

PENERAPAN METODE ALGORITMA K-MEANS UNTUK DATA PENGANGGURAN DI JAWA BARAT

Aan Ansen Andryadi¹, Nisa²

Ansen25@gmail.com

¹ STMIK JABAR; ² STMIK JABAR

Abstract

In West Java the unemployment rate changes every year in all districts/cities, so that unemployment becomes a social development problem in realizing the prosperity and welfare of the community. Purpose This research was conducted to get results Clustering districts/cities in West Java based on unemployment rate indicators. This research data comes from the West Java open data dataset for 2007 - 2022. This research uses the Data mining namely the Algorithm K-means, Method K-means is the method Clustering which serves to break the dataset into groups. Application of Algorithmic Methods K-means can be used for data on the percentage of unemployment by district/city. Data will be divided or grouped into 2 Cluster, that is Cluster high and low with RapidMiner testing. The results of this research are as a way to assist the government in expanding employment opportunities to develop and improve the economy in Indonesia, especially in West Java.

Keywords: *Unemployment, Data Mining, Clustering, Algorithm K-means*

Abstrak

Di Jawa Barat tingkat pengangguran mengalami perubahan setiap tahunnya di seluruh kabupaten/kota, sehingga pengangguran menjadi permasalahan pembangunan sosial kemasyarakatan dalam mewujudkan kemakmuran dan kesejahteraan masyarakat. Tujuan Penelitian ini dilakukan untuk mendapatkan hasil Clustering kabupaten/kota di Jawa Barat berdasarkan indikator tingkat pengangguran. Data Penelitian ini bersumber dari dataset open data Jabar Tahun 2007 - 2022. Penelitian ini menggunakan teknik Data mining yakni Algoritma K-means, Metode K-means merupakan Metode Clustering yang berfungsi untuk memecahkan dataset menjadi berkelompok. Penerapan Metode Algoritma K-means dapat digunakan untuk data persentase pengangguran berdasarkan kabupaten/kota. Data akan dibagi atau dikelompokkan menjadi 2 Cluster, yaitu Cluster tinggi dan rendah dengan pengujian RapidMiner. Hasil dari penelitian ini adalah sebagai cara untuk membantu pemerintah dalam memperluas lapangan pekerjaan untuk mengembangkan dan meningkatkan ekonomi di Indonesia khususnya di Jawa Barata.

Keywords: Algoritma K-means; Clustering; Data Mining; Pengangguran

Pendahuluan

Pengangguran adalah suatu kondisi di mana seseorang yang termasuk dalam golongan angkatan kerja ingin mendapatkan pekerjaan tetapi belum dapat memperolehnya. Pengangguran merupakan keadaan dimana orang ingin bekerja namun tidak mendapat pekerjaan [1]. Pengangguran menjadi salah satu bagian dari permasalahan utama dalam Negara berkembang. Pengangguran di negara berkembang seperti Indonesia dapat disebabkan oleh beberapa faktor, diantaranya laju pertumbuhan penduduk yang tinggi, ketersediaan lapangan pekerjaan yang tidak berimbang, ketidaksiapan penduduk usia produktif dalam menghadapi pesatnya kemajuan teknologi, tingkat pendidikan, perubahan struktur perekonomian, dan besaran upah [2]. Jawa Barat salah satu provinsi di Indonesia, masuk dalam kategori pengangguran yang jumlahnya sering menurun, namun masih cukup tinggi. Jumlah pengangguran dapat meningkat setiap tahunnya.



Gambar 1 : Tingkat Pengangguran Terbuka Jawa Barat

Pengangguran yang tinggi mempengaruhi tingkat kemiskinan di beberapa daerah. Akibatnya, karena pengangguran adalah salah satu penentu utama tingkat kemiskinan, penelitian yang lebih mendalam tentang tingkat pengangguran perlu dilakukan. Salah satunya adalah dengan memanfaatkan Data Mining dengan menggunakan teknik clustering untuk mengelompokkan data pengangguran yang ada di Provinsi Jawa Barat. Clustering merupakan salah satu teknik pengelompokkan untuk analisis pada data mining [3]. Metode klasterisasi juga tidak menggunakan label dalam prosesnya dan label dapat diberikan ketika cluster sudah terbentuk. Banyak metode klasterisasi yang diusulkan oleh para ahli, salah satunya adalah metode algoritma k-Means yang termasuk dalam metode berbasis partisi, dimana metode ini bekerja dengan cara membagi data ke dalam sejumlah kelompok. Metode k-Means merupakan algoritma klasterisasi yang paling sering digunakan karena prosesnya sederhana dan cocok untuk diimplementasikan. Berdasarkan uraian diatas, penulis tertarik untuk melakukan penelitian sehingga dapat memberikan masukan terhadap pemerintah berdasarkan cluster tingkat pengangguran dengan bantuan pengujian aplikasi RapidMiner, khususnya

kabupaten/kota yang minim dalam lapangan pekerjaan yang berdampak pada pengangguran. Data yang akan digunakan adalah dataset open data Jabar Tahun 2007 - 2022.

Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah metode deskriptif analisis. Menurut Sugiyono [4] penelitian dengan metode deskriptif analisis merupakan metode penelitian dengan cara mengumpulkan data sesuai dengan yang sebenarnya kemudian data – data tersebut disusun, diolah dan dianalisis untuk dapat memberikan gambaran mengenai masalah yang ada. Dengan kata lain metode deskriptif analisis mengambil masalah atau memusatkan perhatian kepada masalah – masalah sebagaimana adanya saat penelitian dilakukan, hasil penelitian kemudian diolah dan dianalisis untuk diambil kesimpulannya. Pada analisis deskriptif data biasanya ditampilkan dalam bentuk tabel biasa atau tabel frekuensi, grafik, diagram batang, diagram garis, diagram lingkaran, ukuran pemusatan data, ukuran penyebaran data dan sebagainya. Menurut Sugiyono [4] metode analisis deskriptif adalah statistik yang digunakan untuk menganalisis data dengan cara mendeskripsikan atau menggambarkan data yang telah terkumpul sebagaimana adanya tanpa bermaksud membuat kesimpulan yang berlaku untuk umum atau generalisasi. Sedangkan menurut Irwan et al., [5] metode deskriptif yaitu suatu metode dalam meneliti status kelompok manusia, suatu objek, suatu kondisi, suatu pemikiran, ataupun suatu kelas peristiwa pada masa sekarang. Langkah – langkah yang perlu dilakukan pada metode deskriptif analisis adalah sebagai berikut :

1. Merumuskan tujuan penelitian
2. Menentukan unit studi dan menghubungkan hal yang akan dikaji berdasarkan sifat – sifat yang dimiliki dan proses yang menjadi landasan penelitian.
3. Menentukan rancangan dalam memilih unit dan teknik pengumpulan data.
4. Mengumpulkan data.
5. Mengorganisasikan informasi, data yang terkumpul, serta melakukan interpretasi dan generalisasi.
6. Menyusun laporan yang diakhiri dengan menyimpulkan hasil penelitian

Kebutuhan Penelitian

Dalam penelitian ini kebutuhan penelitian yang dibutuhkan untuk kemudian diolah menjadi acuan adalah:

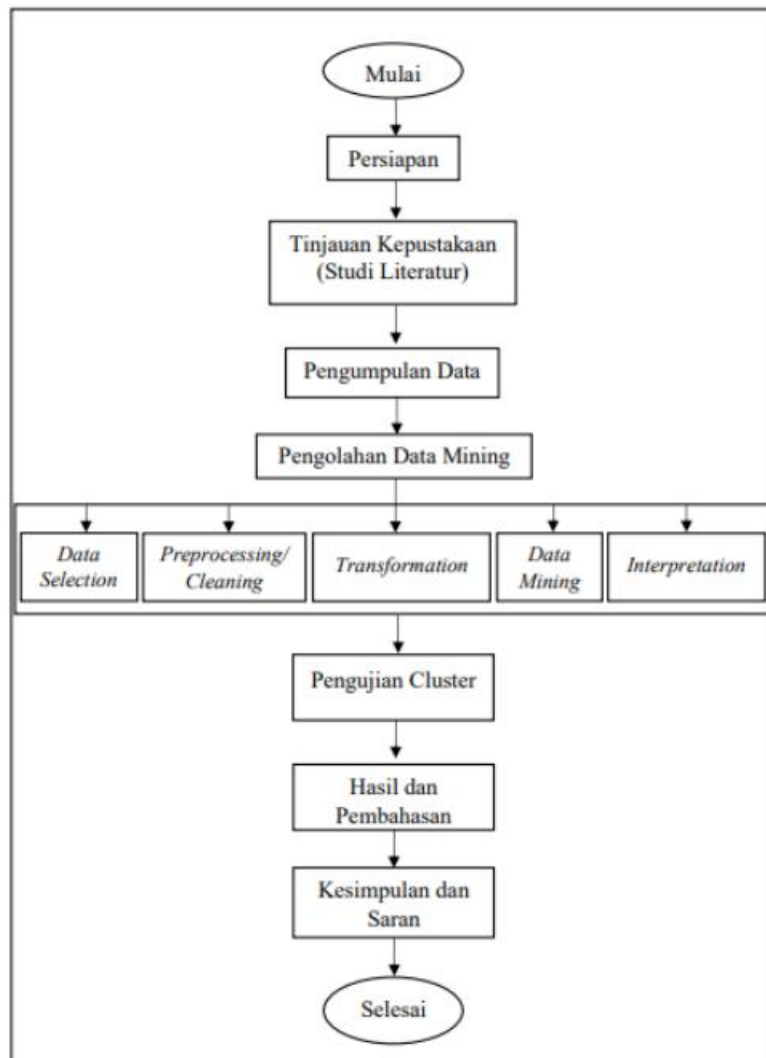
1. Data Pengangguran Terbuka di Jawa Barat.
2. Aplikasi Microsoft excel
3. Aplikasi RapidMiner

Metode Pengumpulan Data

Menurut Sugiyono [4] sumber data sekunder adalah sumber yang tidak langsung memberikan data kepada pengumpul data, misalnya lewat orang lain atau dokumen. Data sekunder dapat diperoleh dengan cara membaca, mempelajari dan memahami melalui media lain yang bersumber pada literatur dan buku – buku perpustakaan atau data – data yang diperoleh dari perusahaan yang berkaitan dengan masalah yang diteliti

Tahap Penelitian

Tahapan Penelitian menggambarkan proses penelitian yang akan ditempuh sekaligus menggambarkan penelitian secara keseluruhan. Tahapan penelitian dapat dilihat pada gambar berikut :



Gambar 2 : Tahapan Penelitian

Gambar di atas, dapat dijelaskan bahwa tahapan penelitian yang akan ditempuh yaitu:

1. Persiapan, tahap ini merupakan tahap menyusun rencana penelitian. Yaitu pencarian tempat penelitian dan pembuatan proposal penelitian.
2. Tinjauan kepustakaan, Dalam tinjauan kepustakaan dilakukan kajian dan studi literatur mengenai data pengangguran.
3. Pengumpulan Data diperoleh dari Dataset Open data Jabar
4. Pengolahan Data Mining, Data yang dikumpulkan diolah sesuai dengan tahapan Knowledge Discovery in Database (KDD).
5. Pengujian cluster, pengujian menggunakan Aplikasi RapidMiner.
6. Hasil dan Pembahasan, Pembahasan pada tahapan ini menjelaskan hasil dari proses data mining yang dilakukan dengan menggunakan algoritma K-Means.
7. Kesimpulan, membuat kesimpulan dari hasil penelitian

Hasil dan Pembahasan

Pengolahan Data Mining

Pengolahan data mining yang dilakukan pada penelitian ini yaitu, mengikuti tahapan dalam Knowledge Discovery in Database (KDD), untuk menghasilkan informasi sesuai dengan urutan yang sudah ditentukan, berikut tahapan – tahapannya:

a. Data Selection

Data yang digunakan dalam penelitian ini adalah data tingkat pengangguran terbuka di Jawa Barat tahun 2007-2022. Data ini berasal dari Dataset Open data Jabar. Atribut yang terdapat pada data tingkat pengangguran terbuka yaitu Kode Provinsi, Nama Provinsi, Kode Kabupaten/Kota, Nama Kabupaten/Kota, Persentase tingkat pengangguran terbuka, Satuan, dan Tahun. Dari semua atribut tersebut hanya tiga atribut yang digunakan untuk proses Knowledge Discovery in Database (KDD). Atribut tersebut yaitu:

1. Nama Kabupaten/Kota merupakan atribut yang berisi nama kabupaten dan kota di Jawa barat.
2. Persentase tingkat pengangguran terbuka merupakan atribut yang menyatakan persentase dari tingkat pengangguran terbuka dengan tipe data numerik.
3. Tahun merupakan atribut yang menyatakan tahun produksi data dengan tipe numerik.

id	kode_provinsi	nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	persentase_tingkat_pengangguran_terbuka	satuan	tahun
1	32	JAWA BARAT	3201	KABUPATEN BOGOR	14,26	PERSEN	2007
2	32	JAWA BARAT	3202	KABUPATEN SUKABUMI	10,85	PERSEN	2007
3	32	JAWA BARAT	3203	KABUPATEN CIANJUR	13,82	PERSEN	2007
4	32	JAWA BARAT	3204	KABUPATEN BANDUNG	17,37	PERSEN	2007
5	32	JAWA BARAT	3205	KABUPATEN GARUT	12,18	PERSEN	2007
6	32	JAWA BARAT	3206	KABUPATEN TASIKMALAYA	8,48	PERSEN	2007
7	32	JAWA BARAT	3207	KABUPATEN CIAMIS	4,39	PERSEN	2007
8	32	JAWA BARAT	3208	KABUPATEN KUNINGAN	10,56	PERSEN	2007
9	32	JAWA BARAT	3209	KABUPATEN CIREBON	13,64	PERSEN	2007
10	32	JAWA BARAT	3210	KABUPATEN MAJALENGKA	7,46	PERSEN	2007

Gambar 3 : Dataset Pengangguran sebelum dilakukan Preprocessing

b. Preprocessing

Sebelum proses data mining dapat dilaksanakan perlu dilakukan tahap preprocessing, pada tahap ini akan dilakukan proses integrasi data untuk penggabungan data dari database yang berbeda, selanjutnya dilakukan data cleaning untuk menghasilkan dataset yang bersih sehingga dapat digunakan dalam tahap berikutnya. Berikut merupakan penjelasan dari kedua proses:

1. Integrasi Data Proses ini merupakan penggabungan data dari berbagai database yang berbeda, sehingga data tersebut saling berintegrasi. Data integrasi dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik. Pada proses ini tidak ada penggabungan data dikarenakan data yang diambil berasal dari satu database.
2. Data Cleaning Pada proses ini data yang tidak relevan, missing value, dan redundant harus dibersihkan. Hal ini dikarenakan data yang bersih merupakan syarat awal dalam melakukan data mining. Suatu data dikatakan missing value jika terdapat atribut dalam dataset yang tidak berisi nilai atau kosong, sedangkan data dikatakan redundant jika dalam satu dataset terdapat record yang berisi nilai yang sama

c. Transformation

Tahapan Transformation merupakan tahap merubah data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Pada penelitian ini akan dilakukan agregasi dengan cara mengkombinasikan beberapa nama kabupaten/kota yang sama menjadi satu nama kabupaten/kota setiap tahunnya dan atribut persentase tingkat pengangguran terbuka dengan tipe numeric akan menjadi record dari atribut tahun

Kesimpulan

Penelitian ini dilakukan untuk menerapkan metode algoritma k-means pada data pengangguran terbuka di Jawa barat tahun 2007 - 2022 (persen) dengan sumber data yang diperoleh dari open data Jabar. Data kemudian diolah dengan menggunakan tahapan Knowledge Discovery in Database (KDD). Proses clustering pada iterasi 1 diawali dengan menentukan jumlah data yang akan dikluster, menetapkan nilai K jumlah kluster yaitu sebanyak 2 kluster, menentukan nilai centroid awal yang telah ditentukan secara random, lalu menghitung jarak setiap data terhadap pusat kluster dengan menggunakan rumus Euclidean Distance. Proses tersebut terus dilakukan pada iterasi berikutnya hingga nilai centroid sudah optimal

serta posisi kluster tidak mengalami perubahan lagi maka proses iterasi berhenti. Perhitungan penelitian ini menggunakan aplikasi Microsoft Excel 2013 dan aplikasi RapidMiner untuk pengujian. Berdasarkan hasil analisis dan pembahasan yang telah dibuktikan, diperoleh kesimpulan sebagai berikut : Penerapan metode algoritma k-means pada data pengangguran terbuka di Jawa barat berhenti pada iterasi 4 yang menghasilkan sebuah informasi clustering pengangguran dengan jumlah record yang digunakan sebanyak 27 kabupaten/kota, menghasilkan 2 cluster yaitu cluster tinggi (C1) sebanyak 13 kabupaten/kota dan cluster rendah (C2) sebanyak 14 kabupaten/kota. Hasil pengujian k-means menggunakan aplikasi RapidMiner memperoleh hasil yang sama dengan analisis perhitungan metode dimana diperoleh 13 kabupaten/kota dengan cluster tinggi yang menjadi perhatian khusus bagi pemerintah untuk memperluas lapangan pekerjaan pada Kabupaten/kota tersebut..

Daftar Pustaka

- [1] R. Franita and A. Fuady, “Analisa Pengangguran Di Indonesia,” *J. Ilmu Pengetah. Sos.*, vol. 1, no. 3, pp. 88–93, 2019.
- [2] F. R. Suhadi and E. Setyowati, “Analisis Pengaruh Jumlah Penduduk, Pendidikan, Upah Minimum, Dan PDRB Terhadap Tingkat Pengangguran Terbuka Di Provinsi Jawa Barat,” *EKOMBIS Rev. J. Ilm. Ekon. dan Bisnis*, vol. 10, no. 2, 2022, doi: 10.37676/ekombis.v10i2.2144
- [3] T. H. Sardar and Z. Ansari, “An analysis of MapReduce efficiency in document clustering using parallel K-means algorithm,” *Futur. Comput. Informatics J.*, vol. 3, no. 2, pp. 200–209, 2018, doi: 10.1016/j.fcij.2018.03.003
- [4] Sugiyono, “Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D,” Bandung Alf., 2013
- [5] I. Irwan, F. N. Ichsan, N. Gistituati, and S. Marsidin, “Analisis Kebijakan Pendidikan Terkait Implementasi Pembelajaran Pada Masa Darurat Covid 19,” *J. Manaj. Pendidik.*, vol. 9, no. 2, pp. 89–95, 2021, doi: 10.33751/jmp.v9i2.4238