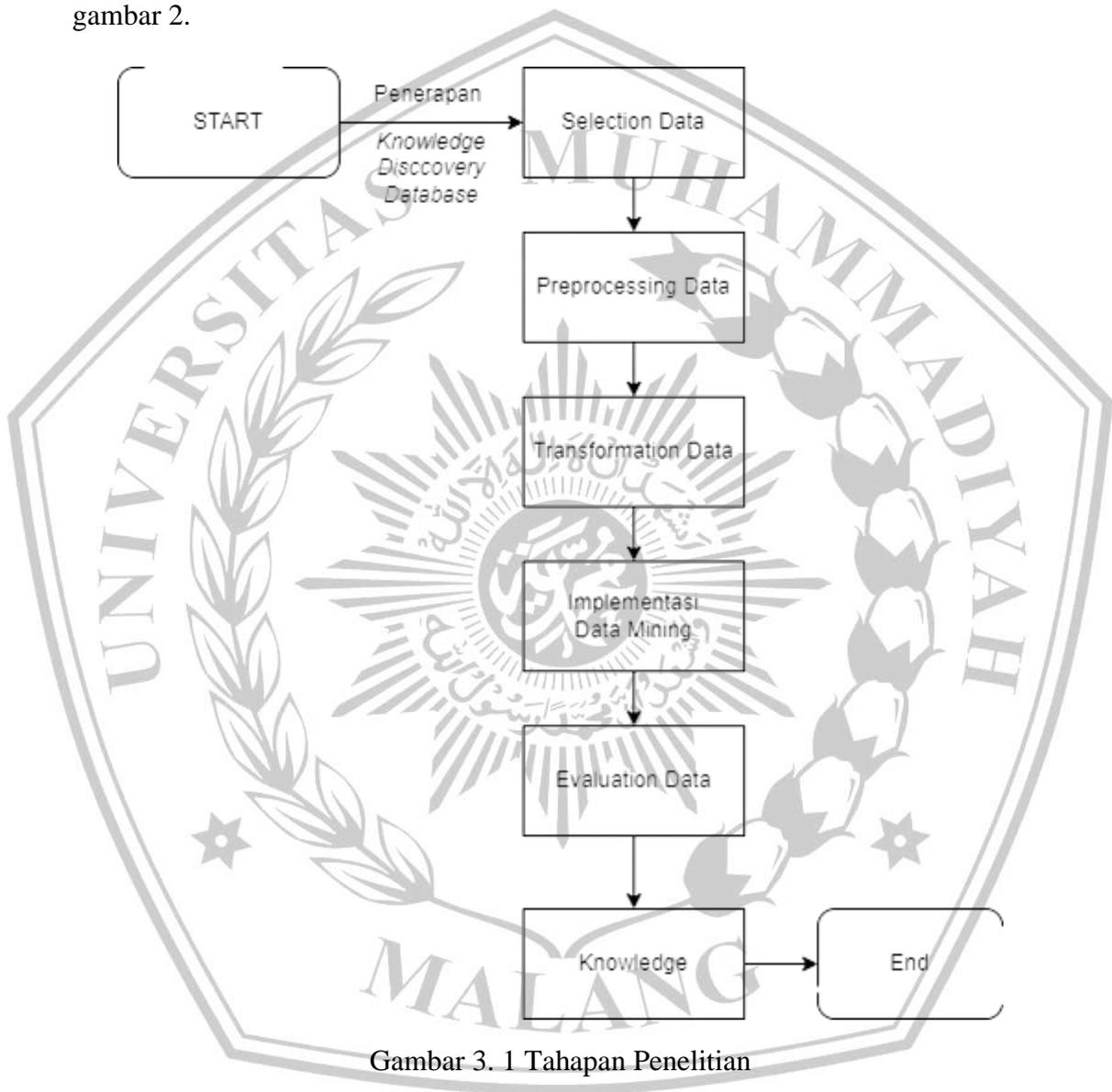


## BAB 3 METODOLOGI PENELITIAN

### 3.1 Tahapan penelitian

Ada beberapa langkah-langkah penelitian yang dilakukan peneliti sehingga dapat memperoleh tujuan atau hasil akhir. Tahapan penelitian dapat dilihat pada gambar 2.



Gambar 3. 1 Tahapan Penelitian

### 3.2 Lingkungan Kerja

Pada sub-bab ini bertujuan untuk menjelaskan lingkungan kerja penulis selama melakukan penelitian ini, seperti menjelaskan perangkat hardware dan software yang digunakan, lingkungan kerja dapat dilihat pada table 2 dan table 3.

1. Perangkat keras (Hardware)

Tabel 3. 1 Perangkat Keras

No	Perangkat	Jumlah	Spesifikasi
1	Laptop	1	Processor : AMD Ryzen 3 5300U RAM : 8 GB HDD : 500GB

2. Perangkat Lunak (software)

Tabel 3. 2 Perangkat Lunak

No	Software
1	OS Windows 11
2	Python
3	Jupyter Notebook
4	Microsoft Excel
5	Microsoft Word

### 3.3 Dataset

Dataset yang digunakan pada penelitian ini adalah data yang diperoleh dari Badan Pusat Statistika (BPS) terkait tentang kemiskinan beserta indikator-indikatornya. Data yang diambil adalah data yang dirilis pada tahun 2022 terkait tentang presentasi kemiskinan pada tahun 2021. Data tersebut dapat diakses pada halaman <https://www.bps.go.id>, yaitu website resmi dari Badan Pusat Statistika.

Total data yang diambil adalah 514 Kota/Kabupaten yang ada di Indonesia. Adapun 10 variabel kemiskinan yang diambil meliputi presentasi penduduk miskin, presentasi angka melek huruf pada usia tertentu, presentasi pengeluaran perkapita untuk makanan pada penduduk miskin, presentasi dibidang ketenagakerjaan penduduk miskin, presentasi angka partisipasi sekolah penduduk miskin.

### 3.4 Implementasi Knowledge Discovery In Database (KDD)

#### 3.4.1 Selection

Pada tahap ini dilakukan pemilihan data yang ingin digunakan, dimana pada penelitian ini data yang digunakan adalah total 514 kota/kabupaten di Indonesia. Adapun 10 indikator atau variable yang diambil seperti :

- a. Presentase penduduk miskin
- b. Presentase penduduk miskin usia 15 tahun keatas tamat pendidikan SD.
- c. Presentase penduduk miskin usia 15 tahun keatas tamat pendidikan SMP.
- d. Presentase penduduk miskin usia 15 tahun keatas tamat pendidikan SMA atau perguruan tinggi.
- e. Presentase angka melek huruf usia 15-55 tahun.
- f. Presentase angka partisipasi sekolah penduduk miskin usia 13-15 tahun.
- g. Presentase penduduk miskin usia 15 tahun keatas yang tidak berkerja.
- h. Presentase penduduk miskin usia 15 tahun keatas yang bekerja di sector informal.
- i. Presentase penduduk miskin usia 15 tahun keatas yang bekerja di sector formal.
- j. Presentase pengeluaran perkapita untuk makanan menurut kabupaten/kota dan status miskin

#### 3.4.2 Pre-processing

Pada tahap ini dilakukan pre-processing pada 10 indikator/variable kemiskinan yang diambil dari 514 Kota/Kabupaten di Indonesia. Dimana bertujuan untuk menghilangkan noise-noise pada data tersebut. Noise tersebut seperti adanya nilai *null* pada kolom atau seperti data yang tidak diinginkan.

#### 3.4.3 Transformation Data

Transformation data juga bagian dari pre-processing bertujuan untuk memudahkan algoritma untuk memproses data, dimana pada penelitian ini nama 10 variabel diganti dan dipersingkat menjadi X1-X10. Lalu dilanjutkan dengan menggunakan metode standarisasi, dimana mengubah dataset sehingga memiliki nilai rata-rata 0 dan deviasi standar 1 sehingga semua data atau fitur memiliki skala yang serupa.

#### 3.4.4 Implementasi Data Mining

Pada tahap ini dilakukan implementasi metode data mining yang digunakan, dimana pada penelitian ini, metode yang digunakan adalah clustering K-means, lalu menguji nilai K menggunakan metode Elbow dan Silhouette Coefficient.

#### 3.4.4.1 Clustering K-Means

Pertama-tama dilakukan cluster K-Means dimana pada penelitian ini menggunakan library *scikit-learn* pada jupyter notebook. K-Means diimplementasikan pada 514 kota/kabupaten dengan 10 indikator atau variable. Pada tahap ini nilai K yang ditentukan adalah nilai K random.

#### 3.4.4.2 Metode Elbow

Setelah cluster terbentuk, akan dilakukan pengujian menggunakan metode elbow. Pengujian elbow dilakukan dengan cara menghitung nilai inerti ( sum squared distance antara pusat dan titik cluster terdekat ) dari beberapa pengelompokkan yang berbeda, dan mengamati penurunan inerti saat jumlah cluster bertambah. Semakin banyak jumlah klaster, inerti rata-rata akan semakin kecil, karena setiap cluster akan memiliki lebih sedikit titik. Namun, penurunan rata-rata inerti ini akan melambat saat jumlah cluster terus bertambah, sehingga membentuk sebuah "siku" pada grafik penurunan rata-rata inerti terhadap jumlah cluster. Maka dari itu jumlah cluster yang optimal terlihat dari siku-siku yang terbentuk.

#### 3.4.5 Evaluation/Interpretation

Setelah terbentuknya cluster akan dilakukan evaluasi, dimana evaluasi ini bertujuan untuk melihat hasil informasi yang diberikan relevan atau tidak. Pada penelitian ini teknik evaluasi yang digunakan adalah teknik silhouette coefficient.

##### 3.4.5.1 Silhouette Coefficient

Setelah nilai K yang optimal diketahui melalui metode elbow, akan diujikan lagi menggunakan silhouette coefficient. Fungsi silhouette coefficient juga hampir sama dengan metode elbow yaitu menentukan nilai K yang optimal, dimana pada penelitian ini peneliti menggunakan library *scikit-learn* untuk menentukan silhouette coefficient. Hasil dari nilai silhouette coefficient dapat berkisar antara -1 hingga 1, dimana nilai yang lebih dekat ke angka 1 menunjukkan jumlah cluster yang terbaik

#### 3.4.6 Knowledge

Pada tahap ini akan dilakukan pemaparan hasil dan analisa dari cluster-cluster yang terbentuk. Sehingga ditemukan pengetahuan baru dari data yang telah diambil

pada tahap pertama yaitu selection data, lalu data yang telah di pre-processing sampai dengan hasil cluster dan evaluasi cluster.

