

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan komunikasi (TIK) telah mengubah cara lembaga pendidikan tinggi mengelola data dan mengambil keputusan strategis. Perguruan tinggi dituntut untuk mengoptimalkan pemanfaatan data dalam mendukung efisiensi administrasi dan transparansi kebijakan, termasuk dalam aspek pembiayaan pendidikan [1]. Penerapan analitik data dan sistem prediktif menjadi langkah penting dalam menciptakan tata kelola pendidikan yang lebih objektif dan berbasis bukti. Dalam konteks digitalisasi ini, pengelolaan data sosial ekonomi mahasiswa menjadi bagian yang krusial, terutama karena berkaitan langsung dengan kebijakan biaya pendidikan yang berimplikasi sosial luas.

Salah satu kebijakan strategis yang masih menimbulkan permasalahan di banyak perguruan tinggi adalah penentuan Uang Kuliah Tunggal (UKT). UKT ditujukan untuk menyesuaikan besaran biaya kuliah dengan kemampuan ekonomi mahasiswa dan keluarganya. Namun, proses penentuan kategori UKT di sejumlah perguruan tinggi masih dilakukan secara manual atau semi-manual, bergantung pada verifikasi dokumen dan penilaian subjektif [2]. Kondisi tersebut sering menyebabkan ketidaktepatan dalam klasifikasi, di mana mahasiswa dengan kondisi ekonomi serupa dapat memperoleh kategori yang berbeda. Akibatnya, muncul persepsi ketidakadilan dan menurunnya kepercayaan terhadap sistem pembiayaan kampus.

Berdasarkan laporan Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (2024), sekitar 63% perguruan tinggi negeri di Indonesia belum sepenuhnya menerapkan sistem digital untuk verifikasi dan klasifikasi UKT. Ketergantungan pada metode manual mengakibatkan proses panjang, potensi kesalahan input data, dan kesulitan mempertahankan konsistensi hasil antarperiode penerimaan mahasiswa. Di

sisi lain, jumlah mahasiswa baru yang terus meningkat menambah kompleksitas pengelolaan data ekonomi keluarga. Untuk itu, dibutuhkan pendekatan berbasis data yang mampu membantu universitas melakukan klasifikasi secara lebih cepat, objektif, dan akurat.

Pendekatan machine learning memberikan peluang besar dalam mendukung pengambilan keputusan administratif berbasis data. Melalui analisis data historis, algoritma dapat mengenali pola sosial ekonomi mahasiswa yang sulit diidentifikasi secara manual dan memprediksi kategori UKT berdasarkan atribut sosial ekonomi mahasiswa yang relevan [3]. Penelitian sebelumnya menunjukkan bahwa penggunaan pembelajaran mesin dalam konteks pendidikan dapat meningkatkan efisiensi dan akurasi prediksi dibandingkan metode konvensional [1]. Namun, sebagian besar studi masih berfokus pada satu algoritma tertentu, tanpa mengombinasikan kelebihan dari beberapa metode untuk meningkatkan kinerja model.

Kesenjangan penelitian (*research gap*) muncul karena masih minimnya kajian yang mengintegrasikan reduksi dimensi, pengelompokan data, dan klasifikasi probabilistik dalam satu sistem prediksi UKT yang utuh. Beberapa penelitian seperti yang dilakukan oleh Yustanti et al. (2025) menunjukkan peningkatan akurasi melalui teknik cluster relabeling, namun belum menggabungkan tahapan reduksi dimensi yang penting untuk menyaring atribut yang redundan. Sementara itu, penelitian Farnoosh et al. (2025) di bidang medis membuktikan bahwa hybrid model yang menggabungkan clustering dan Classifier dapat meningkatkan akurasi prediksi secara signifikan. Berdasarkan temuan tersebut, penelitian ini mengusulkan hybrid model berbasis Principal Component Analysis (PCA), K-Means, dan Naïve Bayes untuk menghasilkan klasifikasi UKT yang lebih efisien dan akurat.

Penelitian ini diharapkan dapat memberikan kontribusi akademik dan praktis bagi sistem pendidikan tinggi di Indonesia. Secara akademik, penelitian ini memperkaya kajian dalam bidang educational data mining melalui penerapan hybrid model machine learning pada kasus sosial ekonomi mahasiswa. Secara praktis, model prediksi yang dihasilkan diharapkan dapat membantu perguruan tinggi menetapkan

kategori UKT secara lebih objektif dan transparan, sehingga mendukung kebijakan pembiayaan yang adil dan berkelanjutan. Dengan demikian, penelitian ini berperan dalam memperkuat sistem pengambilan keputusan berbasis data di lingkungan pendidikan tinggi nasional.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, rumusan masalah penelitian ini adalah sebagai berikut:

1. Bagaimana karakteristik dan pengelompokan mahasiswa berdasarkan kondisi sosial ekonomi dapat dimanfaatkan untuk mendukung prediksi kategori Uang Kuliah Tunggal (UKT)?
2. Bagaimana peran PCA dalam mereduksi dimensi data sosial ekonomi mahasiswa untuk mendukung proses penentuan kategori UKT?
3. Bagaimana pengelompokan mahasiswa berdasarkan kesamaan kondisi sosial ekonomi melalui algoritma K-Means dapat digunakan sebagai dasar relabeling data sebelum proses klasifikasi UKT?
4. Bagaimana kinerja model Naïve Bayes dalam memprediksi kategori UKT mahasiswa berdasarkan hasil klasterisasi?
5. Bagaimana integrasi PCA, K-Means, dan Naïve Bayes dalam hybrid model dapat meningkatkan akurasi dan efisiensi prediksi kategori UKT mahasiswa?

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Menganalisis karakteristik dan pengelompokan mahasiswa berdasarkan kondisi sosial ekonomi sebagai dasar dalam mendukung proses prediksi kategori Uang Kuliah Tunggal (UKT).
2. Menganalisis peran Principal Component Analysis (PCA) dalam mereduksi dimensi data sosial ekonomi mahasiswa dengan tetap mempertahankan

informasi penting yang relevan untuk mendukung proses penentuan atau prediksi kategori UKT.

3. Menganalisis pemanfaatan algoritma K-Means dalam mengelompokkan mahasiswa berdasarkan kesamaan kondisi sosial ekonomi serta memanfaatkan hasil pengelompokan tersebut sebagai dasar relabeling data sebelum proses klasifikasi kategori UKT.
4. Menganalisis kinerja algoritma Naïve Bayes dalam memprediksi kategori UKT mahasiswa berdasarkan data yang telah dilakukan relabeling sesuai dengan hasil klasterisasi K-Means.
5. Menganalisis kinerja hybrid model yang mengintegrasikan PCA, K-Means, dan Naïve Bayes dalam meningkatkan akurasi dan efisiensi prediksi kategori UKT mahasiswa.

1.4 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat secara teoretis dan praktis sebagai berikut:

1. Memberikan kontribusi terhadap pengembangan ilmu di bidang data mining dan machine learning melalui penerapan hybrid model PCA, K-Means, dan Naïve Bayes dalam prediksi kategori UKT mahasiswa.
2. Meningkatkan efisiensi dan akurasi proses penentuan UKT melalui penerapan teknik reduksi dimensi dan pengelompokan data sosial ekonomi mahasiswa secara lebih objektif, sehingga hasil klasifikasi UKT menjadi lebih representatif terhadap kondisi sebenarnya.
3. Mendukung proses pengambilan keputusan yang lebih transparan dan adil dalam administrasi kampus, khususnya dalam kebijakan keuangan dan penetapan biaya pendidikan, dengan memanfaatkan hasil prediksi yang dihasilkan oleh hybrid model.