

# Prediksi Kelulusan Siswa dengan Algoritma Pembelajaran Mesin: Aplikasi Regresi Linear dan Logistik pada Faktor-Faktor Pendidikan

Fenni Aprilia<sup>1</sup>, Rasti Aulia Anggraini<sup>2</sup>, Yunita Dwi Putri<sup>3</sup>  
<sup>1,2,3</sup>Politeknik Negeri Lampung, Teknologi Rekayasa Internet

## INFORMASI ARTIKEL

Diterima 27 Desember 2024  
Direvisi 27 Februari 2025  
Diterbitkan 27 Februari 2025

### Kata kunci:

Regresi linier;  
Regresi logistik;  
Prediksi kelulusan siswa;  
Algoritma pembelajaran;

## ABSTRAK

Permasalahan utama dalam pendidikan adalah memprediksi kelulusan siswa secara akurat dan memahami faktor-faktor yang memengaruhinya. Penelitian ini bertujuan untuk memberikan solusi berbasis data menggunakan algoritma pembelajaran mesin, yaitu regresi linear untuk memprediksi nilai ujian siswa dan regresi logistik untuk mengklasifikasikan kelulusan siswa. Kontribusi penelitian ini meliputi pengembangan model prediksi yang dapat digunakan sebagai alat untuk mendukung pengambilan keputusan strategis di institusi pendidikan. Penelitian ini menggunakan dataset Student Performance Factors yang mencakup 6.607 sampel dengan variabel independen seperti jam belajar, kehadiran, dan keterlibatan orang tua. Data dianalisis melalui tahapan pembersihan, transformasi, dan normalisasi sebelum diterapkan ke model regresi. Hasil analisis menunjukkan regresi linear menghasilkan Mean Squared Error (MSE) sebesar 3.256, yang mencerminkan tingkat akurasi yang baik dalam memprediksi nilai ujian. Sementara itu, regresi logistik memiliki tingkat akurasi 99,85% dalam mengklasifikasikan kelulusan siswa. Kedua model ini saling melengkapi dalam memberikan wawasan strategis untuk meningkatkan kualitas pendidikan.

# Student Graduation Prediction Using Machine Learning Algorithms: Application of Linear and Logistic Regression on Educational Factors

## ARTICLE INFO

Received December 27, 2024  
Revised February 27, 2025  
Published February 27, 2025

### Keyword:

Linear regression;  
Logistic regression;  
Student graduation prediction;  
Learning algorithm;

## ABSTRACT

The main challenge in education lies in accurately predicting student graduation and understanding the factors influencing it. This study aims to provide a data-driven solution using machine learning algorithms, specifically linear regression to predict student exam scores and logistic regression to classify student graduation. The study contributes by developing predictive models that serve as tools to support strategic decision-making in educational institutions. This study utilized the Student Performance Factors dataset, comprising 6,607 samples with independent variables such as study hours, attendance, and parental involvement. Data analysis involved cleaning, transformation, and normalization before applying regression models. The findings showed that linear regression achieved a Mean Squared Error (MSE) of 3.256, indicating high accuracy in predicting exam scores. Logistic regression demonstrated an accuracy of 99.85% in classifying student graduation. These models complement each other by offering strategic insights to enhance educational quality.

This work is licensed under a [Creative Commons Attribution 4.0](https://creativecommons.org/licenses/by/4.0/)



## Corresponding Author:

Fenni Aprilia, Politeknik Negeri Lampung  
Email: [esa71696@gmail.com](mailto:esa71696@gmail.com)

## 1. PENDAHULUAN

Kemajuan teknologi informasi di era digital telah menghadirkan peluang besar dalam pengolahan dan analisis data, termasuk di bidang pendidikan. Dalam konteks pendidikan, pengambilan keputusan yang berbasis data menjadi semakin penting untuk memahami kebutuhan siswa, merancang strategi pembelajaran, dan meningkatkan kualitas hasil pendidikan [1]. Salah satu tantangan utama yang dihadapi oleh institusi pendidikan adalah mengidentifikasi faktor-faktor yang memengaruhi kelulusan siswa serta memprediksi hasil belajar secara akurat. Permasalahan ini menjadi krusial mengingat kelulusan siswa tidak hanya mencerminkan keberhasilan individu, tetapi juga mencerminkan efektivitas sistem pendidikan secara keseluruhan [2].

Pemanfaatan algoritma pembelajaran mesin, seperti regresi linear dan logistik, memberikan peluang untuk menganalisis data pendidikan secara lebih mendalam. Kedua algoritma ini memungkinkan pemrosesan hubungan kompleks antar variabel yang berkontribusi terhadap performa akademik siswa. Faktor-faktor seperti jumlah jam belajar, tingkat kehadiran, keterlibatan orang tua, akses terhadap sumber daya pendidikan, dan skor ujian sebelumnya dapat dijadikan prediktor dalam memodelkan hasil belajar siswa. Dengan menerapkan analisis berbasis data ini, pendidik dan pembuat kebijakan dapat memperoleh wawasan yang lebih akurat dan strategis untuk meningkatkan mutu pendidikan [3].

Penelitian ini bertujuan untuk menerapkan regresi linear dalam memprediksi skor ujian siswa secara kuantitatif dan regresi logistik untuk memprediksi kelulusan siswa dalam bentuk klasifikasi biner (lulus atau tidak lulus). Dataset yang digunakan dalam penelitian ini, yaitu Student

Performance Factors, mencakup berbagai indikator penting yang relevan dalam analisis performa akademik. Melalui pendekatan ini, penelitian ini tidak hanya menyajikan solusi untuk memprediksi hasil belajar siswa, tetapi juga memberikan landasan bagi pengembangan sistem pendidikan yang lebih adaptif dan berbasis data. Dengan memahami hubungan antara berbagai faktor pendidikan dan hasil belajar, penelitian ini diharapkan dapat mendukung pengambilan keputusan yang lebih efektif dalam meningkatkan tingkat kelulusan siswa secara keseluruhan [3].

## 2. METODE

Metode yang digunakan dalam penelitian ini bertujuan untuk memberikan solusi berbasis data yang mampu mendukung pengembangan sistem pendidikan yang lebih adaptif dan terukur. Dengan menerapkan regresi linear dan logistik, penelitian ini memberikan pendekatan yang tidak hanya memprediksi hasil belajar siswa secara akurat tetapi juga memberikan landasan bagi pengambilan keputusan strategis untuk meningkatkan tingkat kelulusan siswa. Adapun langkah-langkah yang dilakukan dalam penelitian ini dijelaskan secara rinci sebagai berikut [4].

### 2.1. Pengumpulan Data

Tahap pertama dalam penelitian ini adalah pengumpulan data yang relevan dengan performa akademik siswa. Dalam penelitian ini, data yang digunakan berasal dari Kaggle (<https://www.kaggle.com/datasets/lainguyn123/student-performance-factors>), yaitu dataset Student Performance Factors, yang terdiri dari 6.607 sampel yang diakses pada tanggal 10 Desember 2024. Dataset ini mencakup berbagai variabel yang relevan dengan performa akademik siswa, termasuk jumlah jam belajar per minggu, tingkat kehadiran di kelas, keterlibatan orang tua dalam mendukung proses pembelajaran, akses siswa terhadap sumber daya pendidikan, dan skor ujian sebelumnya.

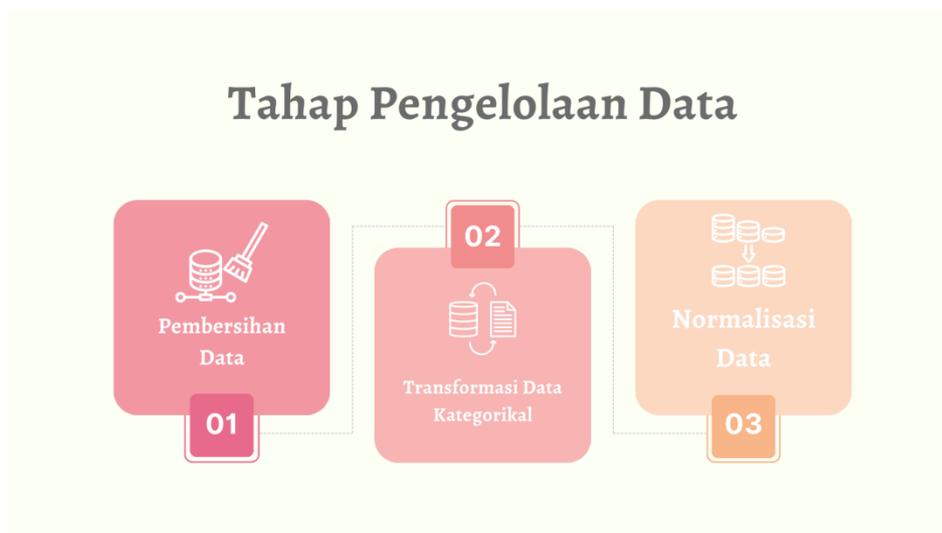
Proses pengumpulan data dilakukan melalui beberapa tahap. Pertama, identifikasi sumber data yang valid dilakukan untuk memastikan bahwa dataset yang digunakan memiliki cakupan yang cukup luas serta representatif terhadap populasi siswa. Kedua, dilakukan verifikasi terhadap kualitas data mentah untuk memastikan bahwa semua informasi yang terkandung dalam dataset bebas dari inkonsistensi atau kesalahan pengukuran. Ketiga, metadata dari dataset dianalisis untuk memahami deskripsi variabel, format data, serta distribusi awal yang akan menjadi dasar dalam langkah-langkah pengolahan data berikutnya.

Variabel-variabel yang terkandung dalam dataset diklasifikasikan menjadi dua kategori utama, yaitu variabel independen dan variabel dependen. Variabel independen meliputi indikator-indikator seperti tingkat kehadiran (dalam persentase), jumlah jam belajar (dalam jam per minggu), keterlibatan orang tua (diukur melalui skala survei), dan akses terhadap sumber daya pendidikan (Seperti akses internet atau materi belajar). Sementara itu, variabel dependen didefinisikan dalam dua bentuk sesuai dengan tujuan penelitian, yaitu nilai ujian untuk analisis regresi linear dan status kelulusan (lulus atau tidak lulus) untuk analisis regresi logistic [5].

Keberagaman dan kelengkapan variabel dalam dataset memungkinkan analisis yang komprehensif terhadap hubungan antara faktor-faktor pendidikan dengan performa akademik siswa. Dengan cakupan data yang luas, penelitian ini tidak hanya bertujuan untuk menghasilkan prediksi yang akurat, tetapi juga memberikan wawasan mengenai faktor-faktor utama yang paling berpengaruh terhadap kelulusan siswa. Pengumpulan data yang dilakukan dengan cermat menjadi fondasi penting untuk memastikan bahwa hasil analisis dapat digunakan sebagai dasar dalam pengambilan keputusan strategis di bidang pendidikan.

### 2.2. Pengolahan Data

Pengolahan data merupakan langkah krusial dalam memastikan kualitas dan kesiapan data untuk analisis regresi yang akurat. Gambar 1 menunjukkan alur proses pengolahan data yang mencakup tiga tahapan utama: pembersihan data, transformasi data kategorikal, dan normalisasi data. Setiap tahapan dilakukan untuk mengatasi potensi masalah pada data dan memastikan konsistensi dalam analisis.



Gambar 1. Tahap pengelolaan data

Tahap pertama adalah pembersihan data, yang bertujuan untuk mendeteksi dan menangani data kosong, duplikasi, serta anomali yang dapat memengaruhi hasil analisis. Data yang bermasalah diidentifikasi dan diperbaiki atau dihapus berdasarkan pedoman standar pengolahan data. Selanjutnya, transformasi data kategorikal dilakukan untuk mengubah variabel-variabel seperti jenis kelamin atau tipe sekolah ke dalam format numerik menggunakan metode *encoding*. Langkah ini penting untuk memastikan bahwa data dapat diproses oleh model regresi, mengingat variabel kategorikal tidak dapat langsung digunakan dalam analisis matematis [6].

Tahapan terakhir adalah normalisasi data, di mana semua fitur numerik dalam dataset disesuaikan skalanya agar seragam. Proses ini bertujuan untuk menghindari bias yang mungkin muncul akibat perbedaan skala antarvariabel, sehingga model regresi dapat bekerja dengan lebih optimal. Ketiga langkah ini memastikan bahwa data yang digunakan dalam penelitian tidak hanya berkualitas tinggi tetapi juga siap untuk mendukung analisis berbasis regresi yang efektif [5].

### 2.3. Penerapan Model Regresi Linear dan Logistik

Tahapan penerapan model regresi linear dan logistik merupakan inti dari penelitian ini, yang bertujuan untuk memberikan solusi prediktif berbasis data guna meningkatkan tingkat kelulusan siswa. Setiap model diterapkan untuk menjawab tujuan analisis yang berbeda, dengan pendekatan yang saling melengkapi dalam mengevaluasi performa akademik siswa [4].

Model regresi linear digunakan untuk memprediksi nilai ujian siswa berdasarkan berbagai faktor seperti jumlah jam belajar, keterlibatan orang tua dalam proses pendidikan, dan skor sebelumnya [5]. Pemilihan regresi linear didasarkan pada kemampuan memodelkannya untuk hubungan linier antara variabel independen dan variabel dependen. Model ini memberikan pemahaman kuantitatif mengenai seberapa besar kontribusi masing-masing faktor terhadap hasil ujian siswa. Sebagai contoh, model ini dapat mengungkap sejauh mana peningkatan jumlah jam belajar siswa dapat memengaruhi nilai ujian mereka. Dengan demikian, regresi linear menjadi alat yang sangat berguna dalam memberikan wawasan strategis bagi institusi pendidikan untuk memprioritaskan intervensi, seperti menambah jam bimbingan belajar atau meningkatkan keterlibatan orang tua.

Sementara itu, regresi logistik diterapkan untuk mengklasifikasikan status kelulusan siswa berdasarkan nilai ujian mereka. Dalam penelitian ini, kelulusan siswa didefinisikan dengan kriteria nilai ujian minimal 60. Model regresi logistik sangat cocok digunakan karena mampu menangani klasifikasi biner (lulus atau tidak lulus) dengan mempertimbangkan hubungan kompleks dan nonlinier antara variabel independen, seperti kualitas pengajaran atau tingkat motivasi siswa, terhadap hasil klasifikasi kelulusan. Hasil dari model ini memungkinkan institusi pendidikan

untuk secara dini mengidentifikasi siswa yang berpotensi tidak lulus, sehingga dapat segera diberikan pendampingan atau intervensi tambahan.

Penerapan kedua model ini memberikan wawasan yang saling melengkapi. Regresi linear membantu dalam memahami faktor-faktor penyebab yang memengaruhi hasil ujian secara kuantitatif, sementara regresi logistik memberikan kemampuan prediktif yang andal untuk mendukung pengambilan keputusan dalam sistem pendidikan. Pendekatan ini menegaskan pentingnya analisis berbasis data dalam mengembangkan kebijakan pendidikan yang lebih adaptif, efektif, dan terukur.

#### 2.4. Evaluasi Model

Tahap evaluasi model merupakan langkah penting untuk memastikan bahwa model yang dibangun memiliki akurasi dan validitas yang memadai dalam menghasilkan prediksi yang andal. Pada penelitian ini, evaluasi dilakukan secara terpisah untuk regresi linear dan regresi logistik, masing-masing dengan pendekatan yang relevan sesuai dengan karakteristik model.

Pada model regresi linear, kinerja diukur menggunakan metrik Mean Squared Error (MSE). Nilai MSE merepresentasikan rata-rata kesalahan antara nilai yang diprediksi oleh model dengan nilai aktual. Semakin kecil nilai MSE, semakin baik model dalam menangkap pola hubungan antara variabel independen dan variabel dependen. Selain itu, visualisasi perbandingan antara nilai aktual dan nilai prediksi disajikan untuk memberikan gambaran visual tentang akurasi model. Visualisasi ini membantu mengidentifikasi pola kesalahan atau anomali yang mungkin tidak terlihat hanya dari nilai MSE, sekaligus memperkuat interpretasi hasil analisis [7].

Evaluasi pada model regresi logistik dilakukan dengan menggunakan metrik akurasi dan Confusion Matrix. Akurasi mengukur proporsi prediksi model yang benar dibandingkan dengan keseluruhan data uji, memberikan indikasi seberapa baik model dalam mengklasifikasikan data. Matriks kebingungan digunakan untuk memberikan analisis yang lebih mendetail mengenai distribusi *true positive* dan *true negative* serta *false positive* dan *false negative*. Informasi ini sangat penting bagi institusi pendidikan, karena memungkinkan pengambil kebijakan untuk memahami pola prediksi dan merancang intervensi yang lebih tepat sasaran bagi siswa yang berpotensi tidak lulus.

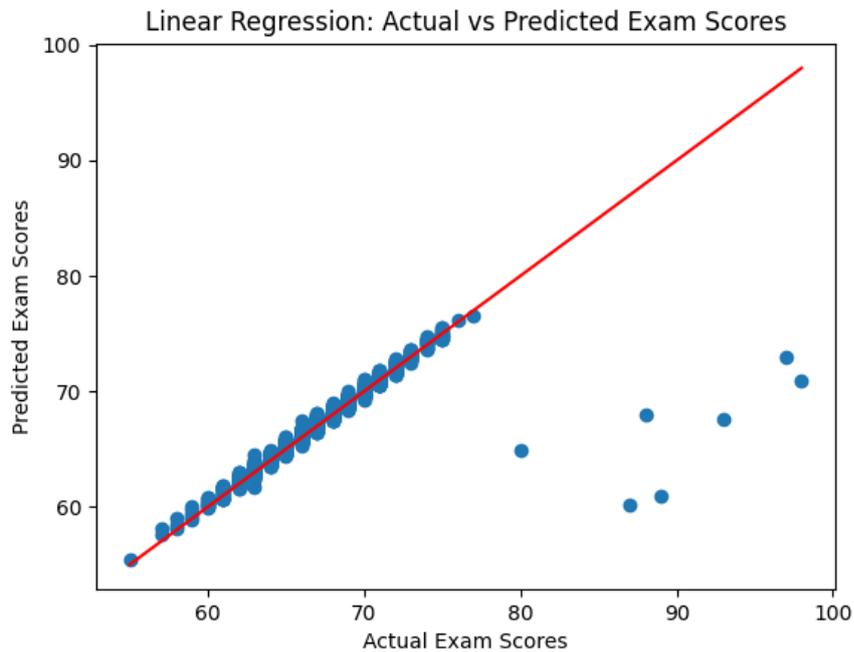
Proses evaluasi ini tidak hanya menilai performa model, tetapi juga memastikan bahwa model yang digunakan mampu memberikan wawasan yang mendukung pengambilan keputusan berbasis data. Dengan evaluasi yang menyeluruh, model regresi linear dan logistik dalam penelitian ini dapat dipercaya untuk menjadi solusi strategis dalam meningkatkan kualitas pendidikan dan tingkat kelulusan siswa [8].

### 3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil analisis yang diperoleh dari penerapan model regresi linear dan regresi logistik terhadap data yang digunakan. Hasil ini mencakup performa prediksi dari masing-masing model, serta visualisasi yang memberikan gambaran lebih mendalam terkait akurasi dan efektivitas model dalam memprediksi nilai dan klasifikasi kelulusan siswa. Selain itu, hasil analisis ini juga dibandingkan untuk mengetahui kelebihan dan keterbatasan masing-masing metode yang digunakan [9].

#### 3.1. Hasil Model Regresi Linear

Model regresi linear digunakan untuk memprediksi nilai ujian siswa berdasarkan variabel-variabel prediktor yang relevan, seperti jumlah jam belajar, kehadiran, dan faktor pendidikan lainnya [10]. Hasil evaluasi menunjukkan bahwa performa model regresi linear dapat diukur melalui nilai Mean Squared Error (MSE) sebesar 3.256, yang menunjukkan tingkat kesalahan prediksi relatif rendah. Meskipun demikian, masih terdapat perbedaan antara nilai aktual dan nilai prediksi pada beberapa data.

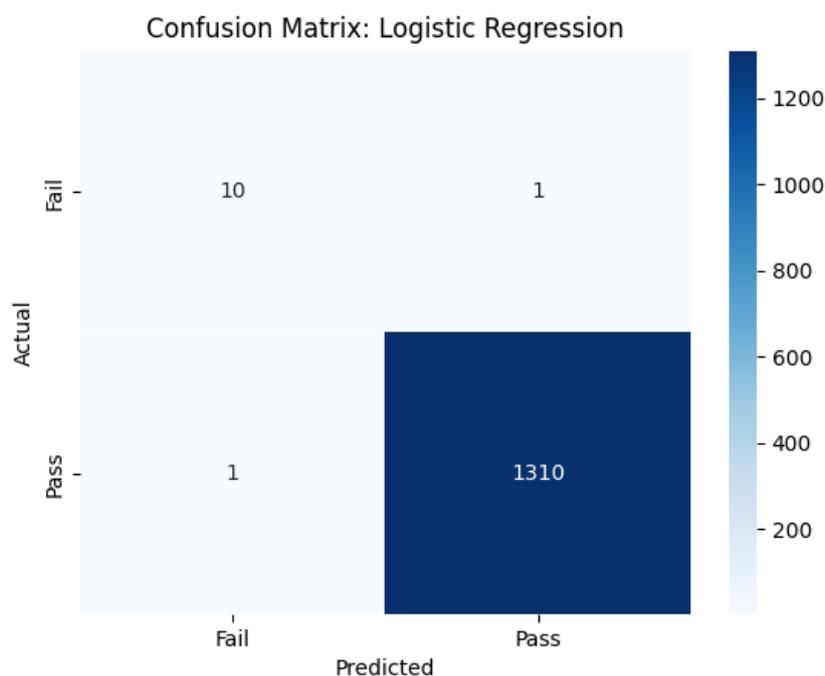


**Gambar 2.** Linear regression

Pada Gambar 2, ditampilkan *scatter plot* yang memperlihatkan perbandingan antara nilai aktual dan nilai prediksi hasil ujian siswa. Sumbu X merepresentasikan nilai aktual ujian siswa, sedangkan sumbu Y menggambarkan nilai prediksi yang dihasilkan model regresi linear. Garis merah diagonal pada grafik mewakili garis ideal, di mana nilai prediksi sama persis dengan nilai aktual. Sebagian besar titik data (ditunjukkan oleh titik-titik biru) tersebar di sekitar garis ideal, yang menandakan bahwa model memiliki tingkat akurasi yang baik dalam memprediksi hasil ujian.

### 3.2. Hasil Model Regresi Logistik

Model regresi logistik digunakan untuk mengklasifikasikan kelulusan siswa berdasarkan nilai ujian yang diperoleh [9]. Dalam penelitian ini, siswa diklasifikasikan ke dalam dua kategori, yaitu "lulus" (nilai  $\geq 60$ ) dan "tidak lulus" (nilai  $< 60$ ). Hasil evaluasi menunjukkan bahwa model regresi logistik memiliki performa yang sangat baik dengan tingkat akurasi mencapai 99.85%. Nilai akurasi ini diperoleh dari perbandingan antara prediksi model dan nilai aktual dalam data uji.



**Gambar 3.** Logistic regression

Pada Gambar 3, ditampilkan confusion matrix yang digunakan untuk mengevaluasi performa model klasifikasi regresi logistik. Matriks ini memperlihatkan distribusi prediksi model terhadap data uji. Terdapat 10 prediksi benar untuk kategori "tidak lulus" (*True Negative*), yang menunjukkan bahwa model dapat mengidentifikasi siswa yang tidak lulus dengan akurat. Selain itu, terdapat 1 prediksi salah untuk kategori "tidak lulus" yang diprediksi "lulus" (*False Positive*) dan 1 prediksi salah untuk kategori "lulus" yang diprediksi "tidak lulus" (*False Negative*). Sementara itu, 1.310 prediksi benar berhasil diklasifikasikan ke dalam kategori "lulus" (*True Positive*), yang menunjukkan dominasi prediksi yang akurat oleh model. Tingginya akurasi dan rendahnya jumlah kesalahan klasifikasi membuktikan bahwa regresi logistik mampu menangani data klasifikasi dengan sangat baik.

### 3.3. Perbandingan Kinerja Model

Hasil perbandingan kinerja model regresi linear dan regresi logistik menunjukkan keunggulan masing-masing model dalam konteks yang berbeda, sesuai dengan tujuan analisis yang ingin dicapai. Model regresi linear menunjukkan kemampuan yang kuat dalam memprediksi nilai ujian siswa secara kuantitatif dengan nilai Mean Squared Error (MSE) sebesar 3.256. Hasil ini menunjukkan bahwa kesalahan prediksi rata-rata relatif kecil, yang mencerminkan kemampuan model untuk memahami hubungan linier antara variabel-variabel prediktor, seperti jam belajar dan kehadiran, dengan hasil nilai ujian siswa. Selain itu, scatter plot pada Gambar 2 menunjukkan bahwa sebagian besar data prediksi mendekati garis ideal, meskipun ada beberapa titik data yang menunjukkan penyimpangan. Penyimpangan ini bisa disebabkan oleh pengaruh variabel yang tidak dimasukkan dalam model atau outlier dalam dataset.

Di sisi lain, model regresi logistik memiliki performa yang sangat baik dalam melakukan klasifikasi biner (lulus atau tidak lulus) dengan akurasi mencapai 99,85%. Tingkat akurasi yang tinggi menunjukkan bahwa model ini efektif dalam memisahkan dua kategori kelulusan berdasarkan nilai ujian siswa. Confusion Matrix pada Gambar 3 menunjukkan bahwa model ini mampu mengidentifikasi siswa yang lulus (*True Positive*) dan yang tidak lulus (*True Negative*)

dengan sangat akurat, meskipun terdapat beberapa kesalahan prediksi (*False Positive* dan *False Negative*) dalam jumlah yang kecil.

Secara umum, perbandingan antara regresi linear dan logistik menyoroti perbedaan fungsi masing-masing model dalam analisis data pendidikan. Regresi linear lebih efektif untuk menghasilkan prediksi numerik yang memberikan wawasan tentang kontribusi relatif setiap variabel terhadap hasil akhir siswa. Sebaliknya, regresi logistik unggul dalam memberikan keputusan klasifikasi yang tajam dan mudah digunakan untuk intervensi praktis. Kelebihan regresi logistik dalam klasifikasi biner menjadikannya alat yang penting untuk mengidentifikasi kelompok siswa yang membutuhkan perhatian khusus, seperti mereka yang berpotensi tidak lulus, sehingga memungkinkan intervensi dini yang lebih terarah.

Namun, kedua model ini juga memiliki keterbatasan yang perlu diperhatikan. Regresi linear cenderung kurang efektif dalam menangkap hubungan yang tidak linier, sementara regresi logistik hanya dapat digunakan untuk analisis klasifikasi biner, sehingga kurang sesuai untuk memprediksi nilai ujian secara detail. Oleh karena itu, pemilihan model harus mempertimbangkan kebutuhan spesifik dari analisis yang dilakukan, dan kombinasi kedua pendekatan dapat memberikan hasil yang lebih holistik.

### 3.4. Interpretasi dan Implikasi Hasil

Hasil penelitian ini menunjukkan bahwa algoritma pembelajaran mesin, seperti regresi linear dan regresi logistik, memiliki potensi besar untuk mendukung pengambilan keputusan berbasis data di sektor pendidikan. Regresi linear memberikan wawasan kuantitatif tentang hubungan antara faktor-faktor pendidikan dengan nilai ujian siswa. Dengan memahami seberapa besar pengaruh variabel independen, seperti jam belajar atau keterlibatan orang tua, terhadap nilai ujian, institusi pendidikan dapat merancang kebijakan yang lebih terarah. Jika jam belajar terbukti memiliki kontribusi signifikan terhadap nilai ujian, sekolah dapat meningkatkan alokasi waktu untuk bimbingan belajar atau memberikan program tambahan bagi siswa dengan nilai rendah [11].

Sementara itu, regresi logistik memberikan solusi praktis dalam klasifikasi kelulusan siswa. Dengan akurasi tinggi sebesar 99,85%, model ini dapat digunakan untuk mengidentifikasi siswa yang berpotensi tidak lulus lebih awal. Informasi ini memungkinkan institusi pendidikan untuk merancang intervensi seperti program remedial, konsultasi dengan orang tua, atau pemberian akses tambahan terhadap sumber daya pendidikan. Tingginya tingkat akurasi juga memberikan kepercayaan bagi pengambil keputusan dalam menggunakan hasil analisis sebagai dasar intervensi strategis [12].

Implikasi dari penelitian ini juga melibatkan pemanfaatan teknologi dan analisis data dalam meningkatkan efisiensi pengelolaan pendidikan. Dengan menggunakan model prediksi yang terintegrasi, sekolah dan institusi pendidikan dapat mengadopsi pendekatan berbasis data dalam mengevaluasi kinerja siswa, mengidentifikasi kebutuhan secara individual, dan merancang kebijakan yang lebih responsif terhadap tantangan pendidikan.

Selain itu, hasil penelitian ini menunjukkan pentingnya memilih model yang sesuai dengan tujuan analisis. Jika fokusnya adalah memahami kontribusi kuantitatif setiap variabel terhadap nilai ujian, maka regresi linear adalah pilihan yang tepat. Sebaliknya, jika tujuannya adalah untuk klasifikasi kelulusan, maka regresi logistik lebih sesuai. Kombinasi kedua pendekatan ini dapat digunakan secara komplementer untuk menghasilkan solusi yang lebih lengkap.

Ke depan, penelitian ini juga membuka peluang untuk pengembangan lebih lanjut, seperti mengintegrasikan metode pembelajaran mesin yang lebih kompleks, algoritma Random Forest atau Neural Network, untuk meningkatkan akurasi prediksi dan klasifikasi. Selain itu, penyempurnaan dataset dengan memasukkan variabel tambahan, seperti motivasi belajar, lingkungan keluarga, atau dukungan psikologis, dapat memberikan pemahaman yang lebih kaya terhadap faktor-faktor yang memengaruhi keberhasilan siswa. Dengan pendekatan yang lebih komprehensif, hasil penelitian ini dapat dioptimalkan untuk mendukung terciptanya sistem pendidikan yang lebih inklusif, efektif, dan adaptif.

#### 4. KESIMPULAN

Kemajuan teknologi dan analisis data berbasis pembelajaran mesin telah membuka peluang besar untuk mendukung peningkatan kualitas pendidikan. Dalam penelitian ini, regresi linear dan logistik diterapkan untuk memprediksi hasil belajar siswa berdasarkan berbagai faktor pendidikan. Hasil analisis menunjukkan bahwa regresi linear efektif dalam memprediksi nilai ujian siswa dengan tingkat kesalahan yang relatif rendah, ditunjukkan oleh nilai Mean Squared Error (MSE) sebesar 3.256. Di sisi lain, regresi logistik mampu mengklasifikasikan kelulusan siswa dengan tingkat akurasi yang sangat tinggi, yaitu 99,85%, menjadikannya alat yang andal untuk analisis klasifikasi biner.

Dengan memahami hubungan antara faktor-faktor pendidikan seperti jam belajar, tingkat kehadiran, dan keterlibatan orang tua terhadap hasil belajar siswa, penelitian ini memberikan dasar bagi institusi pendidikan untuk merancang kebijakan yang lebih adaptif dan berbasis data. Selain itu, kemampuan regresi logistik untuk mengidentifikasi siswa yang berpotensi tidak lulus membuka peluang bagi pemberian intervensi dini yang lebih efektif, seperti program remedial atau dukungan tambahan.

Secara keseluruhan, penelitian ini menegaskan pentingnya pemanfaatan model pembelajaran mesin dalam sistem pendidikan, baik untuk mendukung analisis kuantitatif maupun klasifikasi prediktif. Kombinasi kedua pendekatan ini memberikan solusi yang saling melengkapi, sehingga dapat membantu institusi pendidikan dalam meningkatkan tingkat kelulusan siswa dan mengoptimalkan proses pengambilan keputusan. Penelitian lebih lanjut dapat dilakukan dengan menggunakan algoritma yang lebih kompleks dan variabel tambahan untuk meningkatkan akurasi serta memperkaya pemahaman mengenai faktor-faktor yang memengaruhi keberhasilan akademik siswa.

#### Ucapan Terima Kasih

Ucapan terima kasih yang sebesar-besarnya saya sampaikan kepada Bapak Agiska Ria Supriyatna, S.Si, M.T.I, selaku dosen pengampu mata kuliah Analisis Data untuk Teknik Informatika. Terima kasih atas bimbingan, arahan, dan pengetahuan yang telah diberikan selama proses penulisan karya ilmiah ini. Semoga ilmu yang telah diberikan dapat menjadi bekal berharga dalam pengembangan diri dan profesionalisme saya di masa yang akan datang.

#### DAFTAR PUSTAKA

- [1] A. J. E. Oktavianus, L. Naibaho, and D. A. Rantung, "Pemanfaatan Artificial Intelligence pada Pembelajaran dan Asesmen di Era Digitalisasi," *J. KRIDATAMA SAINS DAN Teknol.*, vol. 5, no. 02, pp. 473–486, Dec. 2023, doi: 10.53863/kst.v5i02.975.
- [2] D. D. S. Fatimah and E. Rahmawati, "Penggunaan Metode Decision Tree dalam Rancang Bangun Sistem Prediksi untuk Kelulusan Mahasiswa," *J. Algoritma*, vol. 18, no. 2, pp. 553–561, Mar. 2022, doi: 10.33364/algoritma/v.18-2.932.
- [3] Y. Huda, P. Rianda, and P. Adi Putra, "Meta-Analisis Pengaruh Model Pembelajaran Problem Based Learning terhadap Hasil Belajar Peserta Didik SMK," *J. PTI Pendidik. DAN Teknol. Inf. Fak. Kegur. ILMU Pendidik. Univ. PUTRA Indones. YPTK PADANG*, vol. 10, no. 1, Apr. 2023, doi: 10.35134/jpti.v10i1.147.
- [4] T. Andayani and F. Madani, "Peran Penilaian Pembelajaran Dalam Meningkatkan Prestasi Siswa di Pendidikan Dasar," *J. Educ. FKIP UNMA*, vol. 9, no. 2, pp. 924–930, Jun. 2023, doi: 10.31949/educatio.v9i2.4402.
- [5] A. M. A. Rusdy, P. Purnawansyah, and H. Herman, "Penerapan Metode Regresi Linear Pada Prediksi Penawaran dan Permintaan Obat Studi Kasus Aplikasi Point Of Sales," *Bul. Sist. Inf. Dan Teknol. Islam*, vol. 3, no. 2, pp. 121–126, May 2022, doi: 10.33096/busiti.v3i2.1130.
- [6] W. Zhu, R. Qiu, and Y. Fu, "Comparative Study on the Performance of Categorical Variable Encoders in Classification and Regression Tasks," 2024, *arXiv*. doi: 10.48550/ARXIV.2401.09682.
- [7] H. Nuha, "Mean Squared Error (MSE) dan Penggunaannya," Apr. 17, 2023, *Social Science Research Network, Rochester, NY*: 4420880. Accessed: Feb. 26, 2025. [Online]. Available: <https://papers.ssrn.com/abstract=4420880>
- [8] R. Darpono and M. Aldi, "SISTEM MONITORING PARKIR MOBIL BERTEMA IoT (INTERNET OF THINGS)," *Power Elektron. J. Orang Elektro*, vol. 9, no. 2, pp. 47–51, Jul. 2020, doi: 10.30591/polektr.v9i2.2012.
- [9] Ronny Susetyoko, Wiratmoko Yuwono, and Elly Purwantini, "Model Klasifikasi Pada Seleksi Mahasiswa Baru Penerima KIP Kuliah Menggunakan Regresi Logistik Biner," *J. Inform. Polinema*, vol. 8, no. 4, pp. 31–40, Aug. 2022, doi: 10.33795/jip.v8i4.914.

- 
- [10] A. Setiawan, "Regresi Linier Sederhana," Smartstat. Accessed: Feb. 26, 2025. [Online]. Available: <https://www.smartstat.info/materi/statistika/regresi/regresi-linier-sederhana.html>
- [11] R. Silaban, F. T. M. Panggabean, H. Simanjuntak, and D. F. P. Sihombing, "Android-Based Learning Media On Students' Learning Outcomes On Thermocymia Materials," *J. Ilmu Pendidik. Indones.*, vol. 10, no. 3, pp. 142–152, Sep. 2022, doi: 10.31957/jipi.v10i3.2411.
- [12] R. Agisna, Z. A. Jauhari, M. S. Zuar, M. Sholihin, and A. K. I, "Evaluasi Pembelajaran," *Soc. Sci. Acad.*, vol. 1, no. 2, pp. 353–362, Aug. 2023, doi: 10.37680/ssa.v1i2.3582.