

# Penerapan Regresi Linear Terhadap Tingkat Perceraian Di Jawa Tengah Menggunakan Bahasa Python

Syarifatul Muhajannah<sup>1,\*</sup>, Angelo Risang Kadisi<sup>2</sup>, Nur Wakhidah<sup>3</sup>

<sup>1,2,3</sup>Fakultas Teknologi Informasi dan Komunikasi, Program Studi Sistem Informasi, Universitas Semarang, Kota Semarang, Indonesia

Email: <sup>1,\*</sup> arifatul02jannah@gmail.com, <sup>2</sup> risangangelo@gmail.com, <sup>3</sup>ida@usm.ac.id

\*Email Penulis Utama

**Abstrak**– Penelitian ini dilakukan untuk mengidentifikasi pola trend hubungan antara jumlah pernikahan dengan tingkat perceraian yang terjadi di 35 kabupaten/kota di Provinsi Jawa Tengah. Metode yang digunakan adalah regresi linier sederhana, dengan data utama yang bersumber dari Badan Pusat Statistik (BPS) Jawa Tengah pada tahun 2023. Dalam proses analisis, penelitian ini memanfaatkan bahasa pemrograman Python karena berbagai keunggulan yang dimilikinya. Python tidak hanya menawarkan sintaks yang sederhana dan mudah dipahami, tetapi juga memiliki berbagai pustaka yang sangat mendukung analisis data, seperti Pandas untuk pengolahan data, Scikit-learn untuk membangun model prediktif, dan Matplotlib untuk visualisasi data. Kemampuannya dalam mengelola data dalam jumlah besar secara efisien membuat Python sangat cocok untuk analisis yang memerlukan kecepatan dan akurasi tinggi. Model regresi linier sederhana yang diterapkan dalam penelitian ini menunjukkan hasil yang cukup signifikan. Nilai R-squared mencapai 87,74%, yang berarti bahwa sebanyak 87,74% variasi tingkat perceraian dapat dijelaskan oleh jumlah pernikahan di kabupaten/kota yang dianalisis. Di sisi lain, nilai Mean Squared Error (MSE) sebesar 79.745,23 menunjukkan tingkat kesalahan prediksi yang wajar, karena terdapat data outlier di beberapa daerah seperti Kabupaten Cilacap, Banyumas, dan Brebes berkontribusi terhadap tingkat kesalahan prediksi. Model ini juga mengungkapkan bahwa setiap peningkatan satu pernikahan rata-rata akan berhubungan dengan kenaikan sebesar 0,3486 pada tingkat perceraian. Beberapa hasil prediksi tingkat perceraian di kabupaten/kota tertentu berdasarkan jumlah pernikahan menghasilkan angka seperti 2.281,39; 3.501,47; 1.518,09; 1.455,32; 1.264,24; 2.240,59; dan 3.046,43 kasus. Hasil prediksi ini menggambarkan estimasi tingkat perceraian yang dihasilkan oleh model di wilayah-wilayah tersebut. Analisis ini memperlihatkan adanya korelasi positif antara jumlah pernikahan dan tingkat perceraian, di mana wilayah dengan jumlah pernikahan yang lebih tinggi cenderung memiliki tingkat perceraian yang juga lebih tinggi. Penemuan ini memberikan bukti empiris yang memperkuat adanya hubungan linier yang signifikan antara kedua variabel tersebut. Dengan demikian, penelitian ini diharapkan dapat menjadi pijakan ilmiah dalam perumusan kebijakan yang lebih efektif. Melalui pendekatan berbasis data ini, intervensi yang dirancang untuk mengurangi angka perceraian di masa depan dapat dilakukan secara lebih terarah dan strategis.

**Kata Kunci:** Tingkat Perceraian, Jumlah Pernikahan, Jawa Tengah, Regresi Linear, Python.

**Abstract**– This study was conducted to identify the trend pattern of the relationship between the number of marriages and the divorce rate in 35 districts/cities in Central Java Province. The method used is simple linear regression, with the main data sourced from the Central Java Statistics Agency (BPS) in 2023. In the analysis process, this research utilizes the Python programming language because of its various advantages. Python not only offers a simple and easy-to-understand syntax, but also has various libraries that greatly support data analysis, such as Pandas for data processing, Scikit-learn for building predictive models, and Matplotlib for data visualization. Its ability to efficiently manage large amounts of data makes Python well suited for analyses that require high speed and accuracy. The simple linear regression model applied in this study showed significant results. The R-squared value reaches 87.74%, which means that 87.74% of the variation in divorce rates can be explained by the number of marriages in the districts/cities analyzed. On the other hand, the Mean Squared Error (MSE) value of 79,745.23 indicates a reasonable level of prediction error, as there are data outliers in several regions such as Cilacap, Banyumas, and Brebes districts that contribute to the level of prediction error. The model also reveals that every increase of one marriage on average will correspond to an increase of 0.3486 in the divorce rate. Some of the predicted divorce rates in certain districts based on the number of marriages resulted in figures such as 2,281.39; 3,501.47; 1,518.09; 1,455.32; 1,264.24; 2,240.59; and 3,046.43 cases. These predicted results illustrate the estimated divorce rates generated by the model in these regions. The analysis shows a positive correlation between the number of marriages and the divorce rate, where regions with a higher number of marriages tend to have a higher divorce rate. This finding provides empirical evidence that reinforces the existence of a significant linear relationship between the two variables. Thus, this research is expected to provide a scientific basis for the formulation of more effective policies. Through this data-driven approach, interventions designed to reduce divorce rates in the future can be more targeted and strategic.

**Keywords:** Divorce Rate, Number of Marriages, Central Java, Linear Regression, Python.

## 1. PENDAHULUAN

Pernikahan adalah ikatan lahir batin antara seorang pria dan seorang wanita sebagai suami istri dengan tujuan membentuk keluarga (rumah tangga) yang bahagia dan kekal berdasarkan pada nilai Ketuhanan Yang Maha Esa, menurut UU Perkawinan No. 1 Tahun 1974. Pernikahan jika dilakukan pada usia yang “tepat” akan membawa kebahagiaan bagi keluarga dan pasangan (suami dan istri) yang menjalankan perkawinan tersebut [1]. Menikah dan hidup bahagia adalah impian yang diinginkan hampir semua orang. Selama menjalani rumah tangga perselisihan

pasti terjadi, namun dari perselisihan tersebut diharapkan tidak sampai berakhir pada perceraian [2]. Retaknya hubungan ikatan suami istri, yang disebabkan karena tidak ada lagi kesamaan visi dan kesamaan prinsip disebut dengan perceraian. Dasar dari perceraian yaitu terputusnya ikatan suci dari pasangan suami istri [3]. Perceraian dapat disebabkan banyak faktor, salah satu diantaranya ada karena masalah finansial, perselingkuhan, atau bahkan perbedaan visi sering kali menjadi penyebab utama perceraian. Perceraian menimbulkan banyak rasa sakit dan kesedihan juga menjadi momok bagi kehidupan anak-anak yang terlibat dalam situasi tersebut [4]. Di Indonesia, angka perceraian terus meningkat setiap tahunnya, termasuk di Jawa Tengah, yang menunjukkan tren kenaikan yang signifikan. Oleh karena itu, penting untuk menganalisis faktor-faktor yang berkontribusi terhadap perceraian, salah satunya adalah jumlah pernikahan yang terjadi di suatu daerah.

Berdasarkan penelitian yang dilakukan oleh Purwaingsih *dkk* (2023) tentang implementasi metode KMeans Clustering dengan Davies Bouldin Index pada analisis faktor penyebab perceraian menyebutkan bahwa pada laporan Statistik Indonesia menunjukkan bahwa terdapat 516.344 perceraian yang terjadi di Indonesia pada tahun 2022[4][5]. Jumlah kasus dari perceraian di negara ini mencapai titik tertinggi dalam enam tahun terakhir. Mayoritas perceraian di dalam negeri pada tahun 2022 adalah cerai gugat yang diajukan oleh pihak istri dan sudah diputuskan oleh Pengadilan. Pada tahun tersebut, total kasus dari perceraian di negara ini mencapai 388.358, dengan kasus tertinggi di daerah Jawa Barat dengan jumlah kasus perceraian sebanyak 113.643 kasus, berdasarkan provinsinya. Jawa Tengah dengan 85.412 kasus dan Jawa Timur dengan 102.065 kasus [5]. Dengan tingginya trend perceraian yang terjadi pada tahun tersebut menimbulkan pertanyaan mengenai faktor-faktor apa saja yang berkontribusi terhadap fenomena ini[6].

Masalah perceraian di Indonesia, salah satunya di provinsi Jawa Tengah, merupakan isu sosial yang memiliki signifikansi tinggi dan memerlukan perhatian serius. Berdasarkan data terkini dari Badan Pusat Statistik (BPS) tahun 2023, tren perceraian di Jawa Tengah tercatat memiliki jumlah kasus sebanyak 23,18 ribu kasus. Angka tersebut mengalami sedikit penurunan dibandingkan dengan tahun 2022 tetapi masih lebih tinggi dibandingkan kasus 2021[7]. Perceraian dapat memicu munculnya berbagai masalah sosial, ekonomi, dan psikologis yang serius, sehingga penting bagi peneliti dan pengambil kebijakan untuk menggali faktor-faktor yang mempengaruhi fenomena ini guna merancang langkah-langkah yang efektif dalam menguranginya serta memperkuat hubungan pernikahan. Penelitian ini bertujuan untuk menganalisis hubungan antara jumlah pernikahan dengan tingkat perceraian, melalui data jumlah pernikahan dan perceraian yang diperoleh dari Badan Pusat Statistika (BPS) yang melibatkan 35 kabupaten dan kota di Jawa Tengah[8], akan dianalisis apa pengaruh jumlah pernikahan terhadap tingkat perceraian di Jawa Tengah pada tahun 2023.

Dengan adanya sistem informasi berbasis data besar (big data) dan analisis statistik yang canggih, peneliti kini dapat mengakses dan memproses informasi dari berbagai sumber, termasuk data demografi, sosial, dan ekonomi, secara lebih efisien[9]. Teknologi ini memungkinkan penggunaan model prediktif yang lebih akurat untuk memahami faktor-faktor yang berkontribusi terhadap perceraian dan pernikahan, serta merumuskan kebijakan yang berbasis data. Di era teknologi yang semakin berkembang ini, Python menjadi salah satu alat yang paling efektif untuk analisis data[9]. Python dipilih dalam penelitian ini karena keunggulannya dalam analisis data dan pemodelan statistik. Sebagai bahasa pemrograman yang fleksibel dengan ekosistem pustaka yang luas, Python memungkinkan pengolahan data secara efisien, dengan pustaka-pustaka seperti Pandas, Scikit-learn, dan Matplotlib memungkinkan untuk pengolahan data dan visualisasi hasil yang lebih mudah dipahami[10][11].

Dalam penelitian ini, pustaka Pandas digunakan untuk mengelola data pernikahan dan perceraian dari Badan Pusat Statistik (BPS), termasuk pembersihan, transformasi, dan manipulasi dataset agar siap untuk analisis lebih lanjut [12]. McKinney (2022) menekankan bahwa Python sangat berguna dalam mengelola data kompleks dan pembuatan model analisis yang mendetail[12]. Dalam menganalisis data menggunakan Python, Scikit-learn digunakan untuk membangun model regresi linier yang menjadi inti penelitian ini. Pustaka ini menyediakan alat intuitif untuk implementasi model prediktif, validasi, dan evaluasi kinerja model, seperti penghitungan nilai  $R^2$  dan Mean Squared Error (MSE). Dengan kemampuannya dalam menangani regresi linier, Scikit-learn memastikan bahwa hubungan antara jumlah pernikahan dan tingkat perceraian dapat dianalisis secara kuantitatif dengan hasil yang dapat diinterpretasikan secara jelas. Untuk mendukung penyajian temuan secara informatif, pustaka Matplotlib dan Seaborn digunakan dalam visualisasi data. Scatter plot, misalnya, memungkinkan pemetaan hubungan antara jumlah pernikahan dan tingkat perceraian secara visual, sehingga pola tren lebih mudah diidentifikasi [10]. Keunggulan Python dalam menangani data kompleks, serta komunitas pengguna yang aktif dan dokumentasi yang luas, menjadikannya pilihan ideal dalam penelitian ini. Dengan memanfaatkan Python dan pustaka pendukungnya, penelitian ini dapat menghasilkan analisis yang akurat dan interpretasi hasil yang lebih baik.

Dalam menganalisis sebuah data, model Regresi merupakan alat statistik yang efektif untuk menganalisis hubungan antara variabel-variabel, sehingga dapat digunakan untuk meramalkan tingkat perceraian berdasarkan karakteristik pernikahan. Metode ini dipilih karena kemampuannya dalam mengidentifikasi dan memodelkan

hubungan antara variabel independen dan variabel dependen secara sederhana namun efektif. Dalam penelitian ini, regresi linier digunakan untuk memahami sejauh mana jumlah pernikahan dapat mempengaruhi tingkat perceraian di Jawa Tengah[13]. Salah satu alasan utama penggunaan regresi linier adalah kemampuannya untuk memberikan interpretasi yang mudah dipahami melalui persamaan matematika sederhana. Regresi linier juga memungkinkan evaluasi kekuatan dan arah hubungan antar variabel menggunakan nilai koefisien dan R-squared ( $R^2$ ) yang tinggi[14]. Pada penelitian sebelumnya yang membahas tentang prediksi jumlah perceraian menggunakan Regresi Linier pernah dilakukan oleh Samosir *dkk.*(2021), penelitian tersebut berhasil memprediksi jumlah perceraian di Pengadilan Agama Simalungun. Metode yang digunakan dalam penelitian tersebut menggunakan metode Regresi Linear Berganda. Berdasarkan penelitian yang telah dilakukan dapat diketahui bahwa metode regresi linear berganda dapat menyelesaikan permasalahan dalam mengestimasi jumlah perceraian. Dengan menggunakan variabel bebas: cerai talak (X1), cerai gugat (X2), dan variabel tidak bebas yaitu jumlah perceraian (Y). Hasil dari perhitungan regresi linear berganda pada periode tahun 2019 adalah terdapat 961 kasus perceraian[13].

Penelitian sebelumnya yang membahas tentang jumlah perceraian juga pernah dilakukan menggunakan metode lain yaitu dengan mengimplementasikan metode K-Means Clustering dengan Davies Boulin Index, yang dilakukan oleh Purwaningsih *dkk.*(2023). Penelitian tersebut dilakukan untuk mengetahui faktor penyebab perceraian di wilayah Jawa Barat, serta menerapkan bagaimana metode K-Means dengan Davies Bouldin Index berfungsi untuk mengkluster faktor penyebab dari perceraian di sekitar wilayah Jawa Barat. Cluster yang dihasilkan dari penelitian metode K-Means dengan nilai  $k=7$  menggunakan Davies Bouldin Indeks yaitu Cluster\_0 terdiri dari "Provinsi Jawa Barat", Cluster\_1 terdiri dari "Kota Tasikmalaya", Custer\_2 terdiri dari "Cirebon" dan "Indramayu", Cluster\_3 terdiri dari "Tasikmalaya", "Kuningan" dan "Subang", Cluster\_4 terdiri dari "Bogor", "Cianjur", "Sumedang"[4].

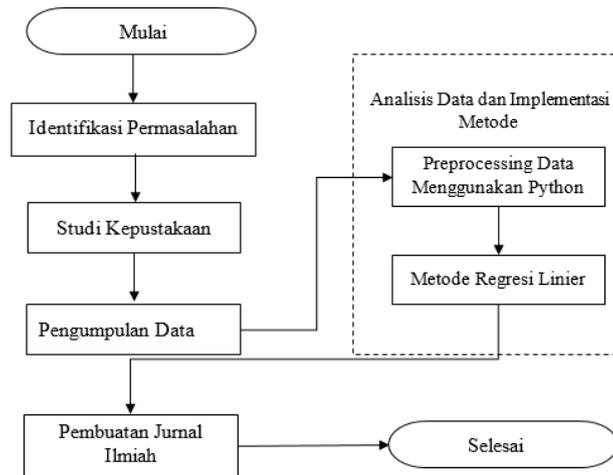
Berdasarkan penelitian terdahulu yang sudah dilakukan, pada penelitian kali ini peneliti mengimplementasikan metode Regresi Linear untuk menganalisis hubungan antara jumlah pernikahan dan jumlah perceraian di Jawa Tengah pada tahun 2023. Meskipun penelitian sebelumnya oleh Samosir *dkk.* (2021) telah menggunakan regresi linier berganda untuk memprediksi jumlah perceraian dengan melibatkan variabel tambahan seperti cerai talak dan cerai gugat, penelitian ini sengaja mengadopsi pendekatan yang lebih sederhana. Ada beberapa alasan mendasar untuk pendekatan ini. Pertama, penelitian ini fokus pada hubungan langsung antara dua variabel utama untuk menyajikan hasil yang lebih spesifik dan mudah diinterpretasikan. Kedua, data tambahan yang mungkin relevan untuk regresi linier berganda tidak tersedia secara memadai dalam konteks Jawa Tengah tahun 2023. Ketiga, penggunaan model sederhana memungkinkan validasi temuan sebelumnya sekaligus memberikan landasan awal bagi penelitian yang lebih kompleks di masa depan. Dengan demikian, regresi linier sederhana dipilih karena kesesuaiannya dengan tujuan dan ruang lingkup penelitian. Untuk mempermudah pemodelan Regresi peneliti menggunakan python untuk mengolah data dan memproses pemodelan tersebut. Adapun variabel yang digunakan pada penelitian ini yaitu menggunakan variabel dependen (pernikahan) dan variabel independen (perceraian). Dengan menggunakan metode regresi linear sederhana, penelitian ini akan mengidentifikasi pola dan tren yang ada serta memberikan prediksi mengenai tingkat perceraian berdasarkan data pernikahan.

Hasil dari penelitian ini dapat menjadi pijakan ilmiah dalam perumusan kebijakan yang lebih efektif untuk mengatasi banyaknya kasus perceraian di Jawa Tengah. Melalui pendekatan berbasis data ini, Pemerintah daerah dan lembaga terkait, seperti Kantor Urusan Agama (KUA), dapat merancang program edukasi pernikahan untuk meningkatkan kesiapan pasangan sebelum menikah seperti melakukan penyuluhan terhadap calon pengantin, melakukan seminar bimbingan pernikahan, konseling pernikahan dan beberapa kebijakan lainnya untuk menurunkan angka perceraian.

## 2. METODE PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini mencakup serangkaian langkah mulai dari identifikasi permasalahan hingga penyusunan laporan hasil. Tahapan-tahapan ini dijelaskan secara rinci dalam Gambar 1 di bawah ini.



**Gambar 1.** Tahapan Penelitian

Dalam sebuah penelitian, diperlukan langkah-langkah yang sistematis untuk memastikan bahwa tujuan yang ingin dicapai pada penelitian dapat terpenuhi[15]. Tahapan-tahapan penelitian seperti yang ditunjukkan pada Gambar 1 di atas dapat diuraikan sebagai berikut:

a. Identifikasi Permasalahan

Langkah pertama dalam penelitian ini adalah mengidentifikasi permasalahan perceraian berdasarkan pernikahan di Jawa Tengah, peneliti mengidentifikasi apa saja faktor yang mempengaruhi tingginya tingkat perceraian di Jawa Tengah mulai dari faktor internal dan eksternal. Dari hasil identifikasi masalah, ditemukan hipotesis apakah adanya hubungan antara tingginya jumlah perceraian dengan jumlah pernikahan yang ada pada tahun 2023, permasalahan yang didapat pada penelitian ini akan dianalisis menggunakan metode Regresi Linier.

b. Studi Kepustakaan

Setelah mengidentifikasi permasalahan, langkah berikutnya adalah mencari sumber-sumber referensi yang relevan dengan permasalahan tersebut dan mengkaji penelitian terdahulu untuk menemukan celah (gap analysis) yang dapat diisi oleh penelitian ini. Melalui kajian pustaka, peneliti dapat memahami konsep-konsep dasar, menemukan celah dalam penelitian sebelumnya, dan merumuskan pertanyaan penelitian yang lebih tajam. Selain itu, identifikasi gap dalam penelitian sebelumnya menjadi dasar penting dalam pemilihan metode yang tepat. Dalam penelitian ini walaupun sudah pernah dilakukan penelitian menggunakan Regresi Linier Berganda peneliti memilih menggunakan metode Linier Sederhana, hal ini didasarkan pada kesenjangan data tambahan yang relevan untuk metode Regresi Linier Berganda. Dengan demikian, kajian pustaka tidak hanya membantu memperkuat validitas akademik, tetapi juga memastikan bahwa pendekatan yang diambil relevan dan sesuai dengan kebutuhan data.

c. Pengumpulan Data

Tahapan selanjutnya adalah melakukan pengumpulan data terkait seperti data jumlah pernikahan dan perceraian pada tiap kabupaten dan kota di Jawa Tengah yang akan dianalisis menggunakan Regresi Linier pada penelitian ini. Teknik pengumpulan data yang dilakukan pada penelitian ini adalah menggunakan analisis data skunder yang di dapat dari website resmi Badan Pusat Statistik Jawa Tengah (<https://jateng.bps.go.id>)[8].

d. Uji Asumsi

Pengujian asumsi dilakukan setelah pengumpulan data. Uji asumsi dilakukan untuk memastikan bahwa data yang digunakan memenuhi asumsi regresi linier. Beberapa uji yang dilakukan meliputi [16][17]:

1. Uji Normalitas

Untuk memastikan bahwa data terdistribusi normal, digunakan uji statistik Shapiro-Wilk. Jika data tidak berdistribusi normal, transformasi data dapat dilakukan.

2. Uji Outlier

Uji ini dilakukan untuk mendeteksi keberadaan data pencilan (outlier) yang dapat mempengaruhi hasil regresi linier. Outlier adalah data yang memiliki nilai ekstrem dan dapat menyebabkan bias dalam estimasi parameter regresi.

Hasil dari kedua uji ini akan digunakan untuk memastikan validitas model sebelum dilakukan analisis regresi linier dan interpretasi hasilnya.

e. Analisis Data

1. Preprocessing Data Dengan Python

Langkah berikutnya adalah mengolah data yang telah dikumpulkan dan menganalisis menggunakan bahasa pemrograman Python. Dalam analisis data, tipe data yang akan diolah adalah data kuantitatif yang mencakup jumlah pernikahan dan tingkat perceraian di setiap kabupaten/kota di Jawa Tengah pada tahun 2023. Data ini berbentuk angka yang menunjukkan total jumlah kejadian pada masing-masing kategori. Pengolahan data bertujuan untuk mengidentifikasi pola hubungan linier antara variabel-variabel tersebut, sehingga model regresi linier dapat digunakan untuk memprediksi tingkat perceraian berdasarkan jumlah pernikahan. Selain itu, analisis ini juga bertujuan untuk mengevaluasi kualitas prediksi model melalui metrik seperti R-squared dan Mean Squared Error (MSE). Dengan demikian, hasil analisis dapat memberikan wawasan yang lebih baik untuk memahami dinamika sosial di Jawa Tengah serta memberikan dasar bagi kebijakan yang lebih efektif.

### 2. Implementasi Metode Regresi

Setelah data diolah menggunakan Python kemudian data tersebut diimplementasikan dengan menggunakan metode Regresi untuk mengetahui nilai prediksi perceraian yang didapat melalui metode penelitian tersebut. Hasil prediksi yang diperoleh dapat digunakan sebagai pijakan ilmiah bagi badan yang berwenang dalam merumuskan kebijakan yang lebih baik sebagai upaya mengurangi tingginya tingkat perceraian di Jawa Tengah.

### 3. Uji Hipotesis

Uji hipotesis dilakukan setelah model regresi linier dibuat dan asumsi-asumsi terpenuhi. Tujuannya adalah untuk menguji signifikansi statistik dari koefisien regresi atau model secara keseluruhan.

Contoh uji hipotesis dalam regresi linier [17][18]:

#### 1. Uji Signifikansi Koefisien Regresi (Uji t)

Pengujian ini dilakukan untuk menilai apakah jumlah pernikahan berpengaruh signifikan terhadap tingkat perceraian.

#### 2. Uji Kesesuaian Model (Uji F)

Pengujian ini dilakukan untuk mengevaluasi apakah model secara keseluruhan dapat menjelaskan hubungan antara variabel independen dan variabel dependen dengan baik. f. Pembuatan Jurnal Ilmiah

Pada tahap ini, hasil penelitian yang telah dilakukan akan diterapkan dalam penulisan jurnal ilmiah.. Pembuatan jurnal ilmiah merupakan langkah penting dalam penelitian ini untuk mendokumentasikan hasil analisis dan menyebarkan temuan kepada komunitas akademik serta pembuat kebijakan. Sebuah jurnal ilmiah harus mencakup beberapa komponen utama, yaitu:

1. Pendahuluan, di dalamnya berisi penjelasan latar belakang penelitian, permasalahan yang diangkat, tujuan penelitian, dan kontribusi terhadap bidang studi terkait.
2. Metode Penelitian, di dalamnya berisi penguraian pendekatan yang digunakan dalam penelitian, termasuk teknik pengumpulan data, metode analisis regresi linier sederhana, serta alat pemrograman seperti Python dan pustaka pendukungnya.
3. Hasil dan Pembahasan, di dalamnya berisi tampilan hasil analisis regresi linier yang diperoleh, termasuk nilai koefisien regresi, R-squared ( $R^2$ ), dan Mean Squared Error (MSE). Data hasil regresi akan disajikan dalam bentuk tabel dan visualisasi grafik untuk memperjelas pola hubungan antara jumlah pernikahan dan tingkat perceraian.
4. Kesimpulan dan Saran, di dalamnya terangkum temuan utama dari penelitian ini dan memberikan rekomendasi untuk penelitian lanjutan maupun kebijakan berbasis data.

Hasil analisis regresi linier akan diintegrasikan ke dalam jurnal dengan menyoroti hubungan antara jumlah pernikahan dan tingkat perceraian. Interpretasi nilai koefisien regresi akan memberikan pemahaman tentang seberapa besar dampak jumlah pernikahan terhadap tingkat perceraian, sementara evaluasi model menggunakan  $R^2$  dan MSE akan menunjukkan sejauh mana model dapat memprediksi fenomena ini secara akurat. Dengan demikian, jurnal ilmiah yang dihasilkan akan memberikan kontribusi penting dalam memahami dan mengatasi fenomena perceraian di Jawa Tengah.

## 2.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari website Badan Pusat Statistik (BPS) dan lembaga pemerintah terkait. Badan Pusat Statistik (BPS) Jawa Tengah merekap jumlah data kasus perceraian dan pernikahan pada tahun 2023. Pada penelitian ini data yang akan diolah mencakup jumlah pernikahan dan perceraian dari 35 kabupaten dan kota di Jawa Tengah pada tahun 2023 [8]. Pemilihan data dari BPS didasarkan pada kredibilitas lembaga ini sebagai sumber data resmi yang diakui secara nasional. Data yang disediakan oleh BPS memiliki keakuratan tinggi, diperoleh melalui survei yang terstruktur, dan mencakup seluruh

wilayah Jawa Tengah, sehingga sangat representatif untuk analisis. Selain itu, data perceraian dan pernikahan di Jawa Tengah penting untuk dipelajari karena memberikan gambaran mengenai dinamika sosial dan keluarga di wilayah ini. Jawa Tengah memiliki jumlah penduduk yang besar dan menjadi salah satu provinsi dengan tingkat perceraian tertinggi di Indonesia, menjadikannya konteks yang relevan untuk mengeksplorasi faktor-faktor yang memengaruhi tingkat perceraian dan pernikahan. Dengan demikian, analisis data dari BPS diharapkan mampu memberikan hasil yang valid dan berguna bagi perumusan kebijakan.

### 2.3 Uji Asumsi

Setelah melakukan pengumpulan data, dilakukan pengujian hipotesis untuk menguji signifikansi model dan variabel. dilakukan uji hipotesis untuk memastikan bahwa model regresi linier sederhana yang digunakan valid dan dapat diinterpretasikan dengan benar. Uji hipotesis ini mencakup beberapa pengujian statistik berikut [16][17]:

#### 1. Uji Normalitas

Uji normalitas dilakukan untuk memeriksa apakah data residual dari model regresi berdistribusi normal. Salah satu metode yang digunakan adalah Uji Shapiro-Wilk, yang dirancang untuk menguji apakah sampel berasal dari distribusi normal.

Hipotesis yang diuji yaitu:

$H_0$ : Data residual berdistribusi normal.

$H_1$ : Data residual tidak berdistribusi normal.

Jika nilai p-value dari uji Shapiro-Wilk lebih kecil dari tingkat signifikansi (misalnya 0,05), maka  $H_0$  ditolak, yang berarti data tidak berdistribusi normal.

Pengujian ini dilakukan menggunakan Python dengan pustaka `scipy.stats.shapiro()` untuk memastikan bahwa asumsi normalitas terpenuhi sebelum melakukan regresi linier.

#### 2. Uji Outlier (Interquartile Range - IQR Method)

Uji ini digunakan untuk mendeteksi adanya nilai pencilan (outlier) dalam dataset yang dapat memengaruhi hasil regresi. Metode IQR bekerja dengan cara [19]:

$$Upper\ Bound = Q3 + 1,5 * IQR \quad (1)$$

Data yang berada di luar rentang tersebut dianggap sebagai outlier.

Uji outlier ini dilakukan menggunakan Python dengan pustaka `pandas` dan `numpy`. Jika ditemukan outlier, maka dapat dilakukan analisis lebih lanjut untuk memutuskan apakah outlier akan dihapus atau tetap dipertahankan dalam model.

Hasil dari uji-uji ini akan menentukan apakah model regresi linier sederhana yang digunakan dapat diandalkan dalam penelitian ini.

### 2.4 Python

Python adalah bahasa pemrograman dengan sintaks sederhana dan mudah dipahami, sehingga cocok bagi pemula maupun profesional. Sebagai bahasa open source, Python dapat digunakan secara gratis dan dimodifikasi sesuai kebutuhan. Selain itu, instalasi yang mudah di berbagai platform seperti Windows, macOS, dan Linux menjadikannya fleksibel dan dapat diakses dalam berbagai lingkungan kerja. Python pertama kali dikembangkan oleh Guido van Rossum pada tahun 1990 di CWI Amsterdam (Enterprise, 2017) [20]. Salah satu keunggulan utama Python adalah kemampuannya dalam analisis data, didukung oleh pustaka seperti `Pandas` untuk manipulasi data, `NumPy` untuk komputasi numerik, serta `Matplotlib` dan `Seaborn` untuk visualisasi [21]. Dalam pembelajaran mesin dan analisis prediktif, Python memiliki pustaka seperti `Scikit-learn` dan `TensorFlow`. Ekosistem yang luas serta komunitas pengguna yang aktif menyediakan dokumentasi lengkap dan dukungan berkelanjutan [12]. Dengan alat seperti `Jupyter Notebook`, Python mempermudah eksplorasi data, pengolahan data besar, serta pengembangan model statistik dan prediktif [11][12].

Dalam penelitian ini, Python dipilih karena keunggulannya dalam pengolahan data dan analisis statistik [10]. Pustaka `Pandas` digunakan untuk mengelola data pernikahan dan perceraian dari Badan Pusat Statistik (BPS), mencakup proses pembersihan, transformasi, dan manipulasi dataset. `Scikit-learn` mempermudah pembangunan model regresi linier, yang digunakan untuk menganalisis hubungan antara jumlah pernikahan dan tingkat perceraian. Pustaka ini juga menyediakan alat untuk validasi dan evaluasi model, seperti penghitungan nilai `R-squared` dan `Mean Squared Error (MSE)`. Untuk mendukung penyajian temuan secara informatif, `Matplotlib` dan `Seaborn` digunakan dalam visualisasi data. `Scatter plot`, misalnya, memungkinkan pemetaan hubungan antara jumlah pernikahan dan tingkat perceraian, sehingga pola tren lebih mudah diidentifikasi. Keunggulan Python dalam menangani data kompleks, serta dokumentasi yang luas dan komunitas yang aktif, menjadikannya pilihan ideal dalam penelitian ini. Dengan memanfaatkan Python dan pustaka pendukungnya, penelitian ini dapat

menghasilkan analisis yang akurat dan interpretasi hasil yang lebih baik dalam memahami hubungan antara jumlah pernikahan dan tingkat perceraian di Jawa Tengah.

## 2.5 Metode Regresi Linier

Regresi linear adalah metode statistik yang digunakan untuk menganalisis hubungan antara variabel independen (prediktor) dan variabel dependen (respon) yang bersifat kuantitatif. Metode ini memungkinkan prediksi nilai variabel dependen berdasarkan variabel independen dengan kesalahan minimal [22][23][24][25]. Regresi linear juga bersifat asimetris, di mana hubungan antara variabel mengikuti pola sebab-akibat.

Dalam penelitian ini, regresi linear digunakan untuk memprediksi tingkat perceraian berdasarkan jumlah pernikahan. Model ini memanfaatkan hubungan linier antara dua variabel, yaitu jumlah pernikahan sebagai variabel independen (X) dan tingkat perceraian sebagai variabel dependen (Y). Dengan data dari BPS Jawa Tengah, model regresi menghasilkan persamaan matematis yang dapat digunakan untuk membuat prediksi. Secara umum, regresi linear sederhana memiliki persamaan dasar berikut [13][25][26]:

$$Y = b + aX \quad (2)$$

Dimana :

Y = Variabel terikat

a = Konstanta (intercept)

b = Koefisien regresi

X = Variabel bebas

Untuk menentukan nilai a dan b, digunakan metode kuadrat terkecil, yang meminimalkan selisih kuadrat antara

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

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

Proses analisis dalam penelitian ini dimulai dengan persiapan data dan pemisahan variabel independen serta dependen. Model regresi dibangun menggunakan pustaka Scikit-learn di Python, yang secara otomatis menghitung nilai koefisien  $a$  dan  $b$  berdasarkan metode kuadrat terkecil. Koefisien ini menunjukkan tingkat pengaruh jumlah pernikahan terhadap tingkat perceraian. Selanjutnya, model dievaluasi untuk menilai akurasi prediksi. Evaluasi dilakukan dengan menghitung R-squared, yang mengukur seberapa besar variasi tingkat perceraian dapat dijelaskan oleh jumlah pernikahan, serta Mean Squared Error (MSE) untuk mengukur rata-rata kesalahan prediksi. Hasil akhir divisualisasikan dalam bentuk scatter plot, di mana data aktual dan garis regresi ditampilkan. Grafik ini memberikan gambaran visual tentang hubungan antara jumlah pernikahan dan tingkat perceraian serta menunjukkan keakuratan model prediktif yang dihasilkan. Dengan langkah-langkah ini, regresi linear sederhana menjadi alat yang efektif untuk memahami dan memprediksi dinamika perceraian berdasarkan data pernikahan di Jawa Tengah.

## 2.6 Uji Hipotesis

Uji hipotesis dilakukan setelah model regresi linier dibuat dan asumsi-asumsi terpenuhi. Tujuan dilakukan uji hipotesis adalah untuk menguji signifikansi statistik dari koefisien regresi atau model secara keseluruhan. Berikut uji hipotesis yang akan digunakan dalam regresi linier [17][18]:

### 1. Uji t (Signifikansi Koefisien Regresi)

Uji ini menentukan apakah variabel independen (jumlah pernikahan) memiliki pengaruh yang signifikan terhadap variabel dependen (tingkat perceraian). Hipotesis yang diuji adalah:

$H_0$  : Koefisien regresi sama dengan nol ( $\beta = 0$ ), yang berarti jumlah pernikahan tidak berpengaruh signifikan terhadap tingkat perceraian.

$H_1$  : Koefisien regresi tidak sama dengan nol ( $\beta \neq 0$ ), yang berarti jumlah pernikahan berpengaruh signifikan terhadap tingkat perceraian.

Pengujian ini dilakukan dengan melihat nilai p-value dari hasil regresi menggunakan Python dengan pustaka statsmodels. Jika p-value lebih kecil dari tingkat signifikansi (misalnya 0,05), maka  $H_0$  ditolak dan variabel independen dianggap signifikan.

### 2. Uji F (Kesesuaian Model)

Uji ini menilai apakah model regresi secara keseluruhan memiliki kemampuan prediksi yang signifikan.

Hipotesis yang diuji adalah:

$H_0$ : Model regresi tidak memiliki kemampuan prediksi yang signifikan.

$H_1$ : Model regresi secara keseluruhan signifikan dalam menjelaskan hubungan antara jumlah pernikahan dan tingkat perceraian. Uji F membandingkan variasi yang dijelaskan oleh model dengan variasi error.

Jika nilai F lebih besar dari nilai kritis atau  $p\text{-value} < 0,05$ , maka model dianggap signifikan.

## 2.7 Evaluasi Model

Dalam tahapan evaluasi model akan dilakukan perhitungan untuk menghitung R-squared ( $R^2$ ) dan Mean Square Error (MSE) untuk mengevaluasi model. Evaluasi model tersebut bertujuan untuk mengetahui seberapa baik model dapat menjelaskan hubungan antara jumlah pernikahan dan tingkat perceraian. Nilai R-squared ( $R^2$ ) digunakan untuk mengukur proporsi variasi dalam tingkat perceraian yang dapat dijelaskan oleh jumlah pernikahan. Nilai  $R^2$  berkisar antara 0 hingga 1, dengan nilai yang lebih tinggi menunjukkan model yang lebih baik dalam menjelaskan variasi data [27]. Sedangkan nilai Mean Squared Error (MSE) digunakan untuk menghitung rata-rata kuadrat selisih antara nilai aktual dan nilai prediksi model. MSE yang lebih kecil menunjukkan bahwa model memiliki kesalahan prediksi yang lebih rendah [28].

Sebagai ilustrasi, jika dalam penelitian ini diperoleh nilai  $R^2$  sebesar 0,85, ini berarti bahwa 85% variasi dalam tingkat perceraian dapat dijelaskan oleh jumlah pernikahan. Dengan kata lain, model memiliki tingkat prediksi yang cukup baik. Sebaliknya, jika MSE sebesar 500, ini menunjukkan bahwa rata-rata kesalahan prediksi model adalah sebesar 500 kasus perceraian. MSE yang tinggi menandakan bahwa model masih memiliki kesalahan prediksi yang signifikan. Visualisasi hasil evaluasi model akan ditampilkan dalam bentuk scatter plot dengan garis regresi untuk menunjukkan sejauh mana model dapat mencocokkan data aktual. Dengan adanya evaluasi ini, keandalan model dalam memprediksi tingkat perceraian berdasarkan jumlah pernikahan dapat dianalisis dengan lebih baik.

## 2.7 Visualisasi Hasil

Setelah melakukan analisis data kemudian data tersebut akan divisualisasikan menggunakan grafik scatter plot untuk menunjukkan hubungan antara jumlah pernikahan dan tingkat perceraian, serta garis regresi yang dihasilkan. Scatter plot dipilih karena mampu menampilkan hubungan dua variabel dalam bentuk titik-titik data, yang memungkinkan analisis pola hubungan linier secara langsung [29]. Dengan menggunakan scatter plot, pola distribusi data dapat diamati dengan jelas, apakah menunjukkan hubungan yang kuat atau terdapat penyimpangan yang signifikan. Selain scatter plot, garis regresi juga akan ditampilkan untuk menunjukkan kecenderungan hubungan antara jumlah pernikahan dan tingkat perceraian. Garis ini memberikan gambaran bagaimana model regresi linier sederhana memprediksi tingkat perceraian berdasarkan jumlah pernikahan. Dengan adanya visualisasi ini dapat mempermudah dalam memahami hubungan antara kedua variabel serta menilai keakuratan model yang digunakan.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan Data

Pada penelitian ini, data diperoleh dari situs resmi Badan Pusat Statistik (BPS) Provinsi Jawa Tengah (<https://jateng.bps.go.id>) [7]. Data yang diambil mewakili berbagai wilayah di Jawa Tengah untuk memberikan gambaran yang representatif tentang hubungan antara pernikahan dan perceraian di tingkat kabupaten/kota, terdapat 35 wilayah kabupaten dan kota di Jawa Tengah dengan jumlah kasus pernikahan dan perceraian yang terjadi pada tiap-tiap daerah tersebut di tahun 2023. Data BPS di Jawa Tengah tahun 2023 dapat dilihat pada Tabel 1. Data pada tabel tersebut akan dibuat menjadi file dengan format .csv agar pemrograman Python dapat membaca data yang akan diolah nantinya

**Tabel 1.** Data Pernikahan dan Perceraian

	Kabupaten / Kota	Pernikahan	Perceraian
0	Kota Magelang	743	229
1	Kota Surakarta	3049	893
2	Kota Salatiga	939	344
3	Kota Semarang	9251	2958
4	Kota Pekalongan	2098	492
5	Kota Tegal	1898	575
6	Kabupaten Cilacap	15156	5922

7	Kabupaten Banyumas	12645	4579
8	Kabupaten Purbalingga	7170	2295
9	Kabupaten Banjarnegara	7147	2407
10	Kabupaten Kebumen	9707	2740
11	Kabupaten Purworejo	5469	1588
12	Kabupaten Wonosobo	6466	2333
13	Kabupaten Magelang	8648	2130
14	Kabupaten Boyolali	6703	1724
15	Kabupaten Klaten	7396	1727
16	Kabupaten Sukoharjo	5324	1264
17	Kabupaten Wonogiri	6579	1680
18	Kabupaten Karanganyar	5821	1302
19	Kabupaten Sragen	6771	2060
20	Kabupaten Grobogan	11012	3377
21	Kabupaten Blora	6583	1959
22	Kabupaten Rembang	4596	1104
23	Kabupaten Pati	9211	2654
24	Kabupaten Kudus	5826	1309
25	Kabupaten Jepara	8658	2080
26	Kabupaten Demak	9034	2396
27	Kabupaten Semarang	6552	2080
28	Kabupaten Temanggung	5144	1404
29	Kabupaten Kendal	7513	2585
30	Kabupaten Batang	5928	1901
31	Kabupaten Pekalongan	7361	1729
32	Kabupaten Pemalang	11811	3823
33	Kabupaten Tegal	12430	3683
34	Kabupaten Brebes	15505	5041

Lanjutan **Tabel 1**

Pada codingan python perlu melakukan import modul Pandas, Matplotlib, Matplotlib.pyplot, dan Seaborn. Pandas digunakan untuk manipulasi dan analisis data berbasis tabel. Matplotlib digunakan untuk membuat visualisasi data digunakan untuk membuat berbagai grafik (seperti garis, batang, scatter, dll.) Matplotlib.pyplot merupakan bagian Matplotlib yang menyediakan interface berbasis state machine seperti MATLAB, digunakan untuk membuat grafik dengan cepat. Seaborn merupakan pustaka berbasis Matplotlib yang digunakan untuk membuat visualisasi data lebih estetik dan informatif. Seaborn menyediakan fungsi bawaan untuk analisis statistik seperti heatmap, pairplot, dan lainnya. Import modul pada python dapat dilihat pada Gambar 2 dibawah ini.

```
import pandas as pd
import matplotlib as mp
import matplotlib.pyplot as plt
import seaborn as sns
```

**Gambar 2.** Import modul

Agar Python dapat memproses data pada Tabel 1, perlu mengimport data dengan bantuan modul pandas. Import data dapat dilihat pada Gambar 3.

```
data = pd.read_csv("pernikahan_perceraian.csv")
```

**Gambar 3.** Import Data

Setelah data pernikahan\_perceraian.csv berhasil diimpor ke dalam program Python, data tersebut dibagi menjadi dua variabel, yaitu variabel independen (x) dan variabel dependen (y). Dalam penelitian ini variabel independennya adalah data pernikahan sedangkan data dependennya adalah perceraian. Proses pembagian variabel dalam program Python dapat dilihat pada Gambar 4 berikut.

```
x = data[['Pernikahan']]
y = data[['Perceraian']]
```

**Gambar 4.** Pembagian Variabel

Tahapan berikutnya setelah melakukan pembagian data membagi data menjadi data uji (*Testing*) dan data latih (*Training*). Data latih digunakan untuk membuat model machine learning, sedangkan data latih digunakan untuk mengevaluasi kinerja model Regresi. Pembagian data dibutuhkan dalam pemodelan untuk mengevaluasi data yang akan diolah menggunakan model Regresi Linear. Pembagian data dalam python dapat dilihat pada Gambar 7 dibawah ini.

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state = 0)
```

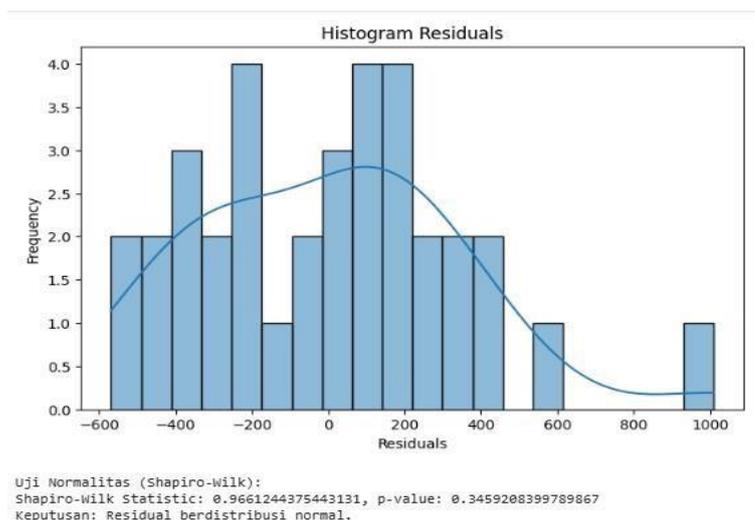
**Gambar 5.** Pembagian Data

Pada Gambar 5 dilakukan proses pembagian data dengan menggunakan fungsi `train_test_split` dari library `scikitlearn` untuk membagi dataset menjadi data latih (*training set*) dan data uji (*test set*). Dalam penelitian ini, `x` merepresentasikan jumlah pernikahan sebagai variabel independen (*input*), sedangkan `y` merepresentasikan jumlah perceraian sebagai variabel dependen (*output*). Parameter `test_size=0.2` berarti 20% data digunakan untuk pengujian, sementara 80% sisanya untuk pelatihan[30]. Ukuran pembagian ini dipilih karena memberikan keseimbangan antara jumlah data yang cukup untuk melatih model dan data yang memadai untuk menguji performa model. Jika proporsi data uji terlalu kecil, hasil evaluasi model dapat menjadi kurang akurat karena tidak mewakili variasi data yang cukup. Sebaliknya, jika proporsi data uji terlalu besar, model mungkin tidak mendapatkan cukup informasi selama pelatihan, yang dapat menyebabkan performa prediksi yang buruk. Dengan pembagian 80:20, model dapat belajar pola dari sebagian besar data dan tetap diuji dengan data yang tidak terlihat sebelumnya, sehingga hasil evaluasi lebih dapat diandalkan[30]. Parameter `random_state=0` memastikan pembagian data tetap konsisten setiap kali kode dijalankan. Hasil pembagian berupa empat variabel yaitu `x_train` dan `y_train` digunakan untuk melatih model regresi, kemudian `x_test` dan `y_test` digunakan untuk mengevaluasi performa model dalam memprediksi tingkat perceraian berdasarkan jumlah pernikahan.

### 3.2 Uji Asumsi

Setelah data dikumpulkan dilakukan uji hipotesis untuk memastikan bahwa data memenuhi persyaratan yang diperlukan agar model regresi linier dapat memberikan hasil yang valid.

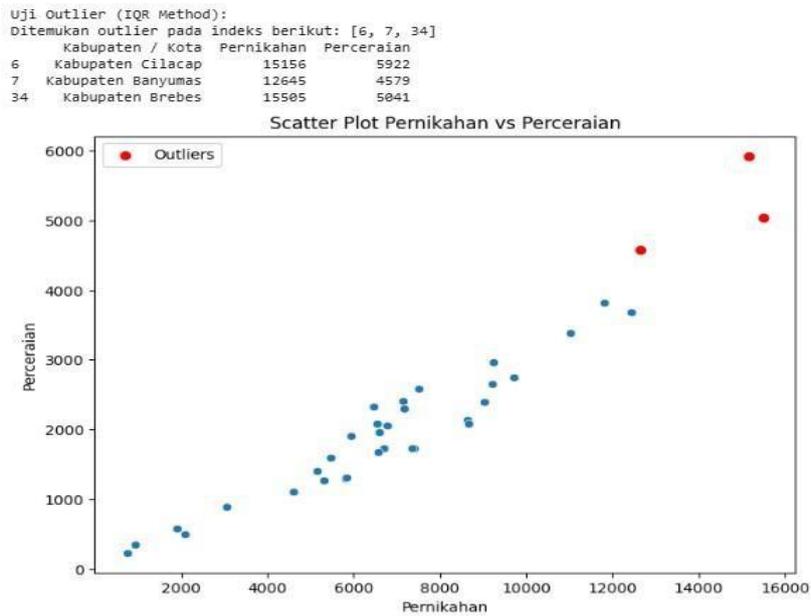
#### 3.2.1 Uji Normalitas



**Gambar 6.** Histogram Residual dan Hasil Uji Normalitas

Pada Gambar 6 terdapat Histogram Residual dan Hasil Uji Normalitas. Histogram residual menunjukkan distribusi residual dari model regresi linier, dengan kurva KDE yang mendekati distribusi normal. Uji Shapiro-Wilk menghasilkan statistik 0.9661 dan p-value 0.3459 ( $> 0.05$ ), sehingga residual dianggap berdistribusi normal. Ini menunjukkan bahwa asumsi normalitas dalam regresi terpenuhi, sehingga hasil estimasi model dapat diandalkan..

### 3.2.2 Uji Outlier



Gambar 7. Hasil Uji Outlier dan Scatter Plot Outlier

Pada Gambar 7 Uji outlier dengan metode IQR mengidentifikasi tiga outlier pada Kabupaten Cilacap, Banyumas, dan Brebes, yang memiliki jumlah perceraian lebih tinggi dibanding daerah lain dengan jumlah pernikahan serupa. Scatter plot menunjukkan titik-titik outlier ini dalam warna merah, menandakan kemungkinan faktor khusus yang memengaruhi angka perceraian di wilayah tersebut. Pada tahapan selanjutnya outlier tidak dihilangkan karena data tersebut mewakili data asli kabupaten dan kota di Jawa Tengah yang berjumlah 35 wilayah.

### 3.3 Analisis Regresi Linier

Setelah uji asumsi terpenuhi, selanjutnya dilakukan pemodelan data menggunakan metode Regresi Linear. Metode Regresi Linear ini yang akan menganalisis prediksi hubungan antara pernikahan dan perceraian. Pada python Regresi Linear dimodelkan dengan menggunakan syntax untuk memunculkan pemodelan tersebut. Visualisasi pemodelan Regresi Linear dapat dilihat pada Gambar 8 dibawah ini.

```

from sklearn.linear_model import LinearRegression

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

LinearRegression

Gambar 8. Pemodelan Regresi Linear

Untuk melakukan prediksi dilakukan dengan cara menampilkan nilai `y_pred`. Angka-angka dalam list Gambar 9 di bawah ini adalah nilai-nilai yang diprediksi oleh model untuk variabel target (perceraian). Hasil nilai prediksi pada penelitian ini dapat dilihat pada Gambar 9.

```

print(y_pred)
[[2281.38561359]
 [3501.4746475 ]
 [1518.08898734]
 [1455.32361835]
 [1264.23793942]
 [2240.58812374]
 [3046.42572231]]
    
```

Gambar 9. Nilai Prediksi

Untuk menganalisis model regresi linier, dilakukan penghitungan nilai Mean Square Error (MSE) dan R-Square ( $R^2$ ). Kedua metrik ini digunakan untuk mengevaluasi kemampuan model dalam memprediksi tingkat perceraian di Jawa Tengah. Hasil perhitungan nilai MSE dan  $R^2$  disajikan pada Gambar 10 berikut.

```

from sklearn.metrics import mean_squared_error, r2_score

mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print ("Mean Square Error (MSE : ", mse)
print ("R-Square (R2 : ", r2)

```

Mean Square Error (MSE : 79745.23459544903  
R-Square (R2 : 0.877434610576003

**Gambar 10.** Nilai MSE dan R-Square

Dari data pada Gambar 10 tersebut setelah diproses menggunakan model regresi, didapatkan hasil bahwa :

1. Mean Square Error (MSE):  
 Nilai MSE sebesar 79745.23 menunjukkan rata-rata kuadrat selisih antara nilai aktual ( $y_{test}$ , data perceraian) dan prediksi ( $y_{pred}$ ). Nilai yang lebih kecil biasanya mengindikasikan model yang lebih baik, tetapi interpretasinya juga bergantung pada skala data. Dalam penelitian ini, jika jumlah perceraian memiliki nilai absolut yang besar, MSE tersebut terhitung wajar, karena terdapat outlier pada data yaitu 3 wilayah dengan tingkat perceraian dan pernikahan yang lebih tinggi dari daerah yang lain yaitu pada Kabupaten Cilacap, Banyumas, dan Brebes.
2. R-Square ( $R^2$ ):  
 Nilai  $R^2$  sebesar 0.8774 menunjukkan bahwa sekitar 87.74% variansi dalam data perceraian (variabel dependen) dapat dijelaskan oleh model yang menggunakan data pernikahan (variabel independen). Ini adalah indikator bahwa model regresi memiliki performa yang sangat baik dalam menjelaskan hubungan antara jumlah pernikahan dan jumlah perceraian. Menurut penelitian yang dilakukan oleh Senggetang *dkk* (2019), apabila nilai  $R^2$  mendekati 1, menunjukkan bahwa variasi variabel dependen dapat dijelaskan oleh variasi variabel independen. Sebaliknya jika nilai  $R^2$  mendekati 0, maka variasi dari variabel dependen tidak dapat dijelaskan oleh variabel independen[27].

Berdasarkan hasil akurasi yang didapat dari model Regresi Linier menunjukkan performa yang sangat baik dengan nilai  $R^2$  yang tinggi, mengindikasikan hubungan kuat antara pernikahan dan perceraian.

### 3.4 Uji Hipotesis

Uji hipotesis dilakukan untuk menguji signifikansi statistik dari koefisien regresi atau model secara keseluruhan. Pada tahap uji hipotesis ini dilakukan uji t dan uji F yang hasilnya dapat dilihat pada Gambar 11 dibawah ini.

```

Uji t:
const          -2.670282
Pernikahan     20.491846
dtype: float64
const          1.167076e-02
Pernikahan     2.432446e-20
dtype: float64
Uji F:
F-statistic: 419.91577129255694
p-value F-statistic: 2.4324460990182995e-20

Uji Hipotesis Regresi Linier Sederhana:
H0: Koefisien regresi = 0 (tidak ada pengaruh signifikan)
H1: Koefisien regresi ≠ 0 (ada pengaruh signifikan)
Keputusan: Tolak H0, terdapat pengaruh signifikan antara jumlah pernikahan dan jumlah perceraian.

```

**Gambar 11.** Hasil Uji t dan Uji F

Pada Gambar 11 Hasil Uji t dan Uji F, pada tahap uji t menunjukkan bahwa variabel jumlah pernikahan memiliki koefisien regresi yang signifikan dengan nilai  $t = 20.49$  dan p-value yang sangat kecil ( $2.43e-20$ ), menunjukkan adanya hubungan signifikan dengan jumlah perceraian. Sedangkan pada tahap Uji F, nilai F-statistic 419.92 dan p-value yang juga sangat kecil mengindikasikan bahwa model regresi secara keseluruhan signifikan. Dengan demikian,  $H_0$  ditolak, yang berarti terdapat pengaruh signifikan antara jumlah pernikahan dan jumlah perceraian.

### 3.5 Interpretasi Hasil

```

Coefficients: [[0.34869649]]
Intercept: [-338.37114886]

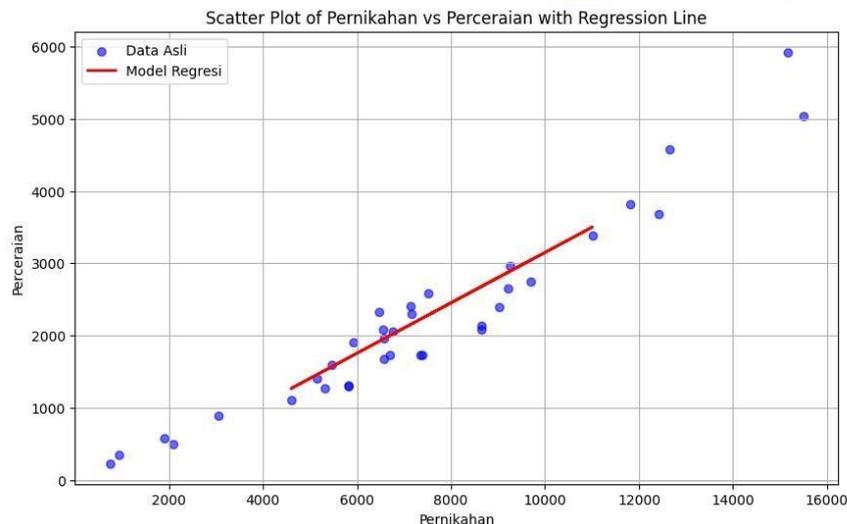
```

**Gambar 12.** Hasil Koefisien dan Intercept

Pada Gambar 12 Hasil Koefisien dan Intercept regresi, terdapat nilai koefisien sebesar 0.3487 menunjukkan bahwa setiap peningkatan satu pernikahan berkorelasi dengan kenaikan rata-rata 0.3487 kasus perceraian, dengan asumsi

faktor lain tetap konstan. Intercept sebesar -338.37 mengindikasikan bahwa jika tidak ada pernikahan yang terjadi ( $X = 0$ ), maka model memprediksi tingkat perceraian sebesar -338, yang secara praktis tidak masuk akal. Hal ini menunjukkan bahwa model ini lebih akurat untuk memprediksi hubungan dalam rentang data yang telah digunakan, tetapi tidak dapat digunakan untuk ekstrapolasi di luar rentang data tersebut.

Visualisasi hasil regresi linier disajikan melalui grafik scatter plot, yang memberikan representasi visual mengenai hubungan antara jumlah pernikahan dan jumlah perceraian di berbagai kabupaten/kota di Jawa Tengah. Pada grafik tersebut, garis merah yang melintasi titik-titik data merepresentasikan garis regresi, yaitu model matematis terbaik yang menggambarkan hubungan antara kedua variabel. Visualisasi ini dapat dilihat pada Gambar 13.



**Gambar 13.** Grafik scatter plot regresi pernikahan vs perceraian

Dari Gambar 13 Grafik scatter plot regresi pernikahan vs perceraian tersebut dapat diamati beberapa hal:

1. Titik Data (Lingkaran Biru):  
Titik-titik biru merepresentasikan data aktual hubungan antara jumlah pernikahan dan jumlah perceraian. Pola distribusi titik-titik menunjukkan bahwa secara umum, jumlah perceraian meningkat seiring dengan bertambahnya jumlah pernikahan.
2. Garis Regresi (Garis Merah):  
Garis merah merupakan hasil dari model regresi linier yang memprediksi hubungan antara jumlah pernikahan dan perceraian. Garis ini dirancang untuk meminimalkan selisih (kesalahan) antara prediksi model dan data aktual.
3. Hubungan Linier Positif:  
Berdasarkan plot, terdapat hubungan linier positif antara jumlah pernikahan dan perceraian, yang berarti bahwa semakin tinggi jumlah pernikahan, semakin tinggi pula jumlah perceraian.
4. Keselarasan Model dengan Data:  
Garis regresi terlihat melewati sebagian besar titik data dengan baik, menunjukkan bahwa model cukup akurat dalam merepresentasikan hubungan tersebut. Hal ini didukung oleh nilai  $R^2 = 0.8774$ , yang menunjukkan model dapat menjelaskan 87.74% variansi data.
5. Terdapat Outlier dalam Scatter Plot  
Terdapat beberapa outlier, yaitu data yang jauh dari pola umum, terutama pada bagian kanan atas grafik, yang menunjukkan daerah dengan jumlah pernikahan dan perceraian yang jauh lebih tinggi dibandingkan tren umum. Outlier ini dapat memengaruhi hasil regresi, pada data yang terdapat outlier tidak dihilangkan karena mewakili data asli kabupaten dan kota di Jawa Tengah yang berjumlah 35 wilayah.

Model regresi linier yang divisualisasikan dalam scatter plot ini memberikan gambaran yang kuat tentang hubungan antara jumlah pernikahan dan perceraian, dengan hubungan yang cenderung linier positif. Namun, interpretasi ini harus dilengkapi dengan analisis terhadap faktor-faktor eksternal yang mungkin tidak ditangkap oleh model.

## 4. KESIMPULAN

Penelitian ini telah berhasil menganalisis dan memprediksi tingkat perceraian di Jawa Tengah pada tahun 2023 berdasarkan data pernikahan dari 35 kabupaten/kota. Hasil analisis menunjukkan adanya hubungan linier yang signifikan antara jumlah pernikahan dan tingkat perceraian di wilayah tersebut. Model regresi linier sederhana yang digunakan menunjukkan performa yang sangat baik dengan nilai R-squared sebesar 87,74%, yang berarti sebanyak 87,74% variasi tingkat perceraian dapat dijelaskan oleh jumlah pernikahan. Menurut Senggetang dkk. (2019), nilai R-squared yang mendekati 1 menandakan bahwa model memiliki kemampuan yang sangat baik dalam menjelaskan hubungan antara variabel independen dan dependen, sementara nilai yang mendekati 0 menunjukkan sebaliknya. Dengan demikian, model yang digunakan dalam penelitian ini memiliki tingkat akurasi yang tinggi dalam menjelaskan fenomena perceraian di Jawa Tengah. Selain itu, nilai Mean Squared Error (MSE) sebesar 79.745,23 menunjukkan bahwa tingkat kesalahan prediksi masih dalam batas wajar. Karena, keberadaan outlier dari beberapa daerah seperti Kabupaten Cilacap, Banyumas, dan Brebes yang memiliki tingkat perceraian lebih tinggi dibanding daerah lain dengan jumlah pernikahan serupa berkontribusi terhadap tingkat kesalahan ini. Model regresi juga menghasilkan koefisien regresi sebesar 0,3486, yang menunjukkan bahwa setiap peningkatan satu pernikahan rata-rata berhubungan dengan kenaikan 0,3486 perceraian. Beberapa wilayah menunjukkan hasil prediksi tingkat perceraian yang bervariasi, dengan estimasi seperti 2.281,39; 3.501,47; 1.518,09; 1.455,32; 1.264,24; 2.240,59; dan 3.046,43 kasus. Hal ini semakin memperkuat adanya hubungan kuat antara jumlah pernikahan dan tingkat perceraian di berbagai wilayah di Jawa Tengah.

Terdapat kekurangan pada model yang digunakan untuk penelitian ini yaitu model hanya mempertimbangkan hubungan linier sederhana antara jumlah pernikahan dan tingkat perceraian, tanpa memasukkan faktor-faktor non-linier atau variabel lain yang dapat memengaruhi perceraian, seperti aspek sosioekonomi, budaya, dan psikologis. Selain itu, penelitian ini terbatas pada data tahun 2023 dari Jawa Tengah, sehingga hasilnya mungkin tidak dapat digeneralisasikan ke wilayah atau periode waktu yang lebih luas.

Hasil penelitian ini menegaskan perlunya kebijakan berbasis data untuk menekan angka perceraian. Pemerintah daerah dan lembaga terkait, seperti Kantor Urusan Agama (KUA), dapat merancang program edukasi pernikahan untuk meningkatkan kesiapan pasangan sebelum menikah seperti melakukan penyuluhan terhadap calon pengantin, melakukan seminar bimbingan pernikahan, konseling pernikahan dan beberapa kebijakan lainnya untuk menurunkan angka perceraian. Selain itu, penelitian ini juga dapat menjadi referensi bagi studi serupa di wilayah lain serta sebagai panduan dalam perencanaan program sosial berbasis data yang lebih efektif.

## UCAPAN TERIMAKASIH

Peneliti menyampaikan penghargaan dan terima kasih yang sebesar-besarnya kepada seluruh pihak yang telah berkontribusi dalam kelancaran penelitian ini. Ucapan terima kasih secara khusus ditujukan kepada para pembimbing atas bimbingan, arahan, serta masukan berharga yang telah diberikan dengan penuh kesabaran. Apresiasi juga disampaikan kepada pihak lembaga yang telah memfasilitasi penyediaan data dan sumber daya yang diperlukan. Kepada rekan-rekan sejawat, peneliti mengucapkan terima kasih atas dukungan, semangat, dan berbagi pengetahuan yang sangat berarti. Tidak lupa, penghargaan setulusnya disampaikan kepada keluarga atas motivasi dan dukungan moril selama proses penelitian ini berlangsung. Peneliti menyadari bahwa penelitian ini merupakan hasil dari kerja sama banyak pihak, dan dengan rendah hati berharap semoga hasilnya dapat memberikan kontribusi positif bagi pengembangan ilmu pengetahuan serta memberi manfaat nyata bagi masyarakat. Terima kasih atas segala bentuk dukungan yang telah diberikan.

## REFERENCES

- [1] Nur'eni and L. Handayani, "Regresi probit untuk analisis variabel-variabel yang mempengaruhi perceraian di Sulawesi Tengah," pp. 13–22, 2016.
- [2] U. D. Rahayu, N. L. Chusna, and M. Fachri, "ANALISIS KASUS PERCERAIAN PADA PENGADILAN NEGERI BEKASI MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING," vol. 6, no. 1, pp. 165–172, 2022.
- [3] Windarman, Sapri, and E. Suryana, "Implementation of the Naïve Bayes Algorithm for Divorce Prediction at the Tais Religious Court Implementasi Algoritma Naïve Bayes Untuk Prediksi Perceraian Pada Pengadilan Agama Tais," vol. 2, no. 2, pp. 501–510, 2022.
- [4] E. Purwaningsih and E. Nurelasari, "Implementasi Metode K-Means Clustering Dengan Davies Bouldin Index Pada Analisis Faktor Penyebab Perceraian," vol. 7, no. 2, pp. 134–143, 2023.
- [5] M. A. Rizaty, "Ada 516.344 Kasus Perceraian di Indonesia pada 2022," *dataindonesia.id*, 2023.

- <https://dataindonesia.id/varia/detail/ada-516344-kasus-perceraian-di-indonesia-pada-2022> (accessed Dec. 08, 2024).
- [6] N. Januari, “MENGKALI AKAR MASALAH : Analisis Kasus Perceraian di Indonesia,” vol. 3, no. 3, pp. 120–130, 2023.
- [7] A. D. Darmawan, “Jumlah Perceraian Jawa Tengah 23,18 Ribu Kasus Data per 2023,” *databoks.katadata.co.id*, 2024. [https://databoks.katadata.co.id/demografi/statistik/0878d7df0f2af26/jumlah-perceraian-jawa-tengah-2318ribu-kasus-data-per-2023#:~:text=Jumlah perceraian di Jawa Tengah tercatat 23%2C18 ribu,di wilayah ini menunjukkan penurunan di angka -0%2C74%25](https://databoks.katadata.co.id/demografi/statistik/0878d7df0f2af26/jumlah-perceraian-jawa-tengah-2318ribu-kasus-data-per-2023#:~:text=Jumlah%20perceraian%20di%20Jawa%20Tengah%20tercatat%2023%2C18%20ribu,di%20wilayah%20ini%20menunjukkan%20penurunan%20di%20angka%20-0%2C74%25). (accessed Feb. 06, 2025).
- [8] Jateng.bps.go.id, “Jumlah Pernikahan dan Perceraian Menurut Kabupaten/Kota di Provinsi Jawa Tengah, 20202023,” *jateng.bps.go.id*, 2024. <https://jateng.bps.go.id/id/statistics-table/2/NDk5IzI=/jumlah-pernikahandanperceraian-menurut-kabupaten-kota-di-provinsi-jawa-tengah.html> (accessed Dec. 08, 2024).
- [9] A. Wang, A. Zhang, E. H. W. Chan, W. Shi, and X. Zhou, “A Review of Human Mobility Research Based on Big Data and Its Implication for Smart City Development,” 2021.
- [10] M. Amien, “Workshop Pengenalan Bahasa Pemrograman Python untuk Data Sains,” pp. 39–42, 2024.
- [11] S. Raschka, J. Patterson, and C. Nolet, “Machine Learning in Python : Main Developments and Technology Trends in Data Science , Machine Learning , and Artificial Intelligence,” 2020, doi: 10.3390/info11040193.
- [12] W. McKinney, *Python for Data Analysis*, 2nd Editio. United States of America: O’Reilly Media, 2018.
- [13] R. A. Samosir, M. F. Rozy, and A. P. Windarto, “Penerapan Algoritma Regresi Linier Berganda dalam Mengestimasi Jumlah Perceraian di Pengadilan Agama Simalungun,” vol. 2, no. 1, pp. 16–20, 2021.
- [14] M. arif abdul & Syukur and M. Faisal, “Penerapan Model Regresi Linear Untuk Estimasi Mobil Bekas Menggunakan Bahasa Python,” vol. 11, no. 2, pp. 182–191, 2023.
- [15] M. Awaluddin *et al.*, *METODE PENELITIAN MANAJEMEN DAN BISNIS*, Cetakan Pe. Yayasan Tri Edukasi Ilmiah, 2024.
- [16] L. S. Febrianto, N. K. Dwiyaniti, and P. Hendikawati, “PERBANDINGAN METODE ROBUST LEAST MEDIAN OF SQUARE ( LMS ) DAN PENDUGA S UNTUK MENANGANI OUTLIER PADA REGRESI LINIER BERGANDA,” vol. 7, no. 1, pp. 83–95, 2018.
- [17] Y. Aryani and D. Gustian, “SISTEM INFORMASI PENJUALAN BARANG DENGAN METODE REGRESI LIEAR BERGANDA DALAM PEMBUATAN PERUSAHAAN,” vol. 2, no. 2, 2020.
- [18] R. Puspitasari and D. Astrini, “Dampak Literasi Dan Inkuisi Keuangan Terhadap Kinerja Pelaku UMKM Di Kota Bogor,” vol. 9, no. 2, pp. 435–446, 2021, doi: 10.37641/jimkes.v9i2.771.
- [19] R. Efendi, A. Junaidi, and A. M. Rizki, “PENENTUAN PUSAT KLASSTER SECARA OTOMATIS PADA ALGORITMA DENSITY PEAKS CLUSTERING BERBASIS METODE ITER QUARTILE RANGE,” vol. 12, no. 3, 2024.
- [20] N. N. Aulia, “ANALISA HASIL PREDIKSI MINYAK SAYUR DENGAN REGRESI LINIER DARI MACHINE LEARNING PYTHON DAN R (STUDI KASUS: DATA KAGGLE),” pp. 6–21, 2021.
- [21] M. Sholeh, Suraya, and D. Andayati, “Machine Linear untuk Analisis Regresi Linier Biaya Asuransi Kesehatan dengan Menggunakan Python Jupyter Notebook,” vol. 8, no. 1, pp. 20–27, 2022.
- [22] Anjeliyani, F. Fatimah, and M. Casmat, “PEMODELAN REGRESI LINEAR UNTUK MEMPREDIKSI NILAI PENJUALAN DI PT GOODIEBAG CUSTOM INDONEIA TANPA MEMPERTIMBANGKAN UJI ASUMSI,” vol. 5, no. September, pp. 1–13, 2024. [23] A. Andiati and S. O. R, “Analisis Kualitas Dan Kepuasan Pengguna Website Istyle.id Dengan Metode Webqual 4.0,” vol. 16, no. 2, pp. 111–123, 2023.
- [24] B. Ardianto, M. T. W. Pangesti, and N. Wakhidah, “Analisis Pengaruh Populasi dan Angkatan Kerja Terhadap Pengangguran di Kabupaten Demak Dengan Regresi Linear Berganda,” pp. 1–9.
- [25] N. Suhandi, E. Ayu, K. Putri, and S. Agnisa, “Analisis Pengaruh Jumlah Penduduk terhadap Jumlah Kemiskinan Menggunakan Metode Regresi Linear di Kota Palembang,” vol. 09, no. 2, pp. 77–82, 2018.
- [26] M. F. Arfa, M. R. Al Fathan, H. B. Lumbantobing, and Rahmadden, “Cryptocurrency Price Prediction with Linear Regression Method Prediksi Harga Cryptocurrency dengan Metode Linier Regresi,” pp. 8–15, 2023.
- [27] V. Senggetang, Silvy L. Mandey, and S. Moniharapon, “PENGARUH LOKASI, PROMOSI DAN PERSEPSI HARGA TERHADAP KEPUTUSAN PEMBELIAN KONSUMEN PADA PERUMAHAN KAWANUA EMERALD CITY MANADO,” vol. 7, no. 1, 2019.
- [28] I. Amansyah, J. Indra, E. Nurlaelasari, and A. R. Juwita, “Prediksi Penjualan Kendaraan Menggunakan Regresi Linear : Studi Kasus pada Industri Otomotif di Indonesia,” vol. 4, pp. 1199–1216, 2024.
- [29] L. Husna and P. E. P. Utomo, “Analisis Dan Visualisasi Data Body Performance Menggunakan Tiga Tools Visualisasi,” vol. 5, no. 01, pp. 32–40, 2023.
- [30] V. R. Joseph and A. Vakayil, “SPLIT : An Optimal Method for Data Splitting,” *Technometrics*, vol. 0, no. 0, pp. 1-23, 2022, doi: 10.1080/00401706.2021.1921037.