

BAB III METODE PENELITIAN

3.1. Objek Penelitian

Objek penelitian yang akan digunakan dalam penelitian Tugas Akhir ini adalah data bantuan sosial tunai (BST) yang berasal dari desa Tanahbaru pada tahun 2022. Lokasi penelitian ini dilakukan di desa Tanahbaru, Kecamatan Pakisjaya, Kabupaten Karawang. Waktu penelitian pada bulan Desember 2021 sampai dengan bulan Juli 2022.

Tabel 3. 1 Waktu penelitian

Kegiatan	Des	Jan	Feb	Mar	Apr	Mei	Jun	Jul
Pengumpulan Data								
Analisis Data								
Pengolahan Data								
Pengujian								
Evaluasi								
Tugas Akhir								

3.1.1. Peralatan Penelitian

- a. *Hardware* yang digunakan dalam penelitian ini menggunakan laptop dengan spesifikasi :
 1. *Processor* : Intel (R) Celeron(R) 2957U @1.40GHz
 2. *Storage* : SSD 128 Giga Byte dan Hardisk 500 Giga Byte
 3. *Memory* : Random Access Memory (RAM) 4 Giga Byte
- b. *Software* yang digunakan dalam penelitian ini sebagai berikut :
 1. Windows 10

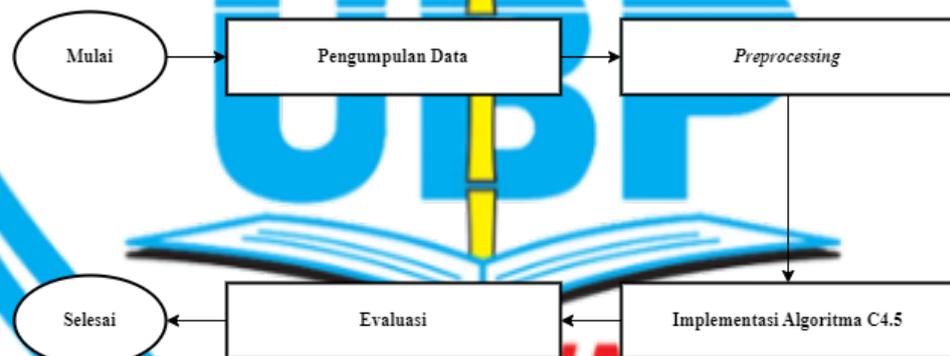
2. Microsoft Word 2016
3. Microsoft Excel 2016
4. *Google Colaburatory*

3.1.2. Bahan Penelitian

Bahan penelitian ini diambil dari data bantuan sosial tunai (BST) di desa Tanahbaru tahun 2022. Desa Tanahbaru merupakan salah satu desa yang berada di Kecamatan Pakisjaya Kabupaten Karawang Provinsi Jawa Barat.

3.2. Prosedur Penelitian

Tahapan penelitian Tugas Akhir yang akan dilakukan dapat digambarkan sebagai berikut :



Gambar 3. 1 Alur prosedur penelitian

3.2.1. Pengumpulan data

Pada tahapan ini dilakukan pengumpulan data dengan dua cara yaitu :

1. Wawancara

Kegiatan ini dilakukan dengan melakukan wawancara dengan pihak pendamping desa dan ketua RT desa untuk memperoleh data penerima bantuan sosial tunai. Data yang diperoleh berjumlah 421 data yang diperoleh dari pihak staff kantor desa dan masing-masing pihak ketua RT.

2. Observasi

Kegiatan ini dilakukan dengan cara mengamati langsung keadaan masyarakat penerima bantuan sosial tunai.

3.2.2. *Preprocessing*

Di dalam tahapan *preprocessing data* yaitu melakukan seleksi atribut yang akan digunakan untuk memilih atribut terbaik. Sehingga mempermudah proses analisis data dalam *machine learning*. Setelah tahap ini selesai maka akan menghasilkan suatu *dataset* yang merupakan sebuah kumpulan data yang diambil dari informasi pada masa lalu dengan dikelola menjadi sebuah informasi untuk melakukan teknik pada ilmu *data mining* (Ginting et al., 2020). Tahapan *preprocessing* dilakukan dengan beberapa tahapan sebagai berikut.

a. *Data Cleaning*

Data Cleaning memiliki tujuan untuk menghapus data duplikat, memeriksa data kosong, dan memperbaiki kesalahan penulisan pada data.

b. *Data Selection*

Data selection merupakan pemilihan atribut dari kumpulan data yang dapat digunakan dalam proses mining.

c. *Data Transformation*

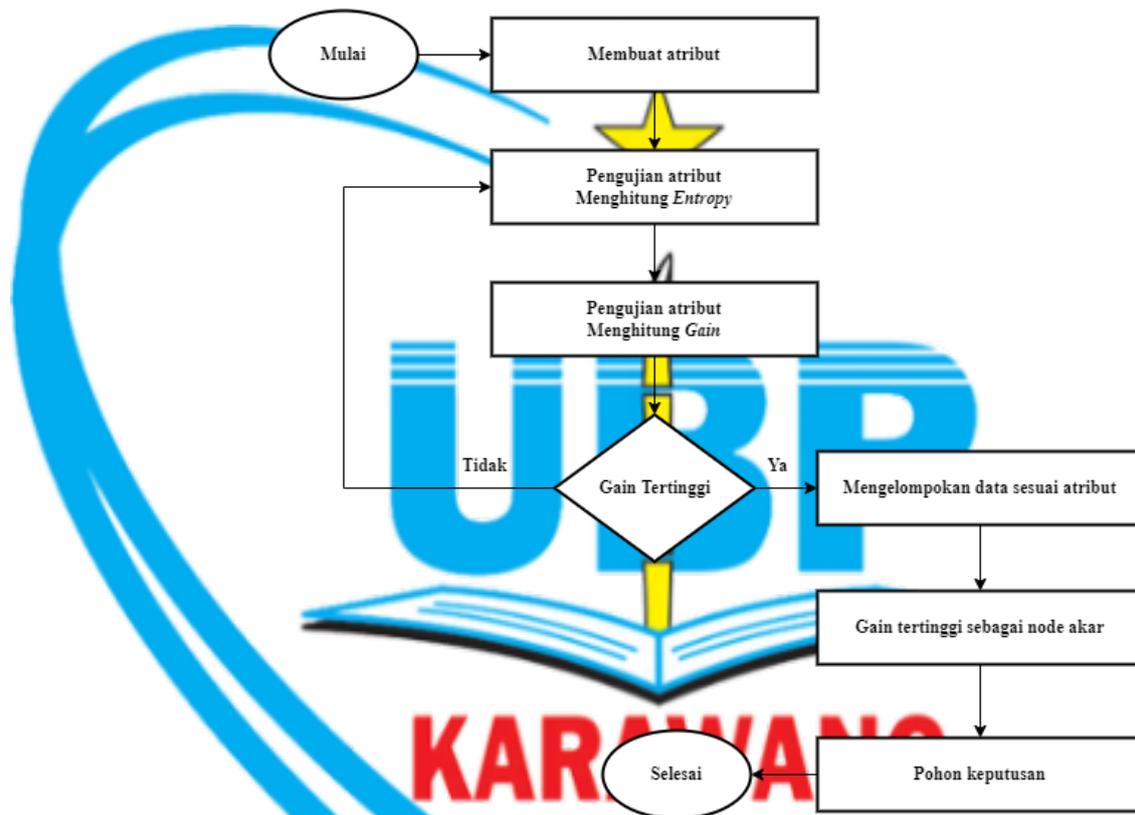
Data Transformation merupakan tahapan mengubah data dengan tujuan untuk mengubah atribut atau variabel sesuai dengan format sehingga bisa diproses dalam program.

3.2.3. Implementasi Algoritma C4.5

Menurut (Marlina et al., 2016) Algoritma C4.5 adalah kelompok algoritma pohon keputusan. Algoritma ini mempunyai beberapa masukan diantaranya sampel pelatihan dan sampel. Sampel pelatihan adalah data sampel yang digunakan untuk mendirikan pohon yang telah dibuktikan kebenarannya. Sedangkan

sampel adalah data lapangan yang akan digunakan untuk parameter dalam klasifikasi data. Algoritma C4.5 merupakan hasil tahap pengembangan dari algoritma ID3.

Pada tahapan ini melakukan perhitungan algoritma C4.5 dengan tahapan-tahapan sebagai berikut.



Gambar 3. 2 Tahapan algoritma c4.5

Berikut ini keterangan pada alur algoritma C4.5 diantaranya sebagai berikut :

1. Membuat atribut terlebih dahulu.
2. Melakukan pengujian dengan menghitung *entropy* dengan persamaan berikut :

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i$$

Keterangan:

S = jumlah kasus

n = total himpunan S

p_i = proporsi ke- i

- Melakukan pengujian dengan menghitung *gain* dengan persamaan berikut :

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

Keterangan:

S : Jumlah kasus

A : Atribut

n : Total partisi atribut A

| S_i |: Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

- Mencari nilai *gain* tertinggi.
- Jika nilai *gain* tertinggi tidak ditemukan maka kembali melakukan pengujian dengan menghitung entropy dan gain.
- Jika nilai *gain* tertinggi ditemukan maka dilakukan pengelompokan data sesuai atribut.
- Kemudian *gain* tertinggi dijadikan sebagai *node* akar atau *root* sampai perhitungan atribut dari nilai *gain* tertinggi tersebut semua cabang telah dikelompokan.
- Setelah semua cabang telah dikelompokan, tahapan selanjutnya adalah pembuatan pohon keputusan.

Pengujian ini memiliki tujuan yaitu untuk mengetahui hasil *accuracy*, *precision*, *recall* dari hasil prediksi klasifikasi.

Pengujian ini dilakukan dengan dua tahapan diantaranya yaitu :

- Melakukan perhitungan manual menggunakan Microsoft Excel.
- Mengimplementasikan algoritma C4.5 pada sistem dengan menggunakan bahasa pemrograman python.

3.2.4. Evaluasi

Setelah melakukan implementasi Algoritma C4.5 maka akan dilakukan evaluasi dengan Confusion Matrix untuk mengetahui tingkat *accuracy*, *precision*, *recall*. *Confusion Matrix* adalah metode

yang dapat digunakan di dalam evaluasi model klasifikasi untuk memperkirakan suatu objek yang benar dan yang salah, hasil prediksi dibandingkan dengan kelas yang asli (Ermawati, 2019).

Tabel 3. 2 *Confusion Matrix*

<i>Classification</i>	<i>Predicted Class</i>	
	<i>Class = Yes</i>	<i>Class = No</i>
<i>Class = Yes</i>	<i>a (true positive - TP)</i>	<i>b (false negative - FN)</i>
<i>Class = No</i>	<i>c (false positive - FP)</i>	<i>d (true negative - TN)</i>

Confusion Matrix ini akan melakukan perhitungan yang menghasilkan keluaran sebagai berikut :

- a. *Recall* merupakan total data positif yang diidentifikasi benar.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Recall dilakukan perhitungan dengan cara pembagian data yang bernilai benar *true positive* (TP) dengan hasil penjumlahan dari data yang bernilai benar *true positive* (TP) dan data salah yang bernilai *False Negative* (FN).

- b. *Precision* merupakan perbandingan kasus dengan hasil positif benar.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Precision dilakukan perhitungan dengan cara pembagian data yang bernilai benar *true positive* (TP) dengan hasil penjumlahan dari data yang bernilai benar *true positive* (TP) dan salah yang bernilai *False Positive* (FP).

- c. *Accuracy* merupakan perbandingan kasus yang teridentifikasi benar dengan total seluruh kasus.

$$Accuracy = \frac{Jumlah\ Prediksi\ Benar}{Jumlah\ Total\ Prediksi}$$

Accuracy dilakukan perhitungan dengan cara pembagian data yang diklasifikasi secara benar dengan total sampel *data testing* yang diuji.

