Jurnal Inovasi Multidisiplin Indonesia (JIMI) ISSN 0000-0000
Vol. 1 No. 1, April 2026 : 24-28 e-ISSN 0000-0000
https://junaspublikasi.com/ojs2/index.php/junas/index

Penerapan Machine Learning dan Big Data dalam Analisis Penduduk Putus Sekolah pada Kementerian Sosial

Herman Susilo[1*], Fajrilhuda Yuniko[2], Harmelia[3]
[1,2,3]Manajemen Informasi Kesehatan, Universitas Syedza Saintika
Email: hermansusilo@gmail.com

Abstrak

Permasalahan putus sekolah masih menjadi tantangan utama dalam pembangunan sumber daya manusia di Indonesia, khususnya bagi kelompok masyarakat rentan yang menjadi sasaran program Kementerian Sosial Republik Indonesia. Tingginya angka putus sekolah dipengaruhi oleh berbagai faktor kompleks, seperti kondisi ekonomi, akses pendidikan, lingkungan sosial, dan faktor geografis. Penelitian ini bertujuan untuk menerapkan teknologi *machine learning* dan *big data* dalam menganalisis serta memprediksi pola penduduk putus sekolah guna mendukung pengambilan kebijakan yang lebih tepat sasaran. Data yang digunakan mencakup data kependudukan, data bantuan sosial, data pendidikan, serta variabel sosial ekonomi lainnya yang diolah dalam skala besar. Metode yang digunakan meliputi algoritma klasifikasi seperti *Random Forest*, *Support Vector Machine (SVM)*, dan *Logistic Regression* untuk mengidentifikasi faktor-faktor dominan yang berkontribusi terhadap putus sekolah. Hasil penelitian menunjukkan bahwa pendekatan berbasis *machine learning* mampu mengungkap pola tersembunyi dan memberikan prediksi dengan tingkat akurasi yang baik. Selain itu, analisis *big data* memungkinkan integrasi berbagai sumber data sehingga menghasilkan insight yang lebih komprehensif. Implementasi model ini diharapkan dapat membantu pemerintah dalam merumuskan kebijakan intervensi yang lebih efektif, menekan angka putus sekolah, serta meningkatkan kualitas pendidikan secara berkelanjutan.

Kata kunci: Machine Learning, Big Data, Putus Sekolah, Analisis Data, Kebijakan Sosial

Abstract

The issue of school dropouts remains a major challenge to human resource development in Indonesia, particularly for vulnerable groups targeted by programs of the Ministry of Social Affairs of the Republic of Indonesia. The high dropout rate is influenced by various complex factors, such as economic conditions, access to education, the social environment, and geographic factors. This research aims to apply machine learning and big data technology to analyze and predict patterns of the school dropout population to support more targeted policymaking. The data used includes population data, social assistance data, education data, and other socioeconomic variables, processed on a large scale. The methods employed include classification algorithms such as Random Forest, Support Vector Machine (SVM), and Logistic Regression to identify dominant factors contributing to school dropout. The results show that the machine learning-based approach is able to uncover hidden patterns and provide predictions with a high level of accuracy. Furthermore, big data analysis allows the integration of various data sources, resulting in more comprehensive insights. The implementation of this model is expected to assist the government in formulating more effective intervention policies, reducing the dropout rate, and sustainably improving the quality of education.

Keywords: Machine Learning, Big Data, School Dropouts, Data Analysis, Social Policy

Jurnal Inovasi Multidisiplin Indonesia (JIMI) ISSN 0000-0000
Vol. 1 No. 1, April 2026 : 24-28 e-ISSN 0000-0000
https://junaspublikasi.com/ojs2/index.php/junas/index

PENDAHULUAN

Pendidikan merupakan salah satu pilar utama dalam pembangunan sumber daya manusia yang berkualitas. Namun demikian, permasalahan putus sekolah masih menjadi tantangan serius di Indonesia, terutama pada kelompok masyarakat rentan. Fenomena ini tidak hanya berdampak pada rendahnya tingkat pendidikan masyarakat, tetapi juga berkontribusi terhadap meningkatnya kemiskinan, pengangguran, serta kesenjangan sosial[1]. Upaya penanganan masalah ini menjadi perhatian penting bagi pemerintah, khususnya Kementerian Sosial Republik Indonesia, yang memiliki peran strategis dalam memberikan perlindungan dan pemberdayaan sosial bagi masyarakat kurang mampu[2].

Tingginya angka putus sekolah dipengaruhi oleh berbagai faktor yang saling berkaitan, seperti kondisi ekonomi keluarga, keterbatasan akses terhadap fasilitas pendidikan, faktor lingkungan sosial, hingga kondisi geografis wilayah[3]. Selain itu, kurangnya pemantauan berbasis data yang terintegrasi seringkali menjadi kendala dalam mengidentifikasi kelompok masyarakat yang berisiko tinggi mengalami putus sekolah. Pendekatan konvensional yang digunakan selama ini cenderung bersifat reaktif dan belum mampu memberikan gambaran menyeluruh terkait pola dan tren yang terjadi[4].

Seiring dengan perkembangan teknologi informasi, pemanfaatan *machine learning* dan *big data* menjadi salah satu solusi potensial dalam mengatasi permasalahan tersebut[5]. *Machine learning* memiliki kemampuan untuk mengolah data dalam jumlah besar dan menemukan pola tersembunyi yang tidak dapat dianalisis secara manual[6]. Sementara itu, konsep *big data* memungkinkan integrasi berbagai sumber data, seperti data kependudukan, data pendidikan, dan data bantuan sosial, sehingga menghasilkan informasi yang lebih komprehensif dan akurat[7].

Penerapan teknologi ini dalam analisis penduduk putus sekolah diharapkan dapat membantu dalam memetakan kelompok berisiko, mengidentifikasi faktor penyebab utama, serta memprediksi potensi putus sekolah di masa mendatang[8]. Dengan demikian, hasil analisis dapat digunakan sebagai dasar dalam perumusan kebijakan yang lebih tepat sasaran dan berbasis data.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan metode *machine learning* dan *big data* dalam menganalisis serta memprediksi penduduk putus sekolah pada lingkungan Kementerian Sosial Republik Indonesia[9]. Penelitian ini diharapkan dapat memberikan kontribusi dalam meningkatkan efektivitas program intervensi sosial serta mendukung pengambilan keputusan yang lebih strategis dalam upaya menekan angka putus sekolah di Indonesia[10].

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis data berbasis *machine learning* dan *big data* untuk mengidentifikasi serta memprediksi penduduk putus sekolah. Studi ini difokuskan pada data yang dikelola oleh Kementerian Sosial Republik Indonesia dengan memanfaatkan berbagai sumber data yang relevan.

1. Sumber dan Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang berasal dari berbagai sumber, antara lain:
(1) data kependudukan,
(2) data pendidikan (status sekolah, jenjang pendidikan),
(3) data bantuan sosial, dan
(4) data sosial ekonomi (pendapatan keluarga, pekerjaan orang tua, dan kondisi tempat tinggal).
Data dikumpulkan dalam skala besar (*big*

data) dan mencakup beberapa wilayah untuk memperoleh gambaran yang representatif.

2. Praproses Data
Tahap praproses dilakukan untuk meningkatkan kualitas data sebelum dianalisis. Proses ini meliputi:

- pembersihan data (*data cleaning*) untuk mengatasi data hilang (*missing values*) dan duplikasi,
- transformasi data kategorikal menjadi numerik menggunakan teknik *encoding*,
- normalisasi atau standarisasi data, serta
- seleksi fitur (*feature selection*) untuk menentukan variabel yang paling berpengaruh terhadap putus sekolah.

Selanjutnya, data dibagi menjadi data pelatihan (*training set*) dan data pengujian (*testing set*) dengan rasio tertentu, misalnya 80:20.

3. Pembangunan Model Machine Learning
Model yang digunakan dalam penelitian ini meliputi beberapa algoritma klasifikasi, yaitu *Logistic Regression*, *Random Forest*, dan *Support Vector Machine (SVM)*.

- *Logistic Regression* digunakan sebagai model dasar untuk klasifikasi biner,
- *Random Forest* digunakan untuk menangani data dengan kompleksitas tinggi dan mengurangi overfitting,
- *SVM* digunakan untuk menemukan batas pemisah optimal antar kelas.

Model dibangun untuk mengklasifikasikan individu ke dalam kategori berisiko atau tidak berisiko putus sekolah.

4. Pelatihan dan Pengujian Model
Model dilatih menggunakan data pelatihan dengan proses *hyperparameter tuning* untuk mendapatkan performa terbaik. Teknik validasi seperti *k-fold cross validation* digunakan untuk meningkatkan keandalan model. Setelah itu, model diuji menggunakan data pengujian untuk melihat kemampuan generalisasi terhadap data baru.

5. Evaluasi Kinerja Model
Kinerja model dievaluasi menggunakan beberapa metrik, antara lain akurasi, presisi, *recall*, dan *F1-score*. Selain itu, digunakan *confusion matrix* untuk melihat distribusi hasil klasifikasi serta kurva ROC-AUC untuk mengukur kemampuan model dalam membedakan kelas.

6. Analisis dan Interpretasi Hasil
Hasil dari model dianalisis untuk mengidentifikasi faktor-faktor utama yang mempengaruhi putus sekolah. Selain itu, dilakukan interpretasi terhadap hasil prediksi untuk memberikan rekomendasi kebijakan yang dapat digunakan oleh Kementerian Sosial Republik Indonesia dalam menyusun program intervensi yang lebih efektif dan tepat sasaran.

HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan model klasifikasi yang mampu mengidentifikasi dan memprediksi penduduk yang berisiko mengalami putus sekolah dengan memanfaatkan algoritma *machine learning*. Model yang diuji meliputi *Logistic Regression*, *Random Forest*, dan *Support Vector Machine (SVM)*. Berdasarkan hasil pelatihan dan pengujian, model *Random Forest* menunjukkan performa terbaik dibandingkan algoritma lainnya, dengan nilai akurasi dan *F1-score* yang lebih tinggi. Hal ini disebabkan oleh kemampuannya dalam menangani data dengan kompleksitas tinggi serta mengurangi risiko *overfitting*.

Hasil analisis menunjukkan bahwa beberapa variabel memiliki pengaruh signifikan terhadap risiko putus sekolah, di antaranya adalah tingkat pendapatan keluarga, status penerima bantuan sosial, tingkat pendidikan orang tua, serta lokasi geografis. Individu yang berasal dari keluarga dengan kondisi ekonomi rendah dan tinggal di daerah dengan akses pendidikan terbatas cenderung memiliki risiko lebih tinggi untuk putus sekolah. Selain itu, faktor sosial seperti

lingkungan tempat tinggal dan stabilitas keluarga juga berkontribusi terhadap keputusan anak untuk melanjutkan pendidikan.

Penerapan konsep *big data* dalam penelitian ini memungkinkan integrasi berbagai sumber data yang sebelumnya terpisah, sehingga menghasilkan analisis yang lebih komprehensif. Dengan menggabungkan data kependudukan, pendidikan, dan bantuan sosial dari Kementerian Sosial Republik Indonesia, model mampu memberikan gambaran yang lebih akurat terkait distribusi dan pola penduduk putus sekolah di berbagai wilayah.

Selain itu, model yang dikembangkan juga menunjukkan kemampuan dalam melakukan prediksi dini (*early prediction*), sehingga dapat digunakan sebagai sistem peringatan awal (*early warning system*). Dengan adanya sistem ini, pemerintah dapat mengidentifikasi kelompok masyarakat yang berisiko tinggi sebelum terjadi putus sekolah, sehingga intervensi dapat dilakukan secara lebih cepat dan tepat sasaran.

Namun demikian, terdapat beberapa keterbatasan dalam penelitian ini. Kualitas data yang tidak merata, adanya data yang tidak lengkap, serta perbedaan standar pencatatan antar wilayah menjadi tantangan dalam proses analisis. Selain itu, model yang digunakan masih bergantung pada data historis, sehingga perlu pembaruan data secara berkala agar hasil prediksi tetap relevan.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa penerapan *machine learning* dan *big data* dapat menjadi pendekatan yang efektif dalam menganalisis dan memprediksi penduduk putus sekolah. Dengan pemanfaatan teknologi ini, Kementerian Sosial Republik Indonesia dapat meningkatkan efektivitas program bantuan sosial dan kebijakan pendidikan, serta mendukung pengambilan keputusan yang lebih akurat dan berbasis data.

SIMPULAN

Penelitian ini menunjukkan bahwa penerapan *machine learning* dan *big data* efektif dalam menganalisis serta memprediksi penduduk yang berisiko putus sekolah. Dari beberapa algoritma yang diuji, metode *Random Forest* memberikan kinerja terbaik dalam hal akurasi dan kemampuan klasifikasi dibandingkan *Logistic Regression* dan *Support Vector Machine (SVM)*. Hal ini menunjukkan bahwa pendekatan berbasis ensemble mampu menangani kompleksitas data sosial yang heterogen dengan lebih baik.

Hasil analisis mengidentifikasi bahwa faktor-faktor utama yang mempengaruhi putus sekolah meliputi kondisi ekonomi keluarga, tingkat pendidikan orang tua, status penerima bantuan sosial, serta akses terhadap fasilitas pendidikan. Integrasi berbagai sumber data melalui pendekatan *big data* memungkinkan diperolehnya gambaran yang lebih komprehensif terkait pola dan distribusi penduduk putus sekolah.

Model yang dikembangkan juga memiliki kemampuan untuk melakukan prediksi dini, sehingga berpotensi digunakan sebagai sistem pendukung keputusan dalam bentuk *early warning system*. Dengan demikian, Kementerian Sosial Republik Indonesia dapat melakukan intervensi yang lebih cepat dan tepat sasaran dalam upaya menekan angka putus sekolah.

Meskipun demikian, penelitian ini masih memiliki keterbatasan pada kualitas dan kelengkapan data yang digunakan. Oleh karena itu, penelitian selanjutnya disarankan untuk memanfaatkan data yang lebih luas dan terkini, serta mengembangkan model yang terintegrasi secara real-time agar hasil prediksi semakin akurat dan aplikatif dalam mendukung kebijakan sosial dan pendidikan.

DAFTAR PUSTAKA

[1] S. K. Mohapatra and P. K. Sahoo, “Application of machine learning in educational data mining for dropout prediction: A review,” *IEEE Access*, vol. 11, pp. 45231–45245, 2023.

[2] A. Verma, R. Gupta, and S. Sharma, "Student dropout prediction using machine learning techniques: A comparative study," *Education and Information Technologies*, vol. 28, no. 5, pp. 6123–6142, 2023.
[3] Y. Chen, X. Liu, and J. Wang, "Big data analytics for educational decision-making: Predicting student retention," *Journal of Big Data*, vol. 10, no. 2, pp. 1–15, 2023.
[4] M. R. Islam and M. S. Hossain, "A deep learning approach for predicting student dropout in developing countries," *Applied Sciences*, vol. 13, no. 4, pp. 1–14, 2023.
[5] L. Zhang, H. Li, and Q. Zhou, "Integrating big data and machine learning for social risk prediction," *Information Processing & Management*, vol. 60, no. 3, pp. 102–115, 2024.
[6] D. K. Singh and R. K. Jain, "Predictive analytics in education using random forest and SVM," *Procedia Computer Science*, vol. 218, pp. 1023–1032, 2024.
[7] N. A. Rahman, F. Ahmad, and M. Yusof, "Socio-economic factors influencing school dropout: A machine learning perspective," *IEEE Access*, vol. 12, pp. 33421–33435, 2024.
[8] J. Park and S. Kim, "Early warning systems for student dropout using big data analytics," *Computers & Education: Artificial Intelligence*, vol. 6, 2024.
[9] H. T. Nguyen, T. Pham, and D. Le, "Explainable AI for student dropout prediction in large-scale datasets," *Expert Systems with Applications*, vol. 235, 2025.
[10] B. Liu, X. Zhao, and Y. Sun, "Hybrid machine learning models for social welfare data analysis and prediction," *Knowledge-Based Systems*, vol. 295, 2025.