

1. Pendahuluan

Pendidikan merupakan salah satu pilar penting dalam pembangunan sumber daya manusia yang berkualitas. Di era digital saat ini, perkembangan teknologi informasi dan komunikasi telah memberikan dampak signifikan terhadap dunia pendidikan. Salah satu perubahan besar adalah kemampuan untuk mengumpulkan dan menganalisis data aktivitas siswa secara lebih efektif dan efisien. Data ini mencakup berbagai aspek, seperti kehadiran, partisipasi dalam diskusi, hasil tugas, dan aktivitas online lainnya yang dapat memberikan wawasan berharga tentang perilaku belajar siswa dan faktor-faktor yang mempengaruhi kelulusan mereka.

Dengan meningkatnya ketersediaan data ini, teknik data mining telah berkembang pesat sebagai alat untuk mengidentifikasi pola-pola penting dalam data pendidikan. Salah satu algoritma yang sering digunakan dalam Data Mining adalah Decision Tree C4.5, yang terkenal dengan kemampuannya dalam menangani data yang kompleks dan memberikan hasil yang mudah dipahami. Algoritma ini dapat membantu dalam mengklasifikasikan siswa berdasarkan kemungkinan kelulusan mereka dengan menganalisis catatan aktivitas mereka.

Penelitian ini berfokus pada penggunaan algoritma Decision Tree C4.5 untuk mengembangkan model klasifikasi kelulusan siswa berdasarkan catatan aktivitas menggunakan dataset OULAD (Open University Learning Analytics Dataset). Dataset ini menyediakan data komprehensif tentang aktivitas belajar mahasiswa di Open University, sebuah institusi pendidikan jarak jauh. Dengan menggunakan data ini, penelitian ini bertujuan untuk mengidentifikasi faktor-faktor kunci yang mempengaruhi kelulusan dan mengembangkan model prediksi yang dapat membantu institusi pendidikan dalam mendeteksi siswa yang berisiko gagal lebih awal.

Latar Belakang

Dalam dunia pendidikan, keberhasilan akademik mahasiswa menjadi salah satu indikator penting dalam menilai efektivitas sistem pembelajaran. Seiring dengan perkembangan teknologi, data aktivitas mahasiswa yang dihasilkan dari sistem pembelajaran daring maupun luring dapat dianalisis untuk mengidentifikasi faktor-faktor yang berkontribusi terhadap kelulusan. Masalah utama dalam penelitian ini adalah bagaimana mengolah dan menganalisis 29.199 baris data aktivitas mahasiswa agar dapat memprediksi kelulusan dengan akurat. Salah satu masalah yang sering dihadapi adalah menentukan apakah seorang mahasiswa lulus atau tidak berdasarkan data aktivitas yang tercatat. Oleh karena itu, Penelitian ini bertujuan untuk memprediksi kelulusan mahasiswa berdasarkan data aktivitas mereka di Virtual Learning Environment (VLE) menggunakan algoritma Decision Tree C4.5.

Penggunaan metode klasifikasi telah berkembang pesat dan menawarkan berbagai metode untuk menganalisis data pendidikan guna meningkatkan proses belajar-mengajar. Metode tersebut terkenal dengan kemampuannya dalam mengelompokkan data untuk mempermudah proses prediksi kelulusan mahasiswa menggunakan Algoritma C4.5. Salah satu penerapannya dalam dunia pendidikan adalah untuk memprediksi kelulusan mahasiswa berdasarkan berbagai faktor, seperti catatan aktivitas akademik, nilai dan kehadiran dalam suatu dataset.

Penelitian sebelumnya telah menggunakan berbagai metode untuk menyelesaikan masalah serupa. Misalnya, Suriani menggunakan algoritma Decision Tree C4.5 untuk memprediksi tingkat kelulusan mahasiswa [1]. Sementara itu, Aprilia dan Amanda memanfaatkan teknik data mining seperti klasifikasi untuk memprediksi kelulusan santri Tahfidz Qur'an [11]. Meskipun metode-metode ini memberikan hasil yang signifikan, mereka seringkali memerlukan interpretasi yang kompleks dan tidak selalu memberikan wawasan yang mudah dipahami oleh pengelola pendidikan.

Decision Tree C4.5 mampu mengidentifikasi pola-pola penting dari data aktivitas mahasiswa yang dapat digunakan untuk memprediksi kelulusannya. Beberapa penelitian telah menunjukkan bahwa faktor-faktor seperti kehadiran, partisipasi aktif, dan kinerja tugas memiliki pengaruh signifikan terhadap hasil akademik siswa [3]. Penelitian ini mengusulkan penggunaan metode Decision Tree C4.5 untuk mengklasifikasikan kelulusan mahasiswa berdasarkan data aktivitas mahasiswa, menggunakan dataset OULAD (Open University Learning Analytics Dataset), dengan kolom-kolom seperti jumlah klik (`sum_click`) pada Virtual Learning Environment dan nilai hasil dari modul yang diikuti (`final_result`) yang dipakai dalam metode Decision Tree C4.5 dipilih karena kemampuannya dalam menangani data kualitatif dan kuantitatif serta memberikan hasil yang mudah diinterpretasikan. Berdasarkan penelitian yang dibuatkan oleh Mubarak, Cao dan Zhang dengan judul "*Prediction of students' early dropout based on their interaction logs in online learning environment*", dikatakan bahwa interaksi mahasiswa adalah sumber utama saat belajar secara online, seperti dalam Virtual Learning Environment [17]. Beberapa dataset dari dalam OULAD yang dipakai Alasan mengapa memakai metode ini, karena memecah proses pengambilan keputusan yang kompleks menjadi penilaian yang lebih mudah dipahami untuk mencapai solusi yang mendekati hasil yang diinginkan [14]. Keunggulan utama dari Decision Tree C4.5 dibandingkan dengan metode lain seperti regresi logistik atau neural networks adalah kemampuannya untuk menghasilkan struktur pohon yang jelas dan mudah dipahami, yang memungkinkan pengelola pendidikan untuk dengan cepat mengidentifikasi faktor-faktor kunci yang mempengaruhi kelulusan mahasiswa [2].

Topik dan Batasannya

Penelitian ini bertujuan untuk memprediksi kelulusan mahasiswa berdasarkan data aktivitas belajar mereka,

seperti kolom bernama `sum_click`, yang berisi jumlah klik dari setiap mahasiswa dalam Virtual learning Environment dan `final_result`, yang berisi nilai akhir dari setiap mahasiswa untuk modul yang diikuti. Kolom tersebut dipakai karena mengetahui sikap mahasiswa dalam Virtual Learning Environment berdasarkan apakah mahasiswa aktif atau tidak.. Tantangan utama adalah bagaimana mengolah dan menganalisis data aktivitas sebesar 29.199 baris untuk memprediksi kelulusan mahasiswa, sehingga institusi pendidikan dapat memberikan intervensi tepat waktu kepada mahasiswa yang berisiko gagal dengan menggunakan metode Decision Tree C4.5 dari dataset OULAD (Open University Learning Analytics Dataset).

Untuk memperjelas cakupan penelitian dan memastikan fokus yang tepat, berikut adalah beberapa batasan. Pertama, jenis data yang diambil hanyalah data-data dari VLE atau *Virtual Learning Environment*. Data yang diluar VLE seperti catatan fisik dari kelas tatap muka atau secara daring (Zoom meeting) tidak digunakan. Kedua, metode klasifikasi pada penelitian ini terbatas hanya pada penggunaan Algoritma C4.5. Ketiga, untuk evaluasi model dalam penelitian ini dilakukan menggunakan dataset yang tersedia, yang dibatasi oleh variasi aktivitas pada VLE..Metode evaluasi yang dipakai adalah perhitungan akurasi, *precision*, *F1-score* dan *recall*.

Tujuan

Tujuan dari tugas akhir ini adalah untuk menerapkan dan mengimplementasikan model klasifikasi yang mampu memprediksi kelulusan mahasiswa berdasarkan aktivitas mereka selama perkuliahan. Algoritma C4.5 dipilih karena kemampuannya dalam menghasilkan pohon keputusan yang dapat diinterpretasikan dengan baik, sehingga memudahkan dalam memahami faktor-faktor yang mempengaruhi kelulusan. Pengukuran keberhasilan dari penelitian ini dapat dilakukan dengan mengevaluasi kinerja model menggunakan metrik-metrik seperti akurasi, presisi, recall, dan F1-score pada data uji. Selain itu, validasi model juga dapat dilakukan melalui cross-validation untuk memastikan model tidak overfitting.

Organisasi Tulisan

Bagian-bagian selanjutnya adalah tinjauan pustaka, yang isinya adalah referensi-referensi dari data mining dan *Decision Tree* terkait dengan penelitian ini. Kedua adalah Sistem yang dibangun, yang isinya adalah sistem yang dibangun dalam penelitian ini. Ketiga adalah evaluasi, yang isinya adalah analisis dan penjelasan dari hasil penelitian ini. Dan terakhir adalah kesimpulan, yang isinya adalah kesimpulan dan saran dari hasil penelitian dan analisis penelitian.