

BAB I

PENDAHULUAN

1.1. Latar Belakang

Tingkat putus kuliah di perguruan tinggi, khususnya di wilayah Jawa Timur, masih cukup tinggi. Menurut data terbaru dari Pangkalan Data Pendidikan Tinggi (PDDIKTI), dari sekitar 9,6 juta mahasiswa aktif di seluruh Indonesia, ribuan di antaranya mengalami putus studi setiap tahun. Di Jawa Timur, persentase mahasiswa yang drop-out mencapai sekitar 14,84%, menjadikannya salah satu yang tertinggi di tanah air. Kondisi ini tidak hanya berdampak negatif pada mahasiswa yang gagal menyelesaikan pendidikan, tetapi juga dapat merusak reputasi institusi serta menurunkan peringkat akreditasi perguruan tinggi. Oleh sebab itu, pengelolaan angka putus kuliah menjadi fokus utama bagi para pengelola pendidikan tinggi.

Prediksi dini terhadap potensi *drop-out* menjadi sangat penting agar institusi dapat melakukan mitigasi dan pencegahan yang efektif. Saat ini, banyak perguruan tinggi masih mengandalkan metode manual dalam mengidentifikasi mahasiswa berisiko, misalnya dengan melihat nilai Indeks Prestasi Kumulatif (IPK) tertentu atau catatan akademik lainnya. Metode ini seringkali tidak cukup akurat dan terlambat, sehingga intervensi yang diberikan pun tidak optimal. Dengan prediksi yang lebih cepat dan tepat, institusi dapat memberikan perhatian khusus, seperti bimbingan akademik dan dukungan finansial, kepada mahasiswa yang membutuhkan sehingga dapat mengurangi angka *drop-out*.

Berbagai penelitian sebelumnya telah menguji metode klasifikasi untuk memprediksi *drop-out* maupun kelulusan mahasiswa. Contohnya, algoritma K-Nearest Neighbor (KNN) telah banyak digunakan karena kesederhanaannya dalam mengklasifikasikan data berdasarkan kedekatan jarak antar sampel. Namun, KNN memiliki kelemahan seperti sensitif terhadap *outlier* dan kurang efektif saat data berdimensi tinggi atau tidak seimbang. Penelitian lain menggunakan Principal Component Analysis (PCA) sebagai metode reduksi dimensi untuk menyederhanakan data, sehingga meningkatkan kinerja algoritma klasifikasi seperti KNN. PCA berperan dalam menghilangkan fitur-fitur yang kurang signifikan sekaligus mempertahankan informasi utama yang terkandung dalam data. Meski begitu, algoritma KNN dan PCA memiliki keterbatasan tersendiri. KNN yang sensitif terhadap *noise* dan PCA yang menghilangkan beberapa informasi minor bisa menyebabkan penurunan akurasi jika tidak dipadukan dengan teknik lain. Untuk mengatasi

kelemahan tersebut, dikembangkan algoritma Modified K-Nearest Neighbor (MKNN) yang menambahkan tahap validasi data latih serta pemberian bobot voting (weight voting). Penelitian ini mengacu pada beberapa studi sebelumnya, salah satunya adalah karya Cholissodin dan kolega (2021) yang mengevaluasi klasifikasi tingkat penyebaran COVID-19 guna mendukung upaya mitigasi dengan menggunakan metode MKNN. Dalam penelitian tersebut, data yang digunakan merupakan rekap perkembangan kasus COVID-19 dari setiap provinsi di Indonesia yang bersumber dari situs resmi Kementerian Kesehatan (KEMENKES). Total sampel yang dipakai berjumlah 510 dengan tiga fitur, serta pembagian data latih dan uji masing-masing sebesar 70% dan 30%. Hasil penelitian menunjukkan akurasi tertinggi mencapai 97%, dengan nilai k pada MKNN sebesar 3 dan 5. MKNN mempertimbangkan validitas tetangga terdekat sehingga menghasilkan klasifikasi yang lebih stabil dan akurat[3]. Kombinasi MKNN dengan PCA diharapkan mampu menciptakan model yang efisien serta memiliki akurasi dan presisi lebih baik dalam memprediksi potensi drop-out mahasiswa..

Namun, masih sedikit penelitian yang mengintegrasikan PCA dan MKNN secara bersamaan untuk kasus klasifikasi kelulusan mahasiswa di Indonesia, khususnya dengan implementasi API. Penelitian ini bertujuan mengembangkan model klasifikasi yang menggabungkan MKNN dengan PCA untuk mengurangi dimensi fitur dan meningkatkan akurasi prediksi *drop-out*. Dengan data masukan berupa atribut akademik, sosial, dan demografis mahasiswa, model ini diharapkan dapat memberikan output klasifikasi risiko *drop-out* yang dapat digunakan oleh perguruan tinggi untuk intervensi tepat waktu. Diharapkan model ini tidak hanya mampu meningkatkan tingkat akurasi dan presisi dalam prediksi, tetapi juga memudahkan proses pengambilan keputusan di bidang akademik serta secara signifikan menurunkan angka putus kuliah.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang yang telah dijelaskan, maka dapat dirumuskan beberapa masalah sebagai berikut :

1. Bagaimana menerapkan metode *Principal Component Analysis* (PCA) untuk mereduksi dimensi dataset mahasiswa guna meningkatkan efisiensi klasifikasi kelulusan?
2. Bagaimana metode *Modified K-Nearest Neighbor* (MKNN) dapat digunakan untuk meningkatkan akurasi klasifikasi kelulusan mahasiswa?

3. Seberapa baik performa kombinasi metode *Principal Component Analysis* (PCA) dan *Modified K-Nearest Neighbor* (MKNN) dalam mengklasifikasikan kelulusan mahasiswa?

1.3. Batasan Masalah

Penulis mengajukan beberapa batasan masalah dalam penelitian ini, yang akan dijelaskan sebagai berikut:

1. Model klasifikasi dalam penelitian ini dikembangkan menggunakan bahasa pemrograman Python.
2. Klasifikasi dilakukan berdasarkan data karakteristik mahasiswa yang dibagi menjadi dua kategori, yaitu mahasiswa yang *drop-out* dan mahasiswa yang lulus (tidak *drop-out*).
3. Dataset yang dipakai dalam penelitian ini merupakan kumpulan data mahasiswa yang diperoleh dari platform *Kaggle*.
4. Dalam algoritma klasifikasi *Modified K-Nearest Neighbor* (MKNN), perhitungan jarak (k) dilakukan dengan menggunakan metode *Euclidean Distance*.
5. Output penelitian ini akan diimplementasikan dalam bentuk API (*Application Programming Interface*) berbasis Python.

1.4. Tujuan Penelitian

Berdasarkan permasalahan yang telah dirumuskan, tujuan utama penelitian ini adalah mengembangkan model klasifikasi kelulusan mahasiswa dengan menggabungkan *Principal Component Analysis* (PCA) dan *Modified K-Nearest Neighbor* (MKNN), yang kemudian diimplementasikan dalam bentuk API (*Application Programming Interface*) berbasis GUI (*Graphical User Interface*). Selain membangun model tersebut, penelitian ini juga bertujuan untuk mengukur kinerjanya secara kuantitatif menggunakan metrik tertentu, sehingga dapat membantu institusi pendidikan dalam mengidentifikasi mahasiswa yang berpotensi mengalami *drop-out* dengan akurat dan efisien untuk langkah-langkah pencegahan.

1.5. Manfaat Penelitian

Penelitian ini memberikan sejumlah manfaat penting. Pertama, dengan melakukan evaluasi terhadap tingkat akurasi algoritma Modified K-Nearest Neighbor (MKNN) yang diperbaiki menggunakan Principal Component Analysis (PCA), penelitian ini mampu meningkatkan efektivitas dalam mengenali mahasiswa yang berisiko mengalami *drop-out*. Hal ini membantu pihak kampus dalam membuat Keputusan yang lebih tepat dan efisien terkait upaya pencegahan *drop-out*. Manfaat lain dari penelitian ini adalah pengembangan Solusi yang lebih efektif dan efisien untuk memprediksi potensi *drop-out* mahasiswa, yang dapat mempercepat proses identifikasi dan intervensi, sehingga mendukung institusi perguruan tinggi dalam memberikan respon yang lebih cepat dan tepat sasaran terhadap mahasiswa yang berisiko *drop-out*. Dengan manfaat ini, penelitian ini berpotensi memberikan dampak positif bagi pengelolaan dan intervensi dini terhadap mahasiswa yang berpotensi *drop-out* dari kampus.

Halaman ini sengaja dikosongkan