

BAB III METODE PENELITIAN

3.1 Objek Penelitian

Objek pada penelitian ini ialah *dataset* berupa Data Terpadu Kesejahteraan Sosial (DTKS) yang berasal dari Desa Wadas. Lokasi penelitian ini dilakukan di desa Wadas, Kecamatan Telukjambe Timur, Kabupaten Karawang. Waktu penelitian pada bulan Oktober 2022 hingga Juli 2023.

Tabel 3.1 Lokasi dan Waktu penelitian

NO	KEGIATAN	Okt 2022	Nov 2022	Des 2022	Jan 2023	Feb 2023	Mar 2023	Apr 2023	Mei 2023	Jun 2023	Jul 2023
1	Pengumpulan Data	■	■								
2	Analisis Data			■							
3	Seleksi Data				■	■					
4	Implementasi						■	■	■		
5	Validasi dan Analisis Hasil									■	■

3.2 Peralatan Penelitian

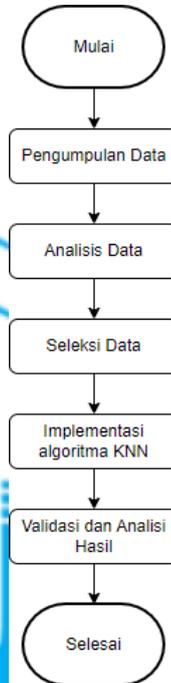
Penelitian yang akan dilakukan menggunakan peralatan yang terdiri dari perangkat keras dan perangkat lunak.

1. Perangkat keras yang digunakan:
 - a. Laptop *Intel(R) Core (TM) i5-10300H CPU @ 2.50GHz* 2.50 GHz
 - b. RAM 8GB
 - c. SSD 512GB
2. Perangkat lunak yang digunakan:
 - a. *64-bit operating system, x64-based processor*
 - b. *Microsoft Word 2019*
 - c. *Microsoft Excel 2019*
 - d. *Google Colaboratory*

3.3 Prosedur Penelitian

Dalam penelitian ini memiliki tahapan-tahapan yang harus dipahami dan dilakukan. Terdapat pada gambar 3.1 Alur penelitian.

1. Pengumpulan Data



Gambar 3.1 Prosedur Penelitian

Pengumpulan data merupakan metode untuk memperoleh data maupun informasi yang berkaitan dengan penelitian. Data yang akan digunakan dalam penelitian ini yaitu *dataset* yang didapat dari hasil observasi dengan aparat desa untuk dapat digunakan pada proses klasifikasi. *Dataset* yang diperoleh dari Desa Wadas Kec. Telukjambe Timur, berupa Data Terpadu Kesejahteraan Sosial (DTKS) dalam format *ms.excel* berjumlah 103 data dan 12 variabel yang dapat dilihat pada gambar 3.2. Maka dari itu, dapat mendukung untuk proses klasifikasi data terpadu kesejahteraan masyarakat.

```

Index(['Nama ', 'Jumlah ART', 'SKDPU', 'ITD', 'SPBTT', 'LL', 'JLT ', 'JDT',
      'JAT', 'DLT', 'BBEM', 'Keterangan'],
      dtype='object')
  
```

Gambar 3.2 Atribut pada data

2. Analisis Data

Analisis data bertujuan untuk memahami data agar dapat dikelola dan diseleksi untuk dapat di analisis dalam menentukan variabel data pada pengklasifikasi, data tersebut dapat dikategorikan menjadi dua yaitu data latih dan data uji, dengan

perbandingan data yang dipakai pada data latih sebesar 80% dan data uji sebesar 20%. Oleh karena itu, dapat memudahkan dalam menentukan kesalahan dalam mengakurasi algoritma KNN terhadap klasifikasi pendataan terpadu kesejahteraan masyarakat.

3. Seleksi Data

Seleksi data yaitu untuk membersihkan atau menghilangkan data dari atribut yang tidak diperlukan. Seleksi data dijalankan untuk mengurangi kesalahan akurasi dalam perhitungan algoritma *K-Nearest Neighbor* (KNN) terhadap klasifikasi kategori kesejahteraan masyarakat.



	Nama	Jumlah ART	SKDPU	ITD	SPBTT	LL	JLT	JDT	JAT	DLT	BBEH	Keterangan
0	Untbox S	5.0	Pekerja Keluarga/tidak dibayar	SMA/ sederajat	Milik Sendiri	60.0	Ubin	Tembok	Asbes	900 watt	Gas > 3 kg	Tidak Mampu
1	Saxx	3.0	Berusaha Sendiri	Tidak Punya Ijazah	Milik Sendiri	50.0	Tanah	Lainnya	Genteng Tanah Liat	450 watt	Gas 3 kg	Tidak Mampu
2	Subxx L	4.0	Berusaha Sendiri	Tidak Punya Ijazah	Milik Sendiri	45.0	Sementara/bata merah	Bambu	Genteng Tanah Liat	450 watt	Gas 3 kg	Tidak Mampu
3	Taxx	2.0	Berusaha Sendiri	SD/ sederajat	Milik Sendiri	28.0	Keramik	Tembok	Genteng Tanah Liat	450 watt	Gas 3 kg	Tidak Mampu
4	Nurxxx	6.0	Berusaha Sendiri	Tidak Punya Ijazah	Milik Sendiri	65.0	Ubin	Tembok	Seng	450 watt	Gas > 3 kg	Tidak Mampu
...
112	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Gambar 3.3 Data sebelum di seleksi

Pada gambar 3.3 merupakan Data Terpadu Kesejahteraan Sosial (DTKS) yang di peroleh dari Desa Wadas dengan jumlah sebelum di seleksi yang mempunyai 103 data dengan 12 atribut.

	Jumlah ART	SKDPU	ITD	SPBTT	LL	JLT	JDT	JAT	DLT	BBEH	Keterangan
0	5	Pekerja Keluarga/tidak dibayar	SMA/ sederajat	Milik Sendiri	60	4	Tembok	Asbes	900 watt	Gas > 3 kg	Tidak Mampu
1	3	Berusaha Sendiri	Tidak Punya Ijazah	Milik Sendiri	50	9	Lainnya	Genteng Tanah Liat	450 watt	Gas 3 kg	Tidak Mampu
2	4	Berusaha Sendiri	Tidak Punya Ijazah	Milik Sendiri	45	6	Bambu	Genteng Tanah Liat	450 watt	Gas 3 kg	Tidak Mampu
3	2	Berusaha Sendiri	SD/ sederajat	Milik Sendiri	28	2	Tembok	Genteng Tanah Liat	450 watt	Gas 3 kg	Tidak Mampu
4	6	Berusaha Sendiri	Tidak Punya Ijazah	Milik Sendiri	65	4	Tembok	Seng	450 watt	Gas > 3 kg	Tidak Mampu
...
98	4	Karyawan Swata	SMA/ sederajat	Milik Sendiri	144	2	Tembok	Genteng	1.300 watt	Gas > 3 kg	Mampu
99	3	Berusaha Sendiri	D4/S1	Milik Sendiri	90	2	Tembok	Asbes	>2.200 watt	Gas > 3 kg	Mampu
100	4	Karyawan Swata	SMA/ sederajat	Milik Sendiri	72	2	Tembok	Genteng	1.300 watt	Gas > 3 kg	Mampu
101	4	Karyawan Swata	SMA/ sederajat	Milik Sendiri	72	2	Tembok	Asbes	1.300 watt	Gas 3 Kg	Mampu
102	4	Karyawan Swata	SMA/ sederajat	Milik Sendiri	72	2	Tembok	Asbes	1.300 watt	Tidak masak dirumah	Mampu

103 rows x 11 columns

Gambar 3.4 Data sesudah di seleksi

Oleh sebab itu, pada gambar 3.4 memperoleh hasil seleksi dengan cara mengurangi variabel secara manual, data yang sebelumnya memiliki 103 data dan 12 variabel menjadi 103 data dan 11 variabel yaitu jumlah anggota rumah tangga, status kedudukan dalam pekerjaan utama, ijazah tertinggi yang dimiliki, status penguasaan tempat tinggal yang ditempati, luas lantai, jenis lantai terluas, jenis dinding terluas, jenis atap terluas, daya listik yang terpasang, bahan bakar/energi

utama untuk memasak, dan keterangan. Setelah tahap seleksi data dilakukan, kemudian data ditransformasi dengan tujuan untuk mengubah data dari suatu atribut atau variabel sesuai dengan format yang dapat di proses dalam program. Kemudian, untuk pengklasifikasi data dibagi menjadi dua yaitu data latih 80% sebanyak 82 data dan data uji 20% sebanyak 21 data uji.

4. Implementasi Algoritma K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* merupakan suatu metode yang digunakan untuk pengklasifikasi sebuah objek atau data berdasarkan jarak tetangga yang paling dekat. *K-Nearest Neighbor* memiliki beberapa tahapan di antara lain:

1. Menentukan parameter K
2. Menetapkan nilai K paling besar
3. Menentukan tetangga terdekat berdasarkan jarak terbesar ke terkecil pada K.
4. Menentukan kelas objek atau data dengan menggunakan kategori *Nearest Neighbor* yang paling mayoritas.

Dengan rumus perhitungan *K-Nearest Neighbor* menggunakan *Eucliden Distance* sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

5. Validasi Dan Analisa Hasil

Evaluasi bertujuan untuk mengetahui nilai akurasi dari pengelompokan yang telah dihasilkan dengan perhitungan *K-Nearest Neighbor* dengan menggunakan perhitungan metode *confusion matrix* sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

Berdasarkan penelitian yang telah dipaparkan diatas dengan menggunakan algoritma *K-nearest Neighbor* (KNN). Penelitian ini bertujuan untuk melakukan pengujian terhadap klasifikasi kategori kesejahteraan masyarakat dengan menggunakan metode algoritma KNN untuk mendapatkan hasil akurasi yang terbaik.

