

BAB 2

TINJAUAN PUSTAKA

2.1. Kinerja Mahasiswa

Kinerja merupakan hasil yang dapat dicapai oleh seseorang maupun sekelompok orang pada suatu organisasi, sesuai akan wewenang serta tanggung jawab masing-masing untuk mencapai tujuan organisasi yang bersangkutan secara legal, tidak melanggar hukum dan sesuai dengan moral maupun etika (Fauzi & Nugroho A, 2020).

Mahasiswa dapat dikatakan orang yang telah terdaftar sebagai siswa pada perguruan tinggi yang memiliki kartu tanda mahasiswa atau KTM, kemudian diakui oleh pemerintah serta mampu mencari ilmu sendiri karena sudah dewasa (Gafur, 2015). Mahasiswa merupakan setiap orang yang secara resmi terdaftar untuk mengikuti pembelajaran di perguruan tinggi yang batas usianya 18 sampai 30 tahun (Gafur, 2015).

Kinerja mahasiswa ini dapat diukur melalui kinerja akademik yang dapat diketahui dari prestasi belajar setiap mahasiswa itu sendiri. Kinerja akademik adalah hasil akhir yang dicapai oleh mahasiswa sebagai keberhasilan selama masa mengikuti Pendidikan dalam sebuah institusi pendidikan (Indra *et al.*, 2021).

2.2. Metode *Data Mining*

Data mining adalah aktivitas yang menggunakan teknik khusus untuk mengekstrak informasi dan pengetahuan dari kumpulan data yang besar, dimana informasi atau pengetahuan yang dihasilkan dari *data mining* dapat digunakan untuk meningkatkan pengambilan keputusan (Santosa & Umam, 2018).

Data mining merupakan serangkaian proses untuk mendapat nilai tambah berupa informasi yang tidak diketahui secara manual pada suatu basis data (Vulandari, 2017). Berdasarkan fungsionalitasnya *data mining* dikelompokkan menjadi 4 bagian yang meliputi (Santosa & Umam, 2018):

a. Klastering (*Clustering*)

Klastering merupakan model *unsupervised* yang mengelompokkan data yang belum diketahui labelnya. Dengan penerapan metode yang tepat akan menghasilkan sebuah klastering yang berkualitas.

b. Klasifikasi (*Classification*)

Klasifikasi merupakan model *supervised* yang diterapkan pada data baru untuk mengelompokkan jenis objek. Pada klasifikasi diperlukannya sampel data, kemudian diprediksi beberapa *class* berdasarkan sampel yang ada.

c. Regresi/Estimasi (*Regression*)

Regresi merupakan suatu fungsi yang digunakan untuk memodelkan data agar dapat meminimalisir hasil kesalahan prediksi.

d. Asosiasi (*Association*)

Aturan asosiasi merupakan pemodelan kebergatungan yang mencari relasi antar variabel.

2.3. Metode Klasifikasi

Klasifikasi merupakan pekerjaan yang menilai objek data untuk dimasukkan ke dalam kelas tertentu dari sejumlah kelas yang tersedia (Prasetyo, 2012). Klasifikasi dapat diartikan sebagai peranan dalam *data mining* yang memiliki pendefinisian sebagai berikut (Muflikhah *et al.*, 2018):

- a. Apabila terdapat sekumpulan *record (training set)* dan setiap *record* nya terdiri dari sekumpulan atribut dan satu atribut, maka dikatakan sebagai kelas.
- b. Suatu model untuk atribut kelas ditentukan sebagai fungsi nilai dari atribut lain.
- c. Adapun tujuannya ialah *record* yang sebelumnya tidak terlihat ditentukan suatu kelas dengan seakurat mungkin.
- d. Kumpulan dari suatu data uji digunakan untuk menentukan keakuratan suatu model. Pada umumnya, data set dibagi ke dalam sekumpulan data latih dan data uji. Dimana data latih digunakan untuk membentuk model dan data uji digunakan untuk mengujinya.

2.4. Naïve Bayes

Metode *Naïve Bayes* atau *Naïve Bayes Classifier* pertama kali ditemukan oleh Thomas Bayes pada tahun 1770, *Naïve Bayes* ini berasal dari Theorem (teorema Bayes) yang menyatakan seberapa jauh derajat kepercayaan subjektif harus berubah secara rasional ketika diberikan petunjuk baru (Primartha, 2021). Dalam buku Daqiqil Id (2021), *Naïve Bayes* merupakan sebuah teknik yang memanfaatkan bayes dalam melakukan klasifikasi data yang diasumsikan semua fitur pada data bersifat independent.

Algoritma *Naïve Bayes* merupakan metode yang menggunakan *probability* untuk membuat model prediksi klasifikasi yang memanfaatkan data mengenai kejadian masa lampau, model ini juga dapat menghitung *probability* suatu kejadian dan dapat berubah jika ada informasi pendukung tambahan yang disediakan (D. Kurniawan, 2020). Algoritma *Naïve Bayes* dapat dirumuskan sebagai berikut (Nofriansyah & Nurcahyo, 2015):

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (2.1)$$

Keterangan:

- X : Merupakan data testing yang kelasnya belum diketahui
- H : Merupakan hipotesis data X yang kelasnya lebih spesifik
- P(H) : Disebut juga dengan *prior probability* yang merupakan probabilitas hipotesis H
- P(X) : Disebut dengan *predictor prior* yang merupakan probabilitas X
- P(X|H) : Disebut juga dengan *likelihood* yang merupakan probabilitas hipotesis X berdasarkan kondisi H.

2.5. Data Preprocessing

Data preprocessing merupakan tahapan dari *data mining* yang digunakan untuk mendapatkan hasil analisis yang lebih akurat, selain itu *preprocessing* bermanfaat untuk pengurangan waktu komputasi dan mampu membuat nilai data menjadi lebih kecil tanpa mengubah informasi (Santosa & Umam, 2018). Langkah-

langkah *data preprocessing* antara lain *data cleaning*, *data integration* dan *data transformation* (Santosa & Umam, 2018).

Data processing merupakan hal yang penting dalam proses data mining yang memiliki empat tahap, diantaranya *data selection*, *data processing* atau *data cleaning*, *data transformation* dan *data reduction* (Nofriansyah & Nurcahyo, 2015). Tahapan persiapan data terdiri dari beberapa langkah diantaranya yaitu (Purwati *et al.*, 2021):

- a. Ekstraksi fitur dan probabilitas (*Feature extraction and probability*).

Pada tahapan ini data mentah seringkali dalam bentuk proses yang tidak semestinya untuk diproses.

- b. Pembersihan data (*Data cleaning*)

Pada tahapan pembersihan data ini dilakukan pada entri data yang hilang, data yang salah dan data yang tidak konsisten dihapus dari data.

- c. Reduksi, Seleksi, dan Transformasi

Pada tahapan reduksi ini dilakukan dengan pemilihan subset data, seleksi subset fitur, atau transformasi data. Terdapat dua keuntungan yang diperoleh dalam melakukan tahapan ini, yaitu pertama ketika ukuran data diperkecil algoritma akan lebih efisien. Kemudian yang kedua ketika fitur yang tidak relevan atau catatan relevan dihapus, kualitas proses data mining akan ditingkatkan.

2.6. *Confusion Matrix*

Confusion matrix merupakan *matrix* berukuran $N \times N$ yang dapat digunakan untuk permasalahan klasifikasi, dimana N merupakan jumlah kelas yang akan diprediksi (Daqiqil Id, 2021). Terdapat empat kolom pada *confusion matrix* yaitu (Daqiqil Id, 2021):

Tabel 2. 1 *Confusion Matrix* Dengan Dua Kelas

<i>Class</i>	<i>Actual = Yes</i>	<i>Actual =No</i>
<i>Predicted = Yes</i>	TP	FP
<i>Predicted = No</i>	FN	TN

Sumber: Daqiqil Id (2021)

- a. TP (*True Positive*) merupakan jumlah data *point* berlabel *yes* yang nilainya diidentifikasi benar.
- b. TN (*True Negative*) merupakan jumlah data *point* berlabel *no* yang nilainya diidentifikasi salah.
- c. FP (*False Positive*) merupakan jumlah data *point* berlabel *yes* yang nilai sebenarnya diidentifikasi salah.
- d. FN (*False Negative*) merupakan jumlah data *point* berlabel *no* yang nilai sebenarnya teridentifikasi benar.

Confusion matrix dapat dihitung dengan menggunakan perhitungan sebagai berikut (Daqiqil Id, 2021):

- a. Akurasi

Akurasi merupakan perbandingan jumlah item yang diprediksi positif pada seluruh prediksi yang dilakukan.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.2)$$

- b. *Sensitivity*

Sensitivity sering disebut dengan *recall* atau TPR (*True Positive Rate*) merupakan perbandingan jumlah item yang diprediksi positif dengan seluruh item yang positif.

$$Recall = \frac{TP}{TP + FN} \quad (2.3)$$

- c. *Specificity*

Specificity atau TNR (*True Negative Rate*) merupakan perbandingan jumlah item yang diprediksi sebagai negatif dengan pada item yang memang negatif.

$$Specificity = \frac{TN}{TN + FP} \quad (2.4)$$

- d. *Precision*

Precision merupakan perbandingan jumlah item yang diprediksi sebagai positif terhadap item positif.

$$Precision = \frac{TP}{TP + FP} \quad (2.5)$$

e. *False Positive Rate (FPR)*

False Positive Rate (FPR) atau *Type I Error* merupakan perbandingan jumlah item yang negatif diidentifikasi sebagai positif dengan jumlah item yang memang negatif.

$$FPR = \frac{FP}{FP + TN} = 1 - Specificity \quad (2.6)$$

f. *False negative Rate (FNR)*

False negative atau *Type II Error* merupakan perbandingan jumlah item yang salah diidentifikasi sebagai negatif dengan jumlah item yang memang positif.

$$FNR = \frac{FN}{FN + TP} \quad (2.7)$$

2.7. Penelitian Terdahulu

Pada penulisan proposal skripsi ini, penulis menggali informasi dari penelitian-penelitian sebelumnya sebagai bahan perbandingan dan acuan untuk penelitian yang akan dilakukan. Studi mengenai penggunaan atau penerapan metode algoritma *Naïve Bayes* telah banyak dilakukan.

Penelitian yang menggunakan algoritma *Naïve Bayes* diantaranya ialah “Kinerja Metode *Naïve Bayes* dalam Prediksi Lama Studi Mahasiswa Fakultas Ilmu Komputer” (Anggreani *et al.*, 2018). Penelitian ini menggunakan data Indeks Prestasi Kumulatif (IPK), dengan mengambil data latih dan data uji yang memiliki kriteria lama studi tepat waktu dan tidak tepat waktu. Berdasarkan hasil prediksi data uji sebanyak 20 didapatkan nilai akurasi 60% dan 30 data uji diperoleh nilai akurasi sebanyak 76%.

Penelitian selanjutnya dengan judul “Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Muhammadiyah Bengkulu Menggunakan Metode *Naïve Bayes*” (Apridiansyah *et al.*, 2021). Dalam penelitian ini mengambil data uji berupa 20 orang mahasiswa yang lulus pada tahun 2019 dengan menggunakan atribut nama mahasiswa, NPM, jalur masuk, nilai IPK, dan jenis kelamin. Berdasarkan hasil prediksi diperoleh nilai persentase *precision* sebesar 90%, *recall* 100% dan akurasi sebesar 90%.

Penelitian yang menggunakan metode *Naïve Bayes* berikutnya adalah “Implementasi Metode Klasifikasi *Naïve Bayes* Dalam Memprediksi Kelulusan Mahasiswa” (Firdaus *et al.*, 2021). Dalam penelitian ini menggunakan data mahasiswa yang diperoleh dari Fakultas Rekayasa Industri program studi Sistem Informasi dengan jumlah 1249 baris, kemudian menggunakan 80% data sebagai *data training*. Berdasarkan hasil penelitian diperoleh akurasi sebesar 90.78%, presisi data sebesar 88% dan *recall* 88.4%.

Penelitian selanjutnya berjudul “Penggunaan Metode *Naïve Bayes Classifier* untuk Mengevaluasi Kinerja Akademik Mahasiswa di Perguruan Tinggi” (Novitalia *et al.*, 2021). Data yang digunakan dalam penelitian ini adalah dataset kelulusan dari 380 mahasiswa. Berdasarkan hasil pengujian diperoleh hasil akurasi 79,82%. Adapun untuk status kelulusan yang terlambat sebesar 76,00% dan status kelulusan tepat sebesar 82,81%.

Penelitian dengan judul “Analisis Keberhasilan Pembelajaran Daring pada Masa Pandemi Covid-19 menggunakan Algoritma C4.5 dan *Naïve Bayes*” (Sihombing *et al.*, 2021). Dalam penelitian ini menggunakan data yang diperoleh dari nilai siswa tahun ajaran 2020/2021 sebanyak 100 data. Berdasarkan hasil pengujian diperoleh hasil bahwa *performance* atau tingkat akurasi *Naïve Bayes* lebih unggul dibanding C4.5, dimana akurasi *Naïve Bayes* sebesar 99% sedangkan C4.5 sebesar 98%.

Dari uraian diatas algoritma *Naïve Bayes* telah banyak digunakan pada berbagai studi kasus. Namun yang membedakan dari penelitian tersebut dengan penelitian penulis adalah indikator yang digunakan, dimana penulis menggunakan indikator dari *platform OpenLearning* dan nilai akhir yang terdiri dari *time spent on course*, *course completed*, tugas, quiz, uts dan simbol. Kemudian ada juga penelitian yang melakukan komparasi algoritma untuk membandingkan algoritma mana yang memiliki akurasi lebih akurat.