

Estimasi Umur Menikah Menggunakan Regresi Linier Berdasarkan Faktor Demografis Di Desa Pringtulis

Chelyn Audya Widany¹, Fatimah Azzahro², Vanessa Maharani³,
Amelia Rahmawati⁴, Muhammad Arifin⁵

¹²³⁴⁵ Sistem Informasi; Universitas Muria Kudus; Jl. Lkr. Utara, Kayuapu Kulon, Gondangmanis, Kec. Bae, Kabupaten Kudus, Jawa Tengah 59327; e-mail: 202353100@std.umk.ac.id, 202353099@std.umk.ac.id, 202353129@std.umk.ac.id, 202353098@std.umk.ac.id, arifin.m@umk.ac.id

* Korespondensi e-mail: 202353100@std.umk.ac.id

Diterima: 29 April 2026 ; Review: 04 Juni 2026; Disetujui: 22 Juni 2026

Cara sitasi: Widany CA, Azzahro F, Maharani V, Rahmawati A, Arifin M. 2026. Estimasi Umur Menikah Menggunakan Regresi Linier Berdasarkan Faktor Demografis Di Desa Pringtulis. *Information Management for Educators and Professionals*. Vol. 11 (1): 35 - 45

Abstrak: Pernikahan ialah suatu peristiwa demografis penting yang dipengaruhi oleh berbagai faktor sosial, ekonomi, dan karakteristik individu. Pergeseran usia pernikahan pertama di Indonesia menunjukkan dinamika yang kompleks, baik dalam bentuk praktik pernikahan dini maupun penundaan pernikahan di masyarakat modern. Penelitian ini bertujuan untuk menganalisis faktor-faktor demografis yang memengaruhi umur menikah serta membangun model prediksi menggunakan pendekatan regresi linier. Metode yang diterapkan ialah pendekatan kuantitatif dengan desain kausal-eksplanatori berbasis data mining terhadap dataset kependudukan Desa Pringtulis yang terdiri dari 359 data. Variabel yang dianalisis meliputi jenis kelamin, pendidikan, pekerjaan, pendapatan, golongan ekonomi, dan wilayah tempat tinggal. Proses analisis meliputi tahap pra-pemrosesan data, uji asumsi klasik, pemodelan regresi linier berganda dengan metode *backward elimination*, serta evaluasi model menggunakan uji statistik dan metrik kesalahan. Hasil penelitian menunjukkan bahwa pendidikan, kondisi ekonomi, dan wilayah tempat tinggal memiliki keterkaitan dengan variasi umur menikah. Pendidikan ialah suatu faktor yang berperan untuk membentuk pola usia pernikahan, sedangkan kondisi ekonomi dan wilayah tempat tinggal memperlihatkan adanya perbedaan kecenderungan umur menikah antarresponden. Model yang dihasilkan memiliki nilai *Root Mean Squared Error (RMSE)* sejumlah 5,19, yang memperlihatkan bahwa rata-rata kesalahan prediksi berada pada kisaran ± 5 tahun. Secara keseluruhan, model regresi linier mampu merepresentasikan hubungan antarvariabel dengan cukup baik, meskipun masih terdapat faktor eksternal seperti budaya dan lingkungan sosial yang belum terakomodasi dalam model.

Kata kunci: data mining, estimasi, faktor demografis, regresi linier, umur menikah

Abstract: Marriage is one of the key demographic events influenced by various social and economic factors, as well as individual characteristics. Shifts in the age of first marriage in Indonesia reveal complex dynamics, ranging from early marriage to delayed marriage in modern society. This study aims to analyze the demographic factors influencing the age at marriage and to develop a predictive model using a linear regression approach. The method employed is a quantitative approach with a causal-explanatory design based on data mining applied to the population dataset of Pringtulis Village, comprising 359 data. The variables analyzed include gender, education, occupation, income, economic class, and residential area. The analysis process includes data preprocessing, classical assumption tests, multiple linear regression modeling using the backward elimination method, and model evaluation using statistical tests and error metrics. The results of the study indicate that education, economic conditions, and region of residence are associated with variations in the age at marriage. Education is one of the factors that influences marriage age patterns, while economic conditions and region of residence reveal

differences in marriage age trends among respondents. The resulting model has a Root Mean Squared Error (RMSE) of 5.19, indicating that the average prediction error is within a range of ± 5 years. Overall, the linear regression model is able to represent the relationship between variables fairly well, although there are still external factors, such as culture and the social environment, that have not been accounted for in the model.

Keywords: *age at marriage, data mining, demographic factors, estimation linear regression*

1. Pendahuluan

Pernikahan merupakan salah satu peristiwa transisi demografis yang memiliki implikasi luas terhadap ketahanan keluarga, kesehatan, serta kondisi ekonomi suatu bangsa. Pergeseran umur perkawinan pertama saat ini menjadi fenomena yang kompleks di Indonesia [1]. Di satu sisi, praktik pernikahan dini masih terjadi di beberapa wilayah dan berkorelasi dengan risiko kesehatan reproduksi serta kerentanan ekonomi [2]. Di sisi lain, tren penundaan usia pernikahan di kawasan urban meningkat seiring perubahan orientasi masyarakat terhadap pendidikan tinggi dan kemampuan finansial [3]. Fenomena ini mencerminkan adanya perpanjangan fase transisi menuju kedewasaan yang selaras dengan modernisasi sosial dan pembangunan ekonomi.

Secara teoretis, keputusan individu dalam menentukan waktu pernikahan dipengaruhi oleh berbagai faktor sosio-demografis [4]. Kajian-kajian terdahulu menunjukkan bahwa variabel seperti tingkat pendidikan, status pekerjaan, tingkat pendapatan, wilayah tempat tinggal (perkotaan atau pedesaan), serta jenis kelamin memiliki korelasi yang kuat dalam membentuk pola perilaku perkawinan [5]. Untuk memahami dan mengukur pola hubungan kausalitas antarvariabel tersebut, diperlukan suatu tinjauan matematis yang terukur. Suatu metode prediksi statistik yang andal guna menguji hubungan antara variabel dependen berskala kontinu, seperti umur menikah, dengan variabel independen adalah Regresi Linier. Menyikapi kompleksitas permasalahan tersebut, diperlukan suatu wawasan pemecahan masalah yang bertumpu pada penggalian informasi (data mining) dari sekumpulan data kependudukan. Rencana pemecahan masalah pada kajian ini ialah dengan mengimplementasikan algoritma Regresi Linier pada dataset demografis penduduk [6].

Pendekatan komputasi prediktif ini dipilih karena mampu mengestimasi umur menikah sekaligus mengukur besarnya pengaruh masing-masing faktor demografis terhadap keputusan pernikahan [5]. Maka demikian, kajian ini tujuannya guna menganalisis faktor-faktor demografis yang memengaruhi usia pernikahan pertama menggunakan metode Regresi Linier Berganda. Adapun faktor-faktor demografis yang dianalisis pada kajian ini termasuk jenis kelamin, tingkat pendidikan, jenis pekerjaan, estimasi pendapatan bulanan, golongan ekonomi, serta wilayah tempat tinggal (dusun) yang tersedia dalam dataset kependudukan. Pemilihan variabel tersebut didasarkan pada peranan faktor sosio-ekonomi dalam memengaruhi keputusan usia pernikahan [8]. Beberapa penelitian sebelumnya juga menunjukkan adanya hubungan antara tingkat pendidikan, kondisi ekonomi, dan kecenderungan terjadinya pernikahan dini [9].

Penelitian mengenai usia pernikahan tidak hanya penting dalam perspektif demografi, tetapi juga memiliki keterkaitan erat dengan aspek pembangunan manusia. Usia pernikahan yang terlalu dini berpotensi meningkatkan risiko putus sekolah, rendahnya kualitas kesehatan ibu dan anak, serta keterbatasan akses terhadap peluang ekonomi [10]. Sebaliknya, penundaan usia pernikahan sering kali dikaitkan dengan peningkatan kualitas sumber daya manusia, terkhusus pada hal pendidikan dan kesiapan finansial. Dalam konteks pembangunan berkelanjutan, pengendalian usia pernikahan menjadi salah satu indikator penting dalam mendukung tercapainya kesejahteraan masyarakat [11]. Berbagai penelitian terdahulu sudah mengkaji faktor-faktor yang memengaruhi pernikahan dini menggunakan metode statistik dan machine learning, seperti Regresi Logistik, Regresi Poisson, Classification Tree, dan Random Forest. Metode-metode tersebut digunakan untuk mengidentifikasi faktor-faktor yang berpengaruh maupun mengklasifikasikan kejadian pernikahan dini berdasarkan karakteristik sosial dan ekonomi. Namun, penelitian tersebut umumnya berfokus pada klasifikasi kejadian pernikahan dini dan belum menghasilkan model estimasi umur menikah dalam bentuk variabel numerik secara langsung.

Maka demikian, kajian ini menawarkan kebaruan yaitu penerapan Regresi Linier Berganda dengan metode Backward Elimination untuk membangun model estimasi umur menikah berdasarkan faktor demografis. Dengan pendekatan tersebut, penelitian ini bukan hanya

mengidentifikasi faktor-faktor yang berpengaruh, namun juga menghasilkan model prediksi umur menikah yang lebih terukur.

2. Metode Penelitian

Bagian ini menjelaskan prosedur penelitian secara sistematis yang diterapkan sebagai pedoman dalam pelaksanaan penelitian. Uraian metode mencakup tahapan mulai dari perencanaan hingga evaluasi hasil, meliputi rancangan penelitian, penentuan populasi dan sampel, teknik pengumpulan data, serta teknik pengolahan dan analisis data kependudukan Desa Pringtulis. Dengan adanya penjelasan yang sistematis ini, diharapkan proses penelitian dapat dipahami dengan jelas serta dapat direplikasi pada penelitian selanjutnya dengan konteks yang serupa.

2.1 Studi Literatur

Studi literatur dilaksanakan sebagai tahap awal penelitian guna mendapatkan landasan teoritis serta memahami hasil penelitian terdahulu yang relevan dengan topik umur menikah [12]. Kajian terhadap berbagai sumber literatur ilmiah nasional dirangkum dalam Tabel 1 berikut.

Table 1. Perbandingan Penelitian Terdahulu

No	Penulis (Tahun)	Metode	Dataset	Hasil
1	Hermambang et al. (2021)	Regresi Logistik Biner	Data SDKI nasional	Pendidikan, wilayah, dan kemiskinan berpengaruh signifikan terhadap pernikahan dini; pendidikan menjadi prediktor paling dominan [6]
2	Nafisah & Perdana (2024)	Analisis korelasi deskriptif	Data BPS tingkat nasional	Tingkat pendidikan dan pertumbuhan ekonomi berkorelasi negatif dengan angka pernikahan dini [7]
3	Pangestika et al. (2024)	Regresi Logistik, <i>Classification Tree</i> , <i>Random Forest</i>	Data Susenas 2022, perempuan usia 20–24 tahun, Provinsi NTB	Regresi Logistik menunjukkan performa terbaik (AUC = 77,86%); pendidikan dan ekonomi menjadi prediktor dominan perkawinan anak [5]
4	Rahmania et al. (2025)	Regresi Poisson	Data BPS dan DP3ACSKB Provinsi Kepulauan Bangka Belitung	APM SMA, Tingkat Pengangguran Terbuka, dan Tingkat Penduduk Miskin berpengaruh signifikan terhadap pernikahan dini (AIC = 964,81) [13]
5	Aprihartha et al. (2025)	Regresi Linier Berganda, <i>Backward Elimination</i>	Data kesehatan Provinsi Jawa Tengah 2023 (35 kabupaten/kota)	Model <i>backward elimination</i> terpilih sebagai model terbaik ($R^2 = 60\%$, RMSE = 19,1%); jumlah penduduk miskin, pengelolaan air minum dan makanan, serta jumlah posyandu aktif berpengaruh signifikan [14]

sumber: Hasil Penelitian (2026)

Berdasarkan Tabel 1, penelitian terdahulu yang membahas pernikahan dini umumnya menggunakan pendekatan klasifikasi seperti regresi logistik dan regresi Poisson dengan data agregat tingkat nasional maupun provinsi. Satu-satunya penelitian yang menerapkan regresi linier berganda dengan *backward elimination* adalah Aprihartha et al. (2025), namun domain penelitiannya berbeda, yaitu kasus gizi buruk balita di Jawa Tengah, bukan pernikahan.

Dengan demikian, terdapat research gap yang jelas yaitu belum ada penelitian yang menggunakan regresi linier berganda untuk memprediksi estimasi umur menikah secara numerik berbasis data kependudukan tingkat desa. Studi ini hadir guna mengisi celah tersebut melalui menerapkan regresi linier berganda dengan metode *backward elimination* pada dataset demografis Desa Pringtulis, sehingga menghasilkan model estimasi yang lebih terukur, berbasis data lokal, dan dapat menjadi acuan bagi perumusan kebijakan di tingkat desa.

2.2 Rancangan Penelitian

Studi ini menerapkan pendekatan kuantitatif dengan desain kausal-eksplanatori, yang tujuannya guna menganalisis hubungan sebab-akibat antara variabel prediktor terhadap variabel target. Dalam konteks ilmu data mining, penelitian ini secara khusus menerapkan peran Estimasi untuk memprediksi nilai variabel kontinu, yaitu umur menikah, berdasarkan pola dari data historis demografis penduduk.

Metode regresi linier berganda dipilih sebab kemampuannya dalam memodelkan hubungan simultan antara beberapa variabel prediktor terhadap satu variabel target untuk menghasilkan nilai estimasi yang akurat [15]. Asumsi linearitas memastikan bahwa setiap kontribusi dari faktor seperti pendidikan dan pendapatan dihitung sebagai penambah atau

pengurang nilai estimasi umur menikah secara konstan dalam sebuah model garis lurus. Secara matematis, model regresi linier berganda dinyatakan yaitu:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \epsilon \quad (1)$$

dengan keterangan:

- Y = umur menikah (variabel dependen, dalam satuan tahun)
 X_1 = wilayah dusun tempat tinggal
 X_2 = jenis kelamin
 X_3 = tingkat pendidikan terakhir
 X_4 = jenis pekerjaan
 X_5 = perkiraan pendapatan bulanan
 X_6 = golongan ekonomi
 β_0 = konstanta (intercept)
 $\beta_1, \beta_2, \dots, \beta_6$ = koefisien regresi masing-masing variabel prediktor
 ϵ = galat (error term), yaitu nilai residual yang mencerminkan pengaruh faktor lain di luar model

Kerangka kerja penelitian disusun secara sistematis meliputi tahap pengumpulan data, pra-pemrosesan (preprocessing), pemodelan statistik, hingga tahap evaluasi akurasi model.

2.3 Populasi dan Sampel

Populasi pada kajian ini ialah seluruh penduduk yang tercatat dalam basis data kependudukan Desa Pringtulis sebanyak 1.000 orang. Teknik pengambilan sampel dilaksanakan secara *purposive sampling*, yakni pengambilan sampel melalui pertimbangan kriteria tertentu untuk mencapai tujuan penelitian yang spesifik [16]. Kriteria inklusi sampel ditetapkan pada penduduk yang telah memiliki status kawin dan bertempat tinggal di wilayah Dusun Pringtulis, Rekesan, dan Gajihan. Sesudah dilaksanakan pembersihan data serta penanganan outlier, total sampel yang memenuhi kriteria dan digunakan pada analisis sejumlah 359 data.

2.4 Teknik Pengumpulan Data

Data penelitian didapatkan dari dokumentasi sekunder yang bersumber dari Sistem Informasi Administrasi Kependudukan (SIAK) Desa Pringtulis dalam format Microsoft Excel. Guna melindungi privasi responden, identitas pribadi seperti nama dan nomor induk kependudukan dianonimkan sebelum digunakan dalam analisis, sehingga data yang diolah hanya mencakup atribut demografis yang relevan dengan variabel penelitian.

Data yang diterapkan ialah data historis penduduk yang sudah berstatus kawin, di mana variabel umur menikah (Y) didapatkan dari selisih antara tahun kelahiran dan tahun pernikahan pertama yang tercatat secara resmi dalam buku induk kependudukan desa. Data tersebut mencakup informasi demografis individu yang telah menikah, meliputi jenis kelamin, tingkat pendidikan, pekerjaan, pendapatan, golongan ekonomi, serta wilayah dusun. Definisi operasional dan skala pengukuran setiap variabel disajikan dalam Tabel 2.

Table 2. Definisi Operasional Variabel

Variabel	Kode	Deskripsi Operasional	Skala
Umur Menikah	Y	Usia responden saat pertama kali menikah	Rasio
Wilayah Dusun	X_1	Lokasi tempat tinggal	Nominal
Jenis Kelamin	X_2	Kategori gender	Nominal
Pendidikan	X_3	ID jenjang pendidikan terakhir	Ordinal
Pekerjaan	X_4	ID jenis mata pencaharian responden	Nominal
Pendapatan	X_5	Nilai nominal estimasi pendapatan bulanan	Rasio
Gol. Ekonomi	X_6	Tingkat kesejahteraan	Ordinal

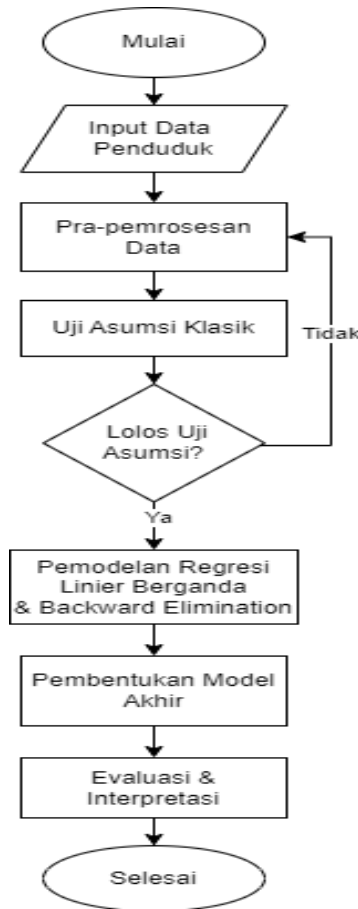
sumber: Hasil Penelitian (2026)

Keterangan: Skala Nominal digunakan untuk pengelompokan tanpa tingkatan, Ordinal untuk data berperingkat, dan Rasio untuk data angka mutlak.

2.5 Teknik Analisis Data

Semua data yang sudah dikumpulkan selanjutnya diproses melalui serangkaian tahapan komputasi secara sistematis untuk menghasilkan model estimasi yang optimal dan dapat dipertanggungjawabkan secara ilmiah.

Setiap tahapan dilakukan secara berurutan guna memastikan kualitas data yang digunakan serta ketepatan model yang dihasilkan dalam memprediksi umur menikah. Dengan adanya tahapan komputasi yang terstruktur, diharapkan hasil analisis yang diperoleh dapat merepresentasikan kondisi data secara akurat dan konsisten.



Sumber: Hasil Penelitian (2026)

Gambar 1. Alur Tahapan Analisis Data

Berdasarkan gambar 1 diagram alur tahapan penelitian, mengidentifikasi proses yang mencakup beberapa tahap utama, yaitu pra-pemrosesan data, uji asumsi klasik, pemodelan regresi linier berganda & *backward elimination*, pembentukan model akhir, hingga evaluasi dan interpretasi.

2.5.1 Pra-pemrosesan Data

Tahap pra-pemrosesan mencakup empat langkah utama. Pertama, pembersihan data melalui menghapus baris yang memiliki nilai kosong (*missing value*). Kedua, penanganan outlier pada variabel umur menikah dengan membatasi nilai antara 17–40 tahun, mengacu pada batas minimal usia pernikahan menurut UU No. 16 Tahun 2019 dan distribusi data aktual yang menunjukkan masih terdapat cukup banyak responden yang menikah pada rentang usia tersebut [17]. Ketiga, transformasi logaritmik pada variabel pendapatan untuk menormalkan distribusi yang bersifat sangat miring (*skewed*). Keempat, kodifikasi variabel kategoris (*dusun*, pekerjaan, dan pendidikan) menggunakan *one-hot encoding* untuk menghindari *dummy variable trap*.

2.5.2 Uji Asumsi Klasik

Untuk memastikan model yang dihasilkan bersifat BLUE (*Best Linear Unbiased Estimator*), yaitu kondisi di mana model regresi memiliki estimasi yang tidak bias, efisien, serta merupakan model terbaik yang mendekati kenyataan [18]. Karenanya, dilakukan serangkaian uji prasyarat sebagai berikut:

Uji Multikolinearitas

Mengevaluasi apakah ada hubungan linear yang terlalu kuat antar variabel independen [19]. Hal ini diukur menggunakan nilai Variance Inflation Factor (VIF) dengan ketentuan $VIF < 10$.

$$VIF = \frac{1}{1-R_j^2} \quad (2)$$

Uji Autokorelasi

Uji autokorelasi bertujuan guna memastikan tidak ada korelasi antara residual pada satu observasi dengan observasi lainnya [20]. Pengujian dilakukan menggunakan uji Durbin-Watson dengan statistik uji (d) yang bernilai antara 0 hingga 4, di mana nilai d yang ada di rentang 1,5 hingga 2,5 mengindikasikan tidak adanya autokorelasi pada model.

Uji Normalitas

Memastikan sisaan (residual) dari model regresi berdistribusi normal menerapkan uji Shapiro-Wilk, yaitu uji statistik yang mengukur seberapa baik data mengikuti distribusi normal melalui membandingkan kuantil data pada kuantil distribusi normal teoritis [21]. Kriteria keputusan yang diterapkan ialah apabila nilai p-value $> 0,05$ maka residual disebut berdistribusi normal.

Uji Homokedastisitas

Uji Homokedastisita bertujuan guna memastikan bahwasanya varians residual bersifat konstan di semua nilai variabel independen [20]. Pengujian dilakukan menggunakan uji Breusch-Pagan, yaitu uji statistik yang mendeteksi heteroskedastisitas dengan cara meregresikan kuadrat residual terhadap variabel independen dalam model [22]. Kriteria keputusan yang diterapkan ialah apabila nilai p-value $> 0,05$ maka model dinyatakan memenuhi asumsi homokedastisitas.

2.5.3 Pemodelan dengan Backward Elimination

Metode backward elimination dipilih karena pendekatannya yang dimulai dari model lengkap dinilai lebih sesuai untuk penelitian eksploratoris seperti ini, di mana seluruh variabel demografis yang tersedia dipertimbangkan terlebih dahulu sebelum disederhanakan. Dibandingkan metode forward selection yang membangun model dari nol, backward elimination lebih efisien dalam mendeteksi variabel yang kehilangan signifikansinya ketika dimodelkan bersama variabel lain secara simultan [23].

Sebelum pemodelan, dataset dibagi menjadi data latih (training data) sebesar 80% serta data uji (testing data) sejumlah 20% menggunakan random state = 42 guna memastikan hasil pembagian data bisa direproduksi secara konsisten. Pembagian ini bertujuan agar evaluasi model dilakukan pada data yang tak pernah terlihat oleh model sebelumnya, sehingga hasil evaluasi menggambarkan kemampuan generalisasi model secara objektif.

Model awal dibangun dengan melibatkan seluruh variabel independen dengan persamaan:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \epsilon \quad (3)$$

Penyederhanaan model dilaksanakan menggunakan metode eliminasi mundur (backward elimination). Metode ini bekerja dengan cara memasukkan seluruh variabel ke dalam model, lalu secara bertahap menghapus variabel yang tidak memiliki pengaruh nyata (memiliki nilai p-value $> 0,05$) [14]. Proses iterasi ini berakhir ketika seluruh variabel yang tersisa di dalam model benar-benar memiliki kontribusi signifikan terhadap variabel dependen (Y).

2.5.4 Evaluasi Kelayakan Model

Kelayakan model diukur melalui Uji F (simultan), Uji t (parsial), serta Koefisien Determinasi (R^2). Kinerja prediktif model dievaluasi menggunakan RMSE dan MAE sebagai metrik kesalahan, di mana RMSE memberikan bobot lebih besar pada kesalahan yang besar sementara MAE menilai rata-rata kesalahan secara absolut. Selain itu, dilakukan 5-fold cross validation untuk memastikan model tidak overfitting terhadap data latih, dengan rata-rata RMSE dari kelima fold sebagai indikator kemampuan generalisasi model.

2.6 Spesifikasi Perangkat

Proses pengolahan data diimplementasikan menggunakan bahasa pemrograman Python 3.x pada lingkungan komputasi awan Google Colaboratory (Colab). Pustaka utama yang digunakan adalah Pandas untuk manajemen dan pembersihan dataset, NumPy untuk operasi matematis termasuk transformasi logaritmik, Scikit-learn untuk implementasi model regresi linier

berganda, pembagian data (train-test split), dan evaluasi model (RMSE, MAE, cross validation), serta Statsmodels untuk pengujian statistik uji asumsi klasik.

3. Hasil dan Pembahasan

3.1 Deskripsi Data Penelitian

Populasi penelitian ini ialah seluruh penduduk yang tercatat dalam basis data kependudukan Desa Pringtulis sebanyak 1.000 orang. Teknik pengambilan sampel menerapkan purposive sampling dengan kriteria responden telah berstatus menikah (status_kawin = 2), memiliki data demografis yang lengkap, serta berasal dari Dusun Pringtulis, Rekesan, dan Gajihan. Berdasarkan proses seleksi tersebut diperoleh 359 data yang memenuhi kriteria dan digunakan sebagai sampel penelitian. Data yang dianalisis mencakup sejumlah variabel yang dikategorikan ke dalam variabel dependen dan variabel independen. Variabel dependen dalam kajian ini ialah umur menikah, yang didefinisikan sebagai usia kronologis individu (dalam satuan tahun) pada saat akad nikah pertama kali dilaksanakan. Dataset ini mencakup rentang umur menikah yang bervariasi dari kategori pernikahan usia muda hingga usia dewasa matang, sehingga memungkinkan model regresi untuk menangkap pola distribusi umur menikah yang ada di Desa Pringtulis secara representatif. Sementara itu, variabel independen yang digunakan meliputi jenis kelamin (sex), pendidikan individu (pendidikan_kk_id), pekerjaan (pekerjaan_id), perkiraan pendapatan, golongan ekonomi, serta wilayah dusun. Variabel-variabel tersebut dipilih untuk menganalisis faktor-faktor yang berpotensi memengaruhi umur menikah pada responden.

Sebelum dilakukan analisis, data terlebih dulu melewati tahap pra-pemrosesan (preprocessing), termasuk pengecekan nilai kosong (missing value), serta transformasi variabel kategorik menjadi bentuk numerik menggunakan teknik encoding. Hasil pengecekan memperlihatkan bahwasanya tak ada nilai kosong pada seluruh variabel yang digunakan, sehingga data dapat langsung digunakan dalam proses pemodelan.

Selain itu, dilakukan pula pemisahan data menjadi data latih (training data) serta data uji (testing data) melalui perbandingan tertentu guna mengevaluasi kinerja model secara objektif. Secara umum, dataset ini memberikan gambaran mengenai karakteristik sosial, ekonomi, dan demografis individu yang telah menikah, sehingga memungkinkan dilakukan analisis lebih lanjut terkait hubungan antara variabel-variabel independen dengan umur menikah sebagai variabel utama pada kajian ini.

3.2 Hasil Analisis Regresi Linear

Hasil analisis regresi linier diterapkan guna mengetahui pengaruh variabel independen terhadap umur menikah sebagai variabel dependen. Ringkasan hasil analisis regresi linier ada pada Tabel 3.

Tabel 3. Tabel Koefisien Regresi Linear Variabel Prediktor Umur Menikah

Variabel	Koefisien
Intercept	27,57
Sex	-1,45
Perkiraan Pendapatan	0,35
Golongan Ekonomi	0,06
Dusun Pringtulis	-1,46
Dusun Rekesan	-2,31
pekerjaan_id_2	5,27
pekerjaan_id_3	1,58
pekerjaan_id_4	-3,52
pekerjaan_id_5	-2,28
pekerjaan_id_7	5,29
pekerjaan_id_8	0,61
pekerjaan_id_9	2,13
pekerjaan_id_10	-2,00
pekerjaan_id_12	-0,53
pekerjaan_id_13	0,01
pekerjaan_id_14	3,43
pekerjaan_id_15	-1,31
pekerjaan_id_19	1,65
pekerjaan_id_20	1,32
pekerjaan_id_26	-1,88
pekerjaan_id_27	-2,58
pekerjaan_id_30	1,25
pekerjaan_id_34	0,54

Variabel	Koefisien
pekerjaan_id_65	-1.46
pekerjaan_id_73	0.17
pekerjaan_id_84	1.79
pekerjaan_id_85	-1.18
pekerjaan_id_88	-1.39
pendidikan_kk_id_2	-2.82
pendidikan_kk_id_3	-1.46
pendidikan_kk_id_4	-3.57
pendidikan_kk_id_5	-2.53
pendidikan_kk_id_6	-2.57
pendidikan_kk_id_7	-2.82
pendidikan_kk_id_8	-1.48
pendidikan_kk_id_9	-0.63

Sumber: Hasil Penelitian (2026)

Berlandaskan hasil analisis regresi linear yang sudah dilaksanakan, didapatkan nilai koefisien untuk setiap variabel independen terhadap umur menikah. Variabel jenis kelamin (sex) memiliki koefisien sebesar -1,45 yang menunjukkan adanya perbedaan kecenderungan umur menikah berdasarkan jenis kelamin. Seluruh kategori pendidikan memiliki koefisien negatif terhadap kategori referensi (tidak sekolah). Namun, perbedaan nilai koefisien antar jenjang pendidikan menunjukkan bahwasanya tingkat pendidikan tetap berkorelasi dengan variasi umur menikah. Temuan ini mengindikasikan bahwa pendidikan merupakan salah satu faktor yang berperan dalam membentuk pola usia pernikahan individu. Variabel pekerjaan menunjukkan variasi koefisien antar kategori, yang mengindikasikan bahwa jenis pekerjaan tertentu memiliki pengaruh yang berbeda terhadap umur menikah. Variabel perkiraan pendapatan memiliki koefisien sebesar 0,35 yang menunjukkan adanya kecenderungan bahwa peningkatan pendapatan diikuti dengan peningkatan umur menikah. Sementara itu, variabel golongan ekonomi memiliki koefisien sebesar 0,06 yang menunjukkan pengaruh yang relatif kecil dalam model. Untuk variabel wilayah, digunakan variabel dummy dengan dusun Gajihan sebagai kategori referensi (baseline), sehingga koefisien yang dihasilkan untuk dusun lain merupakan perbandingan terhadap dusun tersebut. Nilai koefisien dusun Pringtulis sebesar -1,46 dan dusun Rekesan sebesar -2,31 menunjukkan bahwa rata-rata umur menikah di kedua dusun tersebut cenderung lebih rendah masing-masing sekitar 1–2 tahun dibandingkan dengan dusun Gajihan sebagai acuan.

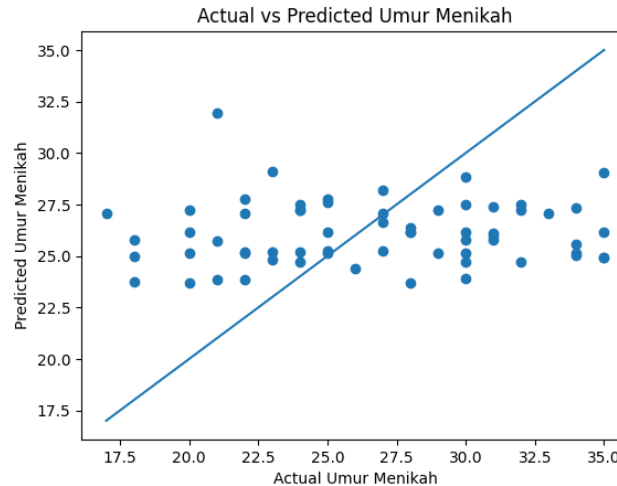
Hasil evaluasi model memperlihatkan nilai Mean Squared Error (MSE) sebesar 26,94, Root Mean Squared Error (RMSE) sebesar 5,19, serta Mean Absolute Error (MAE) sebesar 4,39. Nilai RMSE menunjukkan bahwa rata-rata kesalahan prediksi model berada pada kisaran ± 5 tahun, sedangkan nilai MAE memperlihatkan bahwasanya rata-rata selisih absolut antara nilai prediksi dan nilai aktual sebesar 4,39 tahun. Dalam menguji konsistensi model, dilaksanakan validasi silang menggunakan metode 5-Fold Cross Validation. Hasil pengujian memperlihatkan rata-rata Cross-Validation RMSE sejumlah 5,17 dengan standar deviasi sebesar 0,84. Nilai tersebut relatif mendekati RMSE pada data uji, yang menandakan bahwasanya performa model cukup konsisten pada berbagai pembagian data. Secara umum, hasil evaluasi menandakan bahwa model regresi linear mampu memberikan estimasi umur menikah dengan rata-rata selisih prediksi sekitar 4–5 tahun dari usia menikah yang sebenarnya. Meskipun demikian, masih ada peluang agar meningkatkan akurasi model melalui penambahan variabel lain yang lebih relevan dengan umur menikah.

3.3 Pembahasan Hasil

Hasil analisis menunjukkan bahwa tingkat pendidikan memiliki keterkaitan dengan variasi umur menikah. Perbedaan koefisien pada setiap kategori pendidikan mengindikasikan bahwa pendidikan ialah susunan faktor yang berperan dalam membentuk pola usia pernikahan individu. Temuannya ini selaras dengan berbagai penelitian terdahulu yang menyebutkan bahwasanya pendidikan dapat memengaruhi keputusan seseorang dalam menentukan waktu pernikahan melalui peningkatan pengetahuan, kesempatan kerja, serta pertimbangan kesiapan ekonomi. Variabel golongan ekonomi memiliki pengaruh positif yang relatif kecil terhadap umur menikah, sebagaimana tercermin dari nilai koefisien sebesar 0,06. Hal ini menandakan bahwa peningkatan golongan ekonomi cenderung diikuti oleh peningkatan umur menikah, meskipun pengaruhnya dalam model tidak terlalu kuat. Selain itu, perbedaan wilayah dusun juga menunjukkan adanya variasi umur menikah yang kemungkinan dipengaruhi oleh perbedaan norma sosial, tingkat pendidikan, serta akses terhadap informasi di masing-masing wilayah. Meskipun model mampu

menggambarkan hubungan antarvariabel dengan cukup baik, nilai kesalahan prediksi yang masih relatif tinggi menunjukkan bahwa umur menikah tidak hanya dipengaruhi oleh faktor-faktor yang terdapat dalam model, tetapi juga oleh faktor lain seperti budaya, lingkungan keluarga, dan nilai sosial yang belum dimasukkan dalam penelitian ini.

3.4 Visualisasi Hasil



Sumber: Hasil Penelitian (2026)

Gambar 2. *Scatter Plot* Perbandingan Nilai Aktual dan Prediksi Umur Menikah

Visualisasi hasil dilaksanakan menggunakan scatter plot yang memperlihatkan perbandingan antara nilai aktual serta nilai prediksi umur menikah yang dihasilkan oleh model regresi linear. Pada grafik tersebut, sumbu horizontal memperlihatkan nilai aktual umur menikah, sedangkan sumbu vertikal menunjukkan nilai prediksi model. Garis diagonal yang ditampilkan merepresentasikan kondisi prediksi sempurna, di mana nilai prediksi sama dengan nilai aktual. Berdasarkan grafik yang diperoleh, terlihat bahwasanya mayoritas titik data tersebar di sekitar garis diagonal, yang menandakan bahwa model telah mampu mengikuti pola umum dari data yang dianalisis. Hal ini menandakan bahwasanya model regresi linear yang digunakan memiliki kemampuan yang cukup baik dalam merepresentasikan hubungan antara variabel independen dengan umur menikah sebagai variabel dependen.

Meskipun demikian, penyebaran titik yang masih terlihat cukup bervariasi menunjukkan bahwa adanya selisih antara nilai prediksi dan nilai aktual pada beberapa observasi. Beberapa titik yang berada cukup jauh dari garis diagonal mengindikasikan adanya ketidaktepatan prediksi pada data tertentu, yang dapat dipengaruhi dari kompleksitas faktor-faktor yang tidak sepenuhnya terakomodasi dalam model. Namun secara keseluruhan, pola sebaran titik yang masih mengikuti arah garis diagonal menandakan bahwasanya model memiliki tingkat konsistensi yang baik dalam menangkap kecenderungan data. Maka, visualisasi ini menunjang hasil evaluasi model sebelumnya dan memperkuat bahwasanya model yang dibangun dapat digunakan untuk memberikan gambaran umum terkait prediksi umur menikah berdasarkan variabel yang diteliti.

4. Kesimpulan

Berlandaskan hasil analisis serta pembahasan, kesimpulannya faktor-faktor demografis, yaitu tingkat pendidikan, jenis kelamin, pekerjaan, pendapatan, golongan ekonomi, dan wilayah tempat tinggal digunakan sebagai variabel prediktor dalam model estimasi umur menikah di Desa Pringtulis. Hasil regresi linear menandakan bahwa setiap variabel memiliki hubungan yang berbeda terhadap umur menikah, yang tercermin dari variasi nilai koefisien pada model. Evaluasi model menghasilkan nilai RMSE sebesar 5,19, MAE sebesar 4,39, dan rata-rata Cross-Validation RMSE sebesar 5,17, yang menunjukkan bahwa model memiliki rata-rata kesalahan prediksi pada kisaran 4–5 tahun. Secara keseluruhan, model regresi linear mampu memberikan estimasi umur menikah dengan cukup baik, meskipun akurasi masih dapat ditingkatkan karena terdapat faktor-faktor lain di luar model, seperti budaya, nilai sosial, lingkungan keluarga, dan faktor personal, yang juga berpotensi memengaruhi umur menikah. Maka demikian, penelitian berikutnya disarankan agar menambahkan variabel-variabel tersebut serta mempertimbangkan penggunaan metode regresi non-linear maupun algoritma machine learning lainnya agar

diperoleh hasil prediksi yang lebih optimal. Sementara itu, hasil penelitian ini diharapkan bisa dijadikan suatu sumber informasi untuk pemerintah desa serta pemangku kebijakan dalam menyusun program edukasi mengenai perencanaan usia pernikahan melalui peningkatan akses pendidikan dan pemberdayaan ekonomi masyarakat, sekaligus mendorong masyarakat untuk mempertimbangkan aspek pendidikan dan kesiapan ekonomi dalam menentukan waktu pernikahan guna mewujudkan keluarga yang lebih sejahtera dan berkualitas.

Referensi

- [1] A. Aswad, M. Adilb, And S. Wijaya, "Usia Pernikahan Di Era Modern: Menelusuri Kebijakan Undang-Undang Nomor 16 Tahun 2019 Dan Bkkbn Melalui Perspektif Mashla H Ah Mursalah," *J. Ilmu Syariah Dan Huk.*, Vol. 3, No. 2, Pp. 158–173, 2024.
- [2] A. Angraini, R. Ibrahim, And S. A. Yusuf, "Dampak Pernikahan Dini Terhadap Kejadian Berat Bayi Lahir Rendah Di Rsd Kota Kendari," *J. Pelita Sains Kesehat.*, Vol. 3, No. 3, Pp. 23–28, May 2023.
- [3] S. Khairunnisa And N. Nurwati, "Pengaruh Pernikahan Pada Usia Dini Terhadap Peluang Bonus Demografi Tahun 2030," *J. Ilmu Kesejaht. Sos. Humanit.*, Vol. 3, No. 1, Pp. 45–59, 2021.
- [4] R. A. B. Leo, M. H. Hasan, And Sukmawati, "Pengaruh Tingkat Pendidikan, Pendapatan Keluarga Dan Usia Kawin Pertama Terhadap Jumlah Kelahiran Di Kelurahan Ledeunu Kecamatan Rajua Kabupaten Sabu Rajua," *J. Geogr.*, Vol. 19, No. 1, Pp. 91–103, 2023.
- [5] D. E. Pangestika, D. A. Mustika, And A. A. Rahman, "Perbandingan Performa Metode Klasifikasi Regresi Logistik, Classification Tree, Dan Random Forest: Studi Kasus Perkawinan Anak Pada Perempuan Usia Muda Di Provinsi Nusa Tenggara Barat Tahun 2022," *Semin. Nas. Off. Stat.*, Vol. 2025, No. 1, Pp. 499–510, 2024.
- [6] A. Hermambang, C. Ummah, E. S. Gratia, F. Sanusi, W. M. Ulfa, And R. Nooraeni, "Faktor-Faktor Yang Memengaruhi Pernikahan Usia Dini Di Indonesia Factors Affecting Early Marriage In Indonesia," Vol. 16, No. 1, Pp. 1–12, 2021, Doi: 10.14203/Jki.V16i1.428.
- [7] A. Nafisah And P. Perdana, "Analisis Tingkat Pendidikan Dan Pertumbuhan Ekonomi Terhadap Tingkat Pernikahan Dini," *J. Ilm. Wahana Pendidik.*, Vol. 10, No. 19, Pp. 181–189, Oct. 2024, Doi: 10.5281/Zenodo.14043130.
- [8] S. N. Razania And S. Muhtadin, "Pengaruh Media Sosial Terhadap Penurunan Minat Menikah Di Kalangan Generasi Muda (Studi Kasus Pengguna Aplikasi Media Sosial Tiktok)," *Hujjah J. Ilm. Komun. Dan Penyiaran Islam*, Vol. 9, No. 1, Pp. 123–136, Dec. 2025.
- [9] D. Rahmania, R. Amelia, And I. Sulistiana, "Pemodelan Faktor-Faktor Yang Memengaruhi Pernikahan Dini Di Provinsi Kepulauan Bangka Belitung Menggunakan Regresi Poisson," *J. Ilm. Mat.*, Vol. 13, No. 03, Pp. 165–171, 2025.
- [10] E. Maslukha And M. B. Firmansyah, "Faktor Penentu Pernikahan Dini Dan Dampaknya Terhadap Kelanjutan Pendidikan Dalam Era Wajib Belajar 12 Tahun," *Jiip (Jurnal Ilm. Ilmu Pendidikan)*, Vol. 9, No. 1, Pp. 246–250, 2026.
- [11] N. S. Purwitasari, "Analisis Faktor Penyebab Perceraian Pada Pernikahan Di Bawah Umur (Literature Review)," Vol. 6, No. September, Pp. 9679–9687, 2025.
- [12] A. Febrianto And R. A. Siroj, "Studi Literatur: Landasan Dalam Memilih Metode Penelitian Yang Tepat," *J. Educ. Res. Dev. E-Issn 3063-9158*, Vol. 1, No. 2, Pp. 259–263, 2024.
- [13] I. S. Dita Rahmania, Ririn Amelia, "Pemodelan Faktor-Faktor Yang Memengaruhi Pernikahan Dini Di Provinsi Kepulauan Bangka Belitung Menggunakan Regresi Poisson," *J. Ilm. Mat.*, Vol. 13, No. 2, Pp. 198–199, 2025, [Online]. Available: <https://ejournal.unesa.ac.id/index.php/mathunesa/article/view/64649>
- [14] M. Anjas Aprihartha, S. Paramita Azzahro, And D. Rahmatul Aziza, "Pemilihan Model Regresi Linear Berganda Terbaik Untuk Menentukan Faktor-Faktor Penyebab Kasus Balita Gizi Buruk Di Jawa Tengah," *J. Eurekamatika*, Vol. 13, No. 1, Pp. 1–12, 2025.
- [15] W. C. Ferdinan, M. R. Noerfikri, P. A. Panchadri, And F. Ferawati, "Implementasi Algoritma Regresi Linear Berganda Untuk Memprediksi Prestasi Siswa," *Bit-Tech*, Vol. 7, No. 3, Pp. 853–864, 2025.
- [16] U. P. Ramadani, R. Muthmainnah, N. Ulhilma, A. Wazabirah, R. Hidayatullah, And Harmonedi, "Strategi Penentuan Populasi Dan Sampel Dalam Penelitian Pendidikan : Antara Validitas Dan Representativitas," *Urnal Pendidikan, Sos. Hum. Qosim J.*

- Pendidikan, Sos. Hum.*, Vol. 3, No. 2, Pp. 574–585, 2025.
- [17] D. W. I. F. Saiman, "Pengaruh Faktor Ekonomi, Orang Tua Dan Adat Istiadat Terhadap Pernikahan Usia Dini Di Kecamatan Belopa Kabupaten Luwu," 2023, *Institut Agama Islam Negeri (lain) Palopo*.
- [18] L. Winarti, R. Permadi, And I. Mahfud, "Analysis Of Factors That Influence Production Of Rice Farming Variety Of Siam Epang At Sub District Of Seruyan Hilir," *J. Agribest*, Vol. 3, No. 2, Pp. 105–115, 2019.
- [19] M. D. Irrawati And M. Mukaramah, "Implementasi Metode Regresi Linear Berganda Untuk Mengatasi Pelanggaran Asumsi Klasik," *Stud. Akuntansi, Keuangan, Dan Manaj.*, Vol. 3, No. 2, Pp. 83–94, 2024.
- [20] A. Saputra And M. A. Samsuddin, "Pengaruh Tingkat Kemiskinan Dan Harapan Lama Sekolah Terhadap Pdrb Di Provinsi Bengkulu," *J. Ilmu Ekon. Pendidik. Dan Tek.*, Vol. 2, No. 3, Pp. 91–96, 2025.
- [21] R. Sianturi, "Uji Normalitas Sebagai Syarat Pengujian Hipotesis," *J. Pembelajaran Dan Mat. Sigma*, Vol. 11, No. 1, Pp. 1–14, 2025.
- [22] P. Y. Haliza, R. Rafiza, And J. Simanullang, "Penerapan Regresi Linier Berganda Dalam Memprediksi Ipm Berdasarkan Faktor Ekonomi Dan Sosial Di Sumatera Barat," *Socius J. Penelit. Ilmu-Ilmu Sos.*, Vol. 2, No. 12, 2025.
- [23] Y. D. Khoerunnisa, "Perbandingan Metode Backward Elimination, Forward Selection dan Stepwise Regression dalam Pemilihan Model Regresi Linear Berganda Terbaik," Skripsi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung, Bandar Lampung, Indonesia, 2021. [Online]. Tersedia: digilib.unilla.ac.id (diakses: 13 Juni 2026).