

## 1. Pendahuluan

Persentase Akreditasi yang diberikan oleh BAN PT (Badan Akreditasi Nasional Perguruan Tinggi) sangat berpengaruh untuk suatu Universitas, karena dengan akreditasi tersebut menggambarkan kualitas dan kinerja yang dilakukan oleh universitas tersebut. Beberapa hal yang digunakan sebagai acuan untuk pemberian akreditasi yang baik yaitu salah satunya adalah persentase kelulusan tepat waktu mahasiswa [1]. Kelulusan tepat waktu bagi seorang mahasiswa dan perguruan tinggi adalah hasil yang baik. Agar dapat memiliki persentase yang baik maka dibutuhkan strategi untuk memperbaiki persentase kelulusan tersebut.

Strategi yang dapat dilakukan untuk memperbaiki persentase kelulusan yaitu dengan cara memprediksi kelulusan mahasiswa tersebut, banyak variabel yang mempengaruhi kelulusan seorang mahasiswa, dimulai dari nilai mata kuliah yang kurang mencukupi. Prediksi kelulusan mahasiswa ini dapat digunakan untuk mengetahui berdasarkan nilai tersebut dengan menggunakan metode yang ada di *data mining*.

Prediksi kelulusan mahasiswa ini dapat kita peroleh dengan menggunakan *data mining*. Definisi sederhana dari *data mining* adalah ekstraksi informasi atau pola yang penting dan menarik dari data yang ada di database yang besar [2], di dalam *data mining* terdapat teknik klasifikasi yang berguna untuk mengelompokkan data berdasarkan sampel yang ada, dan dataset yang digunakan adalah berupa *time series data* yaitu data historis dari akademik mahasiswa Informatika Telkom University dari tahun yang berurutan yaitu 2008 - 2011.

Algoritma yang digunakan untuk melakukan klasifikasi sangat beragam, namun setiap algoritma tersebut memiliki kelebihan dan kekurangannya masing-masing, salah satunya adalah algoritma *naive bayes* yang dipilih karena algoritma tersebut biasa digunakan untuk melakukan prediksi, lalu untuk mencoba meningkatkan tingkat akurasi dilakukanlah pengkombinasian algoritma *naive bayes* dengan menggunakan PCA untuk mereduksi dimensi dan mengecek fitur mana yang paling berpengaruh untuk dataset tersebut, dari proses prediksi kita dapat melakukan pencegahan dini untuk mahasiswa yang beresiko lulus tidak tepat pada waktunya dan juga mahasiswa yang beresiko *drop out*, variabel penting untuk prediksi ini adalah Data Mahasiswa dan Data Nilai Telkom University, berdasarkan nilai tersebut kita dapat melakukan prediksi kelulusan mahasiswa dan menganalisis hasil tersebut. Penelitian sebelumnya dengan melakukan PCA sebelum masuk menggunakan algoritma *naive bayes* dapat memiliki akurasi yang lebih optimal, pada penelitian Diagnosa Kerusakan Bearing Menggunakan Principal Component Analysis (PCA) dan Naive Bayes Classifier menggunakan *naive bayes* untuk menjadi algoritma klasifikasi, dan algoritma PCA untuk mereduksi dimensi agar lebih mudah menganalisa lebih jauh dibandingkan yang berdimensi lebih besar, penelitian tersebut memberikan hasil yang sangat optimal untuk akurasi prediksi yang didapatkan, karena tidak ada satupun prediksi yang salah dari 3 skenario yang dilakukan [3].