**PREDIKSI KELULUSAN TEPAT WAKTU MENGGUNAKAN ALGORITMA NAÏVE BAYES BERBASIS WEBSITE**

**PROPOSAL TUGAS AKHIR**



Oleh

**Wahyu Mutiara Rahmadani**

**NIM E31201697**

**PROGRAM STUDI MANAJEMEN INFORMATIKA**

**JURUSAN TEKNOLOGI INFORMASI**

**POLITEKNIK NEGERI JEMBER**

**2022**

**PREDIKSI KELULUSAN TEPAT WAKTU MENGGUNAKAN ALGORITMA NAÏVE BAYES BERBASIS WEBSITE**

# **PROPOSAL TUGAS AKHIR**



sebagai salah satu syarat untuk memperoleh gelar Ahli Madya(A.Md.)

di Program Studi Manajemen Informatika

Jurusan Teknologi Informasi

Oleh

**Wahyu Mutiara Rahmadani**

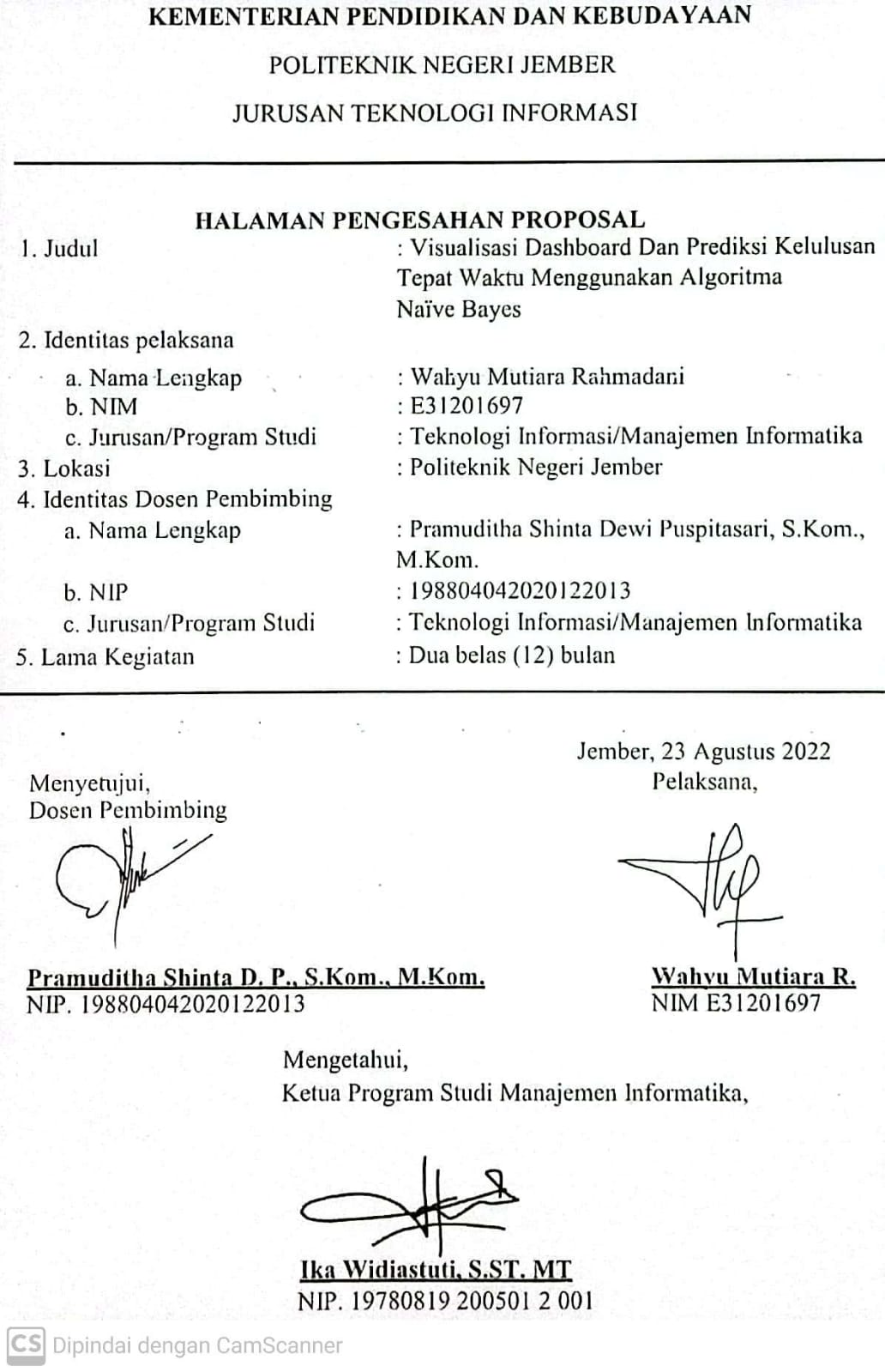
**NIM E31201697**

**PROGRAM STUDI MANAJEMEN INFORMATIKA**

**JURUSAN TEKNOLOGI INFORMASI**

**POLITEKNIK NEGERI JEMBER**

**2022**



**Prediksi Kelulusan Tepat Waktu menggunakan Algoritma Naïve Bayes Berbasis Website**

**Wahyu Mutiara Rahmadani**

Program Studi Manajemen Informatika

Jurusan Teknologi Informasi

# ***ABSTRAK***

*Politeknik Negeri Jember memiliki kendala pada kelulusan mahasiswa. Hal ini ditandai dengan adanya evaluasi tiap tahun. dibutuhkan suatu metode dalam mengatasi permasalahan tersebut. Salah satu teknik yang digunakan adalah data mining menggunakan algoritma naive bayes. Algoritma ini digunakan untuk memprediksi tingkat kelulusan mahasiswa. kemudian dibuat template excel sebagai inputan dalam memvisualisasikan data pada dashboard Power BI. Metode yang digunakan yaitu CRISP-DM dalam mengembangkan data minin yang terdapat enam tahapan, diantaranya, bussiness understanding berisi pemahaman permasalahan yang dihadapi Politeknik Negeri Jember, data understanding yaitu pengumpulan dan pengambilan data, data preparation yaitu tahap mempersiapkan data agar mudah digunakan, modelling yaitu pemodelan naive bayes, evaluation yaitu pengujian model, deployment yaitu pembuatan dashboard analisa. Diharapkan dengan penerapan metode tersebut dapat memberikan manfaat dalam memprediksi kelulusan mahasiswa dan sebagai alat bantu yang mempermudah evaluasi kualitas mahasiswa.*

***Kata kunci:*** *Kelulusan Mahasiswa, Naïve Bayes, visual*

# **DAFTAR ISI**

halaman

[HALAMAN JUDUL ii](#_Toc112010315)

[HALAMAN PENGESAHAN iii](#_Toc112010316)

[*ABSTRAK* iv](#_Toc112010317)

[DAFTAR ISI v](#_Toc112010318)

[DAFTAR GAMBAR vii](#_Toc112010319)

[DAFTAR TABEL viii](#_Toc112010320)

[BAB 1. PENDAHULUAN 1](#_Toc112010321)

[1.1 Latar Belakang 1](#_Toc112010322)

[1.2 Rumusan Masalah 2](#_Toc112010323)

[1.3 Batasan Masalah 3](#_Toc112010324)

[1.4 Tujuan 3](#_Toc112010325)

[1.5 Manfaat 4](#_Toc112010326)

[BAB 2. TINJAUAN PUSTAKA 5](#_Toc112010327)

[2.1 Karya Penelitian Terdahulu 5](#_Toc112010328)

[2.2 State Of The Art 8](#_Toc112010329)

[2.3 Landasan Teori 10](#_Toc112010330)

[2.3.1 Prediksi 10](#_Toc112010331)

[2.3.2 Mahasiswa 10](#_Toc112010332)

[2.3.3 Data Mining 10](#_Toc112010333)

[2.3.4 Naïve Bayes 10](#_Toc112010334)

[2.3.5 KPI 11](#_Toc112010335)

[2.4 Kajian Teknologi 12](#_Toc112010336)

[2.4.1 Software 12](#_Toc112010337)

[2.4.2 Rapidminer 12](#_Toc112010338)

[2.4.3 Power BI 13](#_Toc112010339)

[2.4.4 Dashboard 13](#_Toc112010340)

[BAB 3. METODE KEGIATAN 15](#_Toc112010341)

[3.1 Tempat dan Waktu Pelaksanaan 15](#_Toc112010342)

[3.2 Alat dan Bahan 15](#_Toc112010343)

[3.3 Metode Kegiatan 16](#_Toc112010344)

[3.4 Jadwal Kegiatan 21](#_Toc112010345)

[DAFTAR PUSTAKA 23](#_Toc112010346)

# **DAFTAR GAMBAR**

halaman

[Gambar 3.1 Metode CRISP-DM 16](#_Toc112010089)

[Gambar 3. 2 Flowchart Perhitungan Nive Bayes 20](#_Toc112010090)

[Gambar 3. 3 Pemodelan Naïve Bayes pada RapidMiner 21](#_Toc112010091)

# **DAFTAR TABEL**

halaman

[2.1 State Of The Art 8](#_Toc110932269)

[3.1 Atribut Dan Kategori 18](#_Toc110932257)

[3.2 Jadwal Kegiatan 22](#_Toc110932258)

# **PENDAHULUAN**

## Latar Belakang

Perguruan tinggi merupakan lanjutan dari Pendidikan menengah. Perguruan tinggi bertujuan untuk mempersiapkan peserta didik agar dapat menjadi salah satu masyarakat yang mempunyai kemampuan profesional. Dalam perguruan tinggi juga terdapat tahap akhir untuk persiapan menempuh dunia kerja. Tahap tersebut yaitu kelulusan mahasiswa yang ditandai dengan berakhirnya pembelajaran di perguruan tinggi.

Kelulusan merupakan sebuah momen terpenting dalam menempuh suatu pembelajaran. Jika telah menyelesaikan pembelajaran dari suatu jenjang, maka dibuktikan dengan kelulusan. Hal ini tentu juga dialami oleh mahasiswa tingkat akhir. Setelah menyelesaikan berbagai tahapan hingga menuju sidang skripsi, mahasiswa tingkat akhir melakukan wisuda. Wisuda merupakan kegiatan yang menyatakan bukti kelulusan bahwa mahasiswa telah menyelesaikan studi di perguruan tinggi.

Persyaratan kelulusan tiap jenjang memiliki perbedaan, diantaranya yaitu waktu kuliah, jumlah semester, dan jumlah SKS. Lama masa kuliah dalam tiap jenjang berbeda-beda. Jika D3, maka harus menempuh 6 semester dan sekurang- kurangnya 110 SKS dan sebanyak-banyaknya 120 SKS. Namun, untuk S1, menempuh 8 semester dan sekurang-kurangnya telah menempuh 144 SKS dan sebanyak-banyaknya 160 SKS (Keputusan Menteri Pendidikan Nasional No.232/U/2000 Pasal 5 dan 6). Selain itu, terdapat unsur penilaian dari perguruan tinggi yang berhubungan langsung dengan mahasiswa. Unsur tersebut yaitu nilai ujian masuk, nilai akademik, prestasi yang dicapai, kompetensi yang dimiliki mahasiswa, dan persentase mahasiswa lulus tepat waktu.

Jurusan Teknologi Informasi di Politeknik Negeri Jember memiliki kendala yang berhubungan dengan ketepatan waktu kelulusan mahasiswa. Hal ini ditandai dengan pengadaan evaluasi yang dilakukan jurusan teknologi Informasi terhadap kualitas kelulusan mahasiswa tiap tahunnya. Hingga kini, masih belum ada alat bantu yang mampu memprediksi kelulusan mahasiswa. Akibatnya, pihak kampus cukup kesulitan dalam mengawasi tingkat kelulusan mahasiswanya. Jumlah mahasiswa yang lulus tepat waktu maupun tidak tepat waktu masih belum dapat diprediksi. Hal ini juga dapat mempengaruhi kualitas perguruan tinggi.

Dari beberapa permasalahan yang telah dijelaskan, guna meningkatkan kualitas mahasiswa Jurusan Teknologi Informasi di Politeknik Negeri Jember, dibutuhkan suatu metode yang perlu diimplementasikan dalam menyelesaikan masalah ketepatan kelulusan mahasiswa. Salah satu teknik yang dapat digunakan untuk memprediksi kelulusan mahasiswa tepat waktu adalah teknik data mining. Data mining telah berkembang sehingga dapat digunakan dalam menyelesaikan masalah di perguruan tinggi.

Salah satu algoritma data mining yang digunakan dalam penelitian ini adalah Naïve Bayes. Naïve Bayes merupakan kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi objek. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari “master” tabel keputusan.

## Rumusan Masalah

Berdasarkan latar belakang masalah di atas, peneliti merumuskan permasalahan sebagai berikut:

1. Bagaimana gambaran faktor-faktor pendukung kelulusan tepat waktu mahasiswa Politeknik Negeri Jember?
2. Bagaimana klasifikasi sistem untuk memprediksi mahasiswa lulus tepat waktu menggunakan metode Naïve Bayes?
3. Bagaimana menyajikan informasi mengenai ketepatan kelulusan mahasiswa Politeknik Negeri Jember ke dalam Dashboard?
4. Informasi apa saja yang harus ditampilkan dalam dashboard untuk membantu pihak akademisi JTI Polije?

## Batasan Masalah

Adapun batasan masalah penelitian yang diberikan agar penelitian ini tidak keluar dari pokok permasalahan yang dirumuskan, maka ruang lingkup pembahasan dibatasi sebagai berikut:

1. Data yang digunakan adalah data kelulusan Mahasiswa Prodi Manajemen Informatika Jurusan Teknologi Informasi satu angkatan. Data diperoleh dari database SIM Polije, Politeknik Negeri Jember.
2. Data yang dibutuhkan yaitu data nilai IP mahasiswa dan nilai mata kuliah Prodi Manajemen Informatika Jurusan Teknologi Informasi satu Angkatan selama 4 semester.
3. Atribut yang digunakan dalam penelitian ini diantaranya Nama, NIM, Jenis kelamin, Usia, IP semester 1, IP semester 2, IP semester 3, IP semester 4, dan Alamat tinggal.
4. Pembuatan model peramalan menggunakan metode Naïve Bayes mereferensi pembuatan model pada penelitian (jurnal yg menjadi referensi).

## Tujuan

Adapun tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut:

1. Menjelaskan gambaran faktor-faktor pendukung ketepatan waktu lulus mahasiswa Prodi Manajemen Informatika Jurusan Teknologi Informasi Politeknik Negeri Jember.
2. Membuat sistem yang dapat memprediksi kelulusan tepat waktu mahasiswa Prodi Manajemen Informatika Jurusan Teknologi Informasi Politeknik Negeri Jember menggunakan metode Naïve Bayes untuk memudahkan analisa sebagai dasar pengambilan keputusan.
3. Membuat sebuah dashboard Analisa data yang menampilkan informasi terkait kelulusan mahasiswa guna membantu pihak akademisi JTI Polije.

## Manfaat

Adapun manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Tugas akhir ini diharapkan dapat memberikan manfaat dalam memprediksi kelulusan mahasiswa Prodi Manajemen Informatika di Jurusan Teknologi Informasi, Politeknik Negeri Jember.
2. Analisa data berbasis website yang dihasilkan dari penelitian ini diharapkan dapat membantu pihak manajemen akademisi Jurusan Teknologi Informasi, Politeknik Negeri Jember dalam mengevaluasi kualitas mahasiswa dalam pengambilan keputusan.

# **TINJAUAN PUSTAKA**

## Karya Penelitian Terdahulu

Penelitian terdahulu merupakan sebuah penelitian yang dilakukan oleh orang lain pada program atau kegagalan aplikasi serupa seperti kali ini yang dibuat oleh peneliti. Berikut beberapa contoh karya tulisan sebelumnya.

1. Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes : Studi Kasus UIN Syarif Hidayatulloh Jakarta

Proses belajar mengajar sampai proses wisuda di Universitas Islam Negeri (UIN) Syarif Hidayatullah Jakarta telah didukung oleh sistem informasi yang Bernama Academic Information System (AIS). AIS mampu memberikan data dan informasi yang dibutuhkan terkait kegiatan akademik mahasiswa. Melalui sistem ini ditemukan bahwa jumlah mahasiswa baru lebih banyak dibandingkan dengan jumlah mahasiswa yang lulus. Namun, AIS belum dapat menghasilkan pengetahuan (knowledge) yang lebih mendalam terkait kondisi ini. Salah satu cara untuk mencapai kualitas mutu tertinggi dari sistem perguruan tinggi adalah dengan menggali pengetahuan dari data Pendidikan. Sehingga diperlukan adanya pemanfaatan Teknik data mining untuk menggali data pengetahuan dari data mahasiswa yang tersedia di AIS.

Penelitian ini bertujuan untuk menemukan faktor-faktor yang mempengaruhi prediksi kelulusan mahasiswa melalui data kinerja akademik mahasiswa pada semester satu sampai dengan semester empat. Model yang digunakan yaitu Cross Industry Standart Proccess for Data Mining (CRISP-DM) dengan mengimplementasikan algoritma Naïve Bayes untuk klasifikasi data. Melalui tahap data understanding didapatkan 9 atribut predictor yang akan dianalisis terhadap 1 atribut class kelulusan mahasiswa. Dengan perolehan hasilakurasi sebesar 80,72%, dapat disimpulkan bahwa model ini dapat dijadikan acuan dalam meprediksi kelulusan mahasiswa. Pengetahuan ini dapat dimanfaatkan oleh pikah UIN sebagai Langkah preventif untuk menghindari penurunan kelulusan mahasiswa setiap tahunnya.

1. Prediksi Ketepatan Waktu Lulus Mahasiswa Dengan K-Nearest Neighbor Dan Naïve Bayes Classifier (Studi Kasus Prodi D3 Sistem Informasi Universitas Airlangga)

Salah satu aspek pengukuran kualitas dalam menilai keberhasilan universitas adalah keakuratan gelar. Persentase mahasiswa yang menyelesaikan studi tepat waktu merupakan salah satu indikator keberhasilan proses belajar mengajar di program sarjana. Penelitian ini menggunakan teknik data mining untuk memprediksi waktu kelulusan mahasiswa dengan menggunakan dua teknik yaitu k Nearest Neighbor dan Naïve Bayes Classifier. Hasil dari penelitian ini adalah sebuah sistem yang dapat memprediksi ketepatan waktu penyelesaian.

Survei ini dilakukan dengan menggunakan data dari lulusan sistem informasi D3 Universitas Airlangga. Hasil pengujian menunjukkan bahwa metode kNearestNeighbor memberikan akurasi yang lebih baik daripada classifier naive Bayes. Akurasi tertinggi diperoleh dengan menggunakan metode k Nearest Neighbor yaitu 98,7%. Oleh karena itu, dapat disimpulkan bahwa sistem yang dikembangkan dalam penelitian ini dapat memprediksi ketepatan waktu penyelesaian dengan akurasi yang cukup tinggi.

1. Perancangan Dan Pembuatan Visualisasi Data Dana Penelitian Internal Dan Hibah Dikti LPPM Universitas Multimedia Nusantara

Universitas Multimedia Nusantara adalah salah satunya perguruan tinggi swasta yang memiliki divisi Institut untuk Penelitian dan Pengabdian kepada Masyarakat atau disebut LPPM. Bagian LPPM bertugas mengurus perizinan dilakukan di lingkungan universitas, baik dilakukan oleh dosen dan mahasiswa. Untuk dapat mengeluarkan perizinan, terutama yang terkait dengan dana, LPPM harus lihat track record yang ada sehingga izin yang dikeluarkan tidak tersedia dana dan dana yang akan dianggarkan akan cukup untuk kebutuhan dana penelitian menjadi dilakukan pada periode berikutnya. Oleh karena itu, peran Analis Bisnis di divisi LPPM diwajibkan untuk memvisualisasikan data yang digunakan, yaitu penelitian, melakukan berbagai studi yang dilakukan oleh program studi dalam

jangka waktu tertentu periode, dana lengkap, dan detail lainnya seperti penelitian judul, kategori penelitian, dan nama atau melakukan penelitian yang lain. Riset. Visualisasi data akan memudahkan LPPM dalam memantau penelitian yang ada dan menjadi alat yang berguna dalam keputusan perizinan penelitian.

1. Sistem Pendukung Keputusan Pemilihan Jurusan Kuliah menggunakan Metode Analytical Hierarchi Process dan Naïve Bayes

Pemilihan jurusan kuliah sangatlah krusial sehingga perlu dipikirkan kembali dengan baik. Jurusan kuliah menentukan bagaimana masa depan setiap siswa. Sehingga setiap siswa dianjurkan untuk memilih jurusan kuliah dengan teliti berdasarkan potensi akademik, minat, dan bakat diri masing-masing. Menurut kuesioner yang telah disebarkan dengan total 59 responden. Dari 40 responden menjawab mengalami kesulitan dalam menempuh jurusan kuliah yang diambil karena tidak mengetahui minat dan bakatnya dan sekolah tidak dapat melakukan ujian tes minat dan bakat karena masalah biaya yang besar.

Maka dari itu, untuk mepermudah siswa dapat mengetahui rekomendasi jurusan kuliah berdasarkan potensi akademik, minat, dan bakatnya dirancang sistem pendukung keputusan berbasis website untuk mempermudah kinerja bimbingan konseling dalam jurusan kuliah siswanya, dan juga mengurangi faktor subjektivitas dalam menentukan jurusan kuliah.

1. Perbandingan Klasifikasi Ketepatan Waktu Kelulusan Mahasiswa Menggunakan Regresi Logistik Biner Dan Naïve Bayes Classifier

Kualitas suatu pendidikan dapat dilihat melalui pencapaian pembelajaran seseorang. Salah satunya yaitu dari faktor kelulusan mahasiswa dalam menjalankan Pendidikan. Kelulusan merupakan hal terpenting dalam menentukan suatu pencaiapian seseorang atas pendidikannnya yang umumnya ditempuh dalam waktu 8 semester atau 4 tahun. Banyak faktor yang berpengaruh terhadap mahasiswa dalam menempuh pendidikannnya. Penelitian ini ditujukan untuk mengetahui faktor-faktor yang mempengaruhi lama studi mahasiswa Universitas Islam Indonesia secara signifikan. Penelitian ini juga melakukan pengklasifikasian dalam menentukan kelulusan tepat waktu mahasiswa dengan menggunakan Regresi Logistik Biiner dan juga melakukan perbandingan pengklasifikasian menggunakan algoritma Naïve Bayes Classifier dikarenakan data yang digunakan cukup banyak.

Berdasarkan analisis Regresi Logistik Biner terdapat faktor yang sangat berpengaruh terhadap lama studi mahasiswa secara signifikan diantaranya variable Jenis Kelamin, Asal daerah, Jenis SMA, Jurusan SMA, IPK Mahasiswa. Klasifikasi Logistik Biner menghasilkan akurasi sebesar 82,29%, sedangkan hasil klasifikasi dengan Algoritma Naïve Bayes Classifier sebesar 89,10%. Oleh karena itu, dapat dikatakan bahwa pengklasifikasian lama studi mahasiswa menggunakan Algoritma Naïve Bayes Classifier levih akurat disbanding dengan Regresi Logistik Biner.

## State Of The Art

Tabel 2.1 State Of The Art

| Tahun | Judul | Nama | Keterangan |
| --- | --- | --- | --- |
| 2017 | Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes : Studi Kasus UIN Syarif Hidayatulloh Jakarta | Supardi Salmu, Achmad Solichin | RapidMiner Studio |
| 2017 | Prediksi Ketepatan Waktu Lulus Mahasiswa Dengan K-Nearest Neighbor Dan Naïve Bayes Classifier (Studi Kasus Prodi D3 Sistem Informasi Universitas Airlangga) | Wilda Imama Sabila, Tesa Eranti Putri | - |
| 2018 | Perbandingan Klasifikasi Ketepatan Waktu Kelulusan Mahasiswa Menggunakan Regresi Logistic Biner Dan Naïve Bayes Classifier | Erene Fajrila | - |
| 2020 | Penerapan Metode CRISP-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik | Yogasetya Suhanda, Ike Kurniati, Siti Norma | SPSS Modeler dan Tableau |
| 2021 | Sistem Pendukung Keputusan Pemilihan Jurusan Kuliah menggunakan Metode Analytical Hierarchi Process dan Naïve Bayes | Maulana Malik Ibrahim | Sistem pendukung keputusan untuk membantu merekomendasikan jurusan kuliah untuk siswa kelas 12 SMA dengan menggunakan parameternilai akademik, minat, dan bakat yang kemudian diterapkan dengan metode AHP untuk mengetahui besar presentase rekomendasi jurusan kuliah dan menerapkan metode naïve bayes untuk mengetahui probabilitas status diterima atau tidak siswa pada jurusan kuliah. |
| 2022 | Visualisasi Dashboard Dan Prediksi Kelulusan Tepat Waktu Menggunakan Algoritma Naïve Bayes | Wahyu Mutiara Rahmadani | Power BI Desktop |

## Landasan Teori

### Prediksi

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi (Kafil, 2019).

### Mahasiswa

Mahasiswa merupakan masa memasuki mahasiswa dapat didefinisikan sebagai individu yang sedang menuntut ilmu ditingkat perguruan tinggi, baik negeri maupun swasta atau Lembaga lain yang setingkat dengan perguruan tinggi. Mahasiswa dinilai memiliki masa dewasa yang pada umum berada pada rentang usia 18-25 tahun, pada masa tersebut mahasiswa memiliki tanggung jawab terhadap masa perkembangannya, termasuk memiliki tanggung jawab terhadap kehidupannya untuk memasuki masa dewasa (Hulukati & Djibran, 2018).

### Data Mining

Sebagai bidang ilmu yang relative baru, saat ini Data Mining menjadi salah satu pusat perhatian para akademis maupun praktisi. Menurut (Suntoro, 2019) data mining adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan.

### Naïve Bayes

Naïve Bayes adalah salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk machine learning dan data mining. Performa naïve bayes yang kompetitif dalam proses klasifikasi walaupun menggunakan asumsi keidependenan atribut (tidak ada kaitan antar atribut). Asumsi keidependenan atribut ini pada data sebenarnya jarang terjadi, namun walaupun asumsi keidependenan atribut tersebut dilanggar performa pengklasifikasian naïve bayes cukup tinggi, hal ini dibuktikan pada berbagai penelitian empiris (Syarli, 2016).

Algoritma Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema Bayes. Teorema tersebut dikombinasikan dengan Naive dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi Naive Bayes diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya.

Berikut ini adalah rumus persamaan dari metode Naïve Bayes yaitu:

𝑃(𝐻|𝑋) =𝑃(𝑋|𝐻) ∙ 𝑃(𝐻) / 𝑃(𝑋)

Keterangan :

* X : Data dengan kelas yang belum diketahui.
* H : Hipotesis data merupakan suatu kelas spesifik.
* P(H│X): Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas).
* P(H): Probabilitas hipotesis H (prior probabilitas).
* P(X│H): Probabilitas X berdasarkan kondisi pada hipotesis H
* P(X): Probabilitas X

### KPI

Key Performance Indicator (KPI) disini adalah sebuah pengukuran untuk mengukur kinerja perusahaan yang mudah di visualisasikan, dimengerti, tidak bersifat ambigu, dan tidak bertentangan dengan KPI lain dalam memenuhi tujuan yang ingin dicapai (Imaroh & Soleh, 2020). Beberapa penelitian terdahulu antara lain dari Muh Rohim (2017) menyatakan bahwa KPI menjadi alat mengukur kinerja organisasi dan memastikan keselarasan visi dan strategi organisasi. KPI performance juga bisa dijadikan alat pengukur kinerja yang efektif bagi manajer dalam memberikan penilaian kinerja tahunan bagi semua engineer dan menghindarkan dari penilaian subjective seperti saat ini yang sudah berjalan.. Adapun beberapa hal yang dilakukan untuk mengambangkan KPI diantaranya :

1. Indikator yaitu kondisi yang digunakan untuk membantu Dekan dalam menilai secara akurat tentang aspek yang mendukung penilaian.
2. Nilai target yaitu nilai sasaran yang ingin dicapai sebuah sorganisasi. Nilai ini ditentukan oleh Dekan atau penanggung jawab penetapan standar dalam perusahaan
3. Nilai batas yaitu pembatas yang memisahkan nilai minimum dan nilai maksimum. Setiap organisasi memilki batasannilai yang berbeda. Hal ini dikarenakan tujuan yang ingin dicapai suatu organisasi juga berbeda-beda.

## Kajian Teknologi

### Software

Software adalah sesuatu yang bersifat logis, berbeda daripada elemen sistem yang bersifat fisik. Oleh karena itu, Software memiliki karakteristik yang sangat berbeda daripada hardware. Definisi yang biasa dilakukan dalam teksbook, software mempunyai kesatuan dari 3 hal dibawah ini :

1) Software adalah suatu instruksi (program komputer) yang bila di eksekusi atau dijalankan akan membentuk feature, fungsi ataupun kinerja tertentu.

2) Software didalamnya ada struktur data yang memungkinkan program untuk memanipulasi(melakukan perubahan yang diinginkan) terhadap informasi secara memadai atau terukur, dan

3) Didalam software ada dokumen yang menggambarkan operasi dan penggunaan program (Rochmadi, 2013).

### Rapidminer

Rapid Miner merupakan perangkat lunak yang dibuat oleh Dr. Markus

Hofmann dari Institute of Technology Blanchardstown dan Raif Klinkenberg dari rapid- i.com dengan tampilan GUI (Graphical User Interface) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini. Perangkat lunak ini bersifat open source dan dibuat dengan menggunakan bahasa java dibawah lisensi GNU Public License dan Rapid Miner dapat dijalankan disistem operasi manapun. Dengan menggunakan Rapid Miner, tidak dibutuhkan kemampuan koding khusus, karena semua fasilitas sudah disediakan. Rapid Miner dikhususkan untuk penggunaan data mining (Hendrian, 2018).

Pada software ini terdapat banyak metode yang dapat digunakan untuk mempermudah dalam perhitungan algoritma. Selain itu, RapidMiner juga telah menyediakan berbagai macam contoh data yang telah siap untuk diolah. Data juga dapat diimport dalam bentuk excel dan lainnya. Pada pemrosesan data juga terdapat banyak pilihan cara dalam pengolahannya. Aplikasi ini sangat memudahkan peneliti dalam mengelola data dan mengimplementasikannya dalam sebuah sistem informasi.

### Power BI

Power BI merupakan aplikasi analisis data yang dikembangkan oleh Microsoft yang utamanya digunakan untuk melakukan analisis data untuk bisnis (business intelligence) (Microsoft: 2020). Meskipun demikian, Power BI secara umum dapat digunakan untuk melakukan analisis dan visualisasi data apa pun, tidak terbatas untuk data yang berhubungan dengan bisnis.

Power BI menyediakan antarmuka untuk mengimpor, menyimpan, mengolah, dan memvisualisasikan data dari berbagai sumber. Fleksibilitas ini menjadikan Power BI salah satu aplikasi yang marak digunakan untuk melakukan analisis data. Selain itu, terdapat banyak widget yang dapat digunakan untuk melakukan visualisasi, sehingga visualisasi yang dibuat lebih beragam dan interaktif (Fiqri, 2020).

### Dashboard

Dashboard adalah berupa visualisasi yang berasal dari informasi penting yang digunakan untuk mencapai tujuan tertentu. Dashboard ditampilkan dalam bentuk yang sudah dianalisis yaitu dalam bentuk layar. Informasi yang ditampilkan pada dashboard dapat berupa grafik dengan jenis yang beragam, mulai dari grafik batang, grafik line, maupun grafik donat. Hal ini ditujukan agar pengguna dapat lebih mudah memahami informasi yang ditampilkan.

Berikut adalah jenis-jenis dari visualisasi data yang dapat digunakan dalam dashboard:

1) Bar Chart

Diagram ini berbentuk seperti batang atau bar. Diagram batang biasanya digunakan untuk menampilkan hubungan antar variable yang bertujuan untuk perbandingan.

2) Line Chart

Diagram ini berbentuk line atau garis. Diagram ini juga memiliki fungsi yang hampir sama dengan diagram batang yaitu untuk membandingkan suatu nilai variable.

3) Pie Chart

Grafik ini berbentuk lingkaran atau seperti kue pie yang diiris. Tiap irisannya mewakili preosentase variable secara keseluruhan perbandingan. Total dari perbandingan tersebut harus selalu bernilai 100%.

4) Map Chart

Diagram ini berbentuk peta yang biasanya memuat variable yang dikategorikan berdasarkan wilayah. Tiap wilayah biasanya diwakili oleh warna tertentu yang merupakan tanda perbandingan nilai.

### Website

Website adalah halaman yang berisi data, baik teks maupun data gambar, suara dan lainnya, yang dapat digunakan secara online. Ada banyak pola pengembangan sistem yang dapat digunakan untuk membuat website, salah satunya adalah Prototyping.

### Framework

### Codeigniter

### Blackbox

# **METODE KEGIATAN**

## Tempat dan Waktu Pelaksanaan

Tugas Akhir (TA) yang berjudul “Prediksi Kelulusan Tepat Waktu menggunakan Algoritma Naïve Bayes Berbasis Website” dan dilaksanakan selama dua belas bulan, mulai bulan Februari 2022 sampai dengan Januari 2023 di Politeknik Negeri Jember secara Daring dan Luring.

## Alat dan Bahan

Dalam pembuatan program ini memerlukan alat dan bahan pada proses pembuatan atau pengerjaannya. Berikut alat dan bahan yang dibutuhkan.

3.2.1 Alat

a. Perangkat keras (Hardware)

Perangkat keras yang dibutuhkan dalam membuat Tugas Akhir ini yaitu Laptop yang memiliki spesifikasi sebagai berikut:

1. Laptop : HP 14s-dk1515AU
2. Processor : AMD Ryzen 3 3250U with radeon Graphics
3. Memory (RAM) : 12288MB
4. Hard Drive : 1TB HDD

b. Perangkat lunak (Software)

Software yang dibutuhkan dalam membuat Tugas Akhir ini adalah sebagai berikut:

1. Sistem Operasi : Windows 10 Home Single Language 64-bit
2. Microsoft Office : Microsoft Excel
3. Platform : RapidMiner dan Power BI Desktop

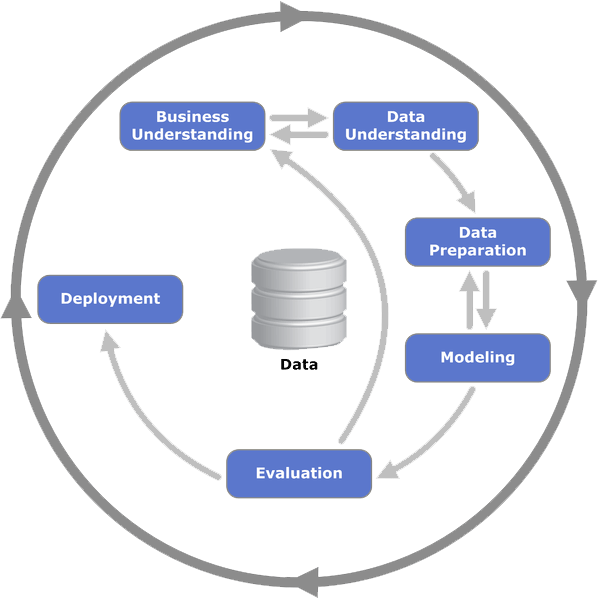
3.2.2 Bahan

Adapun bahan yang dibutuhkan dalam pembuatan Tugas Akhir “Visualisasi Dashboard Dan Prediksi Kelulusan Tepat Waktu Menggunakan Metode Naïve Bayes”, yaitu berupa:

1. Data mahasiswa Prodi Manajemen Informatika Jurusan Teknologi Informasi di Politeknik Negeri Jember satu angkatan.
2. Data mengenai parameter kriteria penilaian mahasiswa agar lulus tepat waktu.
3. Data mengenai KPI yang ditampilkan dalam Dashboard.

## Metode Kegiatan

CRISP-DM merupakan metode yang digunakan untuk pengembangan data mining. Metode ini adalah metode yang diterapkan standar dalam proses data mining yang dibuat oleh Eropa di tahun 1996 (Salmu & Achmad, 2017). Pada metode CRISP-DM memiliki enam tahapan yang akan diimplementasikan pada penelitian ini dengan ilustrasi seperti berikut.

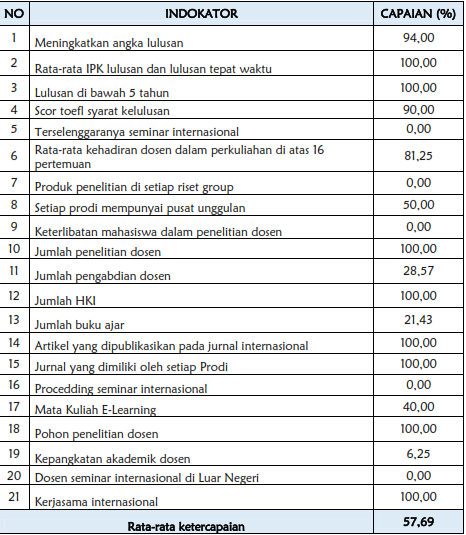


Gambar 3.1 Metode CRISP-DM

1) Business understanding

Permasalahan yang dialami oleh Politeknik Negeri Jember yaitu mempunyai masalah terhadap ketepatan waktu kelulusan terhadap mahasiswa. Untuk itu, peneliti ingin membuat alat bantu dalam memprediksi mahasiswa yang lulus tepat waktu maupun tidak lulus tepat waktu dengan menerapkan metode Naïve Bayes. Pada bagian ini memerlukan pemahaman mengenai pentingnya pemanfaatan data mahasiswa yang dapat digunakan untuk mengetahui ketepatan waktu kelulusan mahasiswa berdasarkan beberapa atribut yang menjadi acuan dalam penilaian. Hal ini ditujukan agar dapat meningkatkan kualitas kelulusan mahasiswa.

Selain itu, terdapat rancangan KPI yang ditampilkan dalam dashboard seperti di bawah ini:



2) Data understanding

Penyelesaian permasalahan terhadap ketepatan waktu kelulusan mahasiswa yaitu dengan pengambilan data dari Sistem Informasi Akademik di Politeknik Negeri Jember (SIM Polije) berupa dokumen berbentuk excel. Data yang dikumpulkan berasal dari profil mahasiswa dan indeks prestasi semester mahasiswa Diploma 3 Manajemen Informatika dengan atribut seperti berikut.

Tabel 3.1 Atribut Dan Kategori

|  |  |
| --- | --- |
| Atribut | Kategori |
| Nama | Text |
| NIM | Text |
| Jenis kelamin | Laki-laki  Perempuan |
| Usia | 18 – 20 tahun  21 – 22 tahun  > 22 tahun |
| IP semester 1 | Numerik |
| IP semester 2 | Numerik |
| IP semester 3 | Numerik |
| IP semester 4 | Numerik |
| Alamat tinggal | Dalam Kabupaten  Luar Kabupaten  Luar Provinsi |
| Status kelulusan | Tepat waktu  Terlambat |

3) Data preparation

Data yang diperoleh dalam penelitian ini berasal dari mahasiswa Diploma 3 Prodi Manajemen Informatika satu angkatan dengan 10 atribut, dimana 9 atribut predictor dan 1 atribut hasil. Dalam mempersiapkan data dilakukan preprocessing data agar kualitas data tetap terjaga. Berikut beberapa tahapan dalam menyiapkan data sebelum dibuat pemodelan data:

1. Proses cleansing data yaitu melakukan pembersihan data dengan menghapus, mengubah, ataupun mengoreksi data yang tidak seusai dengan kebutuhan penelitian. Hal ini ditujukan agar data dapat terhitung lebih akurat.
2. Missing value yaitu dengan mencari nilai yang hilang atau data kosong kemudian diisi dengan nilai nol ataupun NULL. Kolom yang kosong perlu diisi sesuai dengan kebutuhan penulis.

Data yang telah dipersiapkan tersebut, selanjutnya disatukan ke dalam satu file excel yang berisikan data Nama, NIM, Jenis Kelamin, Usia, IP semester 1, IP semester 2, IP semester 3, IP semester 4, Alamat. Kemudian ditambahkan atribut hasil yaitu Status kelulusan. File excel yang telah dibuat dimanfaatkan sebagai acuan sehingga dapat digunakan secara berkelanjutan.

4) Modelling fase

Pada tahap ini, data training diklasifikasi oleh model yang kemudian menghasilkan aturan. Penelitian ini menggunakan algoritma Naïve Bayes dalam pembuatan model untuk mendapatkan insight yang berguna dalam mencapai tujuan penelitian.

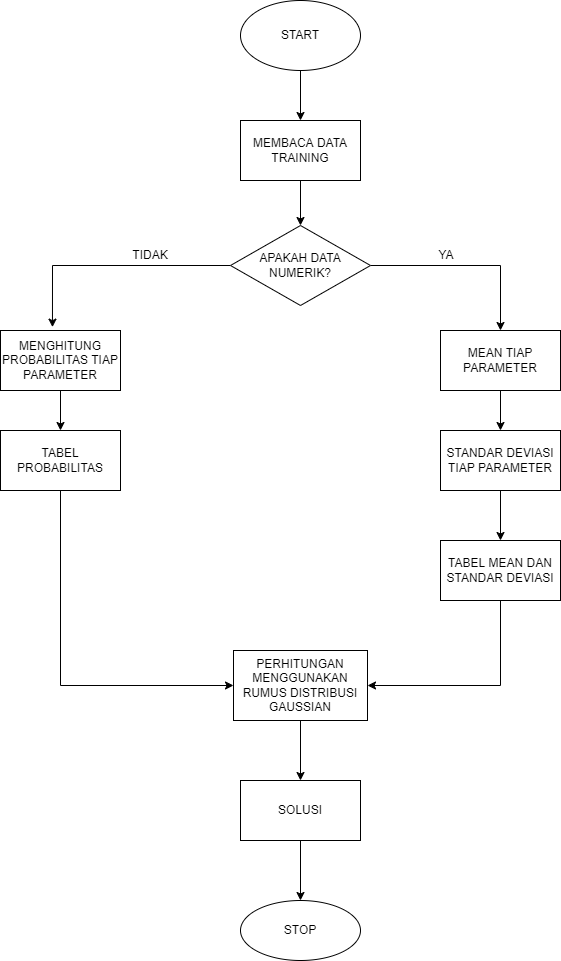
Data penelitian yang diolah terdiri dari data berjenis numerikal, kategorikal, dan juga berupa text sehingga tergolong data campuran. Dalam perhitungan Naïve Bayes, jika data yang digunakan adalah data campuran maka terdapat perbedaan dalam menghitungnya. Berikut tahapan dalam mengolah data menggunakan algoritma naïve bayes.

1. Tahap paling awal yaitu dengan menginputkan data training yang telah dipersiapkan sebelumnya.
2. Jika data tergolong kategorikal atau berupa text, maka dihitung probabilitas tiap atribut pata tabel probabilitas. Berikut rumus dalam menghitung probabilitas:
3. Jika data tergolong numerikal, maka dihitung nilai MEAN dan standar deviasi tiap atribut. Berikut rumus dalam menghitung MEAN:

Berikut rumus perhitungan standar deviasi:

1. Tahap akhir yaitu menggunakan perhitungan menggunakan rumus distribusi Gaussian. Rumus ini digunakan untuk menentukan nilai dari data uji baru apakah tepat atau terlambat berdasarkan nilai probabilitas tertinggi. Berikut rumus perhitungan distribusi Gaussian:

Adapun flowchart mengenai perhitungan data menggunakan algoritma Naïve Bayes seperti di bawah ini agar alur atau tahapan dapat lebih jelas.

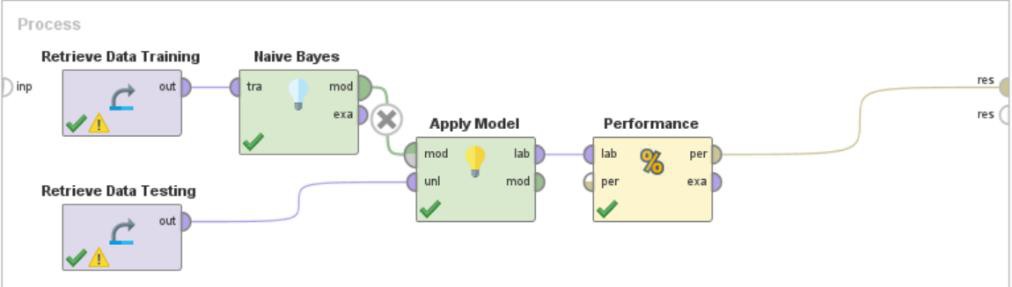


Gambar 3. 2 Flowchart Perhitungan Nive Bayes

5) Evaluation fase

Pada tahap ini dilakukan pengujian terhadap model-model untuk mendapat informasi yang tepat dan akurat. Pada tahap ini berfokus pada model yang dihasilkan sudah sesuai standar Naïve Bayes.

Selanjutnya, melakukan perhitungan tingkat akurasi ketepatan waktu kelulusan mahasiswa menggunakan RapidMiner. Input atau masukan peramalan pada RapidMiner berasal dari template excel yang berisikan data actual nilai mahasiswa Prodi Manajemen Informatika satu Angkatan.



Gambar 3. 3 Pemodelan Naïve Bayes pada RapidMiner

6) Deployment

Pembetukan model dan analisa yang telah dilakukan, selanjutnya pada tahap ini menerapkan model yang terakurat dalam menentukan klasifikasi prediksi mahasiswa yang lulus tepat waktu. Selain itu, pada tahap ini dilakukan perancangan dashboard yang menggunakan tools Misrosoft Power BI. Tahapan yang dilakukan adalah sebagai berikut:

1. Menambahkan data dengan mengimport data yang berformat excel
2. Menghubungkan file excel dengan mengklik connect
3. Memilih file excel yang telah dibuat sebelumnya
4. Mencentang tabel atau sheet diperlukan untuk visualisasi kemudian klik edit
5. Menghapus dan mengedit kolom yang tidak perlu dari tabel
6. Menerapkan perubahan kemudian menyimpan data
7. Setelahnya terlihat fields dari tabel yang telah disimpan sebelumnya
8. Membuat pemodelan dan visualisasi data sesuai dengan kebutuhan

## Jadwal Kegiatan

Jadwal pelaksanaan kegiatan ini dilakukan selama 12 bulan dimulai dari bulan Februari 2022 hingga bulan Januari 2023. Berikut langkah-langkah metode kegiatan yang ditentukan pada tabel berikut.

Tabel 3.2 Jadwal Kegiatan

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No. | Kegiatan | Bulan ke- | | | | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1. | Perencanaan |  |  |  |  |  |  |  |  |  |  |  |  |
| 2. | Analisa |  |  |  |  |  |  |  |  |  |  |  |  |
| 3. | Perancangan |  |  |  |  |  |  |  |  |  |  |  |  |
| 4. | Implementasi |  |  |  |  |  |  |  |  |  |  |  |  |
| 5. | Pengujian |  |  |  |  |  |  |  |  |  |  |  |  |

# **DAFTAR PUSTAKA**

Fajrila, E. (2018). Perbandingan Klasifikasi Ketepatan Waktu Kelulusan Mahasiswa Menggunakan Regresi Logistik Biner Dan Naïve Bayes Classifier. *ULTIMA InfoSys*.

Fiqri, F. N. (2020). Analisis dan Visualisasi Data menggunakan Power BI : Studi

Kasus Open Recruitment Laboratorium Komputer Angkatan 2019.

Hendrian, S. (2018). Algoritma Klasifikasi Data Mining Untuk Memprediksi

Siswa Dalam Memperoleh Bantuan Dana Pendidikan. *Faktor Exacta*.

Hulukati, W., & Djibran, M. R. (2018). Analisis Tugas Perkembangan Mahasiswa

Fakultas Ilmu Pebdidikan Universitas Negeri Gorontalo. *Jurnal Bikotetik*.

Imaroh, T. S., & Soleh, I. A. (2020). Analisis Pencapaian Key Performance Indicator (KPI) Pada System Application And Product (SAP) PT. GMF AeroAsia Tbk. *Jurnal Ilmiah Manajemen*.

Indonesia. (2002). *Undang-undang Nomor 232 Pasal 5 dan 6.* Jakarta: Menteri

Pendidikan Nasional.

Kafil, M. (2019). Penerapan Metode K-Nearest Neighbors Untuk Prediksi Penjualan Berbasis Web Pada Boutiq Dealove Bondowoso. *Jurnal Mahasiswa Teknik Informatika* .

Rochmadi, T. (2013). *Methode Pembangunan Software Pengenalan Software*

*Engineering.* ISBN.

Sabilla, W. I., & Putri, T. E. (2017). Prediksi Ketepatan Waktu Lulus Mahasiswa Dengan K-Nearest Neighbor Dan Naïve Bayes Classifier (Studi Kasus Prodi D3 Sistem Informasi Universitas Airlangga). *Jurnal Komputer Terapan*.

Salmu, S., & Achmad, S. (2017). Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes : Studi Kasus UIN Syarif Hidayatullah Jakarta. Jakarta: ISSN.

Siregar, M. I., Saggaf, A., & Hidayat, M. (2021). Pelatihan Pembuatan Laporan Keuangan Berbasis Microsoft Excel Pada Kerajinan Songket Mayang Palembang. *Jurnal Abdimas Mandiri*.

Suhanda, Y., Kurniati, I., & Norma, S. (2020). Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik. *Jurnal Teknologi Informatika dan Komputer MH Thamrin*

Suntoro, J. (2019). *Data mining : algoritma dan implementasi dengan pemrograman PHP.* Karawang: PT Elex Media Komputindo.

Syarli. (2016). Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus : Data Mahasiswa Baru Perguruan Tinggi). *Jurnal Ilmiah Ilmu Komputer*.

Hailitik, A. H., Djahi, B. S., & Nabuasa, Y. Y. (2017). Klasifikasi Jurusan Menggunakan Metode Naive Bayes pada Sekolah Menengah Atas Negeri (SMAN) 1 FATULEU. *J-ICON*.