

## **PROPOSAL SKRIPSI**

***Data Mining Untuk Memprediksi Kelulusan Mahasiswa Matematika UIN  
Syarif Hidayatullah Jakarta Menggunakan Naive Bayes***



**Disusun oleh :**

**Ade Kurniasih**

**11190940000010**

**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH  
JAKARTA  
2022 M / 1443 H**

## **PROPOSAL SKRIPSI**

### ***Data Mining Untuk Memprediksi Kelulusan Mahasiswa Matematika UIN Syarif Hidayatullah Jakarta Menggunakan Naive Bayes***

**Diajukan oleh :**

**Ade Kurniasih**

**11190940000010**

Usulan Dosen Pembimbing

Dosen Pembimbing I

**Ibu Madona Yunita Wijaya, M.Sc**

NIP 198506242019032007

Memiliki pendidikan terakhir yaitu S2 di Universiteit Hasselt, Belgia. Beliau memiliki keahlian mengajar dalam bidang Statistika.

Dosen Pembimbing II

**Ibu Dr. Nur Inayah, M.Si**

NIP 197401252003122001

Memiliki pendidikan terakhir yaitu S3 di Institut Teknologi Bandung. Beliau memiliki keahlian mengajar dalam bidang Aljabar Graf.

## **BAB I. LATAR BELAKANG**

### ***A. Perumusan Masalah***


Pendidikan adalah usaha untuk menyiapkan seorang manusia melalui berbagai kegiatan bimbingan, pengajaran, dan latihan yang berpengaruh terhadap sikap dan tingkah laku untuk meningkatkan kualitas kehidupan manusia. Pendidikan merupakan hal yang terpenting dalam kehidupan setiap manusia, karena didalamnya terdapat kegiatan yang kompleks, memiliki dimensi yang luas, dan memiliki banyak variabel yang mempengaruhinya [1]. Melalui pendidikan itulah manusia dapat memperluas wawasannya dan memperoleh ilmu pengetahuan. Pendidikan dapat meningkatkan spritual keagamaan, kecerdasan, kepedulian, akhlak mulia serta keterampilan yang sangat diperlukan dirinya dan masyarakat [2]. Pendidikan juga merupakan hal yang sangat penting dalam kehidupan sehingga dapat dilihat profil pribadi muslim yang ideal adalah pribadi yang memiliki wawasan luas. Dan dengan ilmu pula manusia ditinggikan derajatnya oleh Allah SWT.

Sejak awal tahun 1970-an, sistem pendidikan di Indonesia mengalami perubahan terus-menerus, sejalan dengan program pembangunan di bidang pendidikan yang mulai dilaksanakan secara terprogram sejak 40 tahun yang lalu. Sejak saat itu pemerintah mulai merintis program perluasan kesempatan belajar bagi semua warga negara, pada semua jenis jenjang pendidikan [3]. Keberhasilan sistem pendidikan khususnya perguruan tinggi dapat dilihat dari meningkatnya kualitas pendidikan dari tahun ketahun. Salah satu yang menjadi indikator kualitas pendidikan adalah tingkat kelulusan mahasiswa dan banyaknya lulusan yang bekerja disuatu lapang pekerjaan.

Saat ini perguruan tinggi dituntut untuk memiliki keunggulan bersaing dan kualitas yang baik. Dengan memanfaatkan pengelolaan sumber daya yang dilakukan secara baik. Selain sumber daya sarana, prasarana dan manusia. Sumber daya teknologi informasi merupakan salah satu sumber daya yang dapat meningkatkan keunggulan bersaing dan kualitas yang baik. Teknologi informasi yang canggih dapat menghasilkan data yang diinginkan masyarakat baik bidang industri, bidang ekonomi,

bidang pendidikan dan berbagai bidang lainnya. Dalam bidang pendidikan teknologi informasi sangat dibutuhkan untuk menghasilkan data yang berlimpah mengenai data mahasiswa dan dapat menunjang setiap kegiatan pembelajaran yang dihasilkan.

Universitas Islam Negeri Syarif Hidayatullah Jakarta merupakan salah satu perguruan tinggi yang bergerak dalam bidang pengembangan ilmu keislaman integratif-multidisipliner berwawasan lingkungan yang ada di Indonesia. Berdasarkan data yang diperoleh dari BIROPK (Badan Perencanaan dan Keuangan) UIN Syarif Hidayatullah Jakarta bahwa jumlah mahasiswa baru dengan tingkat pendidikan S1 pada tahun 2014 memiliki jumlah sebanyak 4.950, sedangkan mahasiswa dengan tingkat pendidikan S1 yang dapat lulus pada tahun 2018 memiliki jumlah sebanyak 3.822 sehingga diperoleh persentase kelulusan sebesar 77,21% pada periode tersebut. Pada tahun 2015 jumlah mahasiswa baru dengan tingkat pendidikan S1 memiliki jumlah sebanyak 5.400, sedangkan mahasiswa dengan tingkat pendidikan S1 yang dapat lulus pada tahun 2019 memiliki jumlah sebanyak 3.952 sehingga diperoleh persentase kelulusan sebesar 73,18% pada periode tersebut. Berdasarkan persentase diatas dapat disimpulkan selama 2 tahun persentase rata-rata kelulusan mahasiswa Universitas Islam Negeri Syarif Hidayatullah Jakarta sebesar 75,19%.



**JUMLAH DATA MAHASISWA MASUK DAN LULUS**  
**UIN SYARIF HIDAYATULLAH JAKARTA**  
**TAHUN 2014 SD. 2019\***

\* Sumber : ois.uinjkt.ac.id, Tanggal 30 nov 2019

Tahun	Jum.Mahasiswa Baru			Jumlah	Jum.Mahasiswa Lulus			Jumlah
	S1	S2	S3		S1	S2	S3	
2019	6,451	472	86	7,009	3,952	262	54	4,268
2018	6,460	488	90	7,038	3,822	269	38	4,129
2017	5,607	403	123	6,133	3,602	189	63	3,854
2016	5,321	417	111	5,849	4,084	134	25	4,243
2015	5,400	365	96	5,861	4,374	70	5	4,449
2014	4,950	278	82	5,310	4,606	97	28	4,731

Gambar 1. Sumber Biro Perencanaan dan Keuangan UIN Syarif Hidayatullah Jakarta

Berdasarkan Gambar 1. tersebut diketahui bahwa jumlah mahasiswa baru terus bertambah namun tidak semua mahasiswa dapat lulus tepat waktu. Sehingga mengakibatkan jumlah mahasiswa semakin bertambah dan dapat pula mengakibatkan berlimpahnya data mahasiswa. Jumlah mahasiswa baru dan mahasiswa yang telah lulus setiap tahunnya tidak sebanding, untuk itu dibutuhkan sistem yang dapat memprediksi kelulusan mahasiswa. Sistem ini digunakan untuk memprediksi kemungkinan mahasiswa dapat lulus tepat waktu atau tidaknya.

Untuk membantu sistem prediksi kelulusan mahasiswa dalam pencarian suatu informasi berharga dengan berlimpahnya data mahasiswa. Maka penggunaan data mining dapat membantu sistem prediksi kelulusan mahasiswa. Data mining, sering juga disebut sebagai Knowledge Discovery in Database (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, histori untuk menemukan keteraturan, pola atau hubungan dalam dataset yang berukuran besar [4]. Mahmud Yunus dalam penelitian yang telah dilakukannya menyatakan bahwa data mining mampu mengolah data yang sangat besar, semakin banyak data yang dikumpulkan dari tahun ke tahun, data mining dapat mengolah data tersebut menjadi sebuah informasi [5].

Penelitian ini akan digunakan metode yang sesuai dengan permasalahan di atas dengan menggunakan metode Naive Bayes. Naive Bayes adalah sebuah metode klasifikasi yang berakar pada teorema Bayes. Dalam metode pengklasifikasian dengan digunakannya metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes [6].

Keuntungan penggunaan Naive Bayes menurut Effrida Manalu dan Fricles Ariwisanto dalam penelitiannya adalah bahwa metode ini hanya membutuhkan data training yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian dan juga dapat bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan [7].

Berdasarkan permasalahan tersebut, akan dimanfaatkan data mining dengan metode Naive Bayes, diharapkan dapat membantu menemukan informasi dalam

memprediksi kelulusan mahasiswa khususnya pada program stidu matematika Universitas Islam Negeri Syarif Hidayatullah Jakarta sehingga membantu bagian program studi untuk memprediksi status kelulusan mahasiswa dan dapat digunakan sebagai sarana untuk menentukan langkah dan kebijakan bagi mahasiswa untuk yang menargetkan lulus tepat waktu.

## ***B. Manfaat dan Tujuan Penelitian***

Tujuan dari penelitian ini adalah sebagai berikut :

1. Untuk mengetahui penerapan teknik data mining dengan Naive Bayes dan menyajikan informasi dalam memprediksi kelulusan dengan berdasarkan data training menggunakan data alumni mahasiswa matematika.
2. Untuk mengetahui prediksi kelulusan mahasiswa matematika Universitas Islam Negeri Syarif Hidayatullah Jakarta dengan menggunakan hasil metode Naive Bayes.
3. Mendapatkan akurasi yang tepat untuk melakukan klasifikasi kelulusan matematika UIN Syarif Hidayatullah Jakarta menggunakan metode Naive Bayes.

Dengan tercapainya tujuan penelitian di atas, maka diharapkan akan bermanfaat bagi semua pihak, manfaat dari penelitian ini adalah :

1. Memberikan informasi prediksi kelulusan mahasiswa yang lulus tepat waktu maupun yang tidak tepat waktu.
2. Memberikan informasi sebagai dasar pertimbangan pengambilan keputusan dalam melakukan evaluasi tingkat kelulusan mahasiswa.
3. Menjadikan sebagai perbandingan bagi peneliti lain dalam menerapkan teknik-teknik data mining di area pendidikan.

## **BAB II. TINJAUAN PUSTAKA DAN LANDASAN TEORI**

### **A. *Data Mining***

Data mining didefinisikan sebagai sebuah proses untuk menemukan hubungan, pola dan tren baru yang bermakna dengan menyaring data yang sangat besar yang tersimpan dalam penyimpanan, menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data [8]. Istilah data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Data mining dapat disebut sebagai Knowledge Discovery in Database (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, histori untuk menemukan keteraturan, pola atau hubungan dalam set data yang berukuran besar [9].

Data mining bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan data mining adalah kenyataan bahwa data mining mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu. Berawal dari beberapa disiplin ilmu, data mining bertujuan untuk memperbaiki teknik tradisional sehingga bisa menangani:

- a. Jumlah data yang sangat besar
- b. Dimensi data yang tinggi
- c. Data yang heterogen dan berbeda sifat

Menurut para ahli, data mining merupakan sebuah analisa dari observasi data dalam jumlah besar untuk menemukan hubungan yang tidak diketahui sebelumnya dan metode baru untuk meringkas data agar mudah dipahami serta kegunaannya untuk pemilih data [10]. Salah satu teknik yang dibuat dalam data mining adalah bagaimana menelusuri data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam basis

data yang tersimpan. Kebutuhan untuk prediksi juga dapat memanfaatkan teknik ini. Data mining dapat pengelompokan data juga bisa dilakukan. Tujuannya adalah agar kita dapat mengetahui pola universal data-data yang ada. Anomali data transaksi juga perlu dideteksi untuk dapat mengetahui tindak lanjut berikutnya yang dapat diambil. Semua hal tersebut bertujuan mendukung kegiatan akhir perusahaan diharapkan dapat tercapai [11].

## ***B. Pengelompokan Data Mining***

### **a. Deskripsi**

Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pola kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

### **b. Estimasi**

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik dari pada ke arah kategori. Model dibangun menggunakan baris data (record) lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

### **c. Prediksi**

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

### **d. Klasifikasi**

Klasifikasi terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

### **e. Pengklasteran (Clustering)**



Pengklasteran merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas obyek-obyek yang memiliki kemiripan. Klaster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan record dalam klaster yang lain.

Berbeda dengan klasifikasi, pada pengklasteran tidak melakukan klasifikasi mengestimasi, atau memprediksi nilai dari variabel target, akan tetapi, algoritma pengklasteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan record dalam suatu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal.

f. Asosiasi

Tugas asosiasi dalam data mining adalah untuk menemukan atribut yang muncul dalam satu waktu.

### **C. Tahap-tahap Data Mining**

a. Pembersihan data (*data cleaning*)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Proses pembersihan data juga mencakup antara lain membuang duplikasi data dan memperbaiki kesalahan pada data, seperti kesalahan cetak (typografi).

b. Integrasi data

Data yang digunakan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lain-lainya.

c. Seleksi data (*data selection*)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database.

Pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai.

d. Transformasi data (*data transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining.

e. Proses mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Proses mining dapat menemukan hubungan, pola dan tren baru yang bermakna dengan menyaring data yang sangat besar yang tersimpan dalam penyimpanan.

f. Evaluasi pola (*pattern evaluation*)

Untuk mengidentifikasi pola-pola menarik ke dalam knowledge based yang ditemukan.

## **D. Klasifikasi**

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau class data, dengan tujuan dapat memperkirakan class dari suatu objek yang labelnya tidak diketahui. Metode metode klasifikasi antara lain adalah *Bayesian*, *Neural Network*, *Genetic Algorithm*, *Fuzzy*, *Case Based Reasoning* dan *K-Nearest Neighbor* [6].

Klasifikasi data terdiri dari dua langkah proses yang akan digunakan, yaitu proses learning (fase training) dimana algoritma klasifikasi dibuat untuk menganalisa data training lalu direpresentasikan dalam bentuk rule klasifikasi, proses kedua adalah klasifikasi dimana data tes digunakan untuk memperkirakan akurasi dari rule klasifikasi.

## E. Naive Bayes

*Naive Bayes* merupakan sebuah pengklasifikasian probabalistik yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dataset yang diberikan. Algoritma menggunakan teorema bayes dan mengasumsikan semua atribut *independen* atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas.

Definisi lain mengatakan *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Naive Bayes didasarkan pada asumsi penyerhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu.

Keuntungan penggunaan Naive Bayes menurut Effrida Manalu dan Fricles Ariwisanto dalam penelitiannya adalah bahwa metode ini hanya membutuhkan data training yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian dan juga dapat bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan.

Metode Naive Bayes menggunakan prinsip teorema bayes, yaitu menghitung probabilitas suatu kejadian berdasarkan suatu kondisi tertentu, dengan menggunakan persamaan :

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

dengan :

$X$  = Data dengan kelas yang belum diketahui;

$H$  = Hipotesis data  $X$  merupakan suatu label kelas tertentu;

$P(H|X)$  = Probabilistik hipotesis  $H$  berdasarkan kondisi  $X$  (posteriori probability);

$P(H)$ : Probabilistik hipotesis  $H$  (prior probability);

$P(X|H)$  = Probabilitas  $X$  berdasarkan kondisi pada hipotesis  $H$ ;

$P(X)$  = Probabilitas  $X$

### **BAB III. METODOLOGI**

#### **A. Sumber dan tipe data**

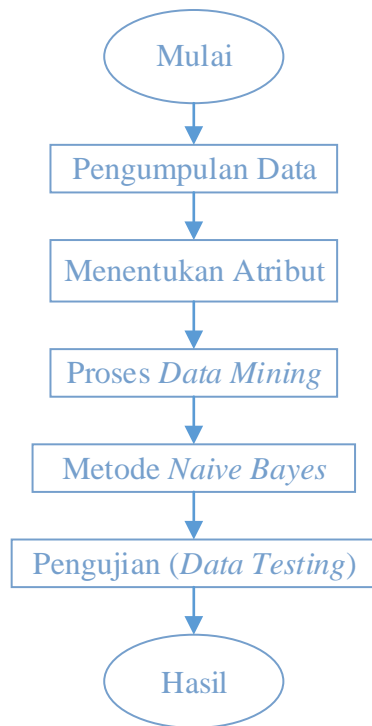
Data yang digunakan dalam penelitian ini adalah data primer dan data sekunder. Data primer diperoleh dengan cara menyebarkan kuesioner kepada alumni mahasiswa matematika UIN Syarif Hidayatullah Jakarta, sedangkan data sekunder diperoleh melalui Pusat Layanan Informasi Publik – PPID UIN Syarif Hidayatullah Jakarta.

#### **B. Usulan Metode**

Tahap-tahap yang dilakukan dalam melakukan penelitian ini secara umum adalah :

1. Mengumpulkan Data
2. Menentukan Atribut
3. Melakukan Proses *Data Mining*
4. Pemilihan *Data Training*
5. Mencari kelayakan metode *Naive Bayes* dengan menghitung probabilitas kelas
6. Menghitung probabilitas masing-masing atribut
7. Menghitung probabilitas akhir untuk setiap kelas
8. Melakukan *Data Testing*

### C. *Diagram alir penelitian*



## BAB IV. JADWAL PENELITIAN

Penelitian direncanakan akan dilaksanakan selama enam bulan. Rincian rencana jadwal penelitian dicantumkan dalam Tabel 1.

**Tabel 1. Jadwal penelitian.**

No.	Keterangan	Bulan					
		1	2	3	4	5	6
1.	Telaah literatur						
2.	Penyusunan Proposal Penelitian						
3.	Pengajuan Proposal Penelitian						
4.	Bimbingan dengan dosen pembimbing						
5.	Pengambilan data						
6.	Kompilasi dan analisis data						
7.	Evaluasi dan perbaikan						
8.	Penulisan Laporan						
9.	Revisi Akhir						

## DAFTAR PUSTAKA

- [1] Mujib dan Mardiyah, “Kemampuan Berpikir Kritis Matematis Berdasarkan Kecerdasan Multiple Intelligences,” *Al-Jabar : Jurnal Pendidikan Matematika* 8, No. 2 (25 Desember, 2017):187–96.
- [2] Pandri Ferdias dan Eka Anis Savitri, “Analisis Materi Volume Benda Putar Pada Aplikasi Cara Kerja Piston Di Mesin Kendaraan Roda Dua,” *Al-Jabar : Jurnal Pendidikan Matematika* 6, No.2 (18 Desember 2 (18 Desember, 2015):177–82.
- [3] Ace Suryadi, *Pendidikan Indonesia Menuju 2025* (Bandung: PT. Remaja Rosdakarya, 2014):54.
- [4] Budi Santoso, “Data Mining Terapan dengan MATLAB,” (Yogyakarta: Graha Ilmu, 2007):23.
- [5] Mahmud Yunus, “Penerapan Konsep Data Mining pada Data Base Akademik STMIK Pradnya Paramita Dengan Delphi,” *Jurnal Dinamika*, No. 1 (Januari, 2010):35-46.
- [6] Bustami, “Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi,” *Jurnal Informatika* 8, No. 1 (1 Januari 2014):1-15.
- [7] Effrida Manalu, Fricies Ariwisanto dan Mamed Rofendy “Penerapan Algoritma Naive Bayes Untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan Dan Jumlah Pesanan Pada CV. Papadan Mama Pastries,” *Jurnal Manajemen Dan Informatika Pelita Nusantara* 1, no. 2 (Desember 2017): 16-21.
- [8] Daniel T,. Larose, “Discovering Knowledge in Data: An Introduction to Data Mining” (United States of America, 2005):2.
- [9] Budi Santosa, “Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis” (Graha Ilmu, 2007):15.
- [10] David Hartanto Kamagi dan Seng Hansun, “Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa,” *Jurnal ULTIMATICS*, No.1 (Juni, 2014):16.
- [11] Siska Haryati, Aji Sudarsono dan “Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dahasen Bengkulu),” *Jurnal Media Infotama* 11, No.2 (September, 2015):130-138.