

## **BAB III METODOLOGI PENELITIAN**

### **3.1. Gambaran Umum Objek Penelitian**

#### **3.1.1 Bank**

Bank adalah sebuah instansi keuangan yang menangani depet dan kredit. Bank bisa juga meminjamkan, menerima serta menyimpan uang, dan membangun kesenjangan antara pemberi pinjaman. Bank tidak hanya berurusan dengan uang tetapi bisa juga menjadi penghasil uang [20]. Sementara menurut professor Kinley mendefinisikan bank sebagai “Bank adalah sebuah pendirian dimana individu menghasilkan sebuah uang muka sebagaimana yang mungkin diperlukan dan dilakukan secara amat, dan kepada siapapun individu yang mempercayakan uang kepada sebuah bank [20].

#### **3.2 Data Collection**

Pengumpulan data penelitian mengambil data kampanye pemasaran langsung dari lembaga bank portugis yang bergerak dibidang perbankan. Dataset yang digunakan yaitu berbentuk daftar keterangan pelanggan yang bersifat anonim atau tidak dicantumkan nama dan alamat.

Sumber data yang digunakan adalah data pelanggan. Data perbankan tersebut diambil dari <https://datahub.io/machine-learning/bank-marketing#readme>. Data diambil dari *datahub* karena sering digunakan, terbuka, dan gratis untuk mengunduh *historical* data sehingga mudah untuk diakses.

Pada tabel 3.1 ditampilkan bentuk atribut dari dataset bank marketing yang didapat sebanyak 20000 data set dengan 16 atribut data dan 1 label. Data perbankan akan diolah menggunakan model algoritma data mining Naïve Bayes, Decision Tree, NN, Random Forest dan K-NN kemudian dibandingkan hasil peramalan dari keempat algoritma tersebut. Pada tabel 3.1 ditampilkan bentuk atribut dari dataset bank marketing.

**Tabel 3.1 Atribut dan Nilai Dataset Bank Marketing**

Atribut	Nilai
Age	Usia para nasabah
Job	('admin.', 'unknown', 'unemployed', 'management', 'housemaid', 'entrepreneur', 'student', 'bluecollar', 'selfemployed', 'retired', 'technician', 's ervices')
Marital	(Categorical: 'divorced', 'married', 'single')
Education	('unknown', 'secondary', 'primary', 'tertiary')
Default	apakah kredit dalam default? (Binary: 'yes', 'no')
Balance	Saldo tahunan rata-rata
Housing	Apakah nasabah punya pinjaman rumah? (Binary: 'yes', 'no')
Loan	Punya pinjaman pribadi? (Binary: 'yes', 'no')
Contact	jenis komunikasi kontak (categorical: 'unknown', 'telephone', 'cellular')
Day	Hari kontak terakhir dalam sebulan
Month	bulan kontak terakhir tahun ini (kategori: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
Duration	Durasi kontak terakhir, dalam detik
Campaign	Jumlah kontak yang dilakukan selama pemasaran ini dan untuk nasabah ini (
Pdays	Jumlah hari yang berlalu setelah nasabah terakhir dihubungi dari pemasaran sebelumnya (-1 berarti klien sebelumnya tidak dihubungi)
Previous	Jumlah kontak yang dilakukan sebelum pemasaran ini dan untuk nasabah ini
Poutcome	Hasil dari pemasaran sebelumnya (categorical: 'unknown', 'other', 'failure', 'success')
Deposit	Variabel output (binary: 'yes', 'no')

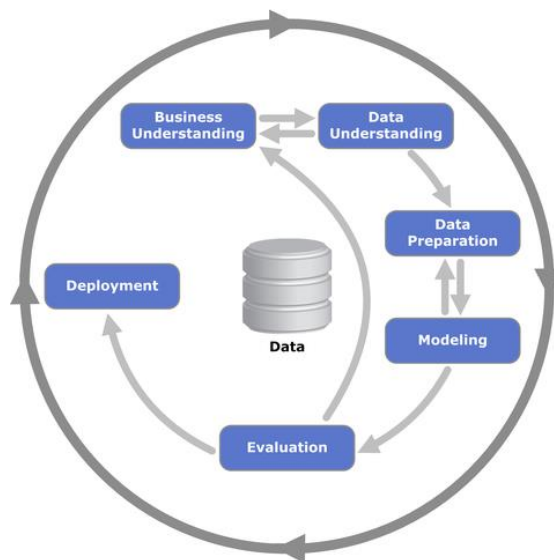
### 3.2.1 Variabel Independent

Dalam pembuatan peramalan model penelitian ini yang menjadi variabel independent yaitu *age*, *job*, *marital*, *education*, *default*, *balance*, *housing*, *loan*, *contact*, *day*, *month*, *duration*, *campaign*, *pdays*, *previous*, dan *poutcome*.

### 3.2.2 Variabel Dependent

Variabel dependen yang terdapat dalam penelitian ini adalah variabel deposit yang berisi 'yes' atau 'no'. Variabel ini digunakan sebagai label karena acuan pertama apakah pelanggan sudah membuka deposito atau belum.

### 3.3. Alur Penelitian



**Gambar 3.1 Siklus hidup dalam CRISP-DM**

Sumber: [3]

Pada gambar 3.1 adalah gambar dari siklus CRISP-DM. Tahap pertama alur penelitian ini yaitu *business understanding* proses yaitu tahap pemahaman sebuah bisnis dengan cara menentukan tujuan proyek dan kebutuhan apa saja yang dibutuhkan lingkup bisnis secara menyeluruh. Kemudian menerjemahkan tujuan batasan penelitian dan menyiapkan strategi untuk mencapai tujuan tersebut [14].

Tahap kedua yaitu *data understanding* tahapan ini dilakukan untuk mengumpulkan data, mengenali data serta membuat analisis. Setelah dilakukan analisis data maka tahap selanjutnya yaitu mengevaluasi kualitas data [14].

Tahap ketiga yaitu *data preparation* tahapan ini dilakukan untuk menyiapkan data yang tersedia kemudian dijadikan satu menjadi sebuah dataset, memilih kasus yang ingin dianalisis, dan melakukan perubahan pada beberapa variabel apabila hal itu diperlukan [14].

Tahap keempat yaitu *data modeling* tahapan ini berfungsi untuk menentukan dan menganalisis persyaratan data yang diperlukan untuk mendukung proses tujuan yang diinginkan. Pada tahapan ini juga harus dilakukannya kecocokan sebuah data dengan model yang ingin diterapkan sehingga mendapatkan hasil yang diinginkan [14].

Tahap kelima yaitu *Evaluation* yang berguna untuk mengevaluasi satu atau lebih model yang telah digunakan untuk penelitian, serta menetapkan apakah terdapat model yang memenuhi tujuan fase awal dan yang terakhir yaitu mengambil keputusan berdasarkan hasil akhir [14].

Tahap keenam yaitu *Deployment* yaitu yang dilakukan tahap akhir ini adalah menggunakan model yang telah dihasilkan kemudian dapat memberikan sebuah *insight* baru dan memberikan laporan model mana yang lebih baik [14].

### **3.3.1 Business Understanding**

Deposit adalah sebuah istilah keuangan yang berarti menyimpan uang di bank yang penarikannya hanya dapat dilakukan pada waktu tertentu. Deposito merupakan hal yang cukup sering digunakan oleh masyarakat untuk berinvestasi namun dengan sedikit resiko [11]. Menurut Undang – Undang Nomor 10 tahun 1998, deposito sendiri memiliki definisi yaitu simpanan yang penarikannya hanya dapat dilakukan pada waktu tertentu berdasarkan perjanjian nasabah penyimpan dengan bank. Hal itu dikarenakan bahwa deposito pada bank memiliki tingkat suku bunga per tahun yang diterima lebih tinggi dibandingkan giro dan simpanan tabungan biasa [14]. Oleh karena itu, suatu bank harus bisa menentukan mana pelanggan yang kemungkinan membuka tabungan deposito dengan cara melakukan promosi dan strategi menawarkan deposito kepada pelanggan secara efisien. Tujuan suatu bank mendapatkan nasabah sebanyak-banyaknya yaitu bank mendapatkan keuntungan dari bunga bersih, biaya perbankan, dan dana pengendapan dari pelanggan. Dengan menggunakan metode data mining yaitu sebuah metode untuk melakukan dan memberikan hasil prediksi kepada kredibilitas dan kemampuan calon nasabah yang ingin mendaftar tabungan deposito.

### **3.3.2 Data Understanding**

Dataset yang digunakan berasal dari <https://datahub.io/machine-learning/bank-marketing#readme> yaitu data Bank Marketing, data tersebut terkait dengan kampanye pemasaran langsung dari lembaga perbankan

Portugis. Kampanye pemasaran didasarkan dari panggilan telepon atau melewati surel dari lembaga perbankan. Selama fase pemahaman data, data diolah dengan cara menganalisis karakteristik utama data serta mengklasifikasi value dari setiap variabel dan menemukan mana yang paling berpengaruh. Penelitian ini menggunakan klasifikasi karena algoritma klasifikasi mencoba menarik beberapa kesimpulan terkait dari nilai input yang diberikan untuk pelatihan, hal ini juga untuk memprediksi label kelas/kategori untuk data baru. Tujuan klasifikasi adalah untuk memprediksi apakah klien akan berlangganan deposito berjangka (variabel  $y$ ). Data tersebut terkait dengan kampanye pemasaran langsung dari lembaga perbankan Portugis. Kampanye pemasaran didasarkan pada panggilan telepon. Seringkali, lebih dari satu kontak ke klien yang sama diperlukan, untuk mengakses apakah produk (deposito bank).

### **3.3.3 Data Preparation**

Data yang didapatkan untuk penelitian ini berjumlah 20000 pelanggan dengan 16 variabel predictor dan 1 variabel label. Dalam penelitian ini, data preparation akan dilakukan beberapa hal yaitu:

- a) Memeriksa Data Kosong Dalam tahap ini yaitu memeriksa apakah terdapat value yang kosong pada setiap variabel.
- b) Mengecek Duplikasi Data Langkah preparation ini yaitu mengecek apakah terdapat duplikasi data.

### **3.3.4 Modeling**

Pada tahap ini dilakukan sebuah pengujian dari setiap model yaitu dengan cara menghitung dan mendapatkan *rules* dari model algoritma yang diterapkan. Pengujian model menggunakan algoritma SVM, Naïve Bayes, Decision Tree, NN, Random Forest dan K-NN. Setelah dilakukan pemodelan maka dapat dilakukan perbandingan dari algoritma satu dan lainnya.

### **3.3.5 Evaluation**

Setelah dilakukan pengujian dan mendapatkan hasil dari model SVM, Naïve Bayes, Decision Tree, NN, Random Forest dan K-NN maka langkah selanjutnya yaitu melakukan evaluasi / validasi hasil dengan cara menggunakan standar umum kemudian dibandingkan hasil akurasi paling besar dari beberapa model yang telah diuji. Untuk membandingkan performa sebuah model dapat dilakukan dengan cara mencari accuracy,

precision, dan f measure. Ketiga parameter tersebut yaitu accuracy, precision, dan f measure, semakin besar hasilnya maka hasil model yang dipilih memiliki tingkat akurat yang bagus.

### **3.3.6 Deployment**

Tahap akhir dari crisp-dm yaitu deployment yang berisikan kesimpulan dan keputusan bisnis yang bisa diambil sehingga dapat menghasilkan insight. Dari modeling yang sudah dilakukan kita dapat mengetahui model mana yang paling baik untuk prediksi deposito untuk customer berdasarkan validasi setiap model yang telah dihasilkan.

### **3.5. Tools**

Penelitian ini secara keseluruhan menggunakan *tools* data mining Rapid Miner, digunakanya perangkat ini karena merupakan *open-source software* yang dapat digunakan secara cuma-cuma dan platform analitik prediktif yang populer. Rapid miner dapat membantu aktifitas data ming dengan cepat serta menyertakan analisis prediktif didalamnya. *Tool* yang digunakan pada penelitian ini yaitu Rapid Miner versi 9.6.000 platform windows.