**UJIAN TENGAH SEMESTER**

**SISTEM CERDAS**

**IMPLEMENTASI METODE *NAÏVE BAYES* UNTUK MEMPREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU BERBASIS WEB**

****

**Disusun Oleh:**

**ADYTIA DWI HERMAWAN 1810817210007**

**PROGRAM STUDI TEKNOLOGI INFORMASI**

**FAKULTAS TEKNIK**

**UNIVERSITAS LAMBUNG MANGKURAT**

**2021**

# DAFTAR ISI

[DAFTAR ISI 2](#_Toc69595667)

[BAB 1 PENDAHULUAN 3](#_Toc69595668)

[1.1. Latar Belakang 3](#_Toc69595669)

[1.1. Rumusan Masalah 4](#_Toc69595670)

[1.2. Tujuan Penelitian 4](#_Toc69595671)

[1.3. Batasan Masalah 4](#_Toc69595672)

[1.4. Manfaat Penelitian 4](#_Toc69595673)

[BAB 2 TINJAUAN PUSTAKA 5](#_Toc69595674)

[2.1. Mahasiswa 5](#_Toc69595675)

[2.2. *Data Mining* 5](#_Toc69595676)

[2.3. Algoritma Naïve Bayes 5](#_Toc69595677)

[2.4. Confusion Matrix 6](#_Toc69595678)

[2.5. Laplace Correction 7](#_Toc69595679)

[2.6. PHP 7](#_Toc69595680)

[2.7. PHPMyAdmin 7](#_Toc69595681)

[2.8. Visual Studio Code 7](#_Toc69595682)

[BAB 3 METODELOGI PENELITIAN 8](#_Toc69595683)

[DAFTAR PUSTAKA 9](#_Toc69595684)

# BAB 1 PENDAHULUAN

## Latar Belakang

Persaingan di dunia pendidikan saat ini sangatlah kompetitif antara perguruan tinggi, baik perguruan tinggi negeri ataupun perguruan tinggi swasta. Mereka saling mengejar satu sama lain untuk menjadi yang terbaik. Perguruan tinggi dapat dikatakan baik bisa diliat dari nilai akreditasi perguruan tinggi, serta akreditasi program studinya. Salah satu faktor yang mempengaruhi nilai akreditasi menurut Badan Akreditas Nasional Perguruan Tinggi (BAN PT) adalah mahasiswa dan lulusan[1].

Melihat kelulusan mahasiswa merupakan faktor yang membantu dalam menaikkan nilai akreditasi, maka diperlukan monitoring dari pihak perguruan tinggi terhadap mahasiswa dalam menempuh pendidikannya. Salah satu caranya bisa dengan memonitoring mahasiswa agar tidak melebihi batas waktu yang sudah ditentukan perguruan tinggi dengan memperhitungkan kelulusan mahasiswa akan terjadi disemester berapa dan mengetahui faktor apa saja yang mempengaruhi kemampuan mereka dalam belajar.

*Naïve Bayes* adalah salah satu algoritma dalam pengklasifikasian data. Algoritma *Naïve Bayes* menggunakan klasifikasi probabilitas yang cukup simple. Algoritma ini memperhitungkan satu set probabilitas dengan menghitung frekuensi dan kombinasi nilai dalam beberapa kumpulan data[2]. Penggunaan metode *Naïve Bayes* ini akan menghitung probabilitas beberapa variable yang diperkirakan akan mempengaruhi cepat lambat lulusnya mahasiswa.

Telah banyak penelitian yang membahas terkait prediksi kelulusan mahasiswa dengan berbagai metode seperti pada penelitian [3] yang menggunakan *Support Vector Machine* yang menghasilkan akurasi sebesar 86,43% dan ada penggunaan metode Algoritma C4.5 pada penelitian [4] dengan hasil analisis kalau kelulusan diliat dari gender akan menghasilkan persentase lebih besar terhadap gender perempuan dibandingkan gender laki-laki. Kemudian pada penelitian [5] yang menggunakan metode *Naïve Bayes* untuk memprediksi kelulusan mahasiswa tepat waktu mendapatkan akurasi diatas 90%.

Dengan mempertimbangkan tingginya akurasi perhitungan menggunakan metode *Naïve Bayes,* maka penelitian ini akan menggunakan metode *Naïve Bayes* dalam memprediksi kelulusan mahasiswa dibantu dengan cara *Laplace Correction* untuk menangani nilai probabilitas 0. Sehingga penelitian ini akan mengangkat judul “Implementasi Metode *Naïve Bayes* untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu Berbasis Web”.

## Rumusan Masalah

Berapa tingkat akurasi metode *Naïve Bayes* dalam memprediksi kelulusan mahasiswa tepat waktu?

## Tujuan Penelitian

Untuk mengetahui tingkat akurasi metode *Naïve Bayes* dalam memprediksi kelulusan mahasiswa tepat waktu.

## Batasan Masalah

1. Atribut yang digunakan masih tergolong sedikit, seperti jenis kelamin, status mahasiswa, status pernikahan, dan IPK akhir.
2. Algoritma yang digunakan hanya *Naïve Bayes*

## Manfaat Penelitian

1. Bisa menjadi referensi untuk membangun sistem prediksi kelulusan mahasiswa yang lebih baik.
2. Untuk meningkatkan kesadaran mahasiswa dalam menempuh pendidikannya agar tidak terlambat lulus.

# BAB 2 TINJAUAN PUSTAKA

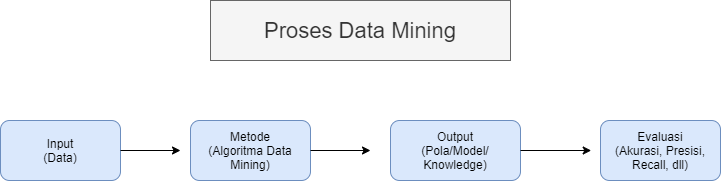
Berikut ini akan dibahas teori yang akan digunakan dalam penelitian ini, antara lain:

## Mahasiswa

Mahasiswa adalah seseorang yang sedang dalam proses menuntut ilmu dan terdaftar sedang menjalani pendidikan di perguruan tinggi. Mahasiswa dalam Kamus Besar Bahasa Indonesia adalah orang yang belajar di perguruan tinggi[6].

## *Data Mining*

*Data mining* merupakan suatu proses dalam penambangan data untuk mendapatkan informasi baru dari sejumlah data yang besar. *Data Mining* memiliki banyak fungsi, tapi fungsi utama *data mining* ada dua, yaitu fungsi deskripsi dan fungsi prediksi. Fungsi deskripsi dalam *data mining* digunakan untuk memahami lebih detil tentang data yang digunakan, sedangkan fungsi prediksi digunakan agar suatu proses dapat menemukan pola tertentu dari suatu data yang diolah.

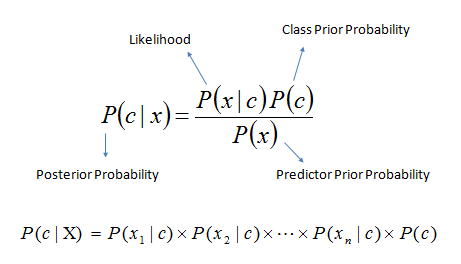


### **Gambar 2.2.1** Proses Data Mining

## Algoritma Naïve Bayes

Algoritma *Naïve Bayes* adalah klasifikasi probabilistic sederhana. *Naïve Bayes* digunakan untuk menghitung satu set probabilitas dengan menghitung frekuensi dan kombinasi nilai dalam suatu kumpulan data tertentu[2]. Kumpulan *data training* adalah bagian yang digunakan untuk melatih algoritma *Naïve Bayes*. Penggunaan nilai pada *data training* yang sudah ada untuk memprediksi nilai yang tidak diketahui.

Rumus *Naïve Bayes*:



Keterangan:

x : data dengan *class* yang belum diketahui  
c : hipotesis data x merupakan suatu *class* spesifik  
P(c|x) : probabilitas hipotesis c berdasar kondisi x  
P(c) : probabilitas hipotesis c  
P(x|c) : probabilitas x berdasar kondisi hipotesis c  
P(x) : probabilitas dari x

## Confusion Matrix

*Confusion matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada pengukuran kinerja menggunakan *confusion matrix* ada 4 variabel yang digunakan untuk memberikan hasil proses klasifikasi, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN).

TP merupakan data positif yang terdeteksi benar, untuk TN merupakan data negatif yang terdeteksi benar. Sementara FP merupakan data negatif yang terdeteksi sebagai data positif, sedangkan FN merupakan data positif yang terdeteksi sebagai data negatif[7]. Tabel *confusion matrix* bisa dilihat dibawah:

**Tabel 2.4.1** Tabel Confusion Matrix

Selanjutnya, untuk menghitung nilai *Accuracy, Recall,* dan *Precision* bisa menggunakan rumus berikut:

## Laplace Correction

*Laplace Correction* adalah teknik perhitungan yang digunakan untuk menghindari nilai probabilitas yang bernilai 0. *Laplace Correction* juga memiliki nama lain, yaitu *Laplace Estimator*. Metode perhitungan ini ditemukan oleh Pierre Laplace pada abad ke-18.

## PHP

PHP adalah bahasa pemrograman yang digunakan untuk mengembangkan situs web. PHP yang merupakan singkatan dari *Hypertext Preprocessor* yang sebelumnya disebut *Personal Home Pages*. Fungsi PHP adalah untuk membangun atau mengembangkan suatu situs web baik yang statis maupun dinamis.

## PHPMyAdmin

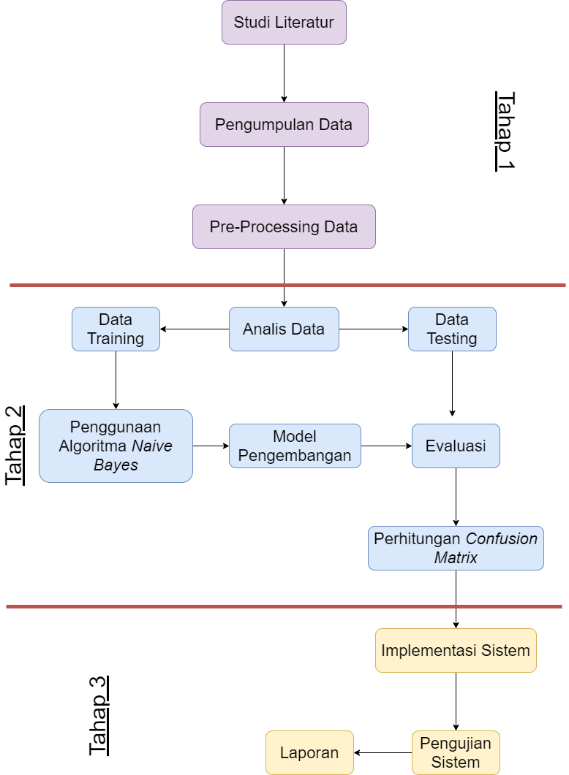
PHPMyAdmin adalah aplikasi berbasis web yang berguna untuk pengelolaan database MySQL melalui jaringan lokal maupun internet.

## Visual Studio Code

Visual Studio Code merupakan perangkat lunak yang ringan untuk editor source code dari Microsoft. Visual Studio Code didukung dengan built-in untuk JavaScript, naskah dan Node.js dan memiliki array beragam ekstensi yang tersedia untuk bahasa lain, termasuk C ++, C # , Python, dan PHP.

# BAB 3 METODELOGI PENELITIAN

## Prosedur Penelitian



### **Gambar 3.1.1** Prosedur Penelitian

Penelitian dimulai dengan studi literatur sebelum melakukan pengumpulan data. Setelah data didapatkan makan akan dilakukan proses *pre-processing* data untuk memilih lagi data yang layak dipakai agar tidak terjadinya permasalahan seperti *missing value, data redundant,* format data yang tidak sesuai, dll. Setelah proses *pre-processing* selesai maka akan dilakukan analis data yang mana dilakukan pembagian antara *data training* dan *data testing*. Pada gambar bisa diliat, data training diuji menggunakan metode *Naïve Bayes* setelah proses uji selesai dilakukan model pengembangan dan lanjut evaluasi. Untuk data training sendiri menggunakan pengujian hasil evaluasi baru dilanjutkan perhitungan menggunakan *confusion matrix*.

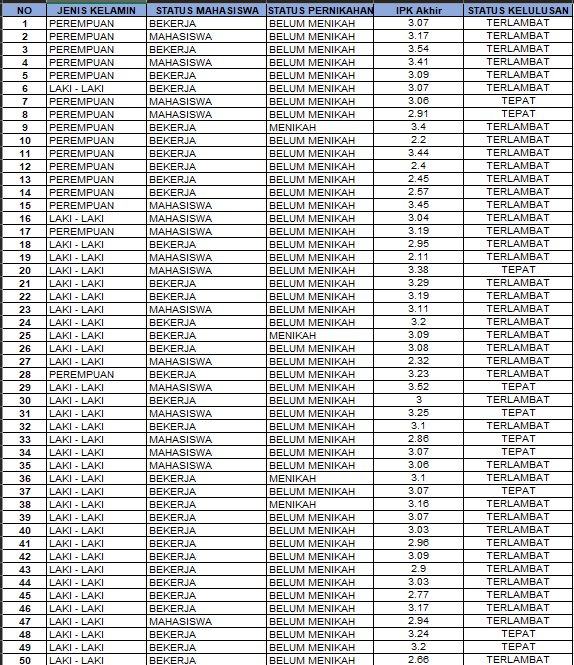
Jika perhitungan manual sudah lumayan akurat, dilanjutkan ke pengimplementasian kedalam sistem lalu dilakukan pengujian dan dibandingkan dengan hasil perhitungan manual sebelumnya.

## Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari salah satu platform *open source* yaitu, github. Berikut link tempat download dataset untuk penelitian ini <https://github.com/achmatim/data-mining/tree/main/Dataset>. Metode yang digunakan adalah *Naïve Bayes* untuk memprediksi ketepatan kelulusan mahasiswa. Data yang didapat akan diklasifikasikan menggunakan metode *Naïve Bayes* dengan total *data training* sebanyak 150 data, sedangkan untuk *data testing* sebanyak 6 data.

Setelah pengklasifikasian selesai, maka dilanjutkan ke bagian pengukuran kinerja metode klasifikasi menggunakan *confusion matrix. Confusion matrix* digunakan untuk menghitung *accuracy, recall,* dan *precision* dari metode yang diterapkan.

## Data Training

**Gambar 3.3.1** *Data Training* 1-50

### **Gambar 3.3.2** *Data Training* 51-99

### **Gambar 3.3.3** *Data Training* 100-150

## Data Testing

### **Gambar 3.4.1** *Data Testing*

## Hasil Perhitungan Manual

Rumus yang digunakan:

Keterangan:  
P = Probabilitas  
Ci = Kelas yang dicari (Tepat / Terlambat)  
Data = *Data Training*Xi = Data variable yang ingin dicari nilai probabilitas

1. Perhitungan Probabilitas Status Kelulusan

### **Gambar 3.5.1.1** Perhitungan Probabilitas Status Kelulusan

1. Perhitungan *Data Testing* 1

### **Gambae 3.5.2.1** Perhitungan *Data Testing* 1

1. Perhitungan *Data Testing* 2

### **Gambar 3.5.3.1** Perhitungan *Data Testing* 2

1. Perhitungan *Data Testing* 3

### **Gambar 3.5.4.1** Perhitungan *Data Testing* 3

1. Perhitungan *Data Testing* 4

### **Gambar 3.5.5.1** Perhitungan *Data Testing* 4

1. Perhitungan *Data Testing* 5

### **Gambar 3.5.6.1** Perhitungan *Data Testing* 5

1. Perhitungan *Data Testing* 6

### **Gambar 3.5.7.1** Perhitungan *Data Testing* 6

1. Perhitungan Menggunakan *Confusion Matrix*

### **Gambar 3.5.8.1** Perhitungan *Confusion Matrix*

Berdasarkan gambar diatas, dengan menggunakan 6 *data testing* hanya dapat menghasilkan akurasi sebesar 67%.

# DAFTAR PUSTAKA

[1] “Akreditasi, Pendorong Peningkatan Kualitas Lembaga Pendidikan? Halaman all - Kompasiana.com.” https://www.kompasiana.com/hferdinando/5b0eab06f133443aa53cabe2/akreditasi-pendorong-peningkatan-kualitas-lembaga-pendidikan?page=all (accessed Apr. 17, 2021).

[2] A. P. Wibawa *et al.*, “Naïve Bayes Classifier for Journal Quartile Classification,” *Int. J. Recent Contrib. from Eng. Sci. IT*, vol. 7, no. 2, p. 91, 2019, doi: 10.3991/ijes.v7i2.10659.

[3] Suhardjono, W. Ganda, and H. Abdul, “Prediksi Kellusan Menggunakan Svm Berbasis Pso,” *Bianglala Inform.*, vol. 7, no. 2, pp. 97–101, 2019.

[4] J. H. J. Universitas, “PREDIKSI KELULUSAN MAHASISWA DENGAN METODE ALGORITMA C4.5,” vol. 2, 2013, doi: 10.1002/jhet.1722.

[5] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, “Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review,” *Fakt. Exacta*, vol. 13, no. 1, pp. 38–47, 2020, doi: 10.30998/faktorexacta.v13i1.5548.

[6] “Arti kata mahasiswa - Kamus Besar Bahasa Indonesia (KBBI) Online.” https://kbbi.web.id/mahasiswa (accessed Apr. 17, 2021).

[7] “Mengukur Kinerja Algoritma Klasifikasi dengan Confusion Matrix – Achmatim.Net.” https://achmatim.net/2017/03/19/mengukur-kinerja-algoritma-klasifikasi-dengan-confusion-matrix/ (accessed Apr. 17, 2021).