* adalah dokumen yang dapat dieksekusi yang dapat digunakan untuk menyimpan, menulis, serta membagikan program yang telah ditulis melalui *Google Drive*. Produk ciptaan *Google Research* ini dikenal sebagai *software* apik dengan menyediakan berbagai keperluan termasuk GPU yang dapat digunakan secara gratis. Perangkat Lunak ini pada dasarnya seperti *Jupyter Notebook* gratis dalam format *Cloud* yang dijalankan menggunakan browser, seperti *Mozilla Firefox* dan *Google Chrome* [11].

3. METODE PENELITIAN

3.1. Sumber Data

Saat melakukan penelitian untuk tugas akhir ini, peneliti akan dihadapkan dengan data. Pada penelitian ini peneliti ingin mengetahui prediksi hasil produksi jagung dan hasil pengukuran *performance* dari hasil produksi jagung di Jawa Barat. Peneliti mengambil data hasil produksi jagung dari tahun 2018 hingga 2022 dari *website* Open Data Jawa Barat yang diperoleh dari Dinas Tanaman Pangan dan Hortikultura. Dengan nama sumber data yaitu Hasil Prediksi Produksi Jagung Kabupaten/Kota di Jawa Barat. Dinas Tanaman Pangan dan Hortikultura beralamat di Jl. Surapati No. 71 Kelurahan Sadang Serang Kecamatan Coblong, Kota Bandung 40133. Peneliti melakukan pengumpulan data melalui link internet <https://opendata.jabarprov.go.id/> pada hari Kamis, 07 Desember 2023 Pukul 08:15 WIB, data penelitian dapat dilihat pada tabel 2 di bawah.

Data yang digunakan dalam penelitian ini adalah data *skunder* yang dikumpulkan dan dicatat dengan berbagai cara, sebagai informasi opsional bagi sebagian besar pembaca dan sebagai bukti yang dapat diverifikasi dalam bentuk catatan atau laporan yang dikumpulkan dari repositori yang diterbitkan atau tidak diterbitkan.

3.2. Teknik Pengumpulan Data

Di bawah ini merupakan teknik pengumpulan data yang dilakukan oleh peneliti dalam tugas akhir :

1) Studi Dokumentasi

Studi Dokumentasi merupakan cara pengumpulan data yang diperoleh dari dokumen atau catatan yang ada, pada penelitian ini data diambil dari *website* resmi Open Data Jawa Barat. Kemudian data yang diperoleh akan peneliti pelajari, dipilih sesuai dengan kriteria, diolah melalui data *mining*, dan selanjutnya akan digunakan untuk penelitian tugas akhir, guna

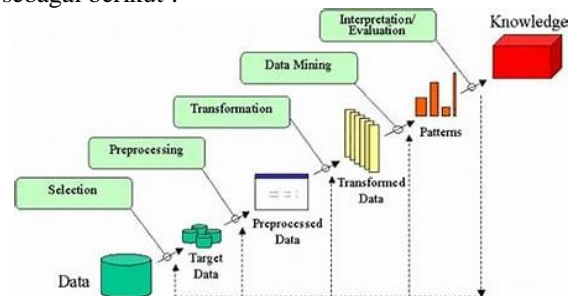
untuk mendapatkan informasi penelitian tugas akhir ini.

2) Studi Literatur

Studi Literatur adalah teori dasar yang akan digunakan untuk menyelesaikan pada penelitian tugas akhir ini menjadi acuan yang kuat dalam melakukan penelitian. Adapun tujuan referensi yang digunakan berupa jurnal nasional yang berkaitan dengan topik penelitian.

3.3. Tahapan Perancangan

Tahapan Perancangan adalah kriteria dan tahapan yang diterapkan pada sebuah penelitian untuk mencapai tujuan penelitian. Tahapan perancangan penelitian ini berdasarkan proses *Knowledge Discovery In Database*, dengan menerapkan langkah demi langkah yang dilaksanakan secara berurutan untuk mencapai tahapan penyelesaian. Tahapan perancangan tugas akhir dapat dilihat gambar 2 sebagai berikut :



Gambar 2. Proses KDD

Berikut penjelasan lebih lanjut mengenai proses *Knowledge Discovery Database* (KDD) berdasarkan gambar 3.1 di atas :

1) Data Selection

Tahapan data *selection* ini peneliti melakukan pemilihan data menggunakan *dataset* hasil produksi jagung di Jawa Barat tahun 2018 hingga 2022 yang sekiranya diperlukan sesuai dengan kebutuhan. Kemudian peneliti melakukan *import* data ke dalam *google colab* untuk mengetahui informasi data.

2) Data Pre-Processing

Setelah dilakukan pemilihan dan *import* data ke dalam *google colab*, selanjutnya peneliti melakukan pembersihan data dengan cara menghapus *missing values* dan duplikat data. Berikut adalah langkah – langkah dalam *pre-processing* data :

Missing Values

Pada tahapan ini peneliti ingin mengetahui, apakah data tersebut memiliki *missing values*. Jika ada maka peneliti melakukan eksekusi untuk menghapus *missing values* tersebut.

Duplikat Data

Pada tahapan ini peneliti ingin mengetahui, apakah data tersebut memiliki duplikat data. Jika ada maka peneliti melakukan eksekusi untuk menghapus data yang memiliki duplikat

3) Data Transformation

Setelah dilakukan pembersihan data, kemudian pada tahapan *data transformation* peneliti ingin mengubah *type* data pada variabel luas panen, dan luas areal tanaman dari *float64* ke *int64*.

4) Data Mining

Selanjutnya pada tahapan Data Mining peneliti ingin mengetahui terlebih dahulu hubungan antara variabel *x* dan *y*, kemudian dicari *korelasi matrix* di mana variabel *x* nanti akan diketahui menggunakan *korelasi matrix*. Apabila hasil *koefisien* 1 atau mendekati 1 yang artinya variabel tersebut berhubungan kuat secara searah. Maka variabel *x* akan digunakan untuk memprediksi hasil produksi jagung di Jawa Barat pada tahapan *evaluation*. Selanjutnya ketika variabel *x* sudah diketahui menggunakan *korelasi matrix*, maka peneliti ingin melihat grafik pengaruh dari variabel *x* dan *y*.

5) Evaluation

Selanjutnya pada tahapan *evaluation* adalah tahapan untuk menyimpulkan hasil dari data *mining* menggunakan *algoritma regresi linear*. Sebelum menyimpulkan, peneliti melakukan perbandingan terlebih dahulu antara 1 variabel *x* dan 2 variabel *x* dengan 1 variabel *y*. Apabila hasil sudah didapatkan maka peneliti mengambil hasil *accuracy* dan hasil prediksi tertinggi dan melakukan pengambilan hasil *RMSE*, *MAE*, dan nilai *performance* serta hasil grafik *regresi linear*.

4. HASIL DAN PEMBAHASAN

Pada tahap pertama peneliti ingin mengetahui hasil prediksi produksi jagung di Jawa Barat. Data yang digunakan dalam penelitian ini menggunakan data asli Hasil Produksi Jagung di Jawa Barat Tahun 2018 hingga 2022. Variabel yang ada pada data ini terdiri dari Nama Kabupaten kota, Tahun, Produktivitas Jagung, Luas Panen, Luas Areal Tanaman, dan Produksi Jagung. Dengan jumlah data sebanyak 140 baris data. Berikut peneliti lampirkan tahapan dalam melakukan Prediksi Produksi Jagung di Jawa Barat menggunakan bahasa pemrograman *python* dalam tahapan KDD :

4.1. Data Collection

Tahapan data *collection* ini peneliti melakukan pemilihan data menggunakan *dataset* Hasil Produksi Jagung di Jawa Barat tahun 2018 hingga 2022 yang sekiranya diperlukan sesuai dengan kebutuhan. Berikut Data Hasil Produksi Jagung di Jawa Barat :

Tabel 2. Dataset data jagung jawa barat 2018 – 2022

2K	T	PJ	LP	LA	PJ
Kabupaten Bogor	2018	80,34	584	800	4688
Kabupaten	2018	66,37	15384	16519	10210

2K	T	PJ	LP	LA	PJ
Sukabumi					
Kabupaten Cianjur	2018	77,75	13321	14398	103566
.....
Kota Tasikmalaya	2022	75,7	4	105	31
Kota Banjar	2022	26,62	222	219	606

Keterangan :

2K = Nama Kabupaten / Kota di Jawa Barat

T = Tahun

PJ = Produktivitas Jagung

LP = Luas Panen

LA = Luas Areal Tanaman

PJ = Produksi Jagung

Kemudian data yang sudah dipilih sesuai dengan kriteria dalam data *mining*, selanjutnya peneliti melakukan *import* data ke dalam *google collab* pada gambar 3 :

	nama_kabupaten_kota	tahun	produktivitasjagung	luaspanen	luasarealtanaman	produksijagung
0	KABUPATEN BOGOR	2018	80.34	584.0	800	4688
1	KABUPATEN SUKABUMI	2018	66.37	15384.0	16519	102104
2	KABUPATEN CIANJUR	2018	77.75	13321.0	14398	103566
3	KABUPATEN BANDUNG	2018	1.00	12014.0	11620	83410
4	KABUPATEN GARUT	2018	81.18	82160.0	82160	666963
...
134	KOTA BANJAR	2022	26.62	222.0	219	606
135	KABUPATEN TASIKMALAYA	2022	68.17	3571.0	5558	23627
136	KABUPATEN CIAMIS	2022	64.03	3068.0	2631	20368
137	KABUPATEN KUNINGAN	2022	71.88	3060.0	2353	21627
138	KABUPATEN CIREBON	2022	74.35	102.0	9290	771

Gambar 3. Data produksi jagung di *import* ke *google collab*

4.2. Data Preprocessing

Selanjutnya pada tahapan data *pre-processing* peneliti melakukan pengecekan *missing values* dan *duplicated data*, hasil tersebut terdapat *missing values* terhadap variabel produktivitas jagung (1), luas panen (3), luas areal tanaman (1), serta terdapat data *duplicated* sebanyak 10 baris gambar 4 dan 5 :

```
[5] # check missing values
df.isna().sum()

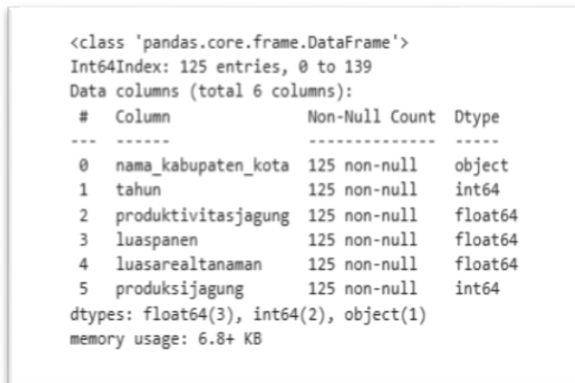
nama_kabupaten_kota    0
tahun                  0
produktivitasjagung    1
luaspanen              3
luasarealtanaman       1
produksijagung         0
dtype: int64
```

Gambar 4. Informasi *missing values*



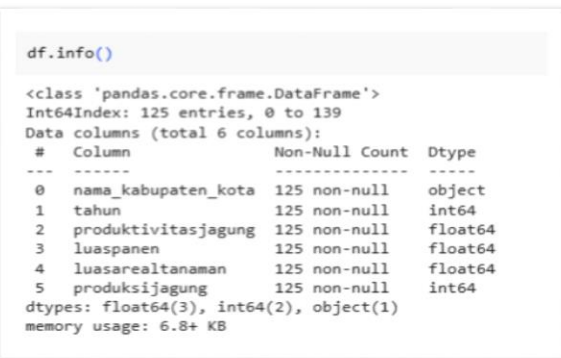
Gambar 5. Informasi data duplicate

Data yang sudah dibersihkan selanjutnya akan dilanjutkan data *transformation*, berikut informasi data yang sudah peneliti bersihkan pada gambar 6

Gambar 6. Informasi data setelah di *preprocessing*

4.3. Data Transformation

Pada tahap *transformation* peneliti ingin mengubah *type data* pada variabel luas panen dan luas areal tanaman dari *float64* ke *int64*, berikut adalah proses pengubahan tipe data gambar 7 dan 8 :



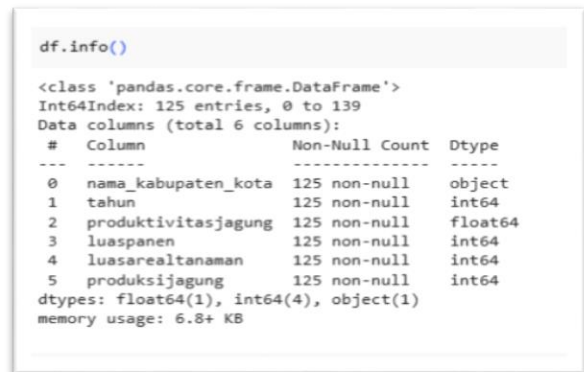
Gambar 7. Informasi data type



Gambar 8. Proses pengubahan data type

Selanjutnya, setelah proses pengubahan tipe data, maka variabel luas panen dan luas areal tanaman akan

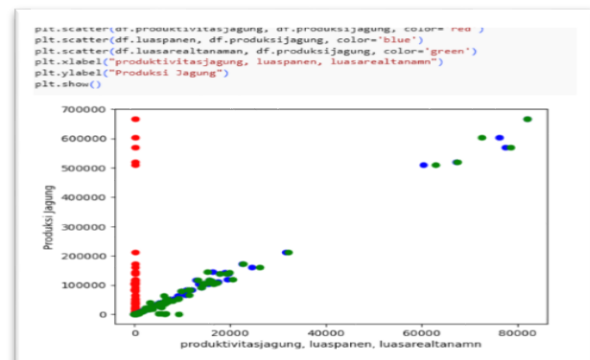
berubah format dari float64 ke int64, berikut adalah hasil transformasi data gambar 9 :



Gambar 9. Informasi data type setelah diubah

4.4. Data Mining

Pada tahapan data *mining* peneliti melakukan terlebih dahulu hubungan antara variabel x dan y, Berikut ini merupakan hubungan variabel x dan y, di mana variabel x adalah data produktivitas jagung yang disimbolkan dengan warna merah, luas panen yang disimbolkan dengan warna biru, dan data luas areal tanaman yang disimbolkan dengan warna hijau, variabel y adalah produksi jagung, berikut adalah

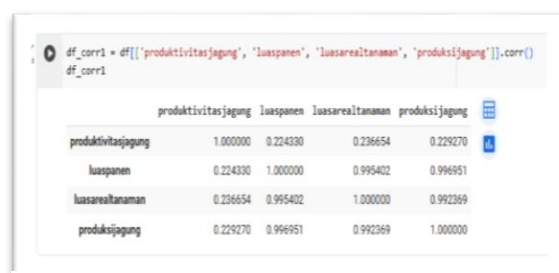


Gambar 10. Hubungan variabel x dan y

grafik hubungan pada gambar 10 :

Kemudian setelah diketahui hubungan variabel x dan y, peneliti melakukan *korelasi matrix* untuk mengetahui variabel (x) mana yang akan digunakan untuk memprediksi hasil produksi jagung di Jawa barat dengan indikator variabel (y) Produksi Jagung gambar 11 :

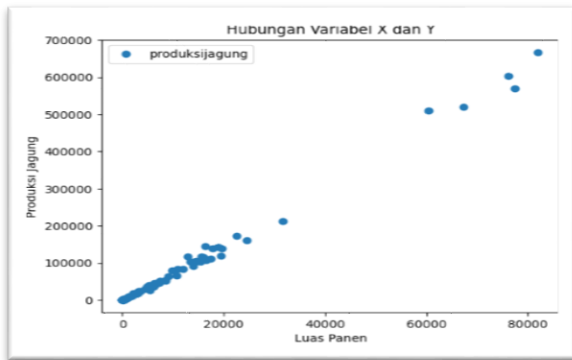
Dari hasil *korelasi matrix* tersebut bahwa variabel x yang akan digunakan untuk memprediksi



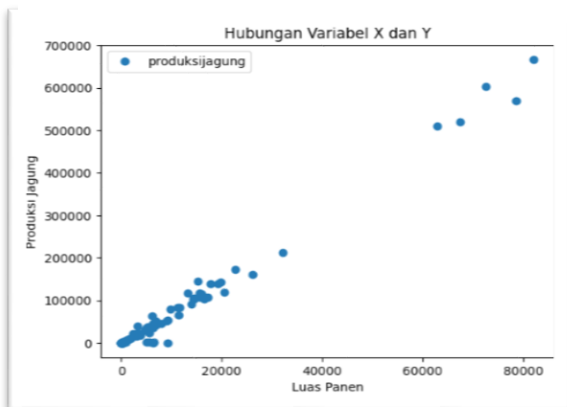
Gambar 11. Correlation matrix

hasil produksi jagung di Jawa Barat menggunakan variabel luas panen dan luas areal tanaman, karena memiliki nilai koefisien mendekati 1 yang artinya variabel luas panen dan luas areal panen berhubungan kuat secara searah dengan variabel y produksi jagung.

Selanjutnya peneliti ingin mengetahui masing-masing hubungan variabel x (luas panen dan luas areal tanaman) dan variabel y produksi jagung gambar 13 dan 14 :



Gambar 13. Pengaruh hubungan luas panen dan produksi jagung



Gambar 14. Pengaruh luas areal tanaman (x) dan produksi jagung (y)

4.5. Evaluation / Interpretation

Pada tahapan *evaluation* ini untuk menyimpulkan hasil dari data mining yang sudah peneliti lakukan. Kemudian peneliti mengambil input variabel x yaitu luas panen dan luas areal tanaman dan variabel y yaitu produksi jagung, perintah yang diinput ke program gambar 15 :

```
x = df.iloc[:, 3:5].values
y = df.iloc[:, 5].values
```

Gambar 15. Mengambil variabel input

Selanjutnya peneliti melakukan perintah untuk melakukan *training* dan *testing* seperti gambar 16 :

```
from sklearn.model_selection import train_test_split
```

Gambar 16. Perintah *split* data *training* dan *testing*

Kemudian peneliti memeriksa jumlah data *training* dan *testing* dapat dilihat gambar 17 :

```
print(X_train.shape)
print(y_train.shape)
print(X_test.shape)
print(y_test.shape)

(100, 2)
(100,)
(25, 2)
(25,)
```

Gambar 17. Memeriksa data *training* dan *testing*

Selanjutnya peneliti melakukan pembuatan objek *linear regresi* dengan nama *regressor* gambar 18 :

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

LinearRegression()
```

Gambar 18. Membuat objek *linear regresi*

Kemudian peneliti mencari nilai *coeficient* dan *intercept* gambar 19 :

```
print('Coefficien: ',regressor.coef_)
print('Intercept: ',regressor.intercept_)

Coefficien: [7.78702592 0.01826113]
Intercept: -3134.7008949365845
```

Gambar 19. Nilai *coefisient* dan *intercept*

Peneliti mencari tahu accuracy dan hasil accuracy scorenya menunjukkan hasil 97.21% gambar 20 :

```
regressor.score(X_test, y_test)

0.9907182026972935
```

Gambar 20. Hasil *accuracy score* 2 variabel x dan 1 variabel y

Peneliti mencari tahu data prediksi hasil produksi jagung di Jawa Barat kedepannya, data prediksi gambar 21 :

```
dataframe = pd.DataFrame({'Data Sebenarnya':y_test, 'Data Prediksi':y_pred})
dataframe
```

	Data Sebenarnya	Data Prediksi
0	2338	-744.633885
1	212097	244256.147009
2	0	-3134.700895
3	45695	52531.093363
4	2192	-541.556004
5	927	-2179.041044
6	91143	105717.467081
7	9497	10104.873794

Gambar 21. Data prediksi

Selanjutnya peneliti ingin mengetahui hasil perhitungan data prediksi, data prediksi untuk hasil produksi jagung di Jawa Barat adalah sebanyak 2.052.496 Ton gambar 22 :

```
jumlah = dataframe['Data Prediksi'].sum()
print('Prediksi Produksi Jagung di Jawa Barat, yaitu :', jumlah)
```

Prediksi Produksi Jagung di Jawa Barat, yaitu : 2052496.9458550413

Gambar 22. Jumlah perhitungan data prediksi 2 variabel x dan 1 variabel y

Kemudian peneliti ingin membandingkan dengan hasil 2 variabel diatas dengan 1 variabel x (luas panen) dan variabel y (produksi jagung) saja, berikut adalah hasil *accuracy* dan hasil prediksi menggunakan 1 variabel saja, gambar 23 :

```
regressor.score(X_test, y_test)

0.9910464945229137
```

Gambar 23. Hasil *accuracy score* 1 variabel x dan 1 variabel y

Selanjutnya peneliti ingin mengetahui hasil perhitungan data prediksi dengan variabel x (luas panen) dan variabel y (produksi jagung), dan prediksi untuk hasil produksi jagung di Jawa Barat adalah sebanyak 2.056.890 Ton pada gambar 24 :

```
jumlah = dataframe['Data Prediksi'].sum()
print('Prediksi Produksi Jagung di Jawa Barat, yaitu :', jumlah)

Prediksi Produksi Jagung di Jawa Barat, yaitu : 2056890.6707767486
```

Gambar 24. Jumlah perhitungan data prediksi 1 variabel x dan 1 variabel y

Berikut adalah tabel 3 perbandingan 2 variabel x dan 1 variabel x dengan 1 variabel y :

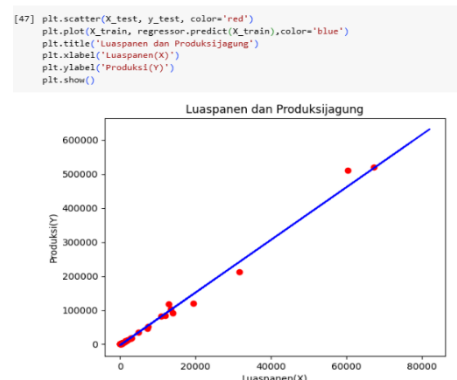
Tabel 3. Perbandingan variabel x dan y

Dua variabel x dan satu variabel y		Satu variabel x dan satu variabel y	
Akurasi	Prediksi	Akurasi	Prediksi
99,07%	2.052.496 Ton	99,10%	2.056.890 Ton

Dari hasil tabel perbandingan tersebut peneliti melakukan kesimpulan bahwa hasil dari dua perbandingan variabel tersebut memiliki hasil yang berbeda walaupun sangat tipis perbedaannya. Selanjutnya peneliti ingin mengetahui hasil *MAE*, *MSE*, *RMSE*, dan *performance* menggunakan hasil prediksi dan akurasi paling tinggi yaitu dengan 1 variabel x (luas panen) dan 1 variabel y (produksi jagung).

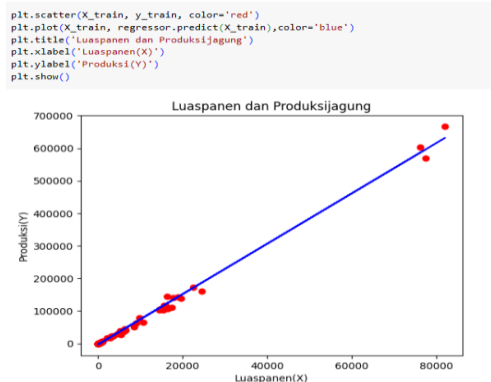
Maka hasil persamaan liner regresi pada 1 variabel x dan 1 variabel y yaitu $Y = -2753.018316202004 * x + [7.72294428]$. Nilai tersebut didapat melalui hasil coefisient dan intercept yang sudah peneliti lakukan.

Selanjutnya peneliti ingin melihasi hasil *visualisasi* dari data *testing* luas panen terhadap produksi jagung gambar 25 :



Gambar 25. Visualisasi data *testing*

Kemudian peneliti ingin melihat hasil dari data *training* luas panen terhadap produksi jagung gambar 26 :



Gambar 26. Visualisasi data training

Setelah diketahui hasil prediksi produksi jagung di Jawa Barat dari perbandingan tersebut, selanjutnya peneliti ingin mengetahui hasil nilai preformance, MAE, MSE, dan RMSE dari hasil *accuracy* dan prediksi paling tertinggi menggunakan 1 variabel x dan 1 variabel y, dan sistem menampilkan nilai MAE: 5741.21, MSE: 124802556, RMSE: 124802556.11, dan *performance*: 0.991046 (99.10%) pada gambar 27 :

```
print('Mean Absolute Error: ', metrics.mean_absolute_error(y_test, y_pred))
print('Mean Squared Error: ', metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error: ', metrics.mean_squared_error(y_test, y_pred))
print('Model R^2 Square value', metrics.r2_score(y_test, y_pred))
```

Mean Absolute Error: 5741.216582851467
Mean Squared Error: 124802556.11750661
Root Mean Squared Error: 124802556.11750661
Model R^2 Square value 0.9910464945229137

Gambar 27. Nilai MAE, MSE, RMSE, performance

5. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis data *mining* yang telah diuraikan pada bagian hasil tugas akhir dapat ditarik kesimpulan bahwa hasil prediksi produksi jagung di Jawa Barat menggunakan regresi linear yaitu sebanyak 2.056.890 Ton, ini menunjukkan bahwa hasil prediksi produksi jagung di Jawa Barat ke depannya akan mengalami kenaikan yang sangat signifikan dari 5 tahun sebelumnya. Hasil prediksi produksi jagung ini bisa jadi lebih atau kurang dari prediksi peneliti. Dilihat dari hasil analisis bahwa hasil MAE: 5741.21, MSE: 124802556, RMSE: 124802556.11, *performance*: 0.991046 (99.10%) dan hasil persamaan regresi linear yaitu $Y = -2753.018316202004 * x + [7.72294428]$. Menunjukkan hasil yang baik. Itu artinya algoritma regresi linear yang peneliti uji sangat cocok dengan data yang dipakai. Adapun saran untuk penelitian selanjutnya adalah silahkan gunakan alat

lain seperti *rapidminer* untuk penelitian selanjutnya, silahkan menggunakan algoritma lain seperti ANN (*Artificial Neural Network*), untuk hasil yang lebih akurat, dan disarankan juga untuk menambahkan jumlah baris data.

DAFTAR PUSTAKA

- [1] M. Adha and E. Utami, "PREDIKSI PRODUKSI JAGUNG MENGGUNAKAN ALGORITMA APRIORI DAN REGRESI LINEAR BERGANDA (STUDI KASUS: DINAS PERTANIAN KABUPATEN DOMPU)." *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, Dec. 2022, doi: 10.47065/bits.v4i3.2704.
- [2] S. Fajri, H. Gunawan, L. R. Batubara, and Z. Sitorus, "Prediksi Hasil Produksi Tanaman Tomat di Indonesia Menurut Provinsi Menggunakan Algoritma Fletcher-Reeves," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, Dec. 2022, doi: 10.47065/bits.v4i3.2704.
- [3] A. Satria, R. Maulida Badri, I. Safitri, and H. Artikel, "Prediksi Hasil Panen Tanaman Pangan Sumatera dengan Metode Machine Learning," *Digital Transformation Technology (Digitech) / e*, vol. 3, no. 2, 2023, doi: 10.47709/digitech.v3i2.2852.
- [4] N. Nurhayati, Mhd. B. Sibuea, D. Kusbiantoro, M. Silaban, and A. Wanto, "Implementasi Algoritma Resilient untuk Prediksi Potensi Produksi Bawang Merah di Indonesia," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 2, Sep. 2022, doi: 10.47065/bits.v4i2.2269.
- [5] N. Moha Lalapa and W. Yunus, "IMPLEMENTASI METODE REGRESI LINEAR SEDERHANA UNTUK PREDIKSI HARGA CABAI RAWIT," *Jurnal Ilmiah Ilmu Komputer Banthayo Lo Komputer*, vol. 2, no. 2, p. 96, 2023.
- [6] S. Oleh, R. Jannah, J. Peternakan, and F. Pertanian, "PENGARUH VARIETAS DAN JARAK TANAM BERBEDA TERHADAP PRODUKSI SEGAR DAN PRODUKSI BAHAN KERING PADA HIJAUAN JAGUNG (*Zea mays L.*)," 2019.
- [7] A. D. Brisa Ayu *et al.*, "Pengembangan Agrowisata Jagung di Desa Giripanggung Kabupaten Gunungkidul," 2023.
- [8] F. O. Lusiana, I. Fatma, and A. P. Windarto, "Estimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Simalungun," 2021. [Online]. Available: <https://hostjournals.com/>
- [9] A. Fitri Boy, "Implementasi Data Mining Dalam Memprediksi Harga Crude Palm Oil (CPO) Pasar Domestik Menggunakan Algoritma Regresi Linier Berganda (Studi Kasus Dinas Perkebunan Provinsi Sumatera Utara)," 2020. [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>

- [10] P. Sari Ramadhan and N. Safitri STMIK Triguna Dharma, “Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang,” vol. 18, no. SAINTIKOM, pp. 55–61, 2019, [Online]. Available: <https://sirusa.bps.go.id/index.php>
- [11] B. Mega Pratiwi and N. Q. Nada, “Penerapan Model Machine Learning dalam Menentukan Rekomendasi Objek Wisata Provinsi Jawa Tengah,” 2022.