

Penerapan Machine Learning Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree

Application of Machine Learning to Predict Student Graduation Using the Decision Tree Algorithm

Berlian Nurul Aisyah¹, Indra Gunawan^{*2}

^{1,2}Program Studi Informatika, STT Ronggolawe Cepu

e-mail: igunsttr@gmail.com

Abstrak – Penelitian ini fokus pada penerapan Machine Learning (ML), sebuah cabang dari Artificial Intelligence (AI), untuk mengetahui prediksi kelulusan mahasiswa. Permasalahan penelitian ini yaitu bagaimana memprediksi masa studi mahasiswa menggunakan Algoritma Decision Tree dengan Bahasa Pemrograman Python dan framework Django. Tujuan penelitian ini adalah untuk melakukan prediksi pada masa studi mahasiswa dengan menggunakan Machine Learning sebagai dasar pertimbangan dalam mengambil keputusan kelulusan mahasiswa. Metode Synthetic Minority Over-sampling Technique (SMOTE) diimplementasikan pada penelitian ini untuk mengatasi ketidakseimbangan jumlah kelas target. Penelitian menggunakan dataset dari UPT TI STT Ronggolawe Cepu dengan 270 record data yang dibagi 243 data training dan 27 data testing. Dataset terdiri dari 14 fitur yaitu Umur, Jenis Kelamin, Jurusan, SKS, IP1, IP2, IP3, IP4, IP5, IP6, IP7, IP8, IPK, Beasiswa. Metode yang digunakan yaitu Algoritma Decision Tree untuk mengetahui akurasi dan efisiensi prediksi kelulusan mahasiswa. Dari hasil eksperimen yang telah dilakukan, dapat disimpulkan bahwa Algoritma Decision Tree menunjukkan akurasi dengan nilai sebesar 0,91, Area Under Curve (AUC) dengan nilai sebesar 0,91, presisi sebesar 0,92, recall sebesar 0,91. Hasil penelitian lainnya didapatkan 5 fitur terpenting yaitu Umur, Jenis Kelamin, Jurusan, SKS dan IP1 yang digunakan sebagai fitur utama pada pendeteksian data baru berbasis web. Dengan fitur terpenting tersebut, maka kelulusan mahasiswa dapat dideteksi semenjak semester pertama perkuliahan, sehingga memudahkan mahasiswa dan pihak pengambil keputusan untuk menentukan langkah strategis selanjutnya.

Kata kunci – komponen ; decision tree, kelulusan mahasiswa, klasifikasi, machine learning.

Abstract – This research applies machine learning (ML), a branch of Artificial Intelligence (AI), to determine student graduation predictions. The problem of this research is how to predict student study periods using the Decision Tree Algorithm with the Python programming language and the Django framework. This research aims to make predictions during a student's study period using machine learning as a basis for considering student's graduation decisions. The Synthetic Minority Over-sampling Technique (SMOTE) method was implemented in this research to overcome the imbalance in the number of target classes. The research uses a dataset from UPT TI STT Ronggolawe Cepu with 270 data records divided into 243 training data and 27 testing data. The dataset consists of 14 features namely Age, Gender, Major, SKS, IP1, IP2, IP3, IP4, IP5, IP6, IP7, IP8, IPK, scholarship. The method used is the Decision Tree Algorithm shows accuracy with a value of 0,91, Area Under Curve (AUC) with a value of 0,91, precision of 0,92, recall of 0,91. Another research result was that the 5 most important features were obtained namely Age, Gender, Major, SKS, and IP1 which were used as the main features in detecting new web-based data. With these most important features, student graduation can be

detected from the first semester of study, making it easier for students and decision-makers to determine the next strategic steps.

Keywords – components; decision tree, student graduation, classification, machine learning.

I. PENDAHULUAN

Machine Learning merupakan salah satu cabang Ilmu Kecerdasan Buatan dari AI yang berfokus pada pengembangan sebuah sistem yang dapat dipelajari secara otomatis tanpa perlu melakukan pemrograman berulang kali. Sebelum *Machine Learning* menghasilkan suatu hasil memerlukan beberapa data *training* dan data *testing* sebagai proses *learning* [1]. Penggunaan *Machine Learning* memiliki beberapa manfaat di bidang pendidikan. Dalam konteks pendidikan, kelulusan mahasiswa menjadi fokus penting karena deteksi dini kunci untuk mengetahui lebih awal mahasiswa lulus tepat waktu atau terlambat. Algoritma *Machine Learning* digunakan untuk memproses data dan memprediksi kelulusan mahasiswa secara akurat. Contoh penerapan machine learning pada peneliti yang telah dilakukan sebelumnya [2] yang berjudul “Optimasi Model Artificial Neural Network Untuk Klasifikasi Paket Jaringan” dan dilakukan [3] yang berjudul “Machine Learning Model For Stunting Prediction”.

Kelulusan mahasiswa merupakan hal penting dan harus dipertimbangkan karena presentase lulusan berdampak pada evaluasi pemerintah dan berdampak pada status akreditasi program studi. Tingkat kelulusan perguruan tinggi sangat penting bagi semua mahasiswa [4]. Mahasiswa dikatakan lulus tepat waktu di perguruan tinggi mampu melewati masa studi kurang dari empat tahun, sedangkan mahasiswa dikatakan tidak lulus tepat waktu apabila masa studi lebih dari empat tahun. Pada kenyataannya hingga saat ini, tidak semua mahasiswa dapat menyelesaikan pendidikan sarjana dalam kurun waktu empat tahun. Permasalahan keterlambatan tersebut menjadi beban Program Studi karena keterbatasan Program Studi untuk melakukan prediksi masa studi mahasiswa dari awal. Dengan adanya prediksi sejak dini pihak Program Studi bisa memberikan bimbingan atau arahan serta dorongan dan juga motivasi belajar agar mahasiswa tersebut bisa lulus sesuai dengan waktu yang diharapkan.

Pada penelitian sebelumnya, belum ada yang menggunakan teknik *machine learning* untuk memprediksi masa studi dengan menggunakan bahasa *Python*. *Python* adalah bahasa pemrograman dinamis dengan manajemen memori otomatis. Seperti bahasa pemrograman dinamis lainnya, *python* biasanya digunakan sebagai bahasa skrip, namun bahasa tersebut sebenarnya cocok untuk mencakup konteks penggunaan yang biasanya tidak dilakukan dalam bahasa skrip [5]. Beberapa penelitian yang telah dilakukan sebelumnya, [6] [7] yang melakukan prediksi klasifikasi kelulusan mahasiswa menggunakan algoritma *Decision Tree* dan algoritma *Naïve Bayes*. Penelitian tersebut menghitung dan membandingkan tingkat akurasi kelulusan mahasiswa dengan data 1739 *record*. Pada penelitian yang lain yang dilakukan [8] yang melakukan analisis klasifikasi kelulusan mahasiswa menggunakan algoritma *Decision Tree*. Penelitian tersebut menghitung tingkat akurasi kelulusan mahasiswa dengan data 1686 *record*.

Dalam hal ini, *Machine Learning* dapat digunakan untuk mengembangkan model klasifikasi yang dapat memprediksi tingkat akurasi kelulusan mahasiswa. Algoritma *Machine Learning* yang digunakan adalah *Decision Tree* untuk klasifikasi kelulusan mahasiswa harus mampu mengenali pola-pola dalam data dengan akurasi yang tinggi. *Decision Tree* adalah suatu struktur yang mirip dengan *flowchart*, di mana setiap node internal (node yang bukan *leaf* atau node terluar) digunakan untuk menguji variabel atribut, dan setiap cabangnya mewakili hasil dari pengujian tersebut. Sementara itu, node terluar (*leaf*) dalam *Decision Tree* menjadi label atau klasifikasi akhir [9].

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i$$

Ket: S : Himpunan kasus.

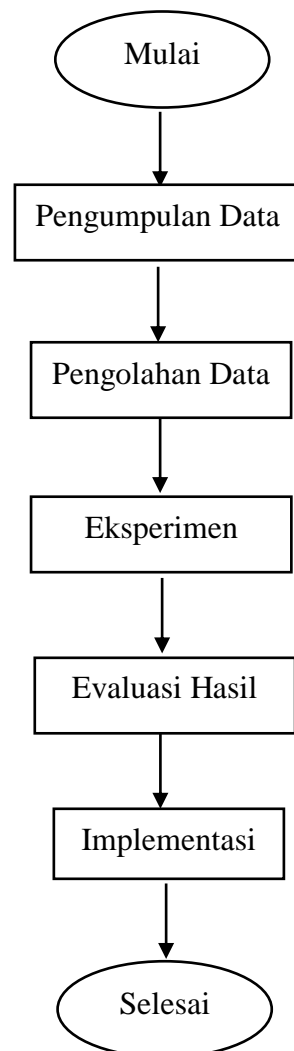
n : Jumlah partisi atribut S.

pi : Proporsi dari Si ke-i [10]

Penelitian ini bertujuan untuk melakukan prediksi awal pada pertengahan masa studi mahasiswa dengan menggunakan teknik *machine learning* sebagai dasar pertimbangan dalam mengambil keputusan kelulusan mahasiswa. Teknik yang digunakan diharapkan bisa membantu dalam mengurangi permasalahan proses prediksi dan hasil prediksi menjadi lebih baik.

II. METODE

Penyusunan penelitian ini tentunya memerlukan alur penelitian berupa kerangka kerja, yang mana kerangka kerja ini akan dijadikan pedoman urutan proses demi proses jalannya penelitian.



Gambar 1. Diagram Alur Penelitian

Pengumpulan data penelitian ini menggunakan dataset yang didapat dari UPT TI STT Ronggolawe Cepu sebagai sumber data utama, dengan 270 record data. Dataset terdiri dari 14 fitur yaitu Umur, Jenis Kelamin, Jurusan, SKS, IP1, IP2, IP3, IP4, IP5, IP6, IP7, IP8, IPK, Beasiswa. Proses pengolahan data dimulai dengan membagi dataset menjadi dua bagian, yakni data latih (90%) dan data uji (10%). Setelahnya, dilakukan langkah preprocessing untuk menyesuaikan skala data agar memiliki rata-rata nol dan deviasi standar satu. Tahap berikutnya, yaitu SMOTE, bertujuan untuk menyeimbangkan kelas data dengan menambah jumlah data. Setelah itu, proses cross-validation dilakukan sepuluh kali untuk menguji model dengan skor AUC, memastikan pengujian data diacak untuk menghindari bias.

Selanjutnya, fokus beralih ke tahap optimasi kernel linear dipilih. Pengaturan probabilitas prediksi diatur menjadi "true" agar dapat menghitung probabilitas prediksi untuk setiap kelas. Langkah berikutnya adalah pembobotan fitur untuk menentukan fitur yang paling berpengaruh terhadap tingkat kelulusan mahasiswa. Akhirnya, implementasi semua optimasi dilakukan pada algoritma *Decision Tree* dan evaluasi hasil sebagai tahap terakhir dari proses ini, dengan tujuan menghasilkan model yang efektif dalam mendeteksi kelulusan mahasiswa.

III. HASIL DAN PEMBAHASAN

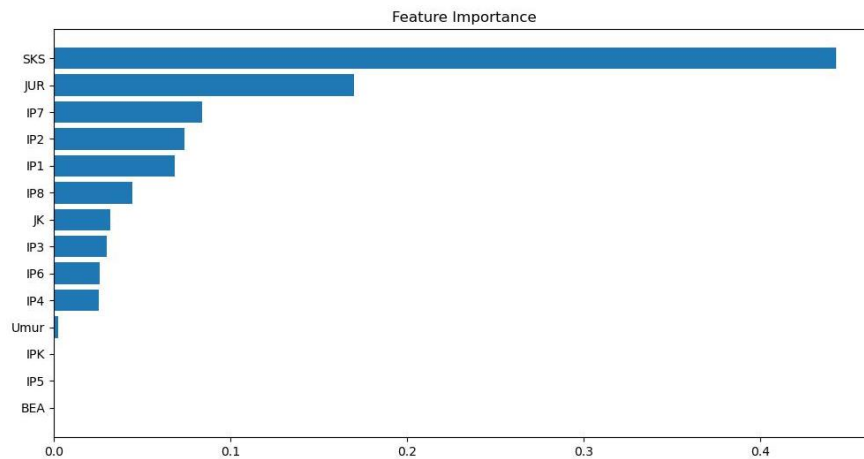
Dalam hasil ini sudah melakukan pengujian *smote* meliputi nilai accuracy, nilai precision, nilai recall dan nilai *fi-score*.

Tabel 1. Hasil Model *Decision Tree*

	<i>precision</i>	<i>recall</i>	<i>fi-score</i>	<i>support</i>
<i>Tepat Waktu</i>	0.96	0.88	0.92	26
<i>Terlambat</i>	0.87	0.95	0.91	21
<i>accuracy</i>			0.91	47
<i>macro avg</i>	0.91	0.92	0.91	47
<i>wighted avg</i>	0.92	0.91	0.92	47
<i>AUC Score</i>	0.91849			

Hasil dari model *Decision Tree* yang digunakan menunjukkan performa yang sangat baik dalam mendeteksi kelulusan mahasiswa dengan akurasi sebesar 91%. Untuk kelas "Tepat Waktu", model memiliki nilai precision sebesar 0.96 dan recall sebesar 0.88, yang berarti model cenderung memberikan prediksi yang benar untuk kasus kelulusan mahasiswa yang sebenarnya, namun mungkin ada sedikit kasus yang terlewat. Sementara untuk kelas "Terlambat", model memiliki *precision* sebesar 0.87 dan *recall* sebesar 0.95, yang menunjukkan bahwa model cenderung memberikan prediksi yang benar untuk semua kasus kelulusan yang sebenarnya tidak terlambat.

Dengan *Fi-score* sebesar 0.92 untuk kedua kelas, model ini memiliki keseimbangan yang baik antara precision dan recall untuk kedua jenis kelulusan mahasiswa. *AUC score* (*Area Under Curve*) yang cukup baik, yaitu 0.91849, menunjukkan bahwa model ini memiliki kinerja yang baik dalam memisahkan antara kelas positif dan negatif dalam pengujian. Dengan demikian, dapat disimpulkan bahwa model *Decision Tree* ini sangat baik dalam deteksi kelulusan mahasiswa dengan kinerja yang baik dan dapat diandalkan.



Gambar 2. Grafik Eksperimen

Hasil atau presentase gambar grafik ini adalah hasil pembobotan fitur yang mencerminkan tingkat pentingnya setiap fitur dalam konteks analisis atau model yang sedang digunakan. Fitur dengan presentase bobot yang lebih tinggi dianggap lebih penting atau memiliki kontribusi yang lebih besar dalam menjelaskan variabilitas data atau dalam melakukan tugas analisis yang sedang dilakukan. Fitur dengan bobot yang rendah mungkin memiliki kontribusi yang lebih kecil atau kurang signifikan. Terdapat 11 fitur yaitu SKS, Jurusan, IP7, IP2, IP1, IP8, Jenis Kelamin, IP3, IP6, IP4, Umur.

IV. KESIMPULAN

Dari hasil eksperimen yang telah dilakukan, dapat disimpulkan bahwa *Algoritma Decision Tree* menunjukkan akurasi dengan nilai sebesar 0,91, *Area Under Curve (AUC)* dengan nilai sebesar 0,91, presisi sebesar 0,92, recall sebesar 0,91. Hasil penelitian lainnya didapatkan 5 fitur terpenting yaitu Umur, Jenis Kelamin, Jurusan, SKS dan IP1 yang digunakan sebagai fitur utama pada pendeteksian data baru berbasis web. Dengan fitur terpenting tersebut, maka kelulusan mahasiswa dapat dideteksi semenjak semester pertama perkuliahan, sehingga memudahkan mahasiswa dan pihak pengambil keputusan untuk menentukan langkah strategis selanjutnya.

DAFTAR PUSTAKA

- [1] C. Chazar and B. Erawan, "Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine," *Inf. (Jurnal Inform. dan Sist. Informasi)*, vol. 12, no. 1, pp. 67–80, 2020, doi: 10.37424/informasi.v12i1.48.
- [2] I. Gunawan, "Optimasi Model Artificial Neural Network untuk Klasifikasi Paket Jaringan," *Simetris*, vol. 14, no. 2, pp. 1–5, 2020, doi: 10.51901/simetris.v14i2.135.
- [3] S. Sutarmi, W. Warijan, T. Indrayana, D. P. P. B, and I. Gunawan, "Machine Learning Model For Stunting Prediction," *J. Heal. Sains*, vol. 4, no. 9, pp. 10–23, 2023, doi: 10.46799/jhs.v4i9.1073.
- [4] R. Thaniket, Kusriani, and E. T. Luthf, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma Support Vector Machine," *J. FATEKSA J. Teknol. dan rekayasa*, vol. 13, no. 2, pp. 69–83, 2019.
- [5] A. N. Syahrudin and T. Kurniawan, "Input dan Output pada Bahasa Pemrograman Python," *J. Dasar Pemrograman Python STMIK*, no. June 2018, pp. 1–7, 2018, [Online]. Available: <https://www.researchgate.net/publication/338385483>
- [6] M. R. Qisthiano, P. P. Armilia, and I. Ruswita, "Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa," *G-Tech J. Teknol. Terap.*, vol. 6, no. 2, pp. 295–305, 2022, [Online]. Available: <https://ejournal.uniramalang.ac.id/index.php/g-tech/article/view/1823/1229>

- [7] M. R. Qisthiano, "PPENERAPAN MODEL KLASIFIKASI NAÏVE BAYES UNTUK PREDIKSI MAHASISWA TEPAT WAKTU," pp. 164–168, 2023.
- [8] U. Suriani, "Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5," *Journalcisa*, vol. 3, no. 2, pp. 55–66, 2023, [Online]. Available: <http://jesik.web.id/index.php/jesik/article/view/91>
- [9] I. Sutoyo, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik," *J. Pilar Nusa Mandiri*, vol. 14, no. 2, p. 217, 2018, doi: 10.33480/pilar.v14i2.926.
- [10] I. N. Kholifah, "Memprediksi Tingkat Kelulusan Peserta Kursus Pertama dengan Algoritma Data Mining C4.5 dan RapidMiner," *J. Inform. dan Sist. Inf.*, vol. 6, no. 1, pp. 1–10, 2020.