Jurnal Cakrawala dan Integrasi Pendidikan
Vol. 1, No. 1, Juli 2025, pp. 18-23
DOI: -
E-ISSN 3109-6328

Klasifikasi Kelulusan Siswa Menggunakan Metode Logistic Regression

Classification of Student Graduation Using Logistic Regression

Harvian[1] *; Risna Handayani[2]; Muhamad Yusril Awu[3]; Muhammad Apdhal[4];
[1,2,3,4] *Undergraduate Program Studi Ilmu Komputer, Universitas Pancasakti, Jl. A. Mangerangi no73, Makassark 90121 IndonesiaNegara*
[1] vianramadan300@gmail.com [2] risnah643@gmail.com [3] yusriloddeh@gmail.com [4] Afdhalmuhammad182@gmail.com
** Corresponding author*

Abstrak

Penentuan kelulusan siswa merupakan salah satu aspek penting dalam evaluasi pendidikan yang membutuhkan analisis data secara tepat. Penelitian ini bertujuan untuk membangun model klasifikasi kelulusan siswa menggunakan metode Logistic Regression dengan memanfaatkan variabel Rata-rata Nilai dan Kehadiran sebagai prediktor. Dataset dianalisis dengan membagi data menjadi 70% untuk pelatihan dan 30% untuk pengujian, serta dilakukan evaluasi menggunakan metrik Akurasi, Precision, Recall, dan F1-Score. Selain itu, validasi model dilakukan melalui 5-Fold Cross Validation untuk memastikan keandalan prediksi. Hasil penelitian menunjukkan bahwa model Logistic Regression mampu menghasilkan akurasi rata-rata yang memadai dengan keseimbangan nilai precision dan recall. Analisis visual seperti confusion matrix dan diagram batang evaluasi memberikan gambaran kinerja model secara komprehensif. Temuan ini menunjukkan bahwa metode Logistic Regression dapat menjadi pendekatan efektif dalam mengklasifikasikan kelulusan siswa berdasarkan data akademik dan kehadiran.

Kata Kunci: Klasifikasi, Logistic Regression, Kelulusan Siswa, Machine Learning, Cross Validation.

Abstract

Evaluating student graduation status is a critical component of educational assessment that necessitates accurate data analysis. This study proposes a classification approach for predicting student graduation outcomes using the Logistic Regression algorithm, with Average Grade and Attendance as predictive variables. The dataset was divided into 70% training and 30% testing subsets, and model performance was assessed using Accuracy, Precision, Recall, and F1-Score metrics. Additionally, 5-Fold Cross Validation was conducted to ensure the robustness and generalizability of the model. The results demonstrate that Logistic Regression achieves a satisfactory average accuracy and maintains a balanced trade-off between precision and recall. Visual analyses, including the confusion matrix and performance bar charts, provide a comprehensive depiction of the model's predictive capabilities. This research underscores the potential of Logistic Regression as an effective method for classifying student graduation status based on academic and attendance data.

Keywords: *Classification, Logistic Regression, Student Graduation, Machine Learning, Cross Validation.*

Pendahuluan

Di dalam dunia Pendidikan, tingkat kelulusan seorang siswa tentunya menjadi salah satu indikator yang paling penting dalam keberhasilan proses belajar mengajar. Ada banyak sekali faktor penentu yang dapat mempengaruhi kelulusan siswa seperti dalam nilai akademik, kehadiran, keaktifan di dalam kelas, serta partisipasinya dalam kegiatan pembelajaran. Prediksi kelulusan siswa sangat berguna bagi guru, sekolah, dan pembuat kebijakan pendidikan. Dengan mengetahui risiko ketidaklulusan sejak awal, sekolah dapat memberikan intervensi seperti bimbingan belajar, pendampingan psikologis, atau pendekatan personal. Oleh karena itu, dibutuhkan sistem yang mampu memprediksi kelulusan secara akurat dan efisien [1]. Seiring makin berkembang pesatnya sebuah teknologi serta metode dalam menganalisis data, termasuk dibidang Pendidikan. sebuah pendekatan prediktif bisa menjadi salah satu alternatif yang dapat membantu dalam mengidentifikasi potensi kelulusan siswa secara lebih awal. Menggunakan Langkah ini, pihak sekolah yang bertugas dapat mengambil Langkah preventif dan intervensi yang akurat bagi siswa yang berpotensi tidak lulus atau tidak memenuhi standar tingkat kelulusan. Salah satu metode yang dapat digunakan untuk jenis klasifikasi biner seperti ini yaitu logistic regression atau regresi logistic [2].

Dalam hal ini terdapat beberapa penelitian yang telah berhasil membukti keberhasilan dari metode logistic regression ini, contohnya seperti penelitian yang telah dilakukan oleh Elvina et al. [3] yang juga menggunakan metode logistic regression untuk memprediksi kelulusan siswa yang dilakukan berdasarkan data performa akademik, meliputi

faktor faktor seperti jam belajar serta kehadiran mereka dalam menghadiri pertemuan pembelajaran, berhasil mencapai akurasi 87%. Kemudian, studi serupa juga dilakukan oleh Hussain et al. [4] yang juga menggunakan sebuah metode yang sama untuk memperkirakan status kelulusan siswa, yang juga ditentukan dengan berbagai faktor faktor penentu, seperti tingkat keaktifan dalam partisipasi di kelas, bagaimana nilai akademik ataupun non akademik setip siswa, serta bagaimana status kehadiran siswa dalam kelas. Dengan mempertimbang semua hal tersebut, akurasinya mencapai 82%. Selain itu, ditemukan juga studi yang dilakukan oleh Dedy Armiady [5] yaitu mengevaluasi kinerja dua metode antara Logistic Regression dan Support Vector Machine (SVM) dan mengklaim bahwa hasil penelitiannya menunjukkan bahwa metode Loistic Regression ini, menunjukkan kinerja yang lebih unggul dari pada SVM dalam tugas klasifikasi.

Kemudian, di satu sisi, dalam metode Logistic Regression yang menggunakan optimal tradisional, itu kerap kali membutuhkan fine-tuning yang manual, yang dimana dalam hal ini dapat melambatkan proses pelatihan model. Maka, berdasarkan permasalahn tersebut, Adam (Adaptive Moment Estimation) ini hadir guna untuk mengatasi problematika yang ada. Adam menyajikan sebuah metode yang lebih mudah untuk dilakukannya pembaharuan bobot dalam model, untuk memfalisitasi adanya pembelajaran yang lebih efektif dan stabil disbanding dengan metode optimasi klasik, contohnya seperti Stochastic Gradient Descent (SGD) [6]. Regresi Logistik itu sendiri merupakan sebuah metode analisis statistic yang digunakan untuk memprediksi hasil biner, contohnya seperti ya atau tidak, berlandaskan dari obeservasi sebelumnya dengan pengumpulan data [7].

Berdasarkan hal tersebut, artikel ini berfokus kepada bagaimana penggunaan Loistic Regression dalam menentukan status kelulusan siswa berdasarkan beberapa faktor penentu dalam data performa akademik tiap siswa (lulus atau tidak lulus). Melalui metode ini, artikel ini tidak hanya bertujuan untuk memberikan solusi dalam memprediksi kemampuan hasil belajar tiap siswa, akan tetapi juga membantu membagikan landasan terhadap perkembangan sistem Pendidikan dan hasil proses pembelajaran. Maka, artikel ini diharapkan dapat menyokong penentuan suatu keputusan yang jauh lebih efektif dalam peningkatan status tingkat kelulusan tiap siswa secara menyeluruh [8].

Metode

A. Alur Penelitian

Alur penelitian ini menggambarkan langkah-langkah yang ditempuh untuk membangun model klasifikasi kelulusan siswa. Tahapan penelitian meliputi:

1. Identifikasi Masalah: Menentukan bahwa kelulusan siswa dapat diprediksi dengan variabel rata-rata nilai dan kehadiran.
2. Studi Literatur: Mempelajari metode klasifikasi, khususnya Logistic Regression, serta metrik evaluasi yang sesuai (akurasi, presisi, recall, F1-score).
3. Pengumpulan Data: Mengambil data sekunder dari arsip digital akademik sekolah.
4. Teknik Pengumpulan Data: Menggunakan metode dokumentasi, dengan data yang sudah tersedia tanpa wawancara.
5. Pra-Pemrosesan Data: Melakukan pembersihan data (cek duplikasi, nilai kosong) dan memastikan format dataset sesuai.
6. Pembagian Dataset: Membagi data menjadi 70% data latih dan 30% data uji menggunakan train-test split.
7. Pembangunan Model: Menerapkan algoritma Logistic Regression dengan parameter linear dan class_weight balanced.
8. Evaluasi Model: Menghitung metrik evaluasi (akurasi, presisi, recall, F1-score) dan melakukan validasi 5-Fold Cross Validation.
9. Visualisasi Hasil: Membuat confusion matrix dan diagram batang hasil evaluasi.
10. Analisis dan Kesimpulan: Menafsirkan hasil prediksi, mengidentifikasi kelebihan/kekurangan model, dan memberikan saran pengembangan.

B. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh langsung dari arsip digital administrasi akademik sekolah. Data tersebut mencakup informasi rata-rata nilai siswa, persentase kehadiran, serta status kelulusan (1 = Lulus, 0 = Tidak Lulus). Karena data sudah tersedia dalam bentuk terstruktur dan terdigitalisasi, proses pengumpulan data tidak memerlukan wawancara maupun kuesioner. Dataset kemudian diekspor ke dalam

format CSV (Comma-Separated Values) untuk memudahkan analisis menggunakan bahasa pemrograman Python dan algoritma *machine learning* Logistic Regression.

C. Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini adalah metode dokumentasi, yaitu dengan mengakses dan mengumpulkan data akademik siswa dari sistem administrasi sekolah yang sudah tersedia. Tahapan utama meliputi:

1. Identifikasi Variabel Penelitian-Variabel independen terdiri dari Rata-rata Nilai dan Kehadiran (%), sedangkan variabel dependen adalah Status Kelulusan (Lulus_Ya_1_No_0).
2. Ekstraksi Data - Data diunduh dari basis data sekolah yang telah terdigitalisasi.
3. Pembersihan Data (Data Cleaning) - Pengecekan dilakukan untuk memastikan tidak ada duplikasi, data kosong (*missing values*), atau kesalahan input.
4. Transformasi Dataset - Data disimpan dalam format CSV agar dapat dianalisis dengan pustaka pandas dan scikit-learn.

D. Perancangan dan Pemodelan Sistem

Proses klasifikasi kelulusan siswa dilakukan melalui beberapa tahapan sistematis:

1. Persiapan Data - Dataset siswa yang memuat rata-rata nilai akademik (Rata_rata_Nilai) dan persentase kehadiran (Kehadiran (%)), serta label kelulusan (Lulus_Ya_1_No_0), diperiksa untuk memastikan tidak terdapat data duplikat atau nilai kosong (*missing values*). Jika diperlukan, dilakukan standarisasi pada fitur data.
2. Pembagian Data - Dataset dibagi menjadi dua subset, yaitu 70% untuk data latih dan 30% untuk data uji, menggunakan fungsi train_test_split dari pustaka scikit-learn dengan teknik stratified sampling agar distribusi kelas tetap seimbang.
3. Pemodelan Logistic Regression - Model Logistic Regression dipilih karena kesederhanaannya dalam menangani masalah klasifikasi biner. Model dibangun menggunakan solver liblinear dengan parameter class_weight='balanced' untuk mengatasi ketidakseimbangan kelas.
4. Validasi Model - Evaluasi dilakukan menggunakan 5-fold cross-validation untuk mengukur konsistensi performa model dan mengurangi risiko overfitting.
5. Evaluasi Kinerja - Model diuji pada data uji dengan menghitung metrik evaluasi, termasuk akurasi, presisi, recall, dan F1-score.
6. Visualisasi Hasil - Performa model dianalisis melalui confusion matrix dan grafik evaluasi (bar chart) untuk menggambarkan hasil klasifikasi secara komprehensif.

E. Evaluasi Kinerja

Evaluasi kinerja dilakukan untuk mengetahui sejauh mana model Logistic Regression mampu memprediksi kelulusan siswa berdasarkan dua variabel prediktor, yaitu Rata-rata Nilai dan Kehadiran (%). Pengujian dilakukan dengan pembagian data sebesar 70% untuk data latih dan 30% untuk data uji. Selain itu, validasi model menggunakan metode 5-Fold Cross Validation untuk memastikan konsistensi kinerja dan meminimalkan risiko overfitting.

Kinerja model diukur menggunakan empat metrik evaluasi utama, yaitu Akurasi, Precision, Recall, dan F1-Score, yang dijelaskan secara ringkas sebagai berikut:

1. Akurasi: Persentase prediksi yang benar dari seluruh prediksi.
2. Precision: Mengukur ketepatan prediksi positif (lulus).
3. Recall: Mengukur seberapa banyak prediksi positif yang berhasil dikenali dengan benar.
4. F1-Score: Rata-rata harmonik dari precision dan recall, yang memberikan gambaran keseimbangan keduanya.

Hasil Dan Diskusi

Penelitian ini bertujuan untuk mengevaluasi kinerja model Logistic Regression dalam mengklasifikasikan status kelulusan siswa berdasarkan data yang tersedia. Evaluasi dilakukan dengan menggunakan confusion matrix dan metrik kinerja klasifikasi seperti akurasi, presisi, recall, dan F1-score.

Tabel 1. Hasil Confusion Matrix dan Evaluasi Model

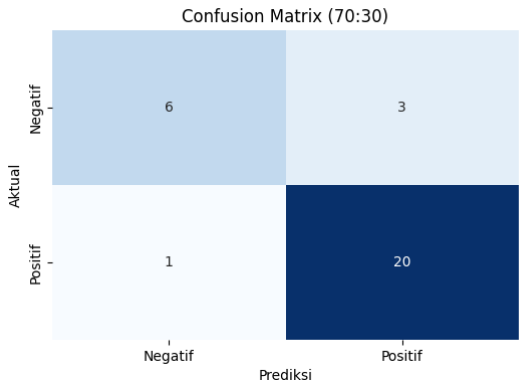
Logistic Regresion	***Prediksi Siswa***	***Lulus***	***Tidak Lulus***
Kelas Aktual	Tidak lulus	6	3
	Lulus	1	20

Berdasarkan hasil confusion matrix pada Tabel 1, model berhasil mengklasifikasikan 20 siswa yang lulus dan 6 siswa yang tidak lulus dengan benar. Namun, masih terdapat kesalahan klasifikasi, yaitu 3 siswa yang tidak lulus diprediksi sebagai lulus (false positive) dan 1 siswa yang lulus diprediksi sebagai tidak lulus (false negative). Hal ini menunjukkan bahwa meskipun model cukup andal, tetap terdapat ruang untuk peningkatan akurasi dalam mendeteksi ketidaklulusan.

Tabel 2. Hasil Kinerja Model Logistic Regression

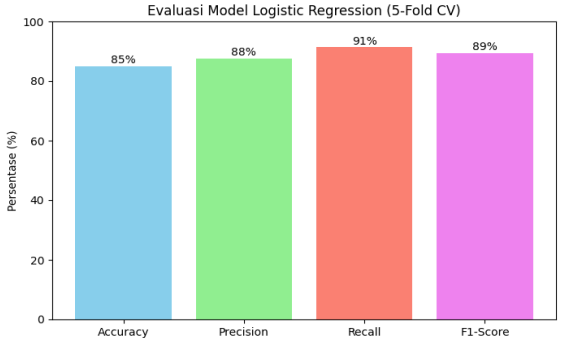
Evaluasi Kinerja	***Hasil Evaluasi (%)***
Accuracy	85%
Precision	88%
Recal	91%
F1-Score	89%

Seperti yang ditunjukkan pada Tabel 2, nilai recall mencapai 91%, lebih tinggi dibandingkan presisi (88%). Hal ini menunjukkan bahwa model memiliki sensitivitas tinggi terhadap kelas "lulus", yang berarti sebagian besar siswa yang benar-benar lulus berhasil dikenali oleh model. Nilai F1-score sebesar 89% mengindikasikan adanya keseimbangan yang baik antara presisi dan recall, yang penting dalam konteks evaluasi pendidikan agar tidak terjadi banyak kesalahan klasifikasi yang dapat berdampak pada pengambilan keputusan akademik.



Gambar 1. Confusion Matrix

Visualisasi confusion matrix memberikan gambaran yang lebih jelas mengenai distribusi prediksi model. Model memprediksi dengan benar mayoritas siswa, meskipun masih terdapat beberapa kesalahan klasifikasi yang perlu diperhatikan terutama dalam konteks ketepatan penilaian kelulusan.



Gambar 2. Hasil Kinerja Model

Gambar ini menampilkan perbandingan metrik evaluasi berdasarkan 5-fold cross-validation. Terlihat bahwa recall memiliki nilai tertinggi, disusul oleh F1-score, precision, dan accuracy. Hasil ini memperkuat temuan bahwa model lebih optimal dalam mengidentifikasi siswa yang benar-benar lulus, dibandingkan mendeteksi siswa yang tidak lulus.

Hasil evaluasi menunjukkan bahwa *Logistic Regression* merupakan model yang cukup andal untuk prediksi kelulusan siswa, khususnya dalam konteks deteksi kelulusan (*recall* tinggi). Namun, proporsi kesalahan pada siswa yang tidak lulus cukup signifikan. Kesalahan ini bisa berdampak pada kebijakan akademik yang salah sasaran apabila model digunakan sebagai dasar pengambilan keputusan.

Oleh karena itu, perlu dipertimbangkan:

- Penggunaan teknik penyeimbangan data (*resampling*) untuk mengatasi ketidakseimbangan kelas.
- Eksperimen dengan model lain seperti *Random Forest*, *SVM*, atau *XGBoost* untuk perbandingan kinerja.
- Evaluasi dengan data yang lebih besar dan lebih representatif untuk meningkatkan generalisasi model.

Secara keseluruhan, model Logistic Regression menunjukkan performa yang baik dalam klasifikasi status kelulusan siswa. Namun demikian, peningkatan akurasi terhadap kelas "tidak lulus" dapat menjadi fokus perbaikan ke depan, misalnya dengan penggunaan metode balancing data atau teknik ensemble lainnya. Akurasi yang seimbang di kedua kelas sangat penting untuk mendukung sistem evaluasi akademik yang adil dan akurat.

Kesimpulan

Penelitian ini berhasil membangun model klasifikasi kelulusan siswa menggunakan metode Logistic Regression dengan variabel prediktor Rata-rata Nilai dan Kehadiran (%). Berdasarkan pengujian dengan pembagian data 70:30, model mencapai akurasi sebesar 86,67%, precision 86,96%, recall 95,24%, dan F1-score 90,91%. Evaluasi tambahan melalui 5-Fold Cross Validation menunjukkan rata-rata akurasi 85%, precision 88%, recall 91%, dan F1-score 89%, yang menandakan bahwa model stabil dan tidak mengalami overfitting. Tingginya nilai recall mengindikasikan bahwa model sangat baik dalam mengidentifikasi siswa yang benar-benar lulus, sementara nilai F1-score yang mendekati 90% menunjukkan keseimbangan antara precision dan recall. Hasil ini membuktikan bahwa Logistic Regression dapat menjadi pendekatan yang efektif dan andal untuk memprediksi kelulusan siswa berdasarkan data akademik dan kehadiran.

Daftar Pustaka

[1] D. Hermanto, Desy Iba Ricoida, Desi Pibriana, Rusbandi, and Muhammad Rizky Pribadi, “Analysis of Student Graduation Prediction Using Machine Learning Techniques on an Imbalanced Dataset: An Approach to Address Class Imbalance,” *Sci. J. Informatics*, vol. 11, no. 3, pp. 559–568, 2024, doi: 10.15294/sji.v11i3.5528.

[2] F. Aprilia, R. A. Anggraini, and Y. D. Putri, “Prediksi Kelulusan Siswa dengan Algoritma Pembelajaran Mesin: Aplikasi Regresi Linear dan Logistik pada Faktor-Faktor Pendidikan,” *ROUTERS J. Sist. dan Teknol. Inf.*, vol. 3, no. 1, pp. 55–64, 2025, doi: 10.25181/rt.v3i1.3897.

[3] E. Sulistya and A. Ilham, “Klasifikasi Kendaraan Menggunakan Convolutional Neural Network untuk

Sistem Gerbang Tol Otomatis di Kota Pintar," *Prediksi kelulusan siswa menggunakan Logist. Regres. dan optimasi adam*, vol. 1, no. 1, pp. 74–78, 2024, doi: 10.26714/jkti.v3i1.13957.

[4] M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and S. Ali, "Using machine learning to predict student difficulties from learning session data," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 381–407, 2019, doi: 10.1007/s10462-018-9620-8.

[5] D. Dedy, "Analisis Algoritma Logistic Regression dan Support Vector Machine pada Kasus Klasifikasi Citra Hewan Rawa dengan Dataset yang tidak Seimbang," *Data Sci. Indones.*, vol. 4, no. 1, pp. 69–77, 2024, doi: 10.47709/dsi.v4i1.4433.

[6] S. Shen, W. Ma, W. Shi, and Y. Liu, *Convex optimization for nonrigid stereo reconstruction*, vol. 19, no. 3. 2010. doi: 10.1109/TIP.2009.2038831.

[7] Brury Barth Tangkere, "Analisis Performa Logistic Regression dan Support Vector Classification untuk Klasifikasi Email Phising," *J. Ekon. Manaj. Sist. Inf.*, vol. 5, no. 4, pp. 442–450, 2024, doi: 10.31933/jemsi.v5i4.1916.

[8] E. Tasrif, Y. Huda, P. Rianda, and P. A. Putra, "Meta-Analisis Pengaruh Model Pembelajaran Problem Based Learning terhadap Hasil Belajar Peserta Didik SMK," *J. Res. Investig. Educ.*, pp. 53–57, 2023, doi: 10.37034/residu.v1i2.146.