

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Salah satu manfaat aplikasi SIAKAD (Sistem Informasi Akademik) yang ada di perguruan tinggi adalah menjabarkan indeks kelulusan hasil belajar setiap mahasiswa untuk menilai keaktifannya dalam menempuh perkuliahan di setiap semester. Perguruan tinggi melalui lembaga khusus pengolah data seperti BAAK (Badan Akademik *Administratif* dan Kemahasiswaan) biasanya memberikan laporan hasil semester yang disajikan ke Lembaga Layanan Pendidikan Tinggi (LLDikti) setiap semesternya, membutuhkan kelengkapan data yang akurat tentang hasil penilaian belajar mahasiswa. Kebutuhan ini dapat dikembangkan melalui penelitian kampus untuk pengambilan keputusan di masa depan.

Setiap kampus mempunyai visi dan misi yang sesuai dengan bidang program studi yang dikelolanya dan merasa sangat puas, jika keaktifan mahasiswa yang tercermin dalam Indeks Kelulusan (IP) setiap semester yang memuaskan. Semakin baik IP, maka capaian pembelajaran mahasiswa setiap semester akan baik. Perkembangan penggunaan aplikasi sistem akademik sejauh ini sudah berbasis *data base cloud computing* dimana dalam aplikasi akademik akan tercipta *fitur* atau fasilitas yang memungkinkan para pengambil keputusan dapat memperoleh informasi tentang kinerja setiap mahasiswa setiap semesternya. Kinerja pembelajaran setiap mahasiswa persemester dapat digunakan untuk

menilai dan memprediksi setiap mahasiswa tentang rentang waktu penyelesaian studinya.

Teknik untuk mengevaluasi data dalam ilmu komputer dikenal sebagai pendekatan pembelajaran mesin. Sampel data yang dikelompokkan dan diklasifikasikan dengan cara berpikir sistematis atau terstruktur merupakan pendekatan *machine learning*. Data disajikan dan dapat diolah sebagai grafik atau hubungan sesama diagram. Khusus untuk mengukur Indeks Kelulusan (IP) mahasiswa, dibutuhkan pendekatan machine learning dengan Algoritma *K-Nearest Neighbors* dan Algoritma *Naive Bayes*.

Menurut Budiharto (2016), dijelaskan bahwa algoritma *K-Nearest Neighbors* adalah algoritma untuk mengklasifikasikan data dari kumpulan data. Melabeli data berdasarkan data yang ada kemudian langsung menempatkannya pada kelas yang sama adalah keunggulan Algoritma *K-Nearest Neighbors*. Hal ini memudahkan konsolidasi nilai-nilai siswa yang ada. Pada algoritma ini harus mencari data secara keseluruhan terlebih dahulu sehingga tidak dapat menemukan data dengan cepat hal ini merupakan satu kelemahan algoritma *K-Nearest Neighbors*. Sedangkan kondisi yang dapat dilaksanakan sesuai dengan kondisi yang terjadi merupakan algoritma klasifikasi data yang naif adalah algoritma *Naive Bayes* (Giuseppe Bonaccorso, 2017). Mengklasifikasikan data menurut kondisi tertentu agar cepat mendapatkan solusi berdasarkan permasalahan yang ada tanpa membandingkan kondisi lainnya merupakan kelebihan algoritma *Naive Bayes*. Sedangkan, kurang mampu dalam menangani masalah yang lebih terkait

dengan masalah lainnya adalah kelemahan daripada algoritma *Naive Bayes* (Giuseppe Bonaccorso, 2017).

Penelitian sebelumnya pernah dilakukan oleh Ihsan A. Abu Amra, Ashraf Yunis Maghari, 2017 yang berjudul *Students performance prediction using K Nearest Neighbour and Naive Bayesian*, pada penelitian yang dipaparkan, membandingkan tiga parameter evaluasi (akurasi, *recall* dan presisi) untuk dua algoritma *K-Nearest Neighbor* dan *Naive Bayes*, algoritma *Naive Bayes* memiliki presisi maksimum sebesar 93,17%, yang berarti hubungan yang kuat antara karakteristik yang mempengaruhi siswa dan akan membantu memprediksi kinerja siswa untuk tahun mendatang. *Naive Bayes* lebih baik dari *K-Nearest Neighbor*, yang berarti hubungan yang kuat antara karakteristik yang mempengaruhi kinerja siswa dan membantu memprediksi kinerja siswa. Terkadang *K-Nearest Neighbor* lebih baik daripada *Naive Bayes* untuk dataset lain dan IDE yang berbeda. Dalam pekerjaan di masa depan, algoritma klasifikasi lain dapat diterapkan pada kumpulan data pendidikan yang berbeda.

Pada tahun 2017 penelitian juga dilakukan dengan judul *Machine Learning Algorithms for Student Employability Prediction Using R*, oleh G.Vadivu, K.Sornalakshmi, 2017, Studi ini memprediksi keterampilan kerja berdasarkan kinerja *reguler* mereka. Hasil dari penelitian ini adalah presisi yang diperoleh setelah analisis untuk K tetangga terdekat adalah 95,33% untuk *Naive Bayes* adalah 97,67%.

Demikian juga yang dilakukan oleh Mohammad Imron Satia Angga Kusumah 2018 mengenai penerapan metode klasifikasi data mining untuk

memprediksi kelulusan siswa menggunakan algoritma *K-Nearest Neighbor*. Penelitian ini bertujuan untuk mengetahui tingkat presisi yang ditransmisikan oleh algoritma *K-Nearest Neighbor* untuk memprediksi tingkat kelulusan mahasiswa STMIK Amikom Purwokerto. Hasil akurasi pada penelitian yaitu 89,04% pada metode *K-Nearest Neighbor*.

Penelitian Abdullah Baz, Fatima Alshareef, Ebtihal Alshareef, Hosam Alhakami, Tahani Alsubait, 2017, yang berjudul *Predicting Students' Academic Performance Using Naïve Bayes* tujuan dari penelitian ini adalah untuk memprediksi kinerja akademik mahasiswa Universitas Umm Al-Qura menggunakan metode *Naive Bayes*.

Berdasarkan penelitian di atas, metode tersebut memiliki keunggulannya masing-masing. Kelebihan dari metode *Naïve Bayes* yaitu mampu mengklasifikasikan data meskipun data yang digunakan berjumlah sedikit untuk estimasi parameternya. *K-Nearest Neighbour*, kelebihan dari metode ini adalah dapat diterapkan secara efektif pada data dengan hasil yang akurat. Melihat penelitian-penelitian sebelumnya dengan keunggulan masing-masing metode, maka penulis memutuskan untuk membandingkan metode klasifikasi *Naive Bayes* dan *K-Nearest Neighbor*, dimana data ini menggunakan mahasiswa yang sudah lulus dan masih aktif pada Sekolah Tinggi Teknologi Nusantara Lampung tahun akademik 2013 sampai 2018 dimana data *training* merupakan mahasiswa yang lulus angkatan 2013 sampai 2017 dan data *testing* menggunakan angkatan 2018 didalam memprediksi kelulusan mahasiswa berdasarkan variabel yaitu Nomor Induk Mahasiswa (NIM), Nama Mahasiswa, Indeks Prestasi Kumulatif (IPK),

Unit Kegiatan Akademik (UKM), Penghasilan Orang Tua dan Test Potensi Akademik, deskripsi untuk melihat bagaimana akurasi prediksi hasil algoritma *K-Nearest Neighbors* dan *Naive Bayes* dalam mengukur tingkat akurasi dengan *cross validation*, di dalam dataset terdapat beberapa karakteristik atau variabel.

1.2. Perumusan Masalah

Adapun rumusan masalah adalah sebagai berikut:

1. Bagaimana hasil prediksi menggunakan algoritma *K-Nearest Neighbors* pada kelulusan mahasiswa Sekolah Tinggi Teknologi Nusantara Lampung ?
2. Bagaimana hasil prediksi menggunakan algoritma *Naive Bayes* pada kelulusan mahasiswa Sekolah Tinggi Teknologi Nusantara Lampung ?
3. Bagaimana hasil evaluasi model yang digunakan?

1.3. Batasan Masalah

Adapun batasan dalam penelitian ini adalah sebagai berikut :

1. Perbandingan algoritma *K-Nearest Neighbors* dan algoritma *Naive Bayes* digunakan dalam mengukur akurasi dengan *confusion matrix*.
2. Variabel yang digunakan adalah 2013 sampai 2018 dalam memprediksi kelulusan mahasiswa berdasarkan variabel yaitu Nomor Induk Mahasiswa (NIM), Nama Mahasiswa, Indeks Prestasi Kumulatif (IPK), Unit Kegiatan Akademik (UKM), Penghasilan Orang Tua dan Test Potensi Akademik, Mahasiswa Program Studi S1 Teknik Elektro dan Program Studi S1 Teknik Industri Sekolah Tinggi Teknologi Nusantara Lampung.

3. Data yang diambil hanya berasal dari data mahasiswa yang lulus (data latih) dan mahasiswa yang aktif (data uji) pada Program Studi S1 Teknik Elektro dan Proram Studi S1 Teknik Industri angkatan 2013 sampai 2017 yang merupakan data latih dan 2018 yang merupakan data uji .

1.4. Tujuan Penelitian

Tujuan yang dilakukan pada penelitian ini adalah

1. Algoritma *K-Nearest Neighbors* dan Algoritma *Naive Bayes* berdasarkan pengukuran tingkat akurasi.
2. Mengetahui algoritma apa yang mampu mencapai akurasi tertinggi.
3. Membantu pihak Sekolah Tinggi Teknologi Nusantara Lampung untuk mengevaluasi kelulusan mahasiswa Proram Studi S1 Teknik Elektro dan Proram Studi S1 Teknik Industri.

1.5. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah:

1. Algoritma mana yang lebih baik dalam mengukur tingkat akurasi antara algoritma *K-Nearest Neighbors* dan algoritma *Naive Bayes*.
2. Memberikan kemudahan kepada pihak Sekolah Tinggi Teknologi Nusantara Lampung untuk mengevaluasi kelulusan mahasiswa Proram Studi S1 Teknik Elektro dan Proram Studi S1 Teknik Industri Sekolah Tinggi Teknologi Nusantara Lampung dalam menganalisis kelulusan mahasiswa.
3. Referensi penelitian data *science* Algoritma *K-Nearest Neighbors*, *Naive Bayes* dan korelasi variabel.

1.6. Sistematika Penulisan

Untuk mempermudah penyusunan laporan penelitian ini, maka sistematika penulisan terbagi atas lima bab seperti terlihat dibawah ini:

BAB I PENDAHULUAN

Bab ini berisi Latar Belakang Masalah, Perumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, dan Sistematika Penulisan.

BAB II LANDASAN TEORI

Bab ini berisi tentang Penelitian Terdahulu, Diagram *Fishbone*, *Machine Learning*, *Data Mining*, *Data Science*, Pengertian Algoritma, Algoritma *Naïve Bayes*, Algoritma *K-Nearest Neighbors*, *Cross Validation*, Analisis Korelasi dan *Confusion Matrix*.

BAB III METODOLOGI PENELITIAN

Bab ini berisi Studi *Literatur*, Teknik Pengumpulan Data, Arsitektur Sistem, Tahapan Penelitian, *Preprocessing* Data, Pemodelan, *K-Nearest Neighbor*, *Naïve bayes* dan Evaluasi.

BAB IV PEMBAHASAN

Bab Ini berisi Analisis Kebutuhan,*Preprocessing* Data, Transformasi data,*Import* data *Rapidminer*,*Cross Validation*,Algoritma *K-Nearest Neighbor*, *Confusion*

Matriks K-Nearest Neighbor, Algoritma Naive Bayes, Confusion Matriks Naive Bayes, dan Evaluation.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi mengenai kesimpulan dan saran, dimana kesimpulan berisi hasil secara keseluruhan dari bab-bab sebelumnya dan saran merupakan hasil yang diperoleh sehingga dapat dilakukan dalam pengembangan selanjutnya.