

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi saat ini, terutama dalam sistem peramalan kelulusan mahasiswa, yang didukung oleh teknik *Data Mining*, memberikan potensi luar biasa dalam memperbaiki proses pendidikan tinggi. Dengan menggunakan teknik *Data Mining* yang canggih, perguruan tinggi dapat mengumpulkan, mengintegrasikan, dan menganalisis beragam data mahasiswa, termasuk hasil ujian, partisipasi kelas, proyek mahasiswa, dan data lainnya. Dalam proses ini, algoritma *Data Mining* dapat mengidentifikasi pola, tren, dan hubungan yang kompleks dalam dataset yang besar, memungkinkan perguruan tinggi untuk meramalkan tingkat kelulusan mahasiswa dengan tingkat akurasi yang lebih tinggi. Dengan pemahaman yang lebih baik tentang faktor-faktor yang memengaruhi kinerja akademik, perguruan tinggi dapat memberikan intervensi yang lebih tepat dan personalisasi pendidikan yang lebih efektif kepada mahasiswa, mengarah pada peningkatan tingkat kelulusan dan kesuksesan akademik secara keseluruhan.

Pertumbuhan data akademik mahasiswa menciptakan pola-pola yang beragam pada atribut data mereka. Potensi ini dapat dieksplorasi untuk memperoleh informasi berharga salah satu teknik yang digunakan adalah *Data Mining* untuk peramalan keputusan berdasarkan pola yang terbentuk dari data yang ada. *Data Mining* merupakan proses pengolahan data yang melibatkan penggalian informasi khusus menggunakan metode dan algoritma yang ditentukan (Sepriansyah et al. 2023). Terdapat berbagai metode untuk menentukan apakah setiap mahasiswa dapat

menyelesaikan studi mereka, entah itu sesuai dengan jadwal dan tepat waktu, atau terlambat, bahkan melewati batas waktu yang ditetapkan untuk setiap mahasiswa (Qisthiano, Prayesy, and Ruswita 2023).

Universitas Adzkie salah satu kampus swasta di Kota Padang dengan jumlah mahasiswa mencapai ribuan dan terus meningkat setiap tahunnya, manajemen kampus masih belum sepenuhnya memahami manfaat dari data akademik yang terus bertambah. Data yang menumpuk tidaklah terlalu memiliki kegunaan data tersebut hanya diperlukan saat universitas membutuhkan suatu informasi tertentu atau saat proses akreditasi dilakukan (Hikmah Tanjung, Yunus, and Widi Nurcahyo 2023). Kondisi ini menunjukkan bahwa masih ada potensi yang belum tereksplorasi secara maksimal dari data akademik mahasiswa yang ada.

Penelitian terdahulu tentang peramalan kelulusan mahasiswa telah banyak dilakukan dengan berbagaimacam algoritma diantaranya penelitian yang berjudul Pola Peramalan Kelulusan Siswa MadrasahAliyah Swasta dengan Support Vector Machine dan Random Forest dipenelitiannya Algoritma RF menunjukkan performa sedikit lebih unggul daripada algoritma SVM dalam hasil pengukuran parameter seperti Accuracy (99,49% vs 98,98%), Precision (99,74% vs 99,23%), Recall (99,74% vs 99,74%). Namun, algoritma SVM memiliki konsumsi waktu yang lebih baik, yaitu 0,04 detik dibandingkan dengan 0,26 detik pada algoritma RF. secara keseluruhan algoritma RF sedikit lebih unggul daripada SVM dalam peramalan kelulusan (Polgan et al. 2023).

Sedangkan penelitian yang berjudul Peramalan Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta Sebagaimana diperoleh Accuracy 75.52%, Precision 75.52%, dan Recall 75.52% (Hasibuan and Mahdiana 2023) dan pada penelitian yang berjudul Peramalan Kelulusan Mahasiswa Dengan Metode Naive Bayes menggunakan 379 data, dengan rincian data training 303 data dan data testing 76 data mendapatkan hasil penelitian yang diperoleh yaitu akurasi = 88,16%, precision = 93,62% dan recall = 88% (Khasanah et al. 2022).

Hasil penelitian yang berjudul Peramalan Tingkat Kelulusan Mahasiswa Menggunakan Machine Learning dengan Teknik Deep Learning memiliki kesimpulan bahwa peramalan tingkat kelulusan mahasiswa menggunakan metode Machine Learning dengan teknik deep learning mendapatkan tingkat akurasi sebesar 72.84%. Hasil tersebut menunjukkan bahwa teknik deep learning tingkat akurasinya masih kurang bagus ditinjau dari hasil penelitian dengan topik yang sama metode yang berbeda (Martanto, Ali, and Mulyawan 2019).

Peramalan kelulusan juga diteliti dengan metode Algoritma *Neural Network* dengan judul Peramalan Tingkat Kelulusan Mahasiswa Program Studi Teknik Informatika Universitas Banten Jaya Menggunakan Algoritma *Neural Network* dengan sample tahun 2018 dan 2019 dengan nilai presentasi 98,27%. Dari penelitian ini didapatkan sebuah hasil yang dapat meningkatkan jumlah dari parameter node input layer penelitian selanjutnya (Rudianto, Kania, and Solihati 2022).

Perbandingan beberapa metode juga dilakukan pada peramalan kelulusan mahasiswa salah satunya pada penelitian yang berjudul Perbandingan algoritma c4.5 dan naive bayes dalam peramalan kelulusan mahasiswa yang mendapatkan hasil implementasi algoritma C4.5 dan Naive Bayes layak digunakan untuk menentukan kelulusan mahasiswa Fakultas Ilmu Sosial dan Ilmu Politik Universitas Andalas. Dari pengujian akurasi menggunakan confusion matrix dengan data 70%:30%, 80%:20% dan 90%:10% dihasilkan akurasi terbaik adalah algoritma Naive Bayes. Hal ini dibuktikan dari nilai akurasi yang lebih tinggi dibandingkan C4.5 pada pengujian data 70%:30%, 80%:20% dengan nilai 79.82%, 81,58% dan pada data 90%:10% akurasi algoritma Naive Bayes tetap sedangkan C4.5 mengalami penurunan (Hikmah Tanjung et al. 2023).

Berpandangan pada penelitian terdahulu penulis tertarik juga melakukan eksplorasi data kelulusan mahasiswa dengan metode yang berbeda yaitu metode *K-Nearest Neighbor* serta menggunakan RapidMiner sebagai aplikasi pendukung dalam menjalankan metodenya, Algoritma KNN ini pada umumnya digunakan dalam mengklasifikasikan objek berdasarkan pada data pembelajaran yang memiliki nilai selisih kecil dan jarak tetangga terdekat dengan objek (Maskuri, Sukerti, and Bhakti

2022). KNN memiliki beberapa kelebihan, antara lain, tidak memerlukan proses pra-pemrosesan data yang kompleks, mudah diimplementasikan dan dipahami, serta efektif ketika digunakan dengan data latih yang berukuran besar (Sukmana, Abdurrahman, and Wicaksono 2020)

Metode ini mudah diimplementasikan dan cukup efektif untuk klasifikasi data dengan jumlah sampel yang kecil hingga menengah (Rismala, Ali, and Rizki Rinaldi 2023). Dalam metode KNN, langkah awalnya adalah menentukan nilai K, misalnya jika $k=3$, berarti kita akan mengambil 3 dokumen terdekat dengan dokumen uji. Kelas yang paling banyak muncul di antara dokumen-dokumen terdekat ini akan menjadi kelas di mana data evaluasi tersebut berada (Supriyanto, Alita, and Isnain 2023). Misalkan untuk mencari waktu kelulusan mahasiswa semester 6, digunakan data dari mahasiswa yang sudah lulus. Kedekatan kasus mahasiswa yang akan lulus dihitung dengan semua kasus mahasiswa yang sudah lulus. Solusi untuk kasus mahasiswa yang akan lulus diambil dari kedekatan terbesar antara kasus tersebut dengan kasus yang sudah lulus.

Algoritma *K-Nearest Neighbor* merupakan algoritma yang digunakan dalam dunia Machine Learning (Faisal, Utami, and Parmica 2023) Dalam penghitungan nilai KNN digunakan rumus perhitungan jarak (euclidean), KNN bekerja dengan cara membandingkan data uji dan data training (Yunus and kadek ari pratiwi 2023). Berdasarkan referensi yang dijadikan penulis sebagai referensi dalam menjalankan penelitian ini maka penelitian dengan menggunakan metode KNN merupakan salah satu metode yang tepat dalam mengeksplorasi data kelulusan mahasiswa untuk peramalan kelulusan. Data yang digunakan dalam penelitian ini adalah data mahasiswa Universitas Adzkie angkatan 2016 yang mencakup atribut seperti indeks prestasi mahasiswa semester 1-6, jumlah sks hingga semester 6, dan keterangan lulus.

Penelitian ini akan menggunakan data training dari mahasiswa Universitas Adzkie Program studi PGSD angkatan 2016 yang sudah memiliki keterangan lulus dan drop out. Sementara itu, data testing menggunakan mahasiswa angkatan 2020 yang dipilih secara acak untuk diolah dan menghasilkan informasi peramalan kelulusan mahasiswa tersebut.

Metode *K-Nearest Neighbor* digunakan karena tidak memerlukan *pre-processing* data yang rumit, mudah diimplementasikan dan dipahami dan efektif apabila data latihnya besar, secara spesifik penelitian ini juga berfokus pada perhitungan menggunakan euclidean untuk menentukan pola data, euclidean merupakan perhitungan jarak terdekat yang mengambil nilai mayoritas untuk menentukan pola dari sebuah data.

Berdasarkan latar belakang masalah dan penelitian yang dilakukan, maka penulis tertarik untuk melakukan penelitian serta mengkaji dan Menyusun penelitian dalam bentuk tesis dengan judul “**Algoritma *K-Nearest Neighbor* Untuk Peramalan Kelulusan Mahasiswa**”.

1.2 Perumusan Masalah

Berdasarkan latar belakang permasalahan yang telah diuraikan sebelumnya, maka didapatkan perumusan masalah sebagai berikut:

1. Bagaimana memahami penerapan metode *K-Nearest Neighbor* dalam meramalkan kelulusan mahasiswa ?
2. Bagaimana data mahasiswa yang selalu bertambah dapat menghasilkan informasi yang berguna bagi ketua program studi maupun pembimbing akademik?
3. Bagaimana tingkat akurasi hasil peramalan kelulusan mahasiswa dari metode *K-Nearest Neighbor*?

1.3 Batasan Masalah

Batasan masalah bertujuan untuk membantu dalam mengidentifikasi masalah, membatasi ruang lingkup dan menjadi gambaran terkait hal yang akan diteliti sehingga penelitian menjadi lebih terarah dan tidak menyimpang dari pokok bahasan dan

permasalahan. Untuk membatasi ruang lingkup penelitian, maka batasan masalah dibatasi sebagai berikut:

1. Data yang digunakan adalah data mahasiswa jurusan PGSD angkatan 2016 Universitas Adzka.
2. Atribut yang diambil meliputi NIM, Nama, Indeks Prestasi Semester 1-6, jumlah SKS hingga semester 6, dan status kelulusan.
3. Penggunaan metode yang diterapkan adalah *K-Nearest Neighbor*.
4. Pengolahan data dilakukan menggunakan aplikasi *RapidMiner*.

1.4 Tujuan Penelitian

Tujuan penelitian merupakan suatu hasil akhir atau sesuatu yang ingin dicapai dalam suatu penelitian. Diperlukan suatu perencanaan dan pelaksanaan penelitian dengan matang supaya tujuan penelitian dapat dicapai. Dalam penelitian ini ada beberapa tujuan yang hendak dicapai, yaitu:

1. Memahami penerapan metode *K-Nearest Neighbor* dalam peralaman kelulusan mahasiswa.
2. Meramalkan data mahasiswa yang selalu bertambah sehingga dapat menghasilkan informasi yang berguna.
3. Menghitung tingkat akurasi peramalan kelulusan mahasiswa menggunakan metode *K-Nearest Neighbor*.

1.5 Manfaat Penelitian

Suatu penelitian yang dilakukan diharapkan dapat memberi manfaat baik bagi kepentingan peneliti, organisasi, maupun ilmu pengetahuan bagi peneliti selanjutnya.

Manfaat penelitian merupakan jawaban atas tujuan penelitian yang dilakukan. Adapun manfaat yang akan didapatkan dalam penelitian ini adalah sebagai berikut:

1. Meningkatkan pemahaman tentang pengaplikasian Algoritma *K-Nearest Neighbor* pada aplikasi RapidMiner.
2. Memberikan informasi peramalan mengenai kelulusan mahasiswa, baik yang lulus tepat waktu maupun yang tidak tepat waktu.
3. Membantu dalam evaluasi kelulusan mahasiswa dalam konteks sistem akademik di institusi pendidikan tinggi.

1.6 Sistematika Penulisan

Sistematika penulisan merupakan sebuah kerangka atau urutan penelitian agar lebih terstruktur. Terdapat enam BAB untuk mekanisme penelitian ini dengan gambaran umum sebagai berikut :

BAB I : PENDAHULUAN

Bab ini berisikan pendahuluan dan latar belakang masalah dalam melakukan penelitian. Adapun yang akan dibahas pada bab ini, yaitu perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian yang diakhiri dengan sistematika penulisan.

BAB II : LANDASAN TEORI

Bab ini menjelaskan tentang beberapa hal yang mendasar tentang masalah yang akan dibahas, yang akan digunakan sebagai landasan dasar dalam melakukan pemecahan masalah pada penelitian yang dilakukan.

BAB III : METODOLOGI PENELITIAN

Pada bab ini berisikan tentang langkah atau kerangka kerja dan prosedur dalam melakukan peramalan kelulusan mahasiswa di Universitas Adzkie menggunakan metode *K-Nearest Neighbor*.

BAB IV : ANALISA DAN PERANCANGAN

Bab ini memuat tentang analisa data, pengolahan data kelulusan mahasiswa. Selain itu bab ini juga memuat hasil dari penggunaan metode *K-Nearest Neighbor*.

BAB V : IMPLEMENTASI DAN PENGUJIAN

Pada bab ini berisi penerapan metode *K-Nearest Neighbor* dengan menggunakan *RapidMiner* sebagai aplikasi pengolahan data mahasiswa yang sudah lulus dan yang masih aktif.

BAB VI : KESIMPULAN DAN SARAN

Bab ini menjelaskan tentang kesimpulan dari hasil penelitian dan saran-saran untuk pihak lain dalam melakukan penelitian dan pengembangan untuk selanjutnya.