

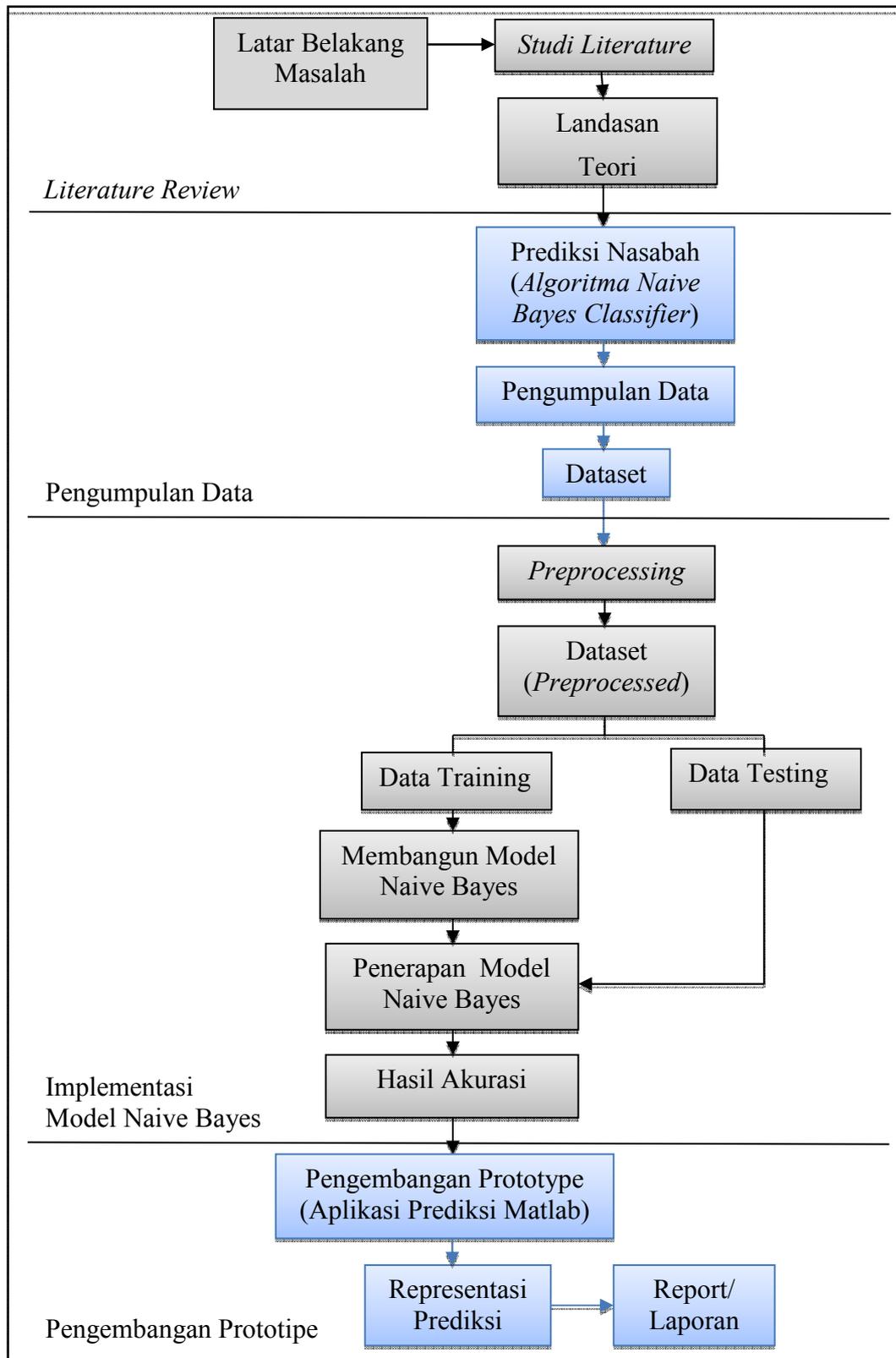
## **BAB 3**

### **METODE PENELITIAN**

Pada proses penelitian ini dilakukan beberapa tahapan mulai dari tahap awal yaitu tahap inisiasi, pengembangan model, dan tahap terakhir pengembangan prototipe. Dalam tahapan inisiasi beberapa kegiatan mulai dari identifikasi permasalahan dan studi literatur untuk menemukan alternatif solusi untuk masalah yang ada. Tahap selanjutnya, yaitu tahap pengembangan model dengan mempersiapkan dataset untuk penelitian dan pemilihan algoritma dari sejumlah alternatif yang sesuai untuk pengolahan dataset. Ditahapan akhir dikembangkan prototipe yang diharapkan dapat membantu memudahkan user untuk mendapatkan informasi prediksi dari data yang dimiliki. Pada gambar 3.1 menunjukkan sistematika penelitian yang digunakan dalam penelitian ini, dengan rincian penjelasan tiap-tiap tahapannya, yaitu:

#### **3.1 *Literature Review***

Tahap awal dalam penelitian ini adalah literature review atau tinjauan pustaka. Dalam melakukan penelitian yang menjadi *literature review* atau tinjauan pustaka yaitu dengan kajian pustaka atau *studi literature*. Suatu penelitian tidak lepas dari kajian pustaka atau *studi literature*. Penelitian ini didukung oleh teori- teori yang mendasari masalah, informasi, dan pemikiran- pemikiran yang relevan, serta bersumber dari buku, jurnal, maupun prosiding.



**Gambar 3.1 Metode Penelitian**

### 3.2 Pengumpulan Data

Data yang digunakan untuk penelitian ini berasal dari data nasabah bagian marketing sebuah bank di portugal, yang didonasi pada tanggal 14 Februari 2012 oleh S.Moro, P.Cortez, dan P.Rita data tersebut termasuk data sekunder dan bersifat public yang diperoleh dari *UCI Machine Learning Repository*. Dataset tersebut akan dibagi menjadi 2 dataset yaitu *data training* dan *data testing*. *Data training* digunakan untuk mengetahui probabilitas guna melakukan klasifikasi untuk pengambilan keputusan sedangkan *data testing* digunakan untuk eksperimen dalam penelitian ini guna pembangunan prototipe.

The screenshot shows the UCI Machine Learning Repository website. The main heading is 'Bank Marketing Data Set'. Below the heading, there are links for 'Download: Data Folder' and 'Data Set Description'. An abstract describes the data as related to direct marketing campaigns (phone calls) of a Portuguese banking institution, with the goal of predicting if a client will subscribe to a term deposit. A table provides key statistics:

Data Set Characteristics:	Multivariate	Number of Instances:	45211	Area:	Business
Attribute Characteristics:	Real	Number of Attributes:	17	Date Donated:	2012-02-14
Associated Tasks:	Classification	Missing Values?	N/A	Number of Web Hits:	309454

Source: [Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014.

Data Set Information: The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

**Gambar 3.2 Dataset Bank Marketing**

Data yang didapatkan berupa data mentah atau data nasabah yang berjumlah 4521 *record*. Data tersebut terdiri 16 atribut dan 1 label kelas. Data tersebut harus melalui tahap seleksi dan *preprocessing* sebelum digunakan dalam penelitian.

age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
30	unemployed	married	primary	no	1787	no	no	cellular	19	oct	79	1	-1	0	unknown	no
33	services	married	secondary	no	4789	yes	yes	cellular	11	may	220	1	339	4	failure	no
35	management	single	tertiary	no	1350	yes	no	cellular	16	apr	185	1	330	1	failure	no
30	management	married	tertiary	no	1476	yes	yes	unknown	3	jun	199	4	-1	0	unknown	no
59	blue-collar	married	secondary	no	0	yes	no	unknown	5	may	226	1	-1	0	unknown	no
35	management	single	tertiary	no	747	no	no	cellular	23	feb	141	2	176	3	failure	no
36	self-employed	married	tertiary	no	307	yes	no	cellular	14	may	341	1	330	2	other	no
39	technician	married	secondary	no	147	yes	no	cellular	6	may	151	2	-1	0	unknown	no
41	entrepreneur	married	tertiary	no	221	yes	no	unknown	14	may	57	2	-1	0	unknown	no
43	services	married	primary	no	-88	yes	yes	cellular	17	apr	313	1	147	2	failure	no
39	services	married	secondary	no	9374	yes	no	unknown	20	may	273	1	-1	0	unknown	no
43	admin.	married	secondary	no	264	yes	no	cellular	17	apr	113	2	-1	0	unknown	no
36	technician	married	tertiary	no	1109	no	no	cellular	13	aug	328	2	-1	0	unknown	no
20	student	single	secondary	no	502	no	no	cellular	30	apr	261	1	-1	0	unknown	yes
31	blue-collar	married	secondary	no	360	yes	yes	cellular	29	jan	89	1	241	1	failure	no
40	management	married	tertiary	no	194	no	yes	cellular	29	aug	189	2	-1	0	unknown	no
56	technician	married	secondary	no	4073	no	no	cellular	27	aug	239	5	-1	0	unknown	no
37	admin.	single	tertiary	no	2317	yes	no	cellular	20	apr	114	1	152	2	failure	no
25	blue-collar	single	primary	no	-221	yes	no	unknown	23	may	250	1	-1	0	unknown	no

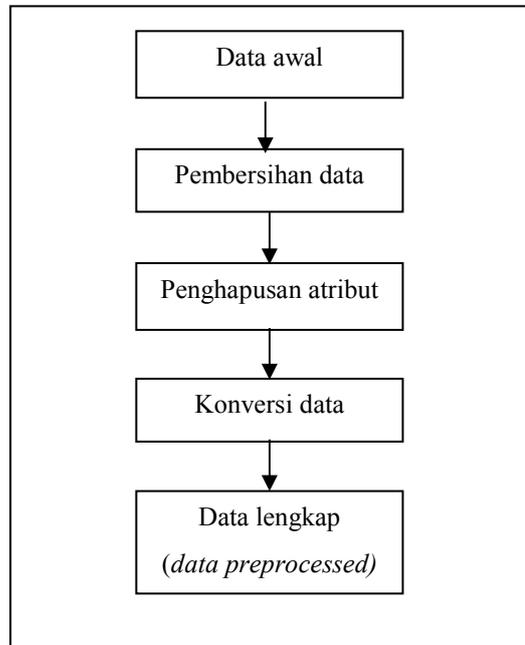
**Gambar 3.3 Data Mentah Data Nasabah**

### 3.3 Implementasi Metode *Naive Bayes Classifier*

Dari kajian pustaka atau *studi literature* penelitian, penerapan metode *Naive Bayes Classifier* berpotensi menentukan probabilitas guna melakukan klasifikasi apakah nasabah tersebut masuk dalam klasifikasi “no” atau “yes”. Berikut ini merupakan langkah penerapan metode *Naive Bayes Classifier*.

#### 3.3.1 *Preprocessing*

Sebelum tahap implementasi diterapkan, tahap *preprocessing* terlebih dahulu dilakukan. Jumlah data awal yang dapat diperoleh dari pengumpulan data yaitu sebanyak 4521 data, tetapi tidak semua data digunakan dan tidak semua atribut dipergunakan karena data tersebut harus melalui tahap pengolahan awal data atau disebut dengan *preparation data*.



**Gambar 3.4 Preprocessing**

Untuk mendapatkan data yang memiliki kualitas bagus, beberapa tahap harus dilakukan, berikut adalah tahapannya:

### **3.3.1.1 Data Cleaning**

Data yang tidak berkualitas akan menghasilkan data mining yang tidak berkualitas. Keputusan yang berkualitas harus didasarkan pada data yang berkualitas (data ganda atau data yang telah hilang dapat menyebabkan ketidakbenaran). Ekstaksi data, pembersihan, dan transformasi data merupakan tugas utama dalam *data warehouse*. Tugas- tugas *data cleaning*, yaitu:

1. Mengisi nilai yang telah hilang pada data yang tidak lengkap (*missing value*).
2. Mengidentifikasi atau menghilangkan *outliers* dan memperhalus *data noise*.
3. Memperbaiki ketidak konsistenan data.
4. Memecahkan redundansi yang disebabkan oleh integrasi data.

### 3.3.1.2 *Data Integration dan Transformation*

Dapat meningkatkan akurasi dan efisiensi algoritma. Algoritma *Naive Bayes Classifier* mampu melakukan proses data yang bernilai nominal, ordinal, atau kontinu. Sehingga nilai dari tiap-tiap atribut yang terdapat pada dataset tidak perlu ditransformasikan.

1. Mengintegrasikan berbagai database atau file- file.
2. Transformasi data (*data transformation*).
3. Normalisasi dan agregasi.

### 3.3.1.3 *Data Reduction*

Proses yang dilakukan dengan mereduksi suatu dataset yaitu dengan mengurangi jumlah atribut dan *record* yang ada sehingga lebih sedikit tetapi masih tetap informatif.

1. Mendapatkan hasil representasi pada volume data yang sebelumnya sudah berkurang namun tetap menghasilkan analisis yang sama.
2. Diskritisasi data adalah bagian dari proses reduksi data, namun pada bagian ini ini adalah penting untuk data-data yang numerik.

### 3.3.2 *Dataset (preprocessed)*

Dataset (*preprocessed*) adalah dataset yang sudah melalui *preprocessing* diantaranya sudah melalui proses *cleaning*, *integration* dan *transformation*, dan *reduction*, sehingga dataset tersebut sudah siap digunakan untuk data training dan data testing.

### 3.3.3 *Data Training dan Data Testing*

Pada *data training* biasanya digunakan sebagai pola data pembentukan model data mining. Sedangkan *data testing* adalah data yang digunakan untuk pengujian dalam menerapkan model klasifikasi.

### 3.3.4 Membangun Model *Naive Bayes Classifier*

Berdasarkan pada *data training* yang sebelumnya sudah melalui *preprocessing* sehingga dapat digunakan untuk membangun model *Naive Bayes Classifier* yaitu menyelesaikan masalah klasifikasi *class* atau atribut dalam data.

### 3.3.5 Penerapan Model *Naive Bayes Classifier*

Penerapan model *Naive Bayes Classifier* dilakukan berdasarkan model yang telah dibangun sebelumnya yang digunakan untuk menentukan atribut atau *class* dari suatu data baru yang atribut atau *class* nya belum diketahui sebelumnya.

### 3.3.6 Hasil Akurasi

Hasil akurasi merupakan hasil nilai akurasi yang didapat dari penerapan algoritma *Naive Bayes Classifier*. Untuk menghitung akurasi dari pola yang didapat, pengujian *Confusion Matrix* dan *Receiver Operating Characteristics* (ROC) merupakan metode akurasi dalam konsep data mining. Perhitungan *Confusion Matrix* menghasilkan output (*Recall*, *Accuracy*, dan, *Error*) sedangkan *Receiver Operating Characteristics* (ROC) menunjukkan penilaian tentang hasil dari prediksi yang dikategorikan kedalam 5 tingkatan nilai akurasi.

## 3.4 Pengembangan Prototipe

Pada pengembangan prototipe terdapat 2 (dua) tahapan yaitu penerjemah pola dan *development prototype*. Penerjemah pola pada proses KDD merupakan proses untuk mengetahui apakah hasil pola dari penerapan metode yang digunakan sesuai atau bertentangan dengan fakta yang ada. Pada pengembangan prototipe dalam penelitian ini dilakukan melalui tahapan antara lain, sebagai berikut:

### 3.4.1 Development Prototype (Aplikasi Prediksi *berbasis web*)

Pengembangan prototipe digunakan untuk merepresentasikan dalam penelitian ini, yaitu dengan mengimplementasikan aplikasi *web*. Hasil yang didapat dari implementasi tersebut dapat digunakan oleh pihak bank untuk melakukan penanganan lebih dini terhadap nasabah yang melakukan simpanan deposito.

### **3.4.2 Representasi Prediksi**

Penerapan algoritma *Naive Bayes Classifier* pada aplikasi *web* yang hasilnya penggambaran dari prediksi nasabah yang melakukan simpanan deposito.

### **3.4.3 Report atau Laporan**

*Report* atau laporan adalah *view* atau tampilan hasil dari prediksi berupa klasifikasi nasabah yang dapat digunakan oleh pihak bank untuk melakukan penanganan lebih dini terhadap nasabah.