

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perguruan tinggi merupakan penyelenggara pendidikan akademik bagi mahasiswa. Lima kelompok perguruan tinggi diantaranya adalah universitas, institut, sekolah tinggi, akademi dan politeknik. Perguruan tinggi diharapkan mampu menyelenggarakan pendidikan yang berkualitas bagi mahasiswa sehingga menghasilkan mahasiswa yang memiliki kompetensi di bidangnya. Di Indonesia kualitas suatu perguruan tinggi ditentukan oleh grade akreditasi yang dikeluarkan oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT). Banyak aspek yang dijadikan tolak ukur kualitas perguruan tinggi, salah satunya dapat dilihat dari banyaknya mahasiswa yang dapat menyelesaikan studi sesuai waktu yang telah ditentukan. Semakin banyak mahasiswa yang lulus tepat waktu maka semakin baik pula kinerja perguruan tinggi tersebut, sehingga tingkat kelulusan mahasiswa tepat waktu menjadi salah satu kriteria penilaian akreditasi bagi suatu perguruan tinggi atau program studi (Legowo & Indiarjo, 2017). Oleh karena itu, kondisi dan status mahasiswa perlu dianalisa dengan seksama sehingga dapat diberikan penanganan yang tepat untuk meningkatkan dan menjaga kualitas perguruan tinggi.

Program studi Pendidikan Teknik Informatika Universitas Pendidikan Ganesha (PTI Undiksha) berada di bawah naungan Jurusan Teknik Informatika, Fakultas Teknik dan Kejuruan, Universitas Pendidikan Ganesha, Singaraja, Bali.

Berdasarkan hasil observasi yang dilakukan oleh peneliti pada data mahasiswa program studi PTI Undiksha angkatan 2014-2017 dengan jumlah data sebanyak 460 data, diketahui bahwa mahasiswa yang mampu menyelesaikan pendidikan tepat waktu adalah sebanyak 50 mahasiswa, dan mahasiswa yang terlambat menyelesaikan pendidikannya adalah sebanyak 410 mahasiswa. Hal ini dapat mempengaruhi kualitas program studi, sehingga para pemangku kepentingan dan mahasiswa perlu melakukan evaluasi sedini mungkin untuk meningkatkan jumlah mahasiswa yang dapat menyelesaikan pendidikan sesuai dengan waktunya.

Tingkat kelulusan tepat waktu mahasiswa dapat diprediksi dengan menggunakan *data mining*. *Data mining* adalah pemecahan masalah dengan cara menganalisis data dan menemukan pola dalam data yang prosesnya bersifat otomatis atau semi otomatis. Metode yang dapat digunakan dalam *data mining* salah satunya adalah klasifikasi data. Klasifikasi merupakan teknik dalam *data mining* untuk mengelompokkan data berdasarkan keterikatan data terhadap data sampel. Beberapa algoritma *machine learning* sudah diimplementasikan pada penelitian-penelitian terkait, untuk melakukan prediksi kelulusan tepat waktu mahasiswa, seperti menggunakan algoritma *Naive Bayes* (Hartatik dkk., 2020), algoritma *Decision Tree* (Hendra dkk., 2020), algoritma C4.5 (Putri dkk., 2019), algoritma *Artificial Neural Networks* (Olalekan dkk., 2020), Hingga menggunakan algoritma *K-Nearest Neighbor* (Salim dkk., 2020).

K-Nearest Neighbor (KNN) merupakan metode yang efektif namun memiliki beberapa kekurangan, yaitu sangat rentan terhadap dimensi data yang tinggi serta fitur yang tidak relevan (Tékouabou Koumético & Toulmi, 2021). Untuk

mengatasi dimensi data yang tinggi dan fitur yang tidak relevan terhadap kasus, diperlukan sebuah teknik yang dikenal dengan seleksi fitur. Ini adalah metode untuk menghilangkan beberapa fitur atau variabel yang *redundant* atau data tidak relevan. Keakuratan algoritma KNN sangat dipengaruhi oleh ada tidaknya fitur yang tidak relevan pada dataset (Puspadini dkk., 2020).

Metode seleksi fitur (*feature selection*) dapat digunakan untuk mengurangi dimensi data pada dataset dan telah terbukti baik secara teori maupun praktik efektif dalam memproses data berdimensi tinggi dan meningkatkan efisiensi pembelajaran (Li dkk., 2017). Pemilihan fitur mengacu pada proses mendapatkan subset dari set fitur asli sesuai dengan kriteria pemilihan fitur tertentu dengan memilih fitur yang relevan dari dataset. Seleksi fitur berperan dalam mengompresi skala pemrosesan data, dimana fitur yang berlebihan dan tidak relevan dihapus (Cai dkk., 2018). Proses pemilihan fitur dapat dilakukan secara manual dengan mengetahui konteks permasalahan dan secara otomatis menggunakan algoritma tertentu. Beberapa algoritma yang dapat digunakan adalah *Univariate Selection*, *Recursive Feature Elimination*, *Principal Component Analysis*, *Feature Importance*, dan lain-lain (Ibnu Daqiqil Id, 2021). Dalam beberapa kasus, meskipun dimensi fitur seringkali tidak terlalu tinggi, pemilihan fitur masih memainkan peran penting seperti meningkatkan kinerja pembelajaran, mencegah *overfitting*, dan mengurangi biaya komputasi (Li dkk., 2017).

Selain mengurangi dimensi data, Hasil seleksi fitur yang baik dapat mengurangi *overfitting* dan meningkatkan akurasi karena dengan berkurangnya data yang tidak relevan berarti mengurangi kemungkinan untuk membuat model

berdasarkan data *noise*, sehingga model yang dihasilkan lebih akurat. Selain itu, seleksi fitur juga mengurangi waktu pelatihan. Lebih sedikit data yang diproses akan mengurangi kompleksitas dan waktu pembelajaran akan lebih cepat (Ibnu Daqiqil Id, 2021). Dalam penelitian ini dilakukan prediksi kelulusan tepat waktu mahasiswa program studi PTI Undiksha serta menemukan faktor yang relevan atau berpengaruh terhadap kelulusan tepat waktu, untuk membantu Mahasiswa, Dosen dan Kaprodi mengevaluasi kinerja mahasiswa sedini mungkin.

Pemilihan algoritma KNN didasari oleh dataset yang digunakan pada penelitian ini adalah mayoritas data *numerical* (tipe data kontinu) sehingga algoritma KNN cocok digunakan sebagai metode klasifikasi dengan menghitung jarak antara *data testing* dengan *data training*, untuk beberapa fitur dengan tipe kategori dibutuhkan penanganan khusus menggunakan metode *feature encoding* sesuai dengan jenis data kategorinya (nominal atau ordinal). Serta dilakukan *hyperparameter tuning* untuk menentukan metrik yang digunakan untuk menghitung jarak dan jumlah tetangga atau nilai K yang menghasilkan tingkat akurasi terbaik. Dimana data dilatih dengan menggunakan metrik jarak dan nilai K yang berbeda-beda untuk menemukan metrik jarak dan nilai K yang menghasilkan tingkat akurasi tertinggi. Tujuan dari memilih nilai K terbaik agar model tidak *overfitting* atau *underfitting* (Pranto dkk., 2020).

Untuk memperoleh fitur yang relevan untuk prediksi kelulusan mahasiswa tepat waktu, dalam penelitian ini peneliti berfokus untuk mengoptimalkan fungsi KNN dengan metode seleksi fitur secara manual berdasarkan hasil studi literatur dan algoritma *feature importance* untuk menunjukkan seberapa besar kontribusi

masing-masing fitur terhadap prediksi model atau seberapa penting sebuah *fitur* berdasarkan hasil pelatihan. Fitur yang tidak relevan atau tidak memiliki kontribusi besar terhadap model lalu dihapus untuk mengurangi dimensi data yang tinggi dan tidak relevan. *Permutation feature importance* (PFI) dipilih dalam penelitian ini sebagai metode *feature importance* karena dapat digunakan pada model *machine learning* apapun, sehingga cocok juga digunakan pada metode KNN. Proses komputasi PFI juga sangat cepat, karena tidak memerlukan proses training model berkali-kali. Selain itu, PFI dalam prosesnya tetap memperhitungkan semua interaksi dengan fitur lainnya (Molnar, 2020). Melalui penelitian ini, diharapkan mampu mendapatkan kombinasi fitur yang relevan, efektif, dan optimal untuk prediksi kelulusan tepat waktu mahasiswa dan mampu mengoptimalkan fungsi dari algoritma KNN.

1.2 Identifikasi Masalah

Berdasarkan penjelasan latar belakang yang telah dipaparkan sebelumnya, dapat diidentifikasi bahwa belum diketahui efektivitas penggunaan PFI sebagai metode seleksi fitur untuk menentukan fitur yang tidak relevan dan mengurangi dimensi data, sehingga mampu mengoptimalkan kinerja algoritma KNN dalam memprediksi kelulusan tepat waktu mahasiswa PTI Undiksha.

1.3 Batasan Masalah

Adapun batasan-batasan masalah pada pengerjaan penelitian ini, antara lain:

1. Penelitian hanya dilakukan pada ruang lingkup program studi PTI Undiksha.

2. Data mahasiswa yang digunakan adalah mahasiswa angkatan 2014 – 2017 sebanyak 460 data.
3. Penelitian yang dilakukan adalah mengenai klasifikasi data.
4. Algoritma klasifikasi yang digunakan adalah KNN.
5. Metode *feature importance* yang digunakan adalah PFI.
6. Evaluasi yang dilakukan adalah untuk mengetahui *Accuracy*, *Precision*, *Recall* dan *F1-score* dari penggunaan metode KNN dalam melakukan prediksi kelulusan tepat waktu mahasiswa.

1.4 Rumusan Masalah

Berdasarkan latar belakang di atas, didapat permasalahan sebagai berikut.

1. Variabel mana yang memiliki kontribusi signifikan terhadap performa klasifikasi menggunakan metode KNN berdasarkan hasil proses PFI?
2. Bagaimana *performance* (*Accuracy*, *Precision*, *Recall* dan *F1-score*) dari metode KNN + PFI untuk prediksi kelulusan tepat waktu mahasiswa PTI Undiksha?

1.5 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah.

1. Mengetahui variabel apa saja yang tepat untuk prediksi kelulusan tepat waktu mahasiswa program studi PTI Undiksha.

2. Melakukan evaluasi dari klasifikasi menggunakan metode KNN untuk prediksi kelulusan tepat waktu mahasiswa program studi PTI Undiksha.

1.6 Manfaat Penelitian

1.6.1 Manfaat teoritis (akademis)

Penelitian ini dapat menjadi referensi yang berguna bagi penelitian-penelitian selanjutnya dalam bidang penggunaan algoritma klasifikasi dan algoritma PFI untuk mendapatkan fitur data yang memberikan kontribusi signifikan terhadap model prediksi. Dengan demikian, peningkatan kinerja algoritma klasifikasi yang digunakan dapat dicapai melalui pendekatan yang efektif. Secara akademis, hasil dari penelitian ini dapat berkontribusi pada pemahaman lebih lanjut tentang bagaimana PFI dapat meningkatkan kualitas dan ketepatan model prediksi, serta menginspirasi penggunaan teknik-teknik baru dalam analisis data dan klasifikasi.

1.6.2 Manfaat praktis

Membantu mahasiswa, dosen dan kaprodi untuk mencegah kelulusan tidak tepat waktu mahasiswa sehingga mahasiswa, dosen, dan kaprodi dapat mengambil tindakan cepat untuk meresponnya.