

Kinerja Metode Naïve Bayes dalam Prediksi Lama Studi Mahasiswa Fakultas Ilmu Komputer

1st *Desi Anggreani
Universitas Muslim Indonesia
Fakultas Ilmu Komputer
Makassar, Indonesia
desi.anggreanis12@gmail.com

2nd Herman
Universitas Muslim Indonesia
Fakultas Ilmu Komputer
Makassar, Indonesia
herman@umi.ac.id

3rd Wistiani Astuti
Universitas Muslim Indonesia
Fakultas Ilmu Komputer
Makassar, Indonesia
wistiani.astuti@umi.ac.id

Abstrak--Lama studi seorang mahasiswa menjadi perhatian khusus oleh program studi. Banyak hal yang menjadi faktor terkait lama studi seorang mahasiswa seperti kehadiran, nilai IPK dan keaktifan mahasiswa. Dalam penelitian ini, lama studi mahasiswa akan di analisa untuk membantu pihak program studi dan fakultas dalam mengetahui lama studi mahasiswa. Dalam studi ini dengan metode Naïve Bayes telah digunakan untuk memprediksi masa atau lama studi mahasiswa. Berdasarkan hasil penelitian, dengan mengambil data latih sejumlah 20 dan data uji yang memiliki kriteria tertentu didapatkan hasil bahwa mahasiswa tersebut dapat menyelesaikan lama studi dengan tepat waktu. Dari pengujian hasil prediksi dengan data uji sebanyak 20 didapatkan nilai akurasi 60% dan 30 data uji didapatkan nilai akurasi sebesar 76%, sehingga semakin banyak data uji yang digunakan maka semakin tinggi nilai akurasi yang didapatkan. Adapun nilai yang dianalisa adalah IPK karena merupakan faktor utama dalam menilai lama masa studi mahasiswa. Metode Naïve Bayes merupakan metode yang dapat digunakan dalam memprediksi karena mampu menghasilkan nilai akurasi yang baik

Kata Kunci—lama studi, Naïve Bayes, mahasiswa, prediksi

I. PENDAHULUAN

Kualitas dari sebuah perguruan tinggi dapat dilihat dari sisi lama lulusan perguruan tinggi tersebut mendapatkan pekerjaan, namun hal lain juga selaku menunjang pengukuran suatu kualitas perguruan tinggi dapat dilihat dari lamanya atau masa studi dari mahasiswa perguruan tinggi tersebut. Semua perguruan tinggi baik negeri maupun swasta memiliki lama studi mahasiswa yang berbeda-beda.

Pihak dari fakultas khususnya pada bagian program studi memiliki kewajiban untuk memantau seluruh mahasiswa yang terdaftar pada program studi tersebut. Fungsi lain dari program studi juga harus bisa memprediksi lama studi dari mahasiswanya. Hal tersebut adalah salah satu tujuan agar dapat menghindari adanya mahasiswa yang tidak hadir dari perkuliahan, yang dapat mengakibatkan program studi menjadi kurang baik atau bahkan dapat menurunkan akreditasi [1].

Lama studi mahasiswa dipengaruhi oleh beberapa faktor di antaranya adalah Indeks Prestasi Kumulatif, Tingkat kesibukan masing-masing mahasiswa itu sendiri, dan lain-lainnya [2].

Untuk dapat meningkatkan kualitas perguruan tinggi dengan cara mengambil kesimpulan dan mengetahui atribut-atribut penting yang dapat meningkatkan kualitas mahasiswa dari perguruan tinggi itu sendiri [3].

Pada peraturan akademik Universitas Muslim Indonesia fakultas ilmu komputer untuk jenjang Program Sarjana (S-1) Reguler yaitu lanjutan dari jenjang menengah atas dijalani selama maksimal 6 tahun dan menjalani mata kuliah dengan jumlah sks sebesar 146 dan dijalani selama 8 semester dan paling lama 12 semester.

Tingkat persaingan di kalangan mahasiswa dalam mencari pekerjaan saat ini sangat ketat, oleh karena itu sebagai seorang mahasiswa haruslah menjadi seorang sarjana yang berkualitas dan memiliki daya saing. Salah satu indikator penentu suatu kualitas mahasiswa adalah informasi dari lama studi mahasiswa tersebut [4].

II. METODOLOGI

A. Metode Naïve Bayes

Metode *Naïve Bayes* dikemukakan pertama kali oleh *Thomas Bayes* dan dikembangkan oleh *Pierre-Simon Laplace* pada tahun 1763 yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Konsep *Naïve Bayes* mengadopsi probabilitas [5]. *Naïve Bayes* berkerja dengan cara mengklasifikasikan dan mengatur data sesuai dengan kategorinya masing-masing yang telah diberikan [6]. Klasifikasi merupakan masalah mendasar dalam penambahan data, Dalam suatu klasifikasi dimisalkan ada suatu klasifikasi yang dinamakan kelas dan bagian kelas itu disebut dengan atribut kelas. Dalam kemampuan pemodelan prediktif *Naïve Bayes* sangat sesuai untuk digunakan [7]. Keuntungan menggunakan *Naïve Bayes* bahwa metode ini hanya membutuhkan jumlah data pelatihan yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian [8]. Dalam mengimplementasikan *Naïve Bayes* digunakan data *mining* yang merupakan suatu proses yang berfungsi untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual oleh basis data, diimplementasikan dengan penggalian pola-pola dari data agar dapat dimanipulasi [9]. Disamping data mining, untuk proses

klasifikasi dikenal juga 2 jenis data yaitu data latih dan data uji yang biasa di sebut Training data dan Testing data. Data latih adalah data yang sudah ada sebelumnya berdasarkan fakta yang sudah terjadi sedangkan data uji adalah data yang sudah memiliki label yang akan digunakan untuk melakukan penghitungan sesuai dengan persamaan metode dan perhitungan akurasi klasifikasi yang ada. Data latih dan data uji memiliki hubungan yang sangat erat dimana data latih akan digunakan sebagai acuan saat menghitung nilai probabilitas data uji.

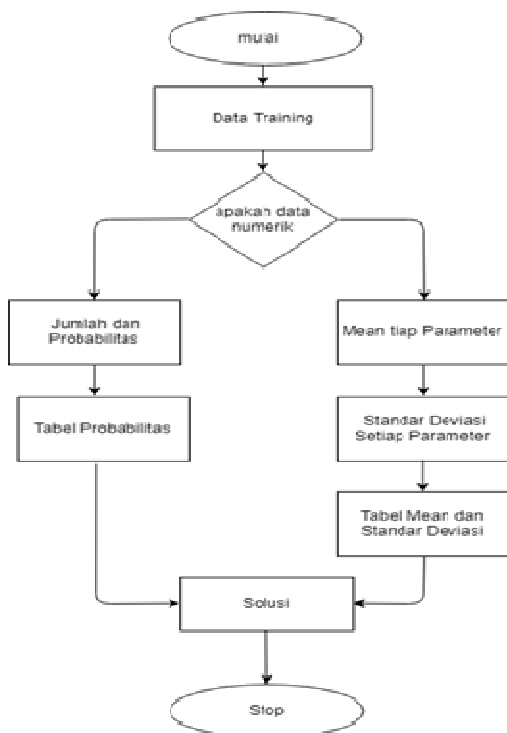
Persamaan (1) metode *Naïve Bayes*

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \tag{1}$$

Keterangan :

- X = Data dengan class yang belum diketahui
- H = Hipotesis data X merupakan suatu class spesifik
- P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X
- P(H) = Probabilitas hipotesis H
- P(X|H) = Probabilitas X berdasarkan kondisi tersebut
- P(X) = Probabilitas dari X

Alur dari metode *Naïve Bayes*, pada Gambar 1.



Gambar. 1. Alur Metode *Naïve Bayes*

Keterangan:

- 1) *Baca Data Training.*
- 2) *Jumlah dan probabilitas.*
- 3) *Jika ada data numeric maka :*
 - Mencari nilai *mean* dan standar deviasi dari masing-masing parameter yang merupakan data numerik.

Persamaan (2) yang digunakan untuk menghitung nilai rata-rata yaitu :

$$\mu = \frac{x_1 x_2 x_3 \dots x_n}{n} \tag{2}$$

Dimana:

- μ = rata-rata nilai
- X_i = nilai sample ke-i
- n = jumlah sample

Persamaan (3) untuk menghitung simpangan baku sebagai berikut:

$$s = \sqrt{\frac{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n-1)}} \tag{3}$$

Dimana, s adalah simpangan baku, x_i adalah nilai x ke- i , \bar{x} adalah rata-rata, dan n adalah ukuran sample

- Mencari probabilitas dengan cara menghitung jumlah data sesuai dengan kategori yang sama dibagi dengan jumlah data pada kategorinya.
- 4) *Nilai telah didapatkan pada table mean.*
- 5) *Mengambil solusi yang dihasilkan*[10].

Setelah melakukan perhitungan persamaan dengan menggunakan persamaan *Naïve Bayes*, selanjutnya akan dilakukan pengujian dan analisa terhadap hasil yang didapatkan untuk mengetahui hasil dari pengujian apakah dikategorikan akurat atau kurang akurat untuk digunakan. Ada beberapa cara untuk melakukan pengukuran seperti menggunakan *Precision* dan *Recall*, *Error*, akurasi, dan lain sebagainya [11]. Dalam penelitian ini metode yang digunakan adalah *Precision*, *Recall*, dan *Accuracy* dengan persamaan (4), (5), dan (6). Untuk klasifikasikan data sesuai kategori ada pada Tabel I.

TABEL I. CONFUSION MATRIX

Nilai Prediksi	Kategori	Nilai Sebenarnya	
		Benar	Salah
	Benar	TP	FP
	Salah	FN	TN

Precision

$$\frac{TP}{TP + FP} \quad (4)$$

Recall

$$\frac{TP}{TP + FN} \quad (5)$$

Accuracy

$$\frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Keterangan:

TP = Klasifikasi bernilai benar menurut prediksi dan benar menurut nilai sebenarnya

FP = Klasifikasi bernilai benar menurut prediksi dan salah menurut nilai sebenarnya

FN = Klasifikasi bernilai salah menurut prediksi dan benar menurut nilai sebenarnya

TN = Klasifikasi bernilai salah menurut prediksi dan salah menurut nilai sebenarnya

III. HASIL DAN PEMBAHASAN

A. Dataset

Dalam proses mining agar dapat dilakukan pengujian dengan mudah maka dilakukan konversi data ke dalam bentuk yang dapat diolah oleh alat bantu data mining. Adapun hasil konversi data yaitu :

1) IPK

Nilai IPK terdiri dari nilai desimal yang akan dikonversi dengan range pada Tabel II.

TABEL II. KONVERSI NILAI IPK

IPK	IPK baru
IPK >= 3.50	3
IPK >= 2.50	2
IPK >= 0.00	1

2) Jenis Kelamin

Tidak dilakukan konversi karna hanya terdiri dari dua jenis yaitu laki-laki dan perempuan.

3) Kota Lahir

Dibagi menjadi dua tempat yaitu lahir di kota Makassar dan lahir di luar kota Makassar.

4) Tipe sekolah

Tipe sekolah dalam hal ini adalah tipe sekolah sebelum bangku perkuliahan yaitu dibagi menjadi dua, sekolah kejuruan komputer atau sekolah biasa dimana sekolah biasa ini yaitu mahasiswa yang lulusan dari sekolah menengah atas yang terdiri dari peminatan IPS atau IPA.

5) Lama studi

Dibagi menjadi dua yaitu tepat waktu dengan 8 semester dan tidak tepat waktu melebihi 8 semester.

Dataset yang digunakan untuk analisa yaitu data dari lulusan Fakultas Ilmu Komputer Universitas Muslim Indonesia sesuai dengan faktor nilai IPK, Jenis Kelamin, Kota Lahir, Tipe Sekolah, Lama Studi. Setelah dilakukan konversi maka dimasukkan ke dalam tabel data training. Data yang diambil sebanyak 20 data mahasiswa, Tabel III.

TABEL III. DATA TRAINING

IPK	Jenis Kelamin	Kota Lahir	Tipe Sekolah	Lama Studi
3	Laki-laki	Luar Makassar	Biasa	Tidak Tepat waktu
3	Laki-laki	Makassar	Biasa	Tepat Waktu
3	Perempuan	Luar Makassar	Biasa	Tepat Waktu
2	Perempuan	Luar Makassar	Biasa	Tepat Waktu
2	Laki-Laki	Makassar	Biasa	Tidak Tepat Waktu
2	Perempuan	Luar Makassar	Biasa	Tepat Waktu
3	Laki-laki	Luar Makassar	Kejuruan	Tepat Waktu
3	Laki-laki	Luar Makassar	Biasa	Tepat Waktu
2	Perempuan	Luar Makassar	Kejuruan	Tepat Waktu
3	Perempuan	Luar Makassar	Biasa	Tepat Waktu
2	Laki-Laki	Luar Makassar	Biasa	Tidak Tepat Waktu
3	Laki-Laki	Makassar	Biasa	Tepat Waktu
2	Laki-Laki	Luar Makassar	Kejuruan	Tidak Tepat Waktu
2	Laki-Laki	Luar Makassar	Kejuruan	Tepat Waktu
2	Laki-Laki	Makassar	Biasa	Tidak Tepat waktu
2	Laki-Laki	Makassar	Kejuruan	Tidak Tepat waktu
2	Laki-Laki	Makassar	Biasa	Tidak Tepat Waktu
3	Perempuan	Makassar	Kejuruan	Tepat Waktu
2	Laki-Laki	Makassar	Biasa	Tidak Tepat Waktu
3	Perempuan	Luar Makassar	Kejuruan	Tepat Waktu

Ada sebuah data yang akan menjadi data testing dengan mengacu pada data *training* yang ada :

Nama : ULHAK
IPK : 3.78
Jenis Kelamin : laki-laki
Kota Lahir : Luar Makassar
Tipe Sekolah : Biasa

Dengan menggunakan persamaan nomor 1 :

Mencari nilai P(H)

$$P(\text{Lama Studi} = \text{"Tepat Waktu"}) = 12$$

$$P(\text{Lama Studi} = \text{"Tidak Tepat Waktu"}) = 8$$

Mencari nilai P(H|X)

$$P(\text{IPK} = \text{"3"} | \text{Lama Studi} = \text{"Tepat Waktu"}) = 8/12$$

$$P(\text{IPK} = \text{"3"} | \text{Lama Studi} = \text{"Tidak Tepat Waktu"}) = 1/8$$

$P(\text{Jenis Kelamin} = \text{"Laki-Laki"} \mid \text{Lama Studi} = \text{"Tepat Waktu"}) = 5/12$

$P(\text{Jenis Kelamin} = \text{"Laki-Laki"} \mid \text{Lama Studi} = \text{"Tidak Tepat Waktu"}) = 8/8$

$P(\text{Kota Lahir} = \text{"Luar Makassar"} \mid \text{Lama Studi} = \text{"Tepat Waktu"}) = 9/12$

$P(\text{Kota Lahir} = \text{"Luar Makassar"} \mid \text{Lama Studi} = \text{"Tidak Tepat Waktu"}) = 3/8$

$P(\text{Tipe Sekolah} = \text{"Biasa"} \mid \text{Lama Studi} = \text{"Tepat Waktu"}) = 7/12$

$P(\text{Tipe Sekolah} = \text{"Biasa"} \mid \text{Lama Studi} = \text{"Tidak Tepat Waktu"}) = 6/8$

$P(H \mid \text{Lama Studi} = \text{"Tepat Waktu"}) = P(\text{IPK} = \text{"3"}, \text{Jenis Kelamin} = \text{"Laki-Laki"}, \text{Kota Lahir} = \text{"Luar Makassar"}, \text{Tipe Sekolah} = \text{"Biasa"})$

$$= 8/12 * 5/12 * 9/12 * 7/12$$

$$= 0.66 * 0.41 * 0.75 * 0.58$$

$$= 0.11$$

$P(H \mid \text{Lama Studi} = \text{"Tidak Tepat Waktu"}) = P(\text{IPK} = \text{"3"}, \text{Jenis Kelamin} = \text{"Laki-Laki"}, \text{Kota Lahir} = \text{"Luar Makassar"}, \text{Tipe Sekolah} = \text{"Biasa"})$

$$= 1/8 * 8/8 * 3/8 * 6/8$$

$$= 0.12 * 1 * 0.37 * 0.75$$

$$= 0.033$$

$P(H \mid \text{Lama Studi} = \text{"Tepat Waktu"}) * P(H) / P(X)$

$$= 0.117 * 12 / 20$$

$$= 0.0702$$

$P(H \mid \text{Lama Studi} = \text{"Tidak Tepat Waktu"}) * P(H) / P(X)$

$$= 0.033 * 8 / 20$$

$$= 0.013$$

TABEL IV. HASIL KESIMPULAN LAMA STUDI DARI DATA TESTING

$P(H \mid \text{Lama Studi} = \text{"Tepat Waktu"})$	0.0702
$P(H \mid \text{Lama Studi} = \text{"Tidak Tepat Waktu"})$	0.013
Kesimpulan Hasil	Tepat Waktu

Berdasarkan Tabel IV, perhitungan yang telah dilakukan dari data testing dengan memperhatikan beberapa faktor maka didapatkan kesimpulan bahwa diprediksi mahasiswa tersebut dapat menjalani lama studi dengan Tepat Waktu.

Selanjutnya akan dilakukan evaluasi pada penggunaan metode *naïve bayes* dalam memprediksi lama studi mahasiswa.

Teknik evaluasi yang digunakan yaitu metode *Precision, Recall, dan Accuracy*. Penghitungan akurasi metode berdasarkan 20 data yang digunakan dapat dilihat pada Tabel V.

TABEL V. PENGUKURAN AKURASI DATA

Mengukur	Rumus	Nilai	Persentasi
Precision	$TP/(TP+FP)$	$10/(10+2)$	85%
Recall	$TP/(TP+FN)$	$10/10+1$	90%
Accuracy	$(TP+TN)/(TP+TN+FP+FN)$	$(10+7)/10+7+2+1$	60%

Berdasarkan Tabel V, nilai akurasi dengan menggunakan 20 data yaitu 60% dengan demikian metode *naïve bayes* dapat dikategorikan layak untuk digunakan untuk memprediksi lama studi mahasiswa. Perbedaan nilai akurasi dari jumlah data training dapat dilihat pada Tabel VI.

TABEL VI. PERBANDINGAN AKURASI DATA

Jumlah Data	Persentasi Akurasi
20	60%
30	76%

Dari pengujian pada Tabel VI di dapatkan bahwa dengan menggunakan data uji sebanyak 20 maka akan didapatkan persentasi akurasi sebesar 60%, dan jika data uji sebanyak 30 maka didapatkan persentasi akurasi sebesar 76%. Sehingga dapat dilihat bahwa jumlah data uji mempengaruhi persentasi akurasi, semakin banyak data uji yang digunakan maka semakin tinggi nilai akurasi dari implementasi metode.

IV. KESIMPULAN

Dari penelitian yang telah dilakukan mengenai prediksi lama studi mahasiswa Fakultas Ilmu Komputer Universitas Muslim Indonesia dengan menggunakan metode *Naïve Bayes* maka di ambil kesimpulan bahwa dengan jumlah data latih sebanyak 20 data dan data uji dengan kriteria nilai IPK = 3.78, Jenis Kelamin = "Laki-Laki", Kota Lahir = "Luar Makassar", Tipe Sekolah = "Biasa" didapatkan hasil prediksi adalah lama studi mahasiswa tersebut masuk dalam kategori lama studi dengan "Tepat waktu".

Prediksi lama studi mahasiswa memiliki nilai *Precision* sebesar 85%, *Recall* sebesar 90%, dan *Accuracy* sebesar 60%. Dapat diambil kesimpulan bahwa prediksi lama studi mahasiswa Fakultas Ilmu Komputer dengan menggunakan metode *Naïve Bayes* layak diimplementasikan pada kasus tersebut.

DAFTAR PUSTAKA

- [1] A. Jananto, "Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa $P(H \mid X) P(X \mid H) P(H)$," vol. 18, no. 1, pp. 9–16, 2013.
- [2] J. Gaussian, "1, 2, 3 1," vol. 3, pp. 173–181, 2014.
- [3] M. M. A. Tair and A. M. El-halees, "Mining Educational Data to Improve Students' Performance: A Case Study," vol. 2, no. 2, pp. 140–146, 2012.

- [4] S. Kasus, U. Dehasen, S. Haryati, A. Sudarsono, and E. Suryana, "IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA C4.5," vol. 11, no. 2, pp. 130–138, 2015.
- [5] D. Kurniawaty, S. K. Sari, I. Cholissodin, F. Ilmu, and K. Universitas, "IMPLEMENTASI METODE NAÏVE BAYES CLASSIFIER UNTUK SELEKSI," vol. 3, no. 4, pp. 273–278, 2016.
- [6] R. Gupta, N. Joshi, and I. Mathur, "Quality Estimation of English-Hindi Outputs using Naïve Bayes Classifier," pp. 3–6.
- [7] C. Science and S. Engineering, "Prediction of Different Dermatological Conditions Using Naïve Bayesian Classification," vol. 4, no. 1, pp. 864–868, 2014.
- [8] A. Saleh and U. P. Utama, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," no. June, 2016.
- [9] D. Teori, "PENGUNAAN METODE NAÏVE BAYES CLASSIFIER PADA APLIKASI PERPUSTAKAAN."
- [10] M. C. Wijanto, "Sistem Pendeteksi Pengirim Tweet dengan Metode Klasifikasi Naive Bayes," *J. Tek. Inform. dan Sist. Inf. Vol.*, vol. 1, no. 2, pp. 172–182, 2015.
- [11] F. Handayani and S. Pribadi, "Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110," vol. 7, no. 1, 2015.