

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1 Metodologi Penelitian**

##### **3.1.1 Waktu Dan Tempat Penelitian**

Penelitian ini dilakukan dengan meneliti data yang didapatkan dari Data Terpadu Kesejahteraan Sosial (DTKS) Kota Tasikmalaya, yakni data penerima bantuan sosial pada tahun 2020. Penelitian ini akan mengklasifikasikan masyarakat yang berhak menjadi ‘Penerima’ ataupun yang ‘Bukan Penerima’ bantuan.

##### **3.1.2 Objek Penelitian**

Objek penelitian yang diteliti adalah Algoritma *K-Nearest Neighbor* dan *Adaptive Boosting*.

##### **3.1.3 Variabel dan Definisi Operasional**

Variabel yang digunakan untuk menguji kinerja dari algoritma *K-Nearest Neighbor* dan *Adaptive Boosting* untuk prediksi penerima bantuan sosial pangan non tunai pada penelitian ini meliputi: status kepemilikan bangunan tempat tinggal, kondisi rumah seperti kondisi atap, kondisi dinding, dan kondisi lantai, penggunaan jamban, sumber air yang digunakan, dan sebagainya.

Pada data kuantitatif terdapat beberapa jenis variabel yang dapat diolah. Sebelum mengolah data, sebaiknya mengenali jenis-jenis variabel yang akan

digunakan. Hal ini bertujuan agar mempermudah pemilihan algoritma yang tepat.

Berikut ini beberapa jenis variabel diantaranya (Sugianto, 2016):

a. Variabel Diskrit

Ini adalah variabel dengan kategori atau bentuk data yang dapat diklasifikasikan ke dalam kategori tertentu. Kategori atau kelas yang terbentuk memiliki nilai yang sama. Misalnya atribut dengan tipe variabel diskrit adalah jenis kelamin. Data jenis kelamin dibagi menjadi dua kategori, yaitu laki-laki dan perempuan. Kedua kategori ini memiliki nilai data yang setara.

b. Variabel Kontinyu

Variabel kontinu adalah data dalam bentuk angka, yang dapat berupa desimal atau bilangan bulat. Dengan cara ini variabel kontinu dapat digunakan untuk operasi aritmatika. Contoh atribut dengan jenis variabel kontinu adalah usia. Karena umur dapat digunakan dalam proses menghitung operasi (seperti menghitung rata-rata).

b. Variabel Ordinal

Variabel ordinal adalah data yang berupa angka-angka dengan tingkatan. Diberikan peringkat, variabel ordinal mungkin tidak sama satu sama lain. Atribut peringkat kelas adalah contoh variabel ordinal. Data peringkat level 1 berbeda dengan data peringkat level 2, dan peringkat kedua angka tersebut tidak sama.

c. Variabel Interval

Jika jarak antar data numerik sudah jelas maka data numerik tersebut juga dapat dikatakan sebagai variabel interval. Contoh atribut dengan jenis variabel interval adalah rentang gaji atau rentang usia. Misalkan ada data kisaran gaji, nilainya 1 juta-3 juta, 5 juta-7 juta, dan seterusnya. Data tersebut memiliki jangkauan dan jarak yang jelas 2 juta.

d. Variabel Rasio

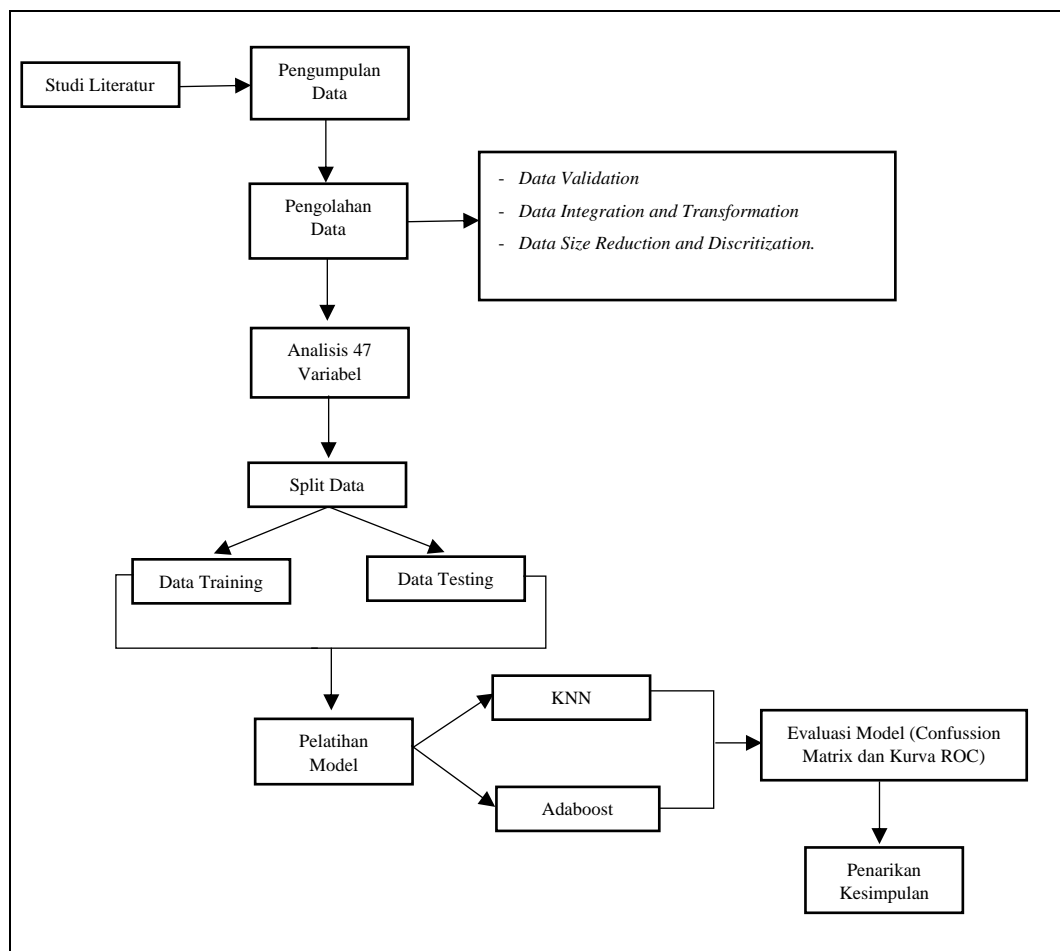
Variabel rasio adalah data numerik yang telah diproses melalui operasi aritmatika yang kompleks. Angka-angka dalam variabel rasio bukanlah tanda atau kategori, tetapi angka sebenarnya. Contoh variabel skala adalah data ketinggian. Dengan asumsi data tinggi A adalah 100 cm, dan data tinggi B adalah 200 cm, maka skala tinggi A adalah setengah dari tinggi B.

### 3.2 Tahapan Penelitian

Proses penelitian dipresentasikan ke dalam bentuk diagram alur. Metode yang digunakan adalah metode kuantitatif yang disesuaikan. Diagram alur penelitian dibuat agar mempermudah penyampaian informasi langkah-langkah yang hendak dilakukan.

Orisinalitas yang ditargetkan pada penelitian yang diusulkan ini adalah membandingkan 2 metode klasifikasi yaitu *K-Nearest Neighbor* dan *Adaptive Boosting* dalam memprediksi penerima bantuan sosial. Hal ini direalisasikan dalam bentuk eksperimen serta pengamatan yang dilakukan untuk membuktikan algoritma mana yang menghasilkan akurasi tertinggi.

Pada Gambar 3.2 terdapat beberapa tahapan penelitian secara keseluruhan. Tahapan penelitian yang dilakukan dimulai dari studi literatur, proses pengumpulan dan analisis data, pengolahan data, Pemodelan algoritma *k-nearest neighbor* dan *adaboost*, Evaluasi Model, sampai proses penarikan kesimpulan.



Gambar 3. 1 Tahapan Penelitian

### 3.2.1 Studi Literatur

Pengumpulan data-data serta sumber yang berhubungan dengan penelitian ini. Studi literatur didapatkan dari jurnal dan *e-proceeding* terkait metode dan algoritma yang digunakan, selain itu diperoleh dari buku dan internet serta dokumen yang berkaitan dengan objek penelitian.

### 3.2.2 Pengumpulan dan Analisis Data

Data yang dibutuhkan untuk penelitian ini adalah data penerima bantuan sosial tahun 2020 yang didapatkan dari DTKS Kota Tasikmalaya yang dapat dijadikan sampel dalam membandingkan kinerja 2 Algoritma klasifikasi terbaik yaitu algoritma *K-Nearest Neighbor* dan *Adaptive Boosting (Adaboost)* untuk melihat algoritma mana yang menghasilkan akurasi yang lebih tinggi. Data penerima bantuan sosial tersebut diambil berdasarkan peraturan yang dibuat oleh Data Terpadu Kesejahteraan Sosial Kota Tasikmalaya

### 3.2.3 Pengolahan Data

Agar data valid dan berkualitas, maka perlu dilakukan beberapa tahap pengolahan data, yaitu: *Data Validation*, *Data Integration and Transformation* dan *Data Size Reduction and Discretization*.

1. *Data Validation*

*Data Validation* merupakan tahapan pengolahan data untuk mengidentifikasi serta menghapus data *outlier/noise*, data yang tidak lengkap (*missing value*) dan data yang tidak konsisten. Pada tahap ini dilakukan *cleaning* data atau pembersihan data dengan membuang data yang tidak diperlukan seperti data yang bernilai kosong atau *null*.

2. *Data Integration and Transformation*

Integrasi dan Transformasi data merupakan tahapan untuk meningkatkan akurasi dan efisiensi dari algoritma yang digunakan.

3. *Data Size Reduction and Discretization*

Tahapan *Data Size Reduction and Discretization* digunakan untuk memperoleh dataset dengan jumlah atribut dan *record* yang lebih sedikit tetapi bersifat informatif.

#### **3.2.4 Analisis 47 Atribut**

Tahapan ini yaitu menganalisis semua variabel yang digunakan dan menentukan variabel mana yang termasuk kedalam variabel independent dan dependent. Pada penelitian ini terdapat data sebanyak 62950 data dari dataset penerima bantuan sosial mempunyai 47 variabel yang terdiri dari 46 variabel independent (X) dan 1 variabel dependent (Y).

#### **3.2.5 Pembagian Data Atau Split Data**

Dalam melakukan analisis klasifikasi untuk membangun model *machine learning*, data perlu dibagi terlebih dahulu menjadi dua bagian yaitu *data training* dan *data testing*.

#### **3.2.6 Pelatihan Model Algoritma *K-Nearest Neighbor* dan *Adaptive Boosting***

Pada tahapan ini data akan digunakan untuk pelatihan model dalam membandingkan kinerja 2 Algoritma klasifikasi terbaik yaitu algoritma *K-Nearest Neighbor* dan *Adaptive Boosting (Adaboost)* untuk melihat algoritma mana yang menghasilkan akurasi yang lebih tinggi. Proses ini dilakukan beberapa skenario dalam pembagian data.

#### **3.2.7 Evaluasi Model**

Dari skenario-skenario pembagian data yang ada pada setiap pemodelan 2 algoritma *K-nearest Neighbor* dan *Adaptive Boosting* akan menghasilkan masing-

masing akurasi dari setiap algoritma dan skenario-skenario pembagian data. Setelah itu akan dibandingkan dan dipilih suatu model dan skenario terbaik berdasarkan hasil perhitungan *Confusion Matrix* dan Nilai AUC.

### **3.2.8 Penarikan Kesimpulan**

Tahapan ini merupakan tahapan akhir dari penelitian dimana hasil penelitian tersebut berupa Analisa perbandingan 2 algoritma klasifikasi dalam prediksi penerima bantuan sosial dengan mengimplementasikan model *Machine Learning* menggunakan algoritma *K-Nearest Neighbor* dan *Adaptive Boosting (Adaboost)*.