

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan salah satu negara berkembang yang tingkat kemiskinannya masih menjadi suatu masalah yang harus diperhatikan oleh negara (Wahyuli et al., 2019). Kemiskinan dalam jangka panjang akan menjadi pengaruh menghambatnya pembangunan negara (Febianto & Palasara, 2019). Menurut data dari Badan Pusat Statistik Jawa Barat pada tahun 2018-2019, Kota Tasikmalaya memiliki persentase tingkat kemiskinan paling tinggi di Jawa Barat sebesar 11,60% dibandingkan dengan Kota Depok dengan persentase tingkat kemiskinan terendah hanya sebesar 2,07% (BPS, 2020). Menurut data dari Badan Pusat Statistik Kota Tasikmalaya, angka kemiskinan di Kota Tasikmalaya meningkat pada Maret 2020 dibandingkan Maret 2019 sebelumnya (BPS, 2021). Pemerintah Kota Tasikmalaya menggunakan Data Terpadu Kesejahteraan Sosial (DTKS) yang merupakan sumber data utama untuk menetapkan sasaran bagi program perlindungan sosial dan penanggulangan kemiskinan. Terdapat status kesejahteraan sosial pada Data Terpadu Kesejahteraan Sosial (DTKS) yang terbagi menjadi 4 jenis yaitu desil 1 yang terdiri dari rumah tangga/individu dengan kondisi kesejahteraan 10% terendah di Indonesia, desil 2 terdiri dari rumah tangga/individu dengan kondisi kesejahteraan antara 11%-20% terendah di Indonesia, desil 3 terdiri dari rumah tangga/individu dengan kondisi kesejahteraan antara 21%-30% terendah di Indonesia, dan desil 4 terdiri dari rumah tangga/individu dengan kondisi

kesejahteraan antara 31%-40% terendah di Indonesia (puspensos.kemensos.go.id, 2021).

Tingginya persentase kemiskinan tentunya pemerintah harus lebih memprioritaskan daerah dengan angka kemiskinan yang tinggi. Oleh karena itu, untuk mengetahui daerah mana saja dengan angka kemiskinan tinggi dan harus diprioritaskan oleh pemerintah dalam menanganinya maka dibutuhkan teknik yang mampu pengelompokan data dengan tepat. Salah satu metode pengolahan data dalam data mining adalah *clustering*/pengelompokan. Penerapan *clustering* dalam permasalahan ini dapat menggunakan algoritma K-Means untuk mengelompokan data yang paling sering digunakan. Penelitian ini menggunakan algoritma K-Means dimana algoritma ini menjadi algoritma yang efisien dan efektif dalam mengolah data dalam jumlah banyak, sehingga dalam penelitian ini akan mencari jumlah *cluster* terbaik. Terdapat banyak cara dalam menentukan hal tersebut, salah satunya dengan melakukan evaluasi *clustering* menggunakan Davies Bouldin Index (DBI) untuk mengetahui jumlah *cluster* yang paling optimal. Evaluasi menggunakan Davies Bouldin Index (DBI) ini memiliki skema evaluasi dari *internal cluster*, dimana baik buruknya hasil *cluster* dilihat dari kuantitas dan kedekatan antar data hasil *cluster*(Muningsih et al., 2021).

Menurut penelitian (Yuli et al., 2016) mengenai kemiskinan yang terjadi di D.I.Yogyakarta yang terbagi menjadi 4 jenis yaitu desil 1, desil 2, desil 3, dan desil 4. Penelitian tersebut menggunakan metode average linked dan software R Menghasilkan 5 cluster yang akan menjadi prioritas 1, prioritas 2, prioritas 3, prioritas 4, dan prioritas 5. Menurut penelitian (Sembiring et al., 2019) mengenai

penerapan metode K-Means *clustering* tentang pengangguran yang terjadi di pulau Jawa dan Bali dengan menggunakan bahasa pemrograman *python* menghasilkan 3 *cluster* yaitu terdapat 1 wilayah tingkat pengangguran rendah (C1), 3 wilayah tingkat pengangguran sedang (C2) dan 3 wilayah lagi dikategorikan sebagai tingkat pengangguran tinggi (C3). Penelitian (Muningsih et al., 2021) mengenai penerapan metode K-Means dan evaluasi *clustering* menggunakan Davies Bouldin Index (DBI) pada potensi Desa didapatkan hasil DBI dengan nilai terkecil pada *cluster* 3 adalah 0,175 dengan menggunakan RapidMiner.

Berdasarkan seluruh uraian tersebut, masih terdapat peluang pengembangan penelitian untuk melakukan pengelompokan daerah kemiskinan di tingkat kelurahan Kota Tasikmalaya menggunakan data kemiskinan berdasarkan desil 1 sampai desil 4 dengan algoritma K-Means *clustering*. Metode tersebut akan di evaluasi *cluster* menggunakan Davies Bouldin Index (DBI). Dengan mempertimbangkan hal tersebut maka akan dilakukan penelitian yang berjudul “Penerapan Algoritma K-Means dan Penentuan Cluster Optimum untuk Mengelompokan Daerah Kemiskinan di Kelurahan Kota Tasikmalaya”.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah disampaikan, maka perlu dirumuskan suatu masalah yang akan diselesaikan pada penelitian ini. Rumusan masalah pada penelitian ini yaitu:

1. Bagaimana mengelompokan daerah kemiskinan menggunakan algoritma K-Means *clustering*?

2. Bagaimana cara menentukan nilai *cluster* optimal pada K-Means *clustering*?

1.3 Batasan Masalah

Terdapat beberapa batasan-batasan masalah yang digunakan sehingga penelitian dapat dilakukan secara spesifik. Batasan masalah pada penelitian ini sebagai berikut:

1. Data yang digunakan diambil dari data kemiskinan pada tahun 2020 dan bersumber dari Data Terpadu Kesejahteraan Sosial (DTKS) Kota Tasikmalaya.
2. Menggunakan atribut desil 1-4 untuk membuat *cluster* daerah kemiskinan berdasarkan kelurahan yang terdapat di Kota Tasikmalaya.
3. Analisis *clustering* menggunakan algoritma K-Means dengan 2 - 5 *cluster*.
4. Pencarian *cluster* optimal menggunakan Davies Bouldin Index (DBI).
5. Pengolahan data menggunakan bahasa pemrograman *python*.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah memecahkan permasalahan yang tergambar dalam latar belakang dan rumusan masalah. Adapun tujuan yang hendak dicapai adalah sebagai berikut:

1. Menentukan kelompok daerah kemiskinan menggunakan algoritma K-Means *clustering*.
2. Menentukan nilai *cluster* optimal K-Means *clustering* menggunakan Davies Bouldin Index (DBI).

1.5 Manfaat Penelitian

Berdasarkan tujuan di atas, maka manfaat yang diharapkan pada penelitian ini adalah sebagai berikut:

1. Tersedianya informasi pengelompokan tingkat kemiskinan berdasarkan kelurahan di Kota Tasikmalaya.
2. Membantu pihak-pihak yang berkepentingan untuk mengetahui *cluster* daerah kemiskinan berdasarkan kelurahan di Kota Tasikmalaya.
3. Menjadi alternatif penelitian untuk mengelompokan kelurahan berdasarkan kemiskinan menggunakan atribut desil 1-4.
4. Menerapkan keilmuan di mata kuliah Data Mining di bidang *Clustering*.

1.6 Metodologi Penelitian

Metodologi penelitian berisi mengenai waktu dan tempat penelitian, tahapan penelitian, pendekatan penelitian, jenis penelitian, variabel penelitian, serta objek penelitian. Metode penelitian yang digunakan untuk menyelesaikan penelitian ini diantaranya:

1. Pendahuluan

Penelitian diawali dengan mengidentifikasi masalah hal apa saja yang menjadi permasalahan mengenai kemiskinan yang terjadi di wilayah Kota Tasikmalaya dan dilakukan studi literatur untuk mengumpulkan referensi dari buku-buku, internet, *e-book*, jurnal atau karya ilmiah yang berkaitan dengan konsep serta teori pada penelitian.

2. Pengumpulan Data

Tahap ini dilakukan pengamatan langsung dengan objek yang diteliti, dalam hal ini akan mengamati bagaimana pengelolaan pada data-data penduduk, sehingga didapatkan *softcopy* dari Data Terpadu Kesejahteraan Sosial (DTKS) tahun 2020 dan.

3. *Preprocessing* / Persiapan Data

Sebelum memulai pengolahan data dilakukan tahapan persiapan data dengan tahapan *Knowledge Discovery* yang harus dilakukan pada dataset sehingga dapat digunakan untuk proses Data Mining terdiri dari *Data Cleaning*, *Data Selection*, dan *Data Transformation*.

4. Pengolahan Data

Menggunakan algoritma K-Means *clustering* dan menentukan jumlah *cluster* paling optimal dalam proses *clustering* menggunakan Davies Bouldin Index (DBI).

5. Kesimpulan

Dilakukan pengambilan kesimpulan yang didapat dari hasil analisis.

1.7 Sistematika Penulisan

Dalam penyusunan laporan ini, sistematika penulisan ini dibagi menjadi beberapa bab sebagai berikut:

BAB I PENDAHULUAN

Bab ini membahas masalah umum dalam laporan meliputi latar belakang masalah, rumusan masalah, tujuan penelitian, batasan

masalah, manfaat penelitian, metode penelitian, dan sistematika penelitian.

BAB II LANDASAN TEORI

Bab ini berisi dasar-dasar teori yang berhubungan dengan penelitian ini. Meliputi teori dari data, teori *Data Mining*, teori *clustering*, teori algoritma K-Means, teori evaluasi *clustering* Davies Bouldin Index (DBI) dan teori uji kolerasi.

BAB III METODOLOGI

Bab ini menjelaskan tentang metode yang digunakan dalam penelitian ini yang terdiri dari tahapan pra penelitian, tahapan persiapan dan proses *Data Mining*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi uraian tentang hasil dari penelitian *Data Mining* dengan dataset kemiskinan yaitu Data Terpadu Kesejahteraan Sosial (DTKS) yang terjadi di Kota Tasikmalaya.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang merupakan rangkuman dari hasil analisis bab-bab sebelumnya dan saran untuk penelitian selanjutnya.