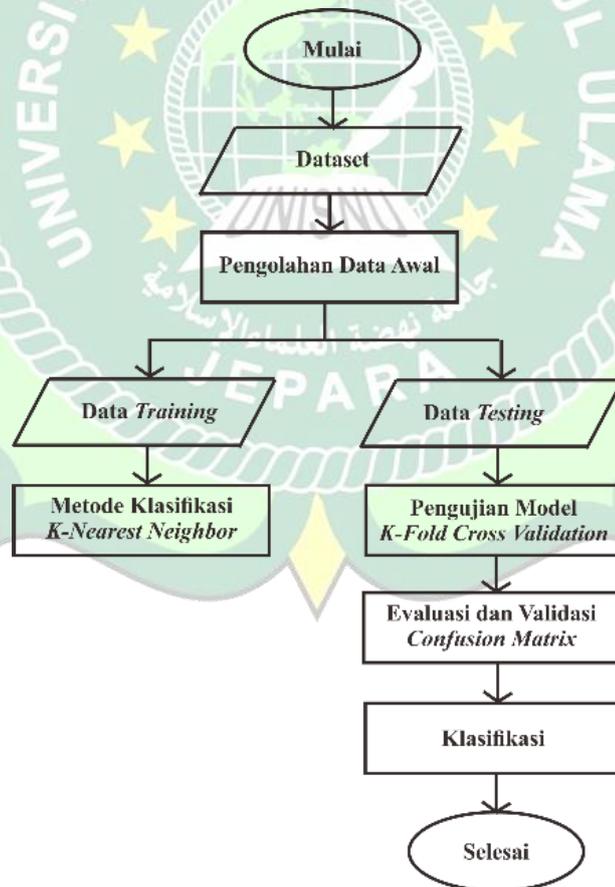


BAB III METODE PENELITIAN

3.1 Desain Penelitian

Desain penelitian yang digunakan peneliti dalam penelitian ini adalah analisis kuantitatif. Metode ini merupakan metode penelitian yang menekankan analisisnya terhadap data yang berupa *numeric* atau angka sebagai alat penunjang tentang apa yang ingin dicapai dalam penelitian. Data yang digunakan adalah data pendaftar yang mengajukan bantuan program rehabilitasi rumah tidak layak huni (RTLH) pada tahun 2019.

Sesuai permasalahan yang ada, pada penelitian ini akan menerapkan *Algoritma K-Nearest Neighbor* untuk memprediksi kelayakan penerima bantuan program rehabilitasi rumah tidak layak huni di Jepara. Adapun langkah-langkah tahapan penelitian diilustrasikan seperti gambar dibawah ini:



Gambar 3.1 *Flowchart* Penelitian

Pada tahapan penelitian membutuhkan dataset yang akan digunakan sebagai data *training* dan data *testing*. Langkah selanjutnya setelah memperoleh dataset yaitu pengelolaan data awal untuk kemudian dilakukan pengujian model menggunakan *K-fold cross validation* pada data *testing* dan Algoritma *K-Nearest Neighbor* (*K-NN*) pada data *training*. Setelah dilakukan pengujian, maka dilakukan klasifikasi terhadap data *testing* yang belum diketahui kategori hasilnya, dan evaluasi serta validasi hasil menggunakan *confusion matrix* sebagai *performance*.

3.2 Pengumpulan Data

Dalam tahapan pengumpulan data, data yang diperoleh adalah data pendaftar yang mengajukan proposal bantuan rehabilitasi rumah tidak layak huni baik yang dinyatakan RTLH maupun Non RTLH pada tahun 2019, yang nantinya digunakan sebagai data *training* dan data *testing*. Adapun teknik pengumpulan data pada penelitian ini adalah sebagai berikut:

1. Observasi

Observasi merupakan metode yang dilakukan untuk mengetahui obyek yang akan diteliti secara langsung guna mendapatkan data atau informasi terkait bantuan program rehabilitasi rumah tidak layak huni (RTLH) di Jepara. Pada tahapan ini, peneliti mendapatkan data sebanyak 2.962 data. Data tersebut digunakan sebagai data *training* dan data *testing* yang diambil dari data pendaftar dan penerima bantuan rehabilitasi RTLH pada tahun 2019, sebanyak 2.932 data sudah diketahui kategori penerimaannya dan 30 data yang belum diketahui masuk kategori mana. Data tersebut memiliki 12 atribut, yaitu nama, kepemilikan rumah dan tanah, jumlah anggota keluarga, ukuran rumah, pondasi, balok, sanitasi, jendela, ventilasi, material atap, material lantai, material dinding dan hasil. Serta 1 *class* target yaitu hasil.

2. Wawancara

Wawancara merupakan metode yang dilakukan dengan cara mengajukan pertanyaan atau tanya jawab kepada bagian tenaga fasilitator lapangan bidang perumahan Dinas Perumahan Rakyat dan Kawasan Permukiman

(DISPERKIM) yang mengetahui masalah penerimaan atau tidaknya bantuan rehabilitasi rumah tidak layak huni (RTLH). Metode ini bertujuan agar data atau informasi yang diperoleh dapat dipertanggungjawabkan.

3. Studi Pustaka

Studi pustaka merupakan usaha yang dilakukan oleh peneliti untuk mengumpulkan informasi yang sesuai dengan topik permasalahan yang sedang diteliti. Informasi ini didapat dari beberapa sumber, seperti jurnal, karya ilmiah, buku, artikel dan lainnya.

3.3 Lokasi Penelitian

Penelitian ini dilaksanakan di Dinas Perumahan Rakyat dan Kawasan Permukiman (DISPERKIM) yang beralamat di Jalan Ki Mangunsarkoro Panggang Kecamatan Jepara Kabupaten Jepara Jawa Tengah. Selain mudah dijangkau, peneliti memilih tempat tersebut dikarenakan belum menemukan metode dalam menentukan penerimaan bantuan rehabilitasi rumah tidak layak huni yang mudah, cepat dan akurat. Mengingat semakin tahun pendaftar bantuan rehabilitasi RTLH semakin meningkat. Pada penelitian ini menggunakan data tahun 2019 sejumlah 2.932 data RTLH dan Non RTLH.

3.4 Pengolahan Data Awal

Pengolahan data awal diperlukan agar data yang akan diolah memiliki kualitas yang baik. Hal ini dilakukan karena tidak semua data dan atribut dapat digunakan. Berikut adalah beberapa teknik pengelolaan data yang dilakukan dalam penelitian ini:

1. Seleksi Atribut

Tahapan ini dilakukan untuk mengambil atribut tertentu dari semua atribut pada data awal. Pada penelitian ini variable independen yang digunakan yaitu nama, kepemilikan rumah dan tanah, jumlah anggota keluarga, ukuran rumah, pondasi, balok, sanitasi, jendela, ventilasi, material atap, material lantai, dan material dinding. Sedangkan variable dependen yang digunakan adalah hasil dengan *output* RTLH atau Non RTLH. Dalam *dataset* terdapat atribut yang

dihapus yaitu NIK, kecamatan, desa, RT, RW, fasilitator, kondisi atap, kondisi lantai dan kondisi dinding. Hal ini dilakukan karena data tersebut tidak akan dipakai dalam perhitungan.

2. Menyusun dan Mengkategorikan Data

Pada tahapan ini atribut yang berupa kategorikal dan nominal dikonversi menjadi data *numeric*. Hal ini dilakukan karena perhitungan jarak menggunakan rumus *Euclidean Distance*. Data yang diubah terdapat pada atribut kepemilikan rumah dan tanah, ukuran rumah, pondasi, balok, sanitasi, jendela, ventilasi, material atap, material lantai, dan material dinding. Sedangkan atribut nama, jumlah anggota keluarga dan hasil masih tetap sama seperti pada data awal. Berikut adalah kriteria variabel yang digunakan untuk konversi data:

Tabel 3.1 Kriteria Variabel

No	Atribut	Keterangan	Nilai Variabel
1	Kepemilikan Rumah dan Tanah	Milik Sendiri	1
		Milik Saudara	2
		Bengkok	3
		Tanah Irigasi	4
		Sewa	5
2	Ukuran Rumah	>100m ²	1
		71m ² - 100m ²	2
		41m ² - 70m ²	3
		21m ² - 40m ²	4
		9m ² - 20m ²	5
3	Pondasi	Layak	1
		Ada	2
		Tidak Layak	3
		Tidak Ada	4
4	Balok	Layak	1
		Ada	2

		Tidak Layak	3
		Tidak Ada	4
5	Sanitasi	Layak	1
		Ada	2
		Tidak Layak	3
		Tidak Ada	4
6	Jendela	Layak	1
		Ada	2
		Tidak Layak	3
		Tidak Ada	4
7	Ventilasi	Layak	1
		Ada	2
		Tidak Layak	3
		Tidak Ada	4
8	Material Atap	Beton	1
		Genteng	2
		Asbes	3
		Seng	4
		Rumbia	5
9	Material Lantai	Keramik	1
		Ubin	2
		Plester	3
		Panggung Kayu	4
		Tanah	5
10	Material Dinding	Bata Plester	1
		Bata Expose	2
		Herbel	3
		Kalsiboard	4
		Kayu	5
		Asbes	6
		Triplek	7

		Bambu	8
11	Hasil	RTLH	
		Non RTLH	

3. Pembersihan Data

Sebelum data diproses, data perlu dilakukan pembersihan data untuk mengidentifikasi dan menghilangkan data yang kosong (*missing value*), data yang sama dan data *outlier*. Hal ini dilakukan untuk mencegah terjadinya *error* pada tahapan data mining ketika dijalankan.

4. Normalisasi Data

Normalisasi data dilakukan untuk memperkecil jarak atau *range* pada data. Penelitian ini menggunakan *min-max normalization* yang merupakan metode normalisasi dengan tranformasi terhadap data *training* sehingga dapat menghasilkan keseimbangan nilai perbandingan antara data sebelum dan data sesudah dinormalisasi [29]. Berikut ini adalah rumus normalisasi data *min-max normalization*

$$V' = \frac{V - \min A}{\max A - \min A} (\text{new max} A - \text{new min} A) + \text{new min} A \dots \dots \dots (5)$$

Dimana:

V' = Nilai dari data baru hasil dari normalisasi

V = Nilai dari data sebelum dinormalisasi

$\text{New max} A$ = Batas nilai maksimum terbaru

$\text{New min} A$ = Batas nilai minimum terbaru

$\text{Max} A$ = Nilai maksimum pada kolom

$\text{Min} A$ = Nilai minimum pada kolom

3.5 Metode yang Diusulkan

Adapun metode yang diusulkan pada penelitian ini adalah metode klasifikasi data mining *K-Nearest Neighbor*. Data diolah menggunakan tool RapidMiner studio versi 9.8. dan akan menghasilkan suatu model. Berikut adalah tahapan pengolahan data menggunakan algoritma *K-Nearest Neighbor* pada RapidMiner:

1. Menentukan parameter k untuk menentukan jumlah tetangga terdekat.
2. Menghitung jarak *Euclidean* antara data *training* dan data *testing*.
3. Mengurutkan jarak hasil perhitungan no 2 dari yang mempunyai jarak *Euclidean* yang terkecil.
4. Mengumpulkan klasifikasi berdasarkan nilai k .
5. Mencari jumlah kelas dari tetangga yang terdekat dan menetapkan kelas tersebut sebagai kelas data yang akan diprediksi berdasarkan hasil kelas yang paling mayoritas [18].

Algoritma klasifikasi yang peneliti gunakan yaitu algoritma *K-Nearest Neighbor* (K -NN). Alasan digunakannya algoritma *K-Nearest Neighbor* maka dapat dilakukan sebuah prediksi berdasarkan jarak terdekat dari data lama (*training*) dengan data baru (*testing*). Sebelumnya dilakukan pengujian data *training* dan data *testing* terhadap metode algoritma klasifikasi yang lain, diantaranya algoritma *Neural Network* mendapat akurasi sebesar 97,28% dan algoritma *Naïve Bayes* mendapat akurasi sebesar 95,26%. Jika dibandingkan maka nilai akurasi tertinggi ketika menggunakan algoritma *K-Nearest Neighbor* sebesar 97,98%. Algoritma *K-Nearest Neighbor* ini salah satu algoritma yang tidak rumit dalam memecahkan masalah klasifikasi dan juga sering menghasilkan nilai akurasi yang tinggi. Hasil dari klasifikasi dalam penelitian ini juga untuk mengetahui algoritma dalam memprediksi berdasarkan nilai *accuracy*, *recall*, dan *precision*. Peneliti menggunakan data *numeric* berdasarkan kategori tiap-tiap parameter.

3.6 Pengujian Model

Data diolah menggunakan algoritma *K-Nearest Neighbor* dan menghasilkan model yang nantinya dilakukan pengujian. Pengujian model berfungsi untuk memperkirakan estimasi akurasi. Pengujian pada penelitian ini menggunakan *K-fold cross validation*, yang akan dilakukan percobaan sebanyak k . Nilai k yang digunakan yaitu 10, atau disebut *10-fold cross validation* untuk kemudian diambil nilai k terbaik dan diaplikasikan ke dalam metode yang diusulkan.

3.7 Evaluasi dan Validasi Hasil

Setelah data diolah menggunakan algoritma *K-Nearest Neighbor* melalui tool RapidMiner dan dilakukan pengujian model menggunakan *10-fold cross validation*, selanjutnya dilakukan evaluasi dan validasi hasil menggunakan *confusion matrix* sebagai *performance* yang menghasilkan nilai *Accuracy*, *Precision*, *Recall*, dan *AUC* (Area Under Curve). Penelitian ini diharapkan mampu memberikan hasil dengan nilai akurasi tinggi sehingga dapat meminimalisir salah sasaran dalam menentukan hasil.

