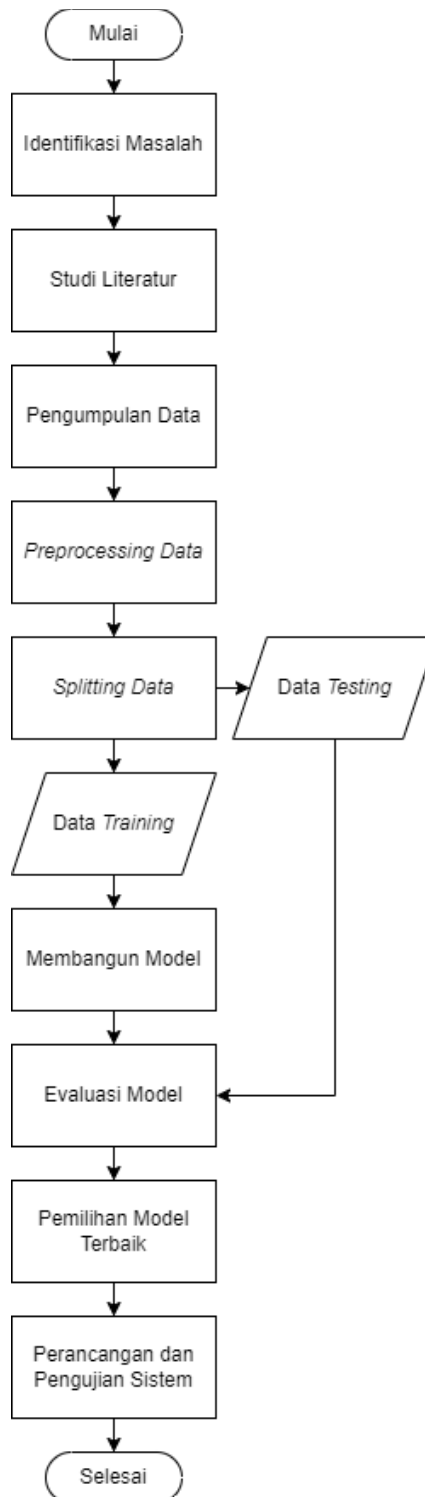


# BAB III

## METODOLOGI PENELITIAN

### 3.1 Alur Penelitian

Berikut merupakan diagram alur penelitian pada penelitian



Gambar 1.1 Diagram Alur Penelitian

### 3.1.1 Identifikasi Masalah

Pada penelitian ini, dilakukan identifikasi masalah pada Universitas Pembangunan Nasional “Veteran” Jakarta, yaitu mengenai permasalahan mahasiswa terindikasi *drop out* yang ingin diidentifikasi berdasarkan data kriteria mahasiswa yaitu “IPK”, “SKS”, dan “Masa Studi”. Penentuan kriteria dilakukan mengacu pada Peraturan Rektor No. 15 Tahun 2020 Tentang Peraturan Akademik Universitas Pembangunan Nasional “Veteran” Jakarta Bagian Kedelapan mengenai “Evaluasi Kemajuan Studi Mahasiswa dan Penghentian Studi Tetap” serta Pedoman Akademik Universitas Pembangunan Nasional “Veteran” Jakarta 2020/2021. Penentuan kriteria juga dilakukan berdasarkan wawancara yang dilakukan oleh pihak AKPK serta UPT TIK. Pemilihan ketiga atribut/kriteria memiliki keterkaitan masing-masing dalam penggunaannya. IPK merupakan atribut yang dapat menentukan derajat keberhasilan akademik mahasiswa serta dapat menentukan prestasi dan predikat kelulusan. IPK merupakan angka yang menunjukkan prestasi atau kemajuan belajar mahasiswa secara kumulatif. Batas minimum IPK padan UPN “Veteran” Jakarta adalah  $\geq 2.00$ . Oleh karena itu, mahasiswa dengan IPK dibawah batas minimum akan terprediksi *drop-out*. Salah satu faktor penentu IPK mahasiswa adalah IPS mahasiswa. IPS merupakan angka yang menunjukkan prestasi atau kemajuan belajar mahasiswa dalam satu semester. Semakin baik IPS mahasiswa, maka IPK yang dihasilkan akan semakin baik juga. Selain itu, IPS mahasiswa juga menentukan berapa banyak SKS yang dapat diambil oleh mahasiswa setiap semester. Semakin besar IPS yang diperoleh mahasiswa maka semakin besar SKS yang dapat diambil sesuai dengan batas maksimal pada ketentuan perguruan tinggi. SKS mahasiswa juga berpengaruh terhadap masa studi mahasiswa. Mahasiswa dengan masa studi minimal pada semester 2 dan maksimal pada semester 14 dapat diidentifikasi berpotensi *drop-out* melalui jumlah pengambilan SKS tempuh. Ketiga atribut yang digunakan untuk melakukan prediksi *dropout* mahasiswa dianggap sebagai nilai akhir dari proses penilaian berbagai aspek yang memengaruhi atribut-atribut tersebut. Berdasarkan faktor-faktor yang telah disebutkan, masing-masing dari ketiga atribut memiliki peran yang penting untuk mendeteksi

mahasiswa terprediksi *dropout*. Keterkaitan atribut satu dengan lainnya memperoleh keputusan bahwa apabila terdapat salah satu dari ketiga atribut tidak terpenuhi, maka mahasiswa akan diidentifikasi sebagai mahasiswa berpotensi *drop-out*. Hal ini dapat membantu pemangku kepentingan yaitu Bidang Kemahasiswaan (AKPK) dalam melakukan pengambilan keputusan dimulai pada saat mahasiswa memulai perkuliahan hingga batas akhir semester mahasiswa. Selanjutnya, referensi-referensi yang juga digunakan pada penelitian ini berisi mengenai *Algoritma Naïve Bayes* dan *Algoritma K-Nearest Neighbor (KNN)* serta berisi mengenai pembelajaran tentang permasalahan pada mahasiswa terindikasi *drop out* yang telah dilakukan oleh penelitian terdahulu.

### **3.1.2 Studi Literatur**

Studi literatur merupakan tahapan dimana dilakukan sebuah pembelajaran serta pemahaman informasi untuk dijadikan sebuah referensi. Penulis menggunakan referensi-referensi yang relevan dengan penelitian-penelitian terdahulu. Penulis menggunakan pendekatan *machine learning* yaitu berisikan algoritma-algoritma seperti *Algoritma Naïve Bayes* dan *Algoritma K-Nearest Neighbor (KNN)*. Penulis mendapatkan teori-teori tersebut dengan melakukan pemahaman melalui jurnal-jurnal, *e-book*, serta penelitian terdahulu yang relevan pada penelitian ini.

### **3.1.2 Pengumpulan Data**

Pengumpulan data merupakan tahapan untuk mencari data yang relevan dan sesuai dengan tujuan penelitian. Penelitian ini dilakukan dengan menggunakan data akademik mahasiswa Universitas Pembangunan Nasional “Veteran” Jakarta. Data-data tersebut mewakili kriteria untuk menentukan mahasiswa terindikasi *drop-out*. Pengumpulan data dilakukan dengan melakukan observasi lapangan, dokumentasi, dan wawancara. Pengumpulan data dilakukan pada LP3M, AKPK, dan UPT TIK Universitas Pembangunan Nasional “Veteran” Jakarta.

### **3.1.3 Preprocessing Data**

Tahapan *preprocessing* merupakan tahapan penting sebelum melakukan *modeling*. Pada tahap *preprocessing* melibatkan data mentah

yang belum di proses dan seleksi. Tahapan *preprocessing* memiliki dampak pada kinerja serta kualitas model. Pada penelitian ini, tahapan yang dilakukan meliputi *data cleaning*, *label encoding*, *feature engineering* dan *data scalling*,

### 3.1.4 Splitting Data

Setelah tahapan *preprocessing*, kemudian dilakukan *splitting data*. Data dibagi menjadi *data training* dan juga *data testing*. *Data training* digunakan untuk proses pelatihan pada *modeling*. Sedangkan *data testing* berguna untuk melakukan prediksi model yang sebelumnya telah dilakukan pelatihan menggunakan *data training*. Hal ini berguna untuk memberikan gambaran mengenai keakuratan model. Penentuan rasio pada *data splitting* akan ditetapkan pada saat proses pembuatan model dilakukan. Rasio dengan *accuracy* tertinggi akan dijadikan acuan dalam pembuatan model.

### 3.1.5 Membangun Model

Pada tahapan ini, akan dilakukan pemodelan dari dua algoritma *machine learning*, seperti *Algoritma K-Nearest Neighbor (KNN)* dan *Algoritma Naïve Bayes*. Hal ini dilakukan untuk mencari model dengan akurasi terbaik. Hal pertama yang dilakukan sebelum membuat model adalah melakukan persiapan data terlebih dahulu. Data-data yang diperlukan dikumpulkan, kemudian dilakukan *preprocessing* data. Setelah data sudah sesuai, dapat dilakukan pembangunan model. Model pertama yang digunakan adalah Naïve Bayes. Pada proses ini, dilakukan *training data* untuk menghitung probabilitas kelas serta probabilitas kondisional fitur-fitur. Prediksi dan evaluasi diterapkan pada model ini dengan menggunakan *metric accuracy*, *precision*, *recall*, dan *F1-score*. Setelah mendapatkan hasil, langkah berikutnya adalah melakukan proses yang sama untuk model kedua. Model kedua yang digunakan adalah *KNN*. Setelah melalui tahap *preprocessing* dan *splitting data*, dilakukan pembangunan model *KNN*. Pembangunan model ini memerlukan parameter *k* yang merupakan jumlah tetangga yang akan digunakan dalam membuat prediksi. Setelah model *KNN* diprediksi, selanjutnya adalah melakukan evaluasi pada model tersebut. Setelah akurasi dari kedua model didapatkan, dapat dilakukan perbandingan antar kedua model. Model

dengan  
20

akurasi terbaik dapat digunakan pada penelitian untuk selanjutnya dilakukan prediksi dan visualisasi.

### 3.1.6 Evaluasi Model

Berdasarkan pembuatan model yang telah dilakukan oleh algoritma *Naïve Bayes dan KNN*, selanjutnya adalah melakukan evaluasi model. Evaluasi model berfungsi untuk mengukur kinerja model serta melakukan pendeteksian apakah model tersebut mengalami *overfitting* atau *underfitting*. Uji model melibatkan melibatkan metrik evaluasi yaitu *accuracy, precision, recall, dan F1-Score*.

### 3.1.7 Pemilihan Model Terbaik

Setelah dilakukan uji model, tahapan berikutnya adalah melakukan pemilihan model terbaik berdasarkan kedua algoritma yang telah dilakukan pemodelan. Model dengan kinerja terbaik dan akurasi yang terbaik adalah model yang akan digunakan dalam penelitian yaitu dalam pembuatan sistem prediksi mahasiswa berpotensi *drop-out* berbasis *web*.

### 3.1.8 Perancangan dan Pengujian Sistem

Dalam perancangan dan pengujian sistem, penulis menggunakan metode *waterfall*. Tujuan dari penggunaan metode *waterfall* yaitu karena metode ini terstruktur dan sistematis. Metode *waterfall* memiliki beberapa ciri seperti persyaratan yang stabil serta estimasi waktu yang akurat sehingga sesuai dengan penelitian ini. Metode *waterfall* berisi *Requirements Analysis* (Analisis Kebutuhan), *System Design* (Desain Sistem), *Implementation* (Implementasi), *Integration and Testing* (Integrasi dan Pengujian), *Deployment* (Penerapan), dan *Maintenance* (Pemeliharaan).

- **Analisis Kebutuhan (*Requirement Analysis*)**

Dalam tahap ini, dilakukan pengumpulan data yang bersumber dari observasi oleh pemangku kepentingan, selanjutnya juga dilakukan persiapan kebutuhan baik dari *software* maupun *hardware*. *Output* dari tahap ini yaitu berupa data API, informasi wawancara, dan kebutuhan fungsional/non-fungsional lainnya.

- **Desain Sistem (*System Design*)**

Pada tahap desain sistem, penulis menggunakan rancangan UML diagram berupa *use case diagram* dan *activity diagram*. Kemudian, penulis membuat rancangan *user interface* berupa *wireframe*.

- **Implementasi (*Implementation*)**

Dalam tahap ini, setelah dilakukan pemodelan, perancangan sistem berbasis *website* dilakukan dengan menggunakan *HTML*, *CSS*, *Javascript* untuk *frontend* dan untuk *backend* menggunakan *Flask*.

- **Pengujian Sistem (*System Testing*)**

Pada tahap ini, dilakukan pengujian sistem. Penulis menggunakan *Black Box Testing* untuk melakukan *testing* dalam rangka memastikan bahwa sistem berjalan sesuai dengan kebutuhan.

- **Penerapan (*Deployment*)**

Setelah dilakukan pengujian, sistem dapat diterapkan dan digunakan oleh pemangku kepentingan untuk membantu dalam melakukan pengambilan keputusan.

- **Pemeliharaan (*Maintenance*)**

Pada tahap ini, dilakukan *maintenance* dengan tujuan untuk memperbaiki *bug*, *update system*, dan penambahan fitur-fitur baru yang dibutuhkan.

## 3.2 Perangkat penelitian

### 3.2.1 Perangkat Keras

1. Perangkat : *HP 14-dq1089wm*
2. *Processor* : *intel(R) Core(TM) i3-1005G1 CPU @ 1.20GHz*  
(*4CPUs*), *~1.2GHz*
3. *RAM* : *8192MB RAM*

### 3.2.2 Perangkat Lunak

1. *Operating System* : *Windows 11 Home 64-bit (10.0, Build 22000)*
2. *Bahasa Pemrograman* : *Python*
3. *Aplikasi* : *Ms Word, Ms Excel, Visual Studio Code, Jupyter, Figma*
4. *Website* : *Google Collaboratory, Google Docs, Draw.io*

### 3.3 Waktu dan Tempat Penelitian

- a. Tempat : Universitas Pembangunan Nasional “Veteran” Jakarta
- b. Waktu : 5 bulan penelitian

### 3.4 Jadwal Penelitian

Tabel 3.1 Jadwal Penelitian

Nama Kegiatan	Bulan-ke																			
	1				2				3				4				5			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Identifikasi Masalah	■	■																		
Studi Literatur		■	■	■																
Pengumpulan Data					■	■	■	■												
<i>Preprocessing Data</i>									■	■	■									
<i>Splitting Data</i>											■									
Membangun Model													■	■	■	■				
Uji Model																			■	
Pemilihan Model Terbaik																			■	
Perancangan dan Pengujian Sistem																			■	■
Dokumentasi	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■