

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Saat ini pendidikan merupakan hal yang penting untuk setiap manusia karena dengan pendidikan, kita bisa mendapatkan ilmu yang berguna untuk menambah wawasan kita (Makkawaru, 2019). Selain itu pendidikan ini juga berkontribusi besar dalam menghasilkan lulusan yang berkompeten agar siap bersaing dalam dunia kerja. Pendidikan umumnya terdiri dari pendidikan usia dini, dasar, menengah, dan tinggi. Perkembangan teknologi juga berkontribusi dalam dunia pendidikan salah satunya seperti pembelajaran online, pelaksanaan ujian nasional berbasis komputer, dan sistem informasi akademik.

Dalam sebuah lembaga pendidikan seperti universitas, kinerja mahasiswa perlu diperhatikan agar mahasiswa tersebut dapat menyelesaikan studinya dengan tepat waktu, walaupun begitu, masih terdapat mahasiswa yang tidak dapat menyelesaikan studinya secara tepat waktu dan bahkan ada yang berhenti dari studinya atau tidak aktif menjadi mahasiswa lagi (Gunawan Sudarsono & Ulan Bani, 2020). Hal ini dibuktikan dengan menurunnya nilai mahasiswa dari angkatan 2020 hingga 2021 pada Mata Kuliah Dasar Umum (MKDU) yaitu Bahasa Indonesia di UMKT. Nilai rata-rata mahasiswa mengalami penurunan pada angkatan tahun 2017 yang awalnya 75,84% lalu pada angkatan 2018 menurun menjadi 74,59%, setelah itu mengalami peningkatan di angkatan 2019 sebesar 78,02% dan pada angkatan 2020 meningkat lagi menjadi 82,46%, lalu menurun drastis pada angkatan 2021 yaitu sebesar 71.58%. Untuk mengatasi masalah tersebut dibutuhkan cara yang dapat mengukur kinerja mahasiswa dalam menyelesaikan studinya.

Data mining merupakan proses memperoleh data tersembunyi dari suatu dataset yang besar. Teknik *data mining* digunakan secara luas dalam berbagai bidang. Dengan teknik *data mining*, kita dapat memprediksi, mengklasifikasi,

memfilter, dan mengelompokkan data (Rady & Anwar, 2019). Salah satu teknik *data mining* yang digunakan dalam klasifikasi adalah algoritma *Random Forest*.

Dalam Penelitian ini, peneliti menggunakan algoritma *Random Forest* dengan seleksi fitur *ANOVA* karena mengambil saran dari penelitian sebelumnya yang menggunakan Algoritma *Random Forest Classifier* dengan *Correlation Based Feature Selection (CFS)*. Dalam penelitian tersebut peneliti terdahulu menyarankan menggunakan seleksi fitur yang lain seperti *Chi-Square*, *Information Gain*, *ANOVA*, dan lain-lain agar diharapkan mendapatkan hasil akurasi yang lebih baik lagi (Priantama & Yoga Siswa, 2022).

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, bagaimana mengetahui atribut-atribut yang berpengaruh dalam menurunnya nilai mahasiswa pada mata kuliah Bahasa Indonesia dan meningkatkan persentase akurasi dari klasifikasi nilai mahasiswa pada algoritma *Random Forest Classifier* dengan menggunakan seleksi fitur *ANOVA*.

1.3. Tujuan Penelitian

Tujuan dari penelitian ini untuk mengetahui atribut-atribut yang memiliki pengaruh dalam menurunnya nilai mahasiswa pada mata kuliah Bahasa Indonesia dan meningkatkan akurasi dari algoritma *Random Forest Classifier* dengan menggunakan seleksi fitur *ANOVA*.

1.4. Batasan Masalah

Agar masalah yang dibahas tidak menjadi lebih luas lagi, penulis membatasi masalah penelitian sebagai berikut:

- a. Data yang digunakan adalah data mahasiswa UMKT yang mengambil mata kuliah “Bahasa Indonesia” angkatan 2020/2021 – 2021/2022.
- b. Nilai Mata kuliah yang digunakan adalah mata kuliah Bahasa Indonesia yang merupakan mata kuliah dasar umum.
- c. Pembagian dataset dilakukan dengan 5-fold Cross Validation.
- d. Melakukan penerapan seleksi fitur *ANOVA* pada algoritma *Random Forest Classifier* untuk meningkatkan akurasi dari algoritma.