

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perguruan tinggi yang unggul dan berkualitas dapat dilihat dari prestasi mahasiswanya salah satunya adalah berupa indeks prestasi mahasiswa semester pertama (Pratama et al., 2018; Setiyani et al., 2020). Indeks prestasi mahasiswa yang berada dibawah standar dapat menyebabkan berbagai masalah yang menimbulkan efek berantai menyebabkan turunnya performa rendah di semester-semester yang akan datang (Alverina et al., 2018). Indeks prestasi mahasiswa rendah disebabkan oleh banyak faktor yang melatar belakang mahasiswa tersebut seperti suasana hati, memenejemen waktu, hubungan dengan keluarga, penjelasan dosen, suasa tempat tinggal, kegiatan selain kuliah, adaptasi lingkungan, perhatian orang tua, pergaulan, suasana kelas, dan kemampuan menangkap materi (Hendikawati, 2011).

Sehingga untuk menjaga prestasi mahasiswa tetap baik maka, diperlukan adanya sebuah pendekatan untuk menggali informasi penting mengenai faktor yang melatar belakang mahasiswa. Untuk mengolah informasi dapat menggunakan sebuah metode baru yang berguna untuk manajemen universitas sebagai monitoring proses pembelajaran, dan mengambil kebijakan yang diperlukan untuk meningkatkan indeks prestasi mahasiswa (Hasudungan & Pranoto, 2021).

Oleh karena itu melakukan prediksi indeks prestasi mahasiswa merupakan hal yang sangat penting, salah satu cara untuk melakukan prediksi adalah menerapkan analisis data menggunakan data mining seperti yang dilakukan oleh Firdaus (2019), Hasudungan (2021), Tommy dan Mahmud (2021), Hasudungan dan Pranoto (2021), Timur dan Beatrix (2020), Rolansa, dkk (2020), Sonang, dkk (2022), Desiani, dkk (2020), Alverina, dkk (2018).

Penjabaran sederhana data mining, bahwa data mining adalah proses mengolah pola agar dapat mengekstrak informasi dan pengetahuan yang akurat serta agar dapat mudah dipahami sehingga berguna untuk pengambilan

keputusan (Setiyani et al., 2020). Salah satu fungsi data mining adalah klasifikasi (classification) yaitu proses untuk mencari model yang mengelompokkannya menjadi kelas-kelas atau konsep data, metode algoritma naïve bayes dan algoritma rough set termasuk didalamnya. Naïve bayes memiliki keunikan bahwa semua atribut dianggap sama sehingga atribut lainnya saling bebas atau indenpenden (Saputro & Sari, 2020). Kelebihan naïve bayes salah satunya dapat dilihat dari tingkat akurasi yang tinggi dan waktu komputasi yang singkat bergantung pada banyak atribut dan data yang digunakan (Desember et al., 2017). Beberapa penelitian terdahulu pernah menggunakan naïve bayes pada bidang pendidikan menganalisa indeks prestasi mahasiswa untuk melakukan sebuah prediksi seperti Widaningsih (2019), Widhi dan Wulan (2020), Mahanggara dan Dwi (2019), Kurniawan, dkk (2018), Etriyanti, dkk (2020), Setiyani, dkk (2020). Meskipun naïve bayes adalah algoritma yang tepat yang dapat digunakan untuk melakukan prediksi, Namun naïve bayes memiliki kekurangan yaitu ketika memakai semua atribut untuk klasifikasi dapat mempengaruhi hasil akurasi, dikarenakan beberapa atribut yang tidak mempengaruhi konsistensi data akan tidak berkontribusi dengan baik pada klasifikasi. Terlebih lagi penggunaan atribut yang berlebihan dapat memakan waktu saat proses klasifikasi (Hasudungan et al., 2020). Sehingga diperlukanya sebuah metode untuk melakukan pemilihan atribut.

Rought set adalah metode matematika yang dikemukakan oleh Zdzislaw Pawlak, yang sering menangani data yang tidak konsisten, tidak memiliki kepastian dan samar (Pawlak, 1998). Rough set masuk dalam salah satu algoritma efisien yang dapat mengungkap pola tersembunyi dan hubungan pada data yang dapat membantu dalam pemilihan atribut (Samaray, 2022). Kegunaan algoritma rough set adalah dapat memilih atribut tanpa harus mengurangi konsistensi data. Pada penelitian sebelumnya metode seperti ini telah digunakan oleh Hasudungan, dkk (2020) untuk memilih atribut terbaik untuk algoritma naïve bayes pada data pendidikan menyatakan peningkatan akurasi dari naïve bayes dari 68% naik menjadi 79%. dan penelitian selanjutnya, Hasudungan (2021) melakukan kombinasi algoritma rough set dan naïve bayes menganalisis data mahasiswa

dengan mengeliminasi atribut yang tidak dibutuhkan. Sehingga meningkatkan nilai akurasi dari 68,09% naik menjadi 80,85%. Oleh karena itu pada penelitian ini kami akan menggunakan algoritma rough set dan naïve bayes untuk melakukan prediksi indeks prestasi mahasiswa.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dibuat maka rumusan masalahnya adalah menerapkan kombinasi algoritma rough set sebagai pemilihan atribut dan algoritma naïve bayes sebagai klasifikasi untuk memprediksi indeks prestasi mahasiswa Universitas Muhammadiyah Kalimantan Timur (UMKT) tahun angkatan tahun 2022 berdasarkan faktor latar belakang mahasiswa.

1.3 Tujuan Penelitian

Dalam langkah menyelesaikan rumusan masalah, terdapat empat tujuan dari penelitian ini yakni:

- 1) Melakukan pemilihan atribut yang berpengaruh terhadap prestasi mahasiswa menggunakan algoritma rough set maximum dependency attributes (MDA)
- 2) Membuat model klasifikasi dengan menggunakan data dengan atribut menggunakan algoritma naïve bayes.
- 3) Evaluasi model yang telah dibuat oleh algoritma naïve bayes dengan menggunakan algoritma rough set dan tanpa algoritma rough set.

1.4 Batasan Penelitian

Data yang diperoleh dalam penelitian ini berdasarkan data yang dikumpulkan dari kuesioner yang diisi oleh mahasiswa Universitas Muhammadiyah Kalimantan Timur (UMKT) Fakultas sains dan teknologi pada jurusan teknik informatika angkatan 2020-2022.