

BAB III METODE PENELITIAN

3.1. Objek Penelitian

Pada objek penelitian dalam penelitian ini membahas penderita penyakit diabetes, memiliki *record* data 768 data dengan 8 atribut dan 1 sebagai label, data ini diambil dari <https://www.kaggle.com/jamaltariqcheema/pima-indians-diabetes-dataset>. yang di akses pada hari sabtu tanggal 26 Desember 2020, jam 01:25 WIB.

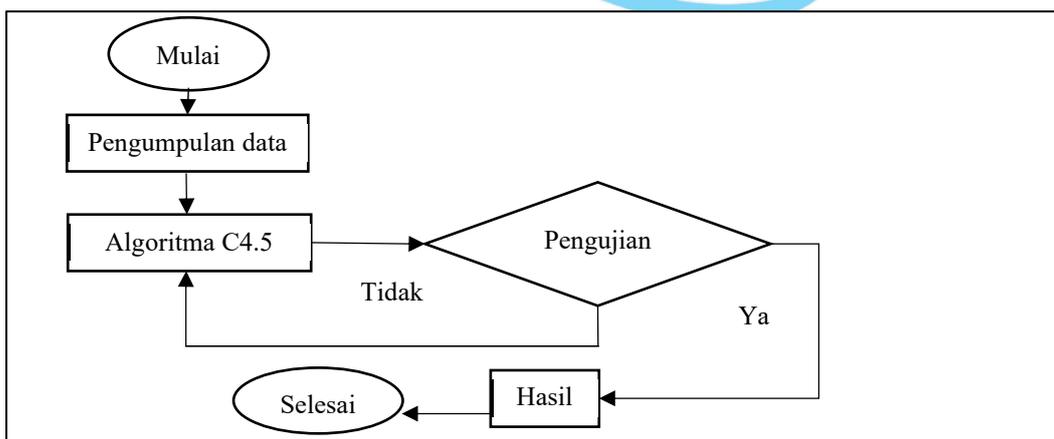
3.2. Peralatan Penelitian

Ada beberapa peralatan yang digunakan untuk penelitian, yaitu:

- a. Perangkat keras
 - 1) Laptop
 - 2) *Flashdisk* 8 GB
 - 3) *Smartphone*
- b. Perangkat lunak
 - 1) Sistem operasi *windows* 10
 - 2) *Microsoft office word*
 - 3) *Microsoft office excel*
 - 4) IDLE python
 - 5) Google chrome.

3.3. Prosedur Penelitian

Berikut adalah tahapan percobaan dari prosedur penelitian ini:



Gambar 3.1 Alur Prosedur Percobaan Percobaan

3.3.1 Pengumpulan Data

Tahapan yang dilakukan dalam pengumpulan data melalui salah satu *online repository* yang diakses pada tanggal 26 Desember 2020 dan mendapatkan dataset tentang penyakit diabetes di situs <https://www.kaggle.com/jamaltariqcheema/pima-indians-diabetes-dataset> yang terdiri dari 768 *records* dan memiliki 8 atribut, diantaranya *Pregnancies* artinya berapa kali kehamilan, *Glucose* (Glukosa), *Blood Pressure* (Tekanan Darah Diastolik), *Skin Thickness* (Ketebalan Kulit), *Insulin*, *Body Mass Index* (BMI) adalah proksi heuristik untuk lemak tubuh manusia, *Diabetes Pedigree Function* adalah informasi tentang riwayat diabetes pada kerabat dan hubungan genetik kerabat tersebut dengan pasien, *Age* atau umur, *Outcome* atau hasil variable target jika 0 menunjukkan pasien bukan diabetes dan jika 1 pasien menunjukkan diabetes. Berikut data set penderita penyakit diabetes yaitu:

Tabel 3.1 Dataset Penderita Diabetes

No	Pregnen	Glucose	Blood P	Skin T	Insulin	BMI	DPF	Age	Outcome
1	6	148	72	35	169.5	33.6	0.627	50	1
2	1	85	66	29	102.5	26.6	0.351	31	0
3	8	183	64	32	169.5	23.3	0.672	32	1
4	1	89	66	23	94	28.1	0.167	21	0
5	0	137	40	35	168	43.1	2.288	33	1
6	5	116	74	27	102.5	25.6	0.201	30	0
...
...
...
...
768	1	93	70	31	102.5	30.4	0.315	23	0

3.3.2 Pengolahan Data

Pada tahapan pengolahan data dalam penelitian ini menggunakan algoritma C4.5. Nilai yang dihasilkan dalam menggunakan algoritma C4.5 berupa tingkat

akurasi dari algoritma C4.5, dan pohon keputusan dari algoritma C4.5 yang menentukan kelas penderita penyakit diabetes.

Berikut tahapan-tahapan *decision tree* algoritma C4.5 (Ayudhitama & Pujiyanto, 2020):

1. Menyiapkan data latih atau uji.
2. Menghitung *entropy*(S) adalah parameter yang digunakan untuk informasi keberagaman setiap nilai atribut kategori atau kriteria terhadap atribut keputusan dalam sebuah dataset. untuk menentukan *entropy* dengan rumus:

$$Entropy(S) = - \sum_{i=1}^n p_i \times \log p_i$$

3. Menghitung nilai *gain* (S, A) adalah untuk mengukur efektivitas masing masing atribut pada *node* tertentu untuk mengklasifikasikan data, nilai terbesar akan menjadi akar pohon utama dengan rumus:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

4. Mengulangi Langkah ke 2 sampai semua *record* data terpartisi.
5. Partisi akan berhenti apabila:
 - a) Semua *record* yang ada pada simpul N mendapat kelas yang sama.
 - b) Tidak adanya atribut pada *record* yang dipartisi.
 - c) Tidak adanya *record* pada cabang yang kosong.

3.3.3 Pengujian

Pada tahapan pengujian untuk menentukan tingkat akurasi dan pohon keputusan algoritma C4.5. Dalam pengujian penelitian ini dilakukan beberapa proses yaitu:

1. Perhitungan secara manual untuk menghasilkan pohon keputusan.
2. Pengujian menggunakan *tool* RapidMiner untuk menghasilkan akurasi.
3. Implementasi algoritma C4.5 pada sistem menggunakan bahasa python.