

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Perguruan tinggi adalah lembaga pendidikan yang menyediakan program studi yang tingkat kompetensinya lebih tinggi dibandingkan dengan sekolah menengah atau sekolah dasar. Perguruan tinggi dapat berupa universitas, institut, atau akademi.

Universitas merupakan lembaga pendidikan tinggi yang bertanggung jawab dalam menyediakan sarana dan prasarana pendidikan yang berkualitas serta menyelenggarakan berbagai program studi yang berkaitan dengan ilmu pengetahuan, teknologi, sosial, dan budaya. Selain menyediakan sarana dan prasarana Pendidikan yang berkualitas, Universitas juga melakukan penelitian yang berkaitan dengan ilmu yang diajarkan di program studi. Penelitian ini berguna untuk pengabdian kepada masyarakat sesuai dengan Tridarma Perguruan Tinggi.

Program studi di Universitas merupakan komponen penting dalam sistem pendidikan tinggi yang menyediakan materi dan kurikulum sesuai dengan bidang yang dipelajari selama 4 tahun atau sekurang-kurangnya 8 semester dan batas waktu maksimal yang diberikan yaitu selama 7 tahun atau sebanyak 14 semester. Hal ini sesuai dengan himbauan Universitas Muhammadiyah Jember kepada mahasiswa program studi Teknik Informatika yang diharapkan dapat menyelesaikan studi mereka sesuai masa yang telah ditentukan. Namun, dalam proses belajar mengajar, beberapa mahasiswa mengalami kesulitan dalam menyelesaikan studi mereka dalam waktu yang telah ditentukan. Padahal, tingkat kelulusan rata-rata menentukan akreditasi setiap program studi. Hal ini dapat disebabkan oleh berbagai faktor seperti kesulitan akademis, masalah keuangan, masalah kesehatan, masalah sosial, atau masalah lainnya.

Karena itu, penting untuk melakukan penelitian tentang masa studi mahasiswa dalam rangka membantu mengidentifikasi faktor-faktor yang mempengaruhi masa studi mahasiswa dan memberikan solusi untuk membantu

mahasiswa lulus tepat waktu.

Penelitian ini menggunakan data IPS (Indeks Prestasi Mahasiswa) dari semester 1 hingga semester 6 sebagai data latih untuk mengetahui faktor-faktor yang mempengaruhi seorang mahasiswa dapat menyelesaikan studi tepat waktu. Data ini digunakan untuk mempermudah dalam mengklasifikasikan mahasiswa yang lulus tepat waktu atau tidak tepat waktu. Model yang digunakan untuk mendukung prediksi masa studi mahasiswa adalah metode *Ensemble* dengan *Base Learner* K-NN dan SVM.

Dataset ini sudah memiliki label 1 sebagai tepat waktu dan 0 sebagai tidak tepat waktu. Jumlah dari setiap label adalah 495 pada kelas 1 dan 127 pada kelas 0. Perbandingan jumlah kelas ini bisa dibilang tidak seimbang atau *imbalanced dataset* dan akan mempengaruhi performa dari model yang akan dibuat. Maka dari itu perlu dilakukan teknik resampling untuk mengatasi masalah ini.

(Indrawati, 2021) menyebutkan, ada 2 teknik *resampling* yaitu *oversampling* dan *undersampling*. *Oversampling* merupakan penambahan sejumlah data pada kelas minoritas. Sebaliknya, *undersampling* adalah pengurangan sejumlah data pada kelas mayoritas. Teknik *resampling* perlu dilakukan agar dimensi dari kedua kelas menjadi seimbang sehingga mempengaruhi performa model. Kedua pendekatan dapat diulang hingga distribusi kelas yang diinginkan tercapai dalam dataset pelatihan, seperti pembagian kelas yang sama rata.

Kedua teknik tersebut dapat digunakan untuk masalah klasifikasi dua kelas (biner) dan masalah klasifikasi multi kelas dengan satu atau lebih kelas mayoritas atau minoritas. Namun, yang utama perubahan distribusi kelas hanya diterapkan pada dataset pelatihan. Tujuannya adalah untuk mempengaruhi kecocokan model. Pengambilan sampel ulang tidak diterapkan pada kumpulan data pengujian atau penyimpanan yang digunakan untuk mengevaluasi performa model.

Menurut (Shatnawi, 2013), algoritma K-NN adalah salah satu metode *Machine Learning* yang dianggap sederhana untuk diterapkan dalam analisis data dengan banyak dimensi variabel. Algoritma K-NN digunakan untuk mengklasifikasikan objek berdasarkan data pembelajaran yang memiliki jarak terdekat dengan objek tersebut. Algoritma K-NN belum jelas nilai K mana yang

mendekati nilai sebenarnya dan parameter mana yang dapat memberikan hasil yang lebih baik ketika menghitung nilai jarak.

Untuk mengoptimalkan kinerja algoritma K-NN, metode *Ensemble* bisa digunakan. Metode ini menggabungkan hasil dari banyak model menjadi satu pendugaan akhir tanpa memilih satu model terbaik, melainkan dengan menggabungkan hasil prediksi dari berbagai model dengan bobot tertentu. Dengan menggunakan metode *Ensemble*, tidak perlu dilakukan pencarian nilai K yang optimal dalam metode K-NN (Satriya, R. H. D., Santoso, E., & Sutrisno, 2018).

Dalam penelitian ini metode *Ensemble* akan dilakukan dengan cara menggabungkan algoritma K-NN dan SVM. Kombinasi dari dua metode tentunya harus memiliki performa lebih baik daripada satu metode saja. Hal ini bisa direalisasikan dengan cara mengevaluasi tiap *Base Learner* menggunakan teknik *Cross Validation*. (Peryanto et al., 2020) mengungkapkan bahwa *K-fold Cross Validation* bisa memberikan solusi untuk penelitian ini. *K-fold Cross Validation* membagi data menjadi beberapa bagian atau *K fold* dimana setiap *fold* digunakan sebagai set pengujian di beberapa titik.

Beberapa penelitian sebelumnya yang dilakukan oleh (Bijaksana & Negara, 2022) menunjukkan bahwa sebagai pendukung proses olah data, penggunaan kombinasi K-NN dan Naïve Bayes dengan fitur seleksi *Information Gain* memberikan hasil performa klasifikasi yang terbaik. Pada evaluasi nilai akademik 4 semester, akurasi mencapai 70,41% dengan *f1-score* 80,68% dan nilai k (jarak antar data) sebesar 17, sedangkan pada evaluasi nilai akademik 7 semester, akurasi mencapai 70,14% dengan *f1-score* 80,68% dan nilai k sebesar 21. Selain itu, penggunaan algoritma Naive Bayes dengan menambahkan fitur seleksi *Information Gain* juga dapat memberikan hasil yang baik, yaitu pada evaluasi nilai akademik 4 semester dengan akurasi sebesar 67,95% dan *f1-score* 72,85%, serta pada evaluasi nilai akademik 7 semester dengan akurasi sebesar 69,32% dan *f1-score* 73,47%. (Saputra dkk, 2021) juga melakukan penelitian tentang penerapan algoritma *Modified K-Nearest Neighbour* (MKNN) pada klasifikasi masa studi mahasiswa Teknik Informatika. Hasil pengujian menunjukkan bahwa akurasi tertinggi yang diperoleh adalah sebesar 84,62%, presisi tertinggi sebesar

100%, dan *recall* tertinggi sebesar 75% dengan menggunakan nilai  $k = 3$ . Sementara itu, penelitian selanjutnya yang dilakukan oleh (Satriya, R. H. D., Santoso, E., & Sutrisno, 2018) adalah tentang implementasi metode *Ensemble* K-Nearest Neighbor untuk prediksi nilai tukar Rupiah terhadap Dollar Amerika. Data yang digunakan dalam penelitian ini terdiri dari 24 data training dan 12 data testing.

### 1.2 Rumusan Masalah

1. Bagaimana penerapan metode *Ensemble* untuk mengklasifikasi masa studi mahasiswa Teknik Informatika Universitas Muhammadiyah Jember?
2. Bagaimana kinerja metode *Ensemble* dalam mengklasifikasi akurasi, presisi, *recall* dan *f1-score* pada masa studi mahasiswa Teknik Informatika Universitas Muhammadiyah Jember?

### 1.3 Batasan masalah

1. Objek penelitian ini adalah mahasiswa Teknik Informatika Universitas Muhammadiyah Jember.
2. Metode yang digunakan dalam penelitian ini adalah metode *Ensemble* dengan *Base Learner* K-NN dan SVM.
3. Data yang digunakan adalah Mahasiswa Jurusan Teknik Informatika Universitas Muhammadiyah Jember dari Angkatan 2010 – 2016.
4. Dataset diambil dari Biro Administrasi Akademik (BAA) Universitas Muhammadiyah Jember yang berjumlah 622 data.
5. Tools yang digunakan adalah *Jupyter Notebook*.
6. Teknik *Syntethic minority oversampling technique* (SMOTE) digunakan untuk *resampling* data.

### 1.4 Tujuan Penelitian

1. Mengetahui metode yang tepat untuk mengklasifikasikan masa studi mahasiswa Teknik Informatika Universitas Muhammadiyah Jember.

2. Mengetahui efektivitas metode *Ensemble* dalam meningkatkan akurasi, presisi, *recall* dan *f1-score* pada klasifikasi masa studi mahasiswa.

### 1.5 Manfaat Penelitian

1. Dapat digunakan sebagai bahan referensi dalam kajian tentang Algoritma *Ensemble* dalam klasifikasi masa studi mahasiswa Teknik Informatika.
2. Dapat digunakan sebagai bahan evaluasi program studi Teknik Informatika Universitas Muhammadiyah Jember.

