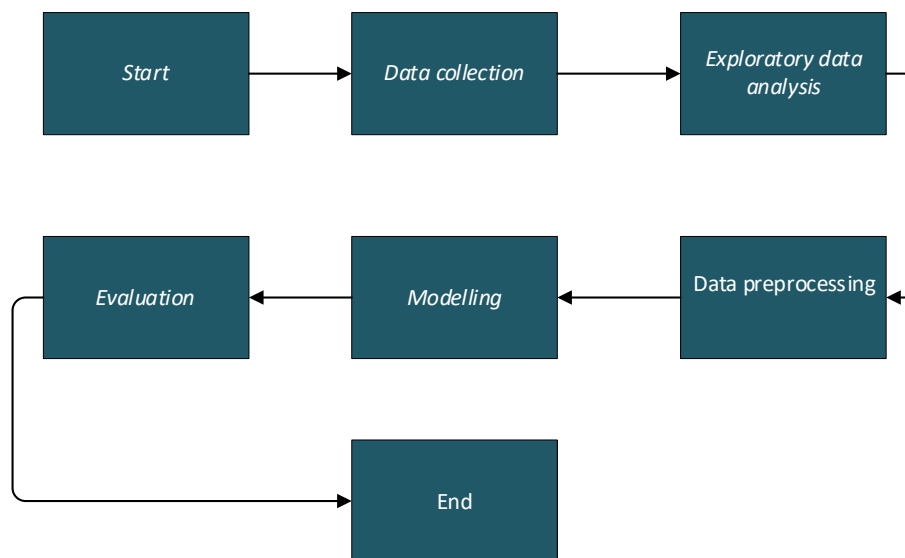


BAB III METODOLOGI PENELITIAN

3.1 Alur Penelitian

Metode penelitian yang digunakan pada penelitian ini menggunakan model proses *data mining*. Adapun tahapan penelitian yang dilakukan sebagai berikut.



Gambar 3. 1 Alur Penelitian

3.1.1 Data collection

Proses yang dilakukan untuk mengumpulkan data atau informasi yang relevan dan akurat.

1. Pengumpulan Data

Penelitian ini menggunakan *dataset* status keputungan peserta BPJS yang diperoleh dari badan penyelenggara jaminan sosial (BPJS). *Dataset* yang digunakan berjumlah 10.000 data dalam bentuk comma separated values (csv). Pada data tersebut terdiri dari 26 *column* untuk *predictor* dan sebuah target. Kolom sebagai *predictor* adalah bobot,

provinsi FKTP, kabupaten kota FKTP, kepemilikan FKTP, jenis FKTP, Tipe FKTP, tingkatan pelayanan FKTP, jenis poli FKTP, Segmen peserta saat akses layanan, provinsi faskes tujuan rujukan, kabupaten kota faskes tujuan rujukan, kepemilikan faskes tujuan rujukan, jenis faskes tujuan rujukan, tipe faskes tujuan rujukan, poli faskes tujuan rujukan, jenis kunjungan_FKTP. Sedangkan, *column* sebagai target adalah kelas status pulang peserta. Format dataset tersebut dapat dilihat pada tabel 3.1

Tabel 3. 1 Meta data

Format Dataset peserta BPJS Kesehatan		Keterangan	Variable
1.	No_peserta	Merupakan nomor kepesertaan pasien BPJS	PSTV01
2.	No_keluarga	Merupakan nomor yang terdiri dari NIK dan Nomor KK	PSTV02
3.	Bobot	Merupakan bobot atau nilai yang digunakan dalam analisis data	PSTV15
4.	Id_kunjungan_FKTP	Id Kunjungan Fasilitas Kesehatan Tingkat Pertama (FKTP) adalah suatu identifikasi atau nomor unik yang diberikan untuk merekam dan melacak kunjungan atau pelayanan kesehatan yang diberikan oleh Fasilitas Kesehatan Tingkat Pertama	FKP02

Format Dataset peserta BPJS Kesehatan		Keterangan	Variable
5.	Tanggal_datang_kunjungan_FKTP	Merupakan suatu atribut yang digunakan untuk mencatat tanggal kedatangan pasien pada kunjungan ke Fasilitas Kesehatan Tingkat Pertama (FKTP)	FKP03
6.	Tanggal_pulang_kunjungan_FKTP	Merupakan suatu atribut yang digunakan untuk mencatat tanggal kepulangan pasien pada kunjungan ke Fasilitas Kesehatan Tingkat Pertama (FKTP)	FKP04
7.	Provinsi_FKTP	Merupakan provinsi fasilitas kesehatan tingkat pertama yang terdiri dari provinsi yang ada di Indonesia.	FKP05
8.	Kabupaten_kota_FKTP	Merupakan kabupaten /kota fasilitas kesehatan tingkat pertama yang terdiri dari kabupaten/kota yang ada di Indonesia.	FKP06
9.	Kepemilikan_FKTP	Merupakan kepemilikan fasilitas kesehatan tingkat pertama yang terdiri dari PemProv, PemKab, POLRI, TNI, swasta, dll.	FKP07

Format Dataset peserta BPJS Kesehatan		Keterangan	Variable
10.	Jenis_FKTP	Merupakan jenis fasilitas kesehatan tingkat pertama yang terdiri dari puskesmas, klinik, dokter umum, dokter gigi, laboratorium dll.	FKP08
11.	Tipe_FKTP	Merupakan tipe fasilitas Kesehatan tingkat pertama yang terdiri dari klinik non rawat inap, rawat inap, klinik rawat inap, dll.	FKP09
12.	Tingkatan_pelayanan_FKTP	Merupakan tingkatan pelayanan fasilitas kesehatan tingkat pertama yang terdiri dari pelayanan rawat jalan tingkat pertama (rjtp), rawat inap tingkat pertama(ritp) dan promotif.	FKP10
13.	Jenis_poli_FKTP	Merupakan jenis poli fasilitas kesehatan tingkat pertama yang terdiri dari poli umum, poli gigi & mulut, poli KIA, UGD, Poli KB, dll.	FKP11
14.	Segmen_peserta_saas_aksas_layanan	Merupakan segmen peserta yang terdiri dari bukan pekerja, penerima bantuan iuran anggaran	FKP12

Format Dataset peserta BPJS Kesehatan		Keterangan	Variable
		pendapatan dan belanja negara (PBI APBN), penerima bantuan iuran anggaran pendapatan dan belanja daerah (PBI APBD), pekerja bukan penerima upah(PBPU), pekerja penerima upah(PPU).	
15.	Kode_dan_nama_diagnosis_ICD_10_3_digit	merujuk pada informasi tentang suatu kode dan nama diagnosis kesehatan	FKP14
16.	Kode _diagnosis_ICD_10_3_digit	merujuk pada informasi tentang suatu kode diagnosis kesehatan	FKP14A
17.	Kode_diagnosis_beragam_3_5_digit	Kode diagnosis beragam 3-5 digit mungkin mengacu pada penomoran diagnosis kesehatan yang dapat memiliki panjang antara 3 hingga 5 digit dalam beberapa sistem atau standar kode	FKP15
18.	Nama_diagnosis_berasal_dari_kode_diagnosis_FKP15	Merupakan Kode dan nama diagnosis	FKP15A

Format Dataset peserta BPJS Kesehatan		Keterangan	Variable
		(beragam 3-5 digit)	
19.	Provinsi_faskes_tujuan_rujukan	Merupakan provinsi fasilitas kesehatan tujuan rujukan yang terdiri dari provinsi yang ada di Indonesia.	FKP16
20.	Kabupaten_kota_faskes_tujuan_rujukan	Merupakan kabupaten/kota fasilitas kesehatan tujuan rujukan yang terdiri dari kabupaten/kota yang ada di Indonesia.	FKP17
21.	Kepemilikan_faskes_tujuan_rujukan	Merupakan kepemilikan fasilitas kesehatan tujuan rujukan yang terdiri dari pemerintah pusat,	FKP18
22.	Jenis_faskes_tujuan_rujukan	Merupakan jenis fasilitas kesehatan tujuan rujukan yang terdiri dari puskesmas, klinik, dokter umum, tidak dirujuk dan <i>missing</i> .	FKP19
23.	Tipe_faskes_tujuan_rujukan	Merupakan tipe fasilitas Kesehatan tujuan rujukan peserta BPJS yang terdiri dari apotik kronis, dokter gigi, khusus bedah, dll.	FKP20
24.	Poli_faskes_tujuan_rujukan	Merupakan poli fasilitas kesehatan tujuan rujukan yang terdiri dari akupuntur medik,	FKP21

Format Dataset peserta BPJS Kesehatan		Keterangan	Variable
		anak, gigi bedah mulut, bedah saraf, dll.	
25.	Jenis_kunjungan_FKTP	Merupakan jenis kunjungan fasilitas kesehatan tingkat pertama yang terdiri dari kunjungan sakit dan kunjungan sehat.	FKP22
26.	Kelas_status_pulang_peserta	Merupakan kelas status pulang peserta BPJS yang terdiri dari sehat dan belum sehat.	FKP13Class

2. Data sample

Tabel 3. 2 Data Sample

157978748	50765420	20.0649833679199	958073632	Sehat
95085112.0	26931775	2.62630701065063	449590620P000054	Belum_Sehat
224470578.0	0,227368233	1.05183005332947	467490619P000001	Belum_Sehat
63871289.0	62313678,364	364.741455078125	254321219Y002368	Belum_Sehat
29915626.0	1.05052304267883	88681119P000020	2019-11-04,2019-11-04,35	Belum_Sehat

3.1.2 Exploratory data analysis (EDA)

Data yang telah dikumpulkan akan dianalisis terlebih dahulu untuk mengetahui karakteristik data yang memiliki nilai yang duplikat, null, jumlah data dan tipe data yang tidak konsisten. [17] fungsi lain EDA adalah mengenali kesalahan dataset dengan menguasai pola suatu data dan menemukan hubungan antar variabel.

3.1.3 Data pre-processing

Sebelum melakukan proses *data mining* yang perlu dilakukan adalah menghilangkan data yang tidak diperlukan seperti menangani *missing value*, *noise data*, serta menangani inkonsistensi data dan ketidak relevan data pada label prediksi. *Pre-processing* diterapkan untuk menambah isi atribut yang hilang atau kosong dan merubah data yang tidak konsisten. [7]

3.1.4 Modelling

Tahap selanjutnya adalah *modelling* menggunakan *decision tree* (pohon keputusan), *decision tree* merupakan salah satu algoritma supervised learning yang melakukan prediksi menggunakan struktur pohon. Komponen utama pada *decision tree* adalah *root node* yang merupakan titik awal, *internal node* atau biasa disebut cabang penghubung suatu pengujian, dan *leaf node* yaitu titik akhir pengujian. Dalam melakukan prediksi *decision tree* melakukan perhitungan dengan mencari *impurity measure* atau pengukuran ketidakmurnian. Berikut perhitungan matematika ketidakmurnian dapat dilihat pada persamaan (1) dan (2). [18]

a. Gini Impurity

$$Gini = 1 - \sum_i^n P_i^2 \quad (1)$$

Keterangan:

n = jumlah dari masing – masing atribut

Pi = jumlah atribut dari masing – masing kelas atau labelnya.

b. Average Gini Impurity

$$AG = \sum \frac{\text{Data point } i}{\text{Jumlah total data point}} \times G_i \quad (2)$$

Gini Impurity melakukan pemisahan optimal simpul akar dan simpul berikutnya yang artinya ukuran seberapa sering elemen yang dipilih secara acak dari kumpulan suatu data. Perhitungan dalam pemilihan atribut sebagai akar yaitu dengan menghitung selisih dari Gini Impurity dan

Average Gini Impurity dalam Decision Tree yang dapat dilihat pada persamaan (3).

$$IG = Gi - AG \quad (3)$$

3.1.5 Evaluation

Pada tahap ini, mengevaluasi data-data yang sudah memenuhi target yang diinginkan sehingga dapat dilakukan evaluasi terhadap performa algoritma c4.5 dengan melakukan klasifikasi data. Setelah hasil klasifikasi didapatkan selanjutnya hasil dilakukan perhitungan menggunakan cross validation (Confusion Matrix) melibatkan perhitungan nilai akurasi, recall dan presisi dari data tersebut. [19]