

Penerapan Algoritma Random Forest dalam Prediksi Status Penerima PIP pada Siswa: Studi Kasus pada SMK Amaliah 1

Risfan Novrian¹, Tia Agustiani², Muhamad Fikri³, Moch Fajar Hikmatulloh⁴,

Muhammad Erlangga Gunawan⁵, Uus Firdaus⁶

¹Universitas Djuanda, risfan.novrian@unida.ac.id

²Universitas Djuanda, tia.agustiani@unida.ac.id

³Universitas Djuanda, muhamad.fikri@unida.ac.id

⁴Universitas Djuanda, m.fajar.hikmatulloh@unida.ac.id

⁵Universitas Djuanda, i.2210161@unida.ac.id

⁶Universitas Djuanda, uus.firdaus@unida.ac.id

ABSTRAK

Penelitian ini bertujuan untuk mengeksplorasi dan mengimplementasikan algoritma Random Forest dalam konteks prediksi status penerima Program Indonesia Pintar (PIP) pada siswa tingkat sekolah menengah. Algoritma Random Forest telah terbukti efektif dalam masalah klasifikasi, dan penelitian ini mengadopsinya untuk menganalisis faktor-faktor yang mempengaruhi kelayakan penerimaan PIP. Studi ini menggunakan dataset siswa yang mencakup informasi seperti nama, kelas, dan beberapa atribut sosio-ekonomi seperti Penerima KPS, Penerima KKS, Miskin, dan Yatim. Metode penelitian melibatkan tahap-tahap persiapan data, pembagian data, dan pelatihan model menggunakan algoritma Random Forest. Hasil eksperimen dievaluasi dengan menggunakan metrik kinerja yang mencakup akurasi prediksi. Analisis hasil memberikan wawasan yang mendalam tentang faktor-faktor yang paling signifikan dalam menentukan status penerima PIP. Selain itu juga membahas keunggulan dan kelemahan penerapan algoritma Random Forest. Penelitian ini tidak hanya memberikan kontribusi pada pemahaman tentang implementasi algoritma machine learning dalam konteks pendidikan, tetapi juga diharapkan dapat memberikan panduan bagi kebijakan sosial di bidang pendidikan untuk lebih efektif mendukung siswa yang membutuhkan bantuan finansial.

Kata Kunci: PIP, Random Forest, prediksi

PENDAHULUAN

Pentingnya pendidikan dalam upaya memberantas kebodohan, penanggulangan kemiskinan, peningkatan kesejahteraan masyarakat, serta pembangunan moral dan identitas bangsa, menjadi fokus serius pemerintah. Upaya ini mencakup penanganan masalah pendidikan dari tingkat dasar hingga perguruan tinggi, dengan alokasi anggaran dan kebijakan yang signifikan. Pemerintah juga

berkomitmen untuk terus berinovasi guna meningkatkan kualitas pendidikan dan menciptakan peluang bagi seluruh lapisan (Yayan, Sri Wulan, Unika, & Nizmah, 2019).

A. Latar Belakang

Pendidikan adalah kunci bagi pembangunan masyarakat yang berkelanjutan dan inklusif. Untuk mendukung siswa dari keluarga kurang mampu, Program Indonesia Pintar (PIP) memberikan bantuan Kartu Indonesia Pintar (KIP) kepada mereka yang memenuhi kriteria tertentu. KIP menjadi solusi signifikan untuk membantu biaya pendidikan dan memberikan akses lebih luas kepada pendidikan berkualitas. Namun, dalam implementasinya, penyaluran KIP seringkali dihadapkan pada kendala penyaluran yang tepat sasaran. Oleh karena itu, analisis prediksi penerima KIP pada siswa menjadi krusial dalam mendukung kebijakan penyaluran yang lebih efisien dan efektif.

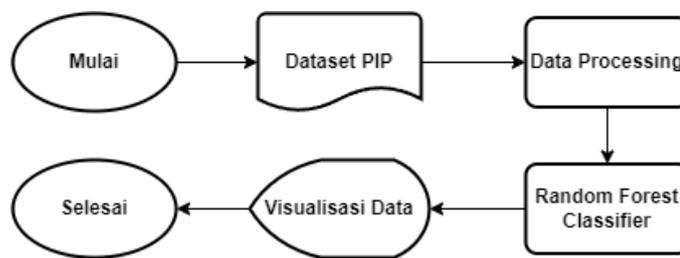
B. Tujuan Penelitian

Tujuan utama penelitian ini adalah mengembangkan model prediksi menggunakan algoritma Random Forest untuk mengidentifikasi siswa yang berpotensi menjadi penerima KIP. Dengan memanfaatkan data historis siswa yang sudah menerima KIP sebelumnya, model ini diharapkan dapat memberikan prediksi yang akurat untuk mendukung kebijakan penyaluran KIP yang lebih tepat sasaran. Selain itu, penelitian ini bertujuan untuk menganalisis faktor-faktor yang paling berpengaruh terhadap keputusan penerimaan KIP. Dengan fokus pada prediksi yang tepat dan analisis faktor-faktor yang mempengaruhi, hasil penelitian ini diharapkan dapat memberikan kontribusi nyata dalam meningkatkan efisiensi dan efektivitas Program Indonesia Pintar (PIP) serta mendukung upaya pemerataan akses pendidikan di Indonesia.

METODE PENELITIAN

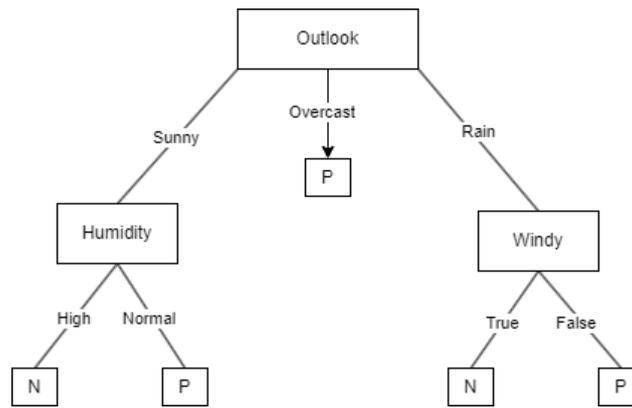
A. Desain Penelitian

Penelitian ini menggunakan desain prediktif dengan memanfaatkan algoritma Random Forest. Random Forest merupakan suatu algoritma machine learning yang menggabungkan beberapa pohon prediksi. Setiap pohon dalam model ini dibentuk secara independen dengan menggunakan sampel data yang diambil secara acak, dan distribusi sampel tersebut seragam untuk semua pohon dalam hutan (Freund & Schapire, 1996). Algoritma ini dipilih karena kemampuannya dalam menangani klasifikasi pada dataset dengan sejumlah besar fitur dan kompleksitas tinggi, serta kemampuannya mengatasi overfitting.



Gambar 1. Tahapan Penelitian

Pada gambar 1. data awal akan melalui serangkaian langkah pra-pemrosesan untuk membersihkan, mengonversi, dan mengatur ulang data. Langkah-langkah ini melibatkan penanganan nilai yang hilang, pengkodean variabel kategorikal, dan normalisasi jika diperlukan. Selain itu, fitur-fitur yang tidak relevan atau redundan juga akan diidentifikasi dan dihapus, Memilih kombinasi metode pra-pemrosesan yang tepat secara signifikan akan meningkatkan akurasi klasifikasi data (Sarab & Mohamed, 2016). Data akan dipartisi menjadi dua kelompok: satu kelompok sebagai set pelatihan (training set) dan yang lainnya sebagai set pengujian (testing set). Set pelatihan berfungsi untuk melatih model, sementara set pengujian digunakan untuk menguji serta mengevaluasi performa dari model tersebut. Pembagian dilakukan secara acak dengan proporsi tertentu (misalnya, 80% untuk pelatihan dan 20% untuk pengujian). Pohon keputusan dibuat dengan menentukan node akar dan berakhir dengan beberapa node daun untuk mendapatkan hasil akhir seperti yang terlihat pada gambar 2.



Gambar 2. Random Forest Proses

B. Data dan Sampel

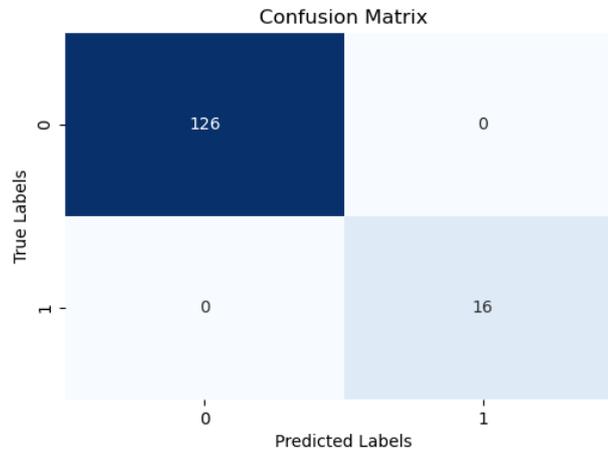
Data yang digunakan dalam penelitian ini merupakan data historis siswa yang telah menerima atau tidak menerima KIP pada tahun-tahun sebelumnya. Data ini mencakup variabel-variabel seperti nama siswa, Penerima KIP, status Penerima KPS, Penerima KKS, Miskin, dan Yatim. Pengumpulan data dilakukan melalui data sekolah terkait. pengolahan data dari kumpulan data mentah yang akan dilakukan menghasilkan informasi yang penting dan bermanfaat (Firdaus & Utama, 2020).

HASIL DAN PEMBAHASAN

Random Forest merupakan gabungan dari berbagai teknik pohon keputusan yang ada, yang kemudian disatukan dan dikombinasikan menjadi suatu model tunggal. Terdapat tiga poin kunci dalam pendekatan Random Forest, yakni (1) menggunakan bootstrap sampling untuk membangun pohon prediksi; (2) setiap pohon keputusan melakukan prediksi dengan menggunakan prediktor secara acak; (3) selanjutnya, Random Forest melakukan prediksi dengan menggabungkan hasil dari setiap pohon keputusan melalui mayoritas suara untuk klasifikasi atau rata-rata untuk regresi (Salamah, Tamazin, Sharkas, & Khedr, 2016).

A. Confusion Matrix

Confusion matrix di random forest memberikan informasi tentang penetapan masukan ke kelas yang berbeda, yang digunakan untuk mengevaluasi kinerja (Faraji, 2022).



Gambar 3. Confusion Matrix kelas pada PIP

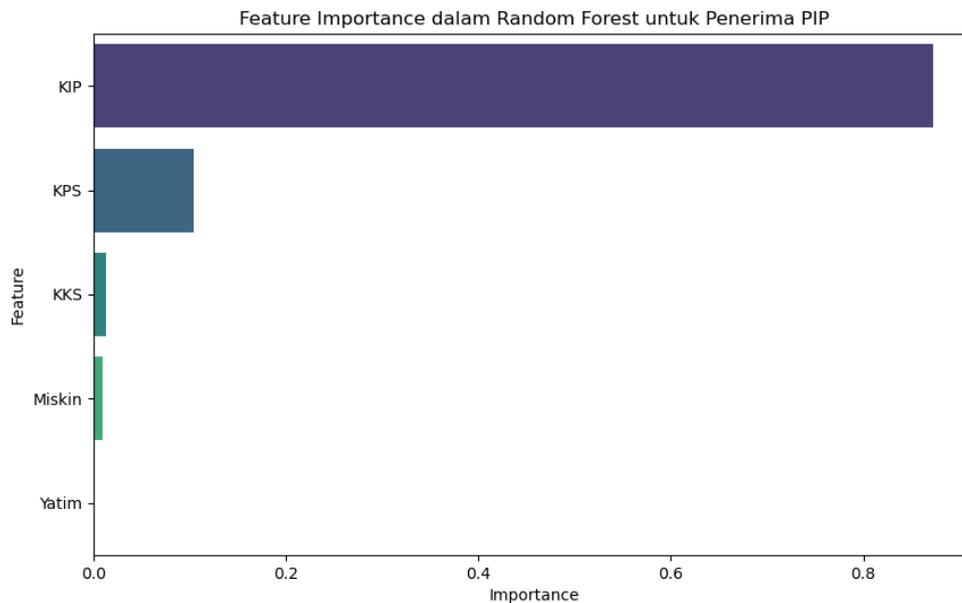
Dari gambar 3 diatas didapatkan bahwa akurasi data menggunakan confusion matriks terdapat True Label dan Predicted label.

- True Label 0 (Tidak Mendapatkan KIP): 126
- Predicted Label 1 (Mendapatkan KIP): 16

Ini menunjukkan bahwa model memiliki 126 True Negative (TN) dan 16 False Positive (FP). Dalam konteks ini, False Positive dapat diartikan sebagai siswa yang diprediksi mendapatkan KIP, tetapi sebenarnya tidak. Evaluasi lebih lanjut, seperti akurasi, presisi, recall, dan F1-score, dapat memberikan pemahaman yang lebih lengkap tentang kinerja model.

B. Feature Importance

Feature importance dalam pengklasifikasi random forest digunakan untuk merancang strategi pemilihan fitur yang penting untuk segmentasi dan akurasi klasifikasi yang tinggi (Mahapatra, 2014).



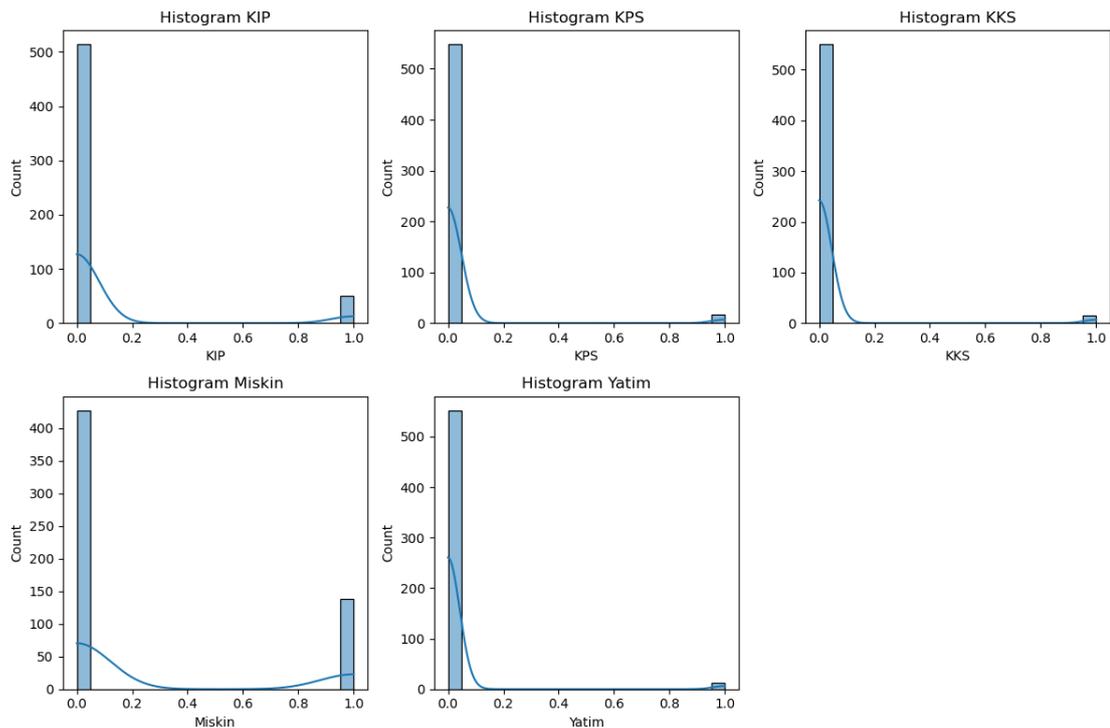
Gambar 4. Feature importance dalam random forest untuk penerima KIP

Gambar 4 menunjukkan bahwa Fitur KIP memiliki tingkat kepentingan sebesar 0.9, sementara fitur KPS, KKS, Miskin, dan Yatim memiliki tingkat kepentingan di rentang 0.0 - 0.2.

Tingginya tingkat kepentingan pada fitur KIP menunjukkan bahwa fitur tersebut adalah faktor yang sangat penting dalam membuat prediksi model. Di sisi lain, fitur KPS, KKS, Miskin, dan Yatim memiliki tingkat kepentingan yang lebih rendah, yang mungkin menunjukkan bahwa kontribusi mereka terhadap prediksi KIP tidak sekuat fitur KIP.

C. Histogram

Random Forest menggunakan Histogram of an Oriented Gradient (HOG) untuk mengekstraksi fitur dari gambar guna mengklasifikasikan daun yang sakit dan sehat (Ramesh, et al., 2018).



Gambar 5. Histogram dalam random forest untuk penerima KIP

Histogram pada gambar 5 menampilkan data dari beberapa variabel antara lain :

- **Histogram untuk Variabel KIP:**

Rentang 0.2 dengan count 100-200, stagnan.

Rentang 0.8 - 1.0 dengan count 0-100.

Histogram ini menunjukkan distribusi variabel KIP. Jika nilai KIP mencakup rentang 0.2, dengan count yang stagnan di kisaran 100-200, ini dapat diartikan bahwa sebagian besar siswa dalam dataset memiliki nilai KIP di kisaran tersebut. Sementara itu, di rentang 0.8 - 1.0, count di kisaran 0-100 menunjukkan bahwa ada sejumlah siswa dengan nilai KIP yang lebih tinggi.

- **Histogram untuk Variabel KPS, KKS, Miskin, dan Yatim:**

Setiap variabel memiliki rentang 0.2 dengan count yang stagnan di kisaran tertentu. Setiap variabel memiliki rentang 0.8 - 1.0 dengan count yang relatif lebih rendah di kisaran tertentu.

Histogram-histogram ini memberikan gambaran distribusi nilai untuk masing-masing variabel. Perhatikan bahwa interpretasi yang tepat akan tergantung pada satuan atau skala yang digunakan untuk setiap variabel tersebut.

KESIMPULAN

Penelitian ini bertujuan untuk mengembangkan model prediksi menggunakan algoritma Random Forest untuk mengidentifikasi siswa yang berpotensi menjadi penerima Kartu Indonesia Pintar (KIP). Hasil analisis dan visualisasi menunjukkan beberapa temuan yang dapat merangkum kesimpulan dari penelitian ini:

1. Kinerja Model

Model memiliki 126 True Negative (TN) dan 16 False Positive (FP) pada confusion matrix. Ini mengindikasikan bahwa model memiliki kecenderungan untuk memprediksi siswa tidak mendapatkan KIP dengan benar, tetapi memiliki sejumlah kasus di mana siswa diprediksi mendapatkan KIP padahal sebenarnya tidak.

2. Feature Importance

Fitur KIP memiliki tingkat kepentingan yang tinggi (0.9), menunjukkan bahwa informasi mengenai status penerimaan KIP sebelumnya sangat memengaruhi prediksi model. Sementara itu, fitur KPS, KKS, Miskin, dan Yatim memiliki tingkat kepentingan yang lebih rendah.

3. Distribusi Nilai Variabel

- Histogram KIP menunjukkan bahwa sebagian besar siswa dalam dataset memiliki nilai KIP di kisaran 0.2 dengan count stagnan di 100-200, sementara count di rentang 0.8 - 1.0 lebih rendah.

- Histogram untuk variabel KPS, KKS, Miskin, dan Yatim menunjukkan distribusi nilai yang beragam dengan rentang 0.2 dan 0.8 - 1.0, di mana nilai cenderung stagnan pada beberapa kisaran.

4. Rekomendasi Pengembangan Model

Evaluasi model lebih lanjut, termasuk perhitungan akurasi, presisi, recall, dan F1-score, dapat memberikan pemahaman lebih mendalam tentang kinerja model dan membimbing langkah-langkah perbaikan. Penyesuaian threshold prediksi atau pengoptimalan hyperparameter model mungkin diperlukan untuk meningkatkan keseimbangan antara False Positive dan True Negative.

Dengan kesimpulan ini, penelitian ini memberikan kontribusi untuk pemahaman lebih lanjut tentang faktor-faktor yang mempengaruhi prediksi penerimaan KIP pada siswa. Rekomendasi pengembangan model dan analisis faktor pengaruh dapat menjadi dasar untuk perbaikan lebih lanjut dalam pengelolaan Program Indonesia Pintar. Kesimpulan ini diharapkan dapat memberikan panduan bagi kebijakan pendidikan yang lebih efisien dan efektif di masa depan.

REFERENSI

- Faraji, Z. (2022). A Review of Machine Learning Applications for Credit Card Fraud Detection with A Case study. *SEISENSE Journal of Management*, 2617-5770.
- Firdaus, U., & Utama, D. N. (2020). Balance as one of the attributes in the customer segmentation analysis method: Systematic literature . *ASTES*, 2415-6698.
- Freund, Y., & Schapire, R. E. (1996). *Experiments with a New Boosting Algorithm*. New Jersey: International Conference on Machine Learning.
- Mahapatra, D. (2014). Analyzing Training Information From Random Forests for Improved Image Segmentation. *IEEE*, 1504-1512.
- Ramesh, S., Hebbar, R., M., N., R., P., N., P. B., N., S., & P.V., V. (2018). Plant Disease Detection Using Machine Learning. *IEEE*.
- Salamah, A. H., Tamazin, M., Sharkas, M. A., & Khedr, M. (2016). An enhanced WiFi indoor localization system based on machine learning. *IEEE*, 4-7.
- Sarab, A., & Mohamed, E. B. (2016). Impact of preprocessing on medical data classification. *Frontiers of Computer Science*, 1082-1102.
- Yayan, A., Sri Wulan, A., Unika, W., & Nizmah, M. S. (2019). PENTINGNYA PENDIDIKAN BAGI MANUSIA. *Jurnal Buana Pengabdian*, 1-19.