



IMPLEMENTASI *DATA MINING* UNTUK PENGELOMPOKAN PENGANGGURAN TERBUKA DI INDONESIA DENGAN METODE *CLUSTERING*

Candra Naya¹, Arif Siswandi² Hemdani Rahendra Herlianto³ Ahmad Fauzi⁴

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa

¹candranaya@pelitabangsa.ac.id, ²arifsiswandi@pelitabangsa.ac.id, ³hemdani.rahendra@pelitabangsa.ac.id,

⁴ahmad_fauzi@pelitabangsa.ac.id

Abstrak

Jumlah pengangguran per agustus 2021 sebanyak 9,1 juta jiwa dari semua jenis kalangan pendidikan di Indonesia, dengan begitu persentase pengangguran di Indonesia cukup besar dan menjadi faktor yang dapat digaris bawahi. Tingkat pengangguran akan semakin meningkat di setiap tahun nya jika tidak ada upaya yang signifikan dalam mengurangi persentase pengangguran di Indonesia. Oleh karena itu, perlu adanya realisasi pemerintah dalam menyediakan lapangan pekerjaan sesuai dengan pengelompokan di wilayah Indonesia agar dapat mengetahui tingkat *cluster* pengangguran penduduk Indonesia. Implementasi *data mining* dengan metode *clustering* dalam penelitian ini bertujuan untuk mengetahui kelompok pengangguran berdasarkan tingkatan pengangguran di wilayah Indonesia. Metode *clustering* menggunakan algoritma *k-means* dapat mengelompokkan wilayah Indonesia ke dalam tiga *cluster*, yaitu *cluster 1* adalah kelompok wilayah dengan tingkat pengangguran rendah yang meliputi 3 wilayah, *cluster 2* adalah kelompok wilayah dengan tingkat pengangguran sedang yang meliputi 10 wilayah, dan *cluster 3* adalah kelompok wilayah dengan tingkat pengangguran tinggi yang meliputi 21 wilayah. Hasil analisa memperoleh persentase pada setiap *cluster*, yaitu *cluster 1* memiliki 9% wilayah, *cluster 2* memiliki 29% wilayah, dan *cluster 3* memiliki 62% wilayah. Hasil tersebut diharapkan dapat memberikan informasi kepada masyarakat dan pemerintah setempat khususnya di wilayah yang memiliki status tingkat pengangguran tinggi untuk dapat melakukan gagasan-gagasan guna mengurangi tingkat persentase pengangguran di Indonesia.

Kata kunci: *Data mining*, *Clustering*, Algoritma *K-means*, Pengangguran.

Abstract

The number of unemployed as of August 2021 is 9.1 million people from all types of education circles in Indonesia, so the percentage of unemployment in Indonesia is quite large and is a factor that can be underlined. The unemployment rate will increase every year if there is no significant effort to reduce the percentage of unemployment in Indonesia. Therefore, it is necessary to realize the government in providing job opportunities in accordance with the grouping in the territory of Indonesia in order to know the unemployment cluster rate of the Indonesian population. The implementation of data mining with the clustering method in this study aims to determine the unemployment group based on the unemployment rate in the territory of Indonesia. The clustering method using the *k-means* algorithm can group Indonesia into three clusters, namely cluster 1 is a group of areas with a low unemployment rate covering 3 regions, cluster 2 is a group of regions with a moderate unemployment rate covering 10 regions, and cluster 3 is a cluster

regions with high unemployment rates covering 21 regions. The results of the analysis obtained a percentage in each cluster, namely cluster 1 has 9% area, cluster 2 has 29% area, and cluster 3 has 62% area. These results are expected to provide information to the community and local government, especially in areas that have a high unemployment rate status to be able to carry out ideas to reduce the percentage rate of unemployment in Indonesia.

Keywords: *Data mining*, *Clustering*, *K-means* Algorithm, *Unemployment*.

1. Pendahuluan

Indonesia merupakan negara dengan jumlah penduduknya mencakup nilai yang cukup besar, informasi yang diperoleh dari situs Badan Pusat Statistik (BPS) menyatakan bahwa perkembangan penduduk di setiap tahun mencakup persentase kenaikan 1%. Melihat dari perkembangan remaja, pada tahun 2021 Indonesia memiliki jumlah Sekolah Menengah Atas (SMA) sebanyak 13.495 sekolah dan Sekolah Menengah

Kejuruan (SMK) sebanyak 13.710 sekolah negeri maupun swasta. Pada tahun 2020/2021 Indonesia memiliki jumlah siswa SMA dan SMK sebanyak 10,26 juta siswa dari lulusan tersebut yang mendapatkan kesempatan untuk berkuliah hanya sekitar 1,9 juta siswa, dengan begitu persentase siswa yang siap bekerja cukup besar.

Melihat dari jumlah perguruan tinggi di Indonesia memiliki 4.517 perguruan tinggi negeri maupun swasta. Persentase lulusan mahasiswa pada tahun 2020/2021 sebesar 1,36 juta jiwa dan yang melanjutkan ke jenjang berikutnya hanya sebagian kecil dari persentase lulusan tersebut. Menurut data Badan Pusat Statistik (BPS) melaporkan jumlah pengangguran per agustus 2021 sebanyak 9,1 juta jiwa dari semua jenis kalangan pendidikan di Indonesia, dengan begitu persentase pengangguran di Indonesia cukup besar dan menjadi faktor yang dapat di garis bawah.

Mata pencarian masyarakat Indonesia sangat beragam mulai dari Petani, Nelayan, Pegawai Negeri, Buruh dan masih banyak lainnya, dengan persentase perkembangan penduduk dan sektor formasi pekerjaan sangat jauh berbeda dari jumlah lowongan pekerjaan dengan jumlah pengangguran yang ada di Indonesia. Selain faktor lowongan pekerjaan, faktor dari kompetensi calon pekerja sangatlah penting karena mematok dari persyaratan wajib dari sebuah instansi yang membuka lowongan pekerjaan.

Pemerintah Indonesia sudah banyak mengeluarkan kebijakan-kebijakan yang berupaya memutus dan mengurangi persentase pengangguran di Indonesia, salah satu upaya dari pemerintah ialah dengan adanya pelatihan-pelatihan kerja yang dapat mendongkrak soft maupun hard skill pada calon pekerja. Selain hal tersebut, pemerintah mengeluarkan kebijakan dengan kartu pra kerja yang dapat dikatakan berhasil dalam membantu penduduk yang belum bekerja dengan menerima bantuan perbulannya. Akan tetapi hal tersebut belum begitu berdampak besar menurunkan persentase pengangguran di Indonesia, 7 dari 10 orang lebih memilih bekerja di bawah sektor instansi dibandingkan dengan memilih bekerja sebagai wirausaha.

Tingkat pengangguran akan semakin meningkat di setiap tahunnya jika tidak ada upaya yang signifikan dalam mengurangi persentase pengangguran di Indonesia, faktor pekerja sangat berperan penting dalam pertumbuhan ekonomi di Indonesia oleh hal itu perlu adanya realisasi pemerintah dalam menyediakan lapangan pekerjaan sesuai dengan pemetaan, pengelompokan dan karakteristik di wilayah Indonesia agar dapat mengetahui tingkat *cluster* pengangguran penduduk Indonesia.

Implementasi Metode K-means Dalam Perbandingan Tingkat Pengangguran dengan Tenaga Kerja berdasarkan Provinsi. Penelitian ini dilakukan oleh Abdi Rahim Damanik, dkk pada tahun 2020 dengan menggunakan data pada rentang tahun 2015 sampai dengan 2018. Tenaga kerja mempunyai peranan dan kedudukan yang sangat penting sebagai pelaku dan tujuan pembangunan. Dalam melaksanakan kegiatan pembangunan dengan tujuan mensejahterahkan masyarakat. Perlu dilakukannya pengelompokan antara

persentase pengangguran dan tenaga kerja untuk mempermudah dalam pembangunan ketenagakerjaan. Untuk hasil dari pengelompokan akan dilakukan perbandingan antara persentase pengangguran dan persentase tenaga kerja. Pengelompokan antara persentase pengangguran dan tenaga kerja dilakukan menggunakan data mining dengan metode k-means. Berdasarkan hasil yang diperoleh jumlah provinsi dengan tingkat pengangguran lebih besar 57% daripada jumlah provinsi dengan tenaga kerja dengan persentase 43%. Hasil yang diperoleh diharapkan dapat menjadi masukan bagi pemerintah dan penelitian selanjutnya dalam menggunakan metode k-means.

Penentuan Cluster Optimum untuk mengelompokkan Kabupatendi Jawa Timur Berdasarkan Tingkat Pengangguran dan Tingkat Kemiskinan menggunakan Metode K-means [6]. Penelitian ini dilakukan oleh Eva Alfariani pada tahun 2019. Di Jawa Timur pengangguran dan tingkat kemiskinan mengalami perubahan setiap tahunnya di beberapa kabupaten/kota. Pengetahuan tentang kelompok wilayah yang terdapat dalam masing-masing kelompok dapat membantu pengambilan kebijakan berbeda-beda setiap tahunnya yang akan dibuat oleh pemerintah provinsi. Dari hasil penelitian yang dilakukan penerapan metode algoritma k-means berdasarkan tingkat pengangguran terbuka dan tingkat kemiskinan untuk mengelompokkan wilayah kabupaten/kota di Jawa Timur diperoleh hasil pengelompokan 3 cluster yang optimum adalah 3 cluster dengan nilai Davies bouldin index (DBI).

Hasil pengelompokan 3 kelompok pada 3 cluster C1 terdapat 11 wilayah kabupaten/kota, C2 terdapat 22 wilayah kabupaten/kota dan C3 terdapat 5 wilayah kabupaten/kota. Pada wilayah C1 tingkat pengangguran terbuka berada di atas rata-rata dan tingkat kemiskinannya berada di bawah rata-rata, wilayah C2 berada di bawah rata-rata dan tingkat kemiskinannya berada di atas rata-rata sedangkan C3 tingkat pengangguran terbuka berada di bawah rata-rata dan tingkat kemiskinannya berada di atas rata-rata.

Penerapan Clustering K-means untuk pemetaan tingkat pengangguran terbuka Provinsi Jawa Timur [7]. Penelitian ini dilakukan oleh Fajar Sarasati pada tahun 2019. Provinsi Jawa Timur merupakan salah satu provinsi di Indonesia dengan angka pengangguran yang masih tinggi meskipun jumlahnya turun. Berdasarkan data dari bappeda.jatimprov.go.id pada tahun 2018 tingkat pengangguran terbuka di Jawa Timur sebesar 3,85% meskipun jumlahnya turun namun nilainya hanya 0,15%. Namun program tersebut dinilai masih belum bisa mengatasi angka pengangguran di Jawa Timur.

Hasil dari pengolahan data pengangguran dengan menggunakan Clustering K-means dapat menghasilkan informasi kelompok wilayah mana yang paling memuat perhatian khusus dari pemerintah dalam melakukan penciptaan dan pemerataan lapangan pekerjaan serta melakukan penanganan dan mengoptimalkan program 3 in 1 yaitu pelatihan, sertifikasi serta penempatan pemagangan dan kerja sama dengan sejumlah perguruan tinggi terhadap kabupaten atau wilayah yang termasuk kelompok pengangguran tinggi. Pada tahap evaluasi dengan

menggunakan metode Davies bouldin index (DBI) pada RapidMiner Studio menghasilkan nilai Davies bouldin index (DBI) sebesar 0,676 artinya hasilnya baik karena hasilnya 0 (non-negatif).

Analisis Algoritma K-means dan K-Medoids Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional [8]. Penelitian ini dilakukan oleh Akhmad Upi Fitriyadi dan Ana Kurniawati pada tahun 2021. Penelitian ini dilakukan untuk mengukur kinerja karyawan pada perusahaan perumahan nasional setiap enam bulan sekali dengan menggunakan delapan parameter, yaitu strategi, job desc, pekerjaan, kehadiran, penampilan, agresifitas, problem solving, dan hasil kerja. Parameter yang digunakan dalam penelitian ini kemudian dikelompokkan menjadi empat cluster, yaitu cluster dengan tingkat kinerja sangat baik, cluster dengan tingkat kinerja baik, cluster dengan tingkat kinerja cukup, dan cluster dengan tingkat kinerja kurang baik.

Pada penelitian ini, penulis melakukan analisa terhadap algoritma k-means dan k-medoids. Hasil yang diperoleh dari penelitian yang dilakukan adalah algoritma k-means mempunyai tingkat accuracy 56%, presision 25% dan recall 60%, sedangkan algoritma k-medoids mempunyai tingkat accuracy 14%, presision 25% dan recall 25%. Dari hasil yang diperoleh maka dapat disimpulkan bahwa algoritma k-means memiliki performa lebih baik jika dibandingkan dengan algoritma *k-medoids*, karena algoritma *k-means* memiliki tingkat akurasi lebih tinggi daripada algoritma *k-medoids*.

2. Landasan Pemikiran

2.1. Data Mining

Data mining atau yang sering disebut juga *Knowledge Discovery in Database (KDD)* merupakan proses pengumpulan dan pengolahan data yang bertujuan untuk mengekstrak informasi penting pada kumpulan data. Proses pengumpulan dan ekstraksi informasi tersebut dapat dilakukan menggunakan perangkat lunak dengan bantuan perhitungan statistika, matematika, ataupun teknologi *Artificial Intelligence (AI)*. Jenis-jenis data mining berdasarkan tugas yang dapat dilakukan, yaitu:

a. Descriptive

Merupakan suatu fungsi bertujuan untuk mengidentifikasi pola yang muncul secara berulang pada suatu data dan mengubah pola tersebut menjadi aturan dan kriteria yang mudah dimengerti oleh para ahli pada domain aplikasinya. Aturan yang dihasilkan harus mudah dimengerti agar dapat dengan efektif meningkatkan tingkat pengetahuan (*knowledge*) pada sistem. Tugas deskriptif merupakan tugas data mining yang sering dibutuhkan pada teknik *postprocessing* untuk melakukan validasi dan menjelaskan hasil dari proses data mining. *Postprocessing* merupakan proses yang digunakan untuk memastikan hanya hasil yang valid dan berguna yang dapat digunakan oleh pihak yang berkepentingan.

b. Classification

Fungsi ini bertujuan untuk menyimpulkan beberapa definisi karakteristik dari sebuah grup. Contoh:

pelanggan-pelanggan perusahaan yang telah berpindah akibat kesaignan perusahaan yang lain.

c. Predictive

Fungsi ini memiliki kemiripan dengan klasifikasi, akan tetapi data diklasifikasikan berdasarkan perilaku atau nilai yang diperkirakan pada masa yang akan datang. Contoh dari tugas prediksi misalnya untuk memprediksikan adanya pengurangan jumlah pelanggan dalam waktu dekat dan prediksi harga saham dalam tiga bulan yang akan datang.

d. Clustering

Merupakan identifikasi kelompok dari produk-produk atau barang-barang yang memiliki kesamaan karakteristik khusus.

e. 5. Association

Merupakan identifikasi hubungan dari kejadian-kejadian yang sudah terjadi di suatu waktu.

f. Sequencing

Sebetulnya hampir sama dengan association tetapi untuk *sequencing* berfungsi untuk identifikasi hubungan-hubungan berbeda di sebuah periode waktu tertentu. Contohnya, para pelanggan yang berkunjung di supermarket secara berulang.

g. Forecasting

Fungsi ini bertujuan untuk memperkirakan nilai di suatu masa di masa mendatang sesuai dengan pola-pola dengan kumpulan data dalam jumlah besar. Contohnya, peralaman permintaan pasar.

2.2. Pengertian Clustering

Clustering merupakan metode untuk mengelompokkan data dengan cara mengekstrak pola yang menarik dari sejumlah *cluster* data yang besar. Metode ini sering digunakan untuk keperluan *intelijen* bisnis, pengenalan pola gambar, pencarian web, ilmu kehidupan, dan keamanan. Cara kerja *clustering* memiliki beberapa proses pengelompokan data menjadi beberapa jumlah *cluster*, sehingga data dalam *cluster* tersebut memiliki kemiripan yang paling besar yang juga memungkinkan untuk menentukan dan mengambil data antara *cluster* berbeda yang menunjukkan kesamaan minimal, dapat juga digunakan untuk menentukan dalam mengidentifikasi kelompok data *cluster* yang dihasilkan dari pengelompokan item-item kecil berdasarkan kesamaannya.

Macam-macam pendekatan pada metode *clustering*:

a. Partitioning clustering

Merupakan pengelompokan menggunakan pendekatan partisi hierarki, atau sering disebut sebagai pengelompokan berbasis partisi, mengelompokkan data dengan menyortir data yang dianalisis ke dalam beberapa cluster, yang ada hierarki data tidak dipertimbangkan. Pada metode *split clustering*, setiap cluster memiliki titik pusat (*center of gravity*), dan metode ini secara umum ditujukan untuk menentukan dalam meminimalkan jarak (*dissimilarity*) dari semua data dengan point ke pusat setiap cluster. Terdapat beberapa algoritma yang sering digunakan untuk metode ini seperti *K-Means*, *K-Medoids*, dan *Fuzzy C-Means Clustering*.

b. Hierarchical clustering

Merupakan pengelompokan menggunakan pendekatan hirarki, membuat hierarki dalam bentuk diagram dimana data yang sejenis ditempatkan pada hierarki yang berdekatan dan bukan pada hierarki yang terpisah. Pengelompokan hierarki mengelompokkan data dalam grafik hierarki dengan mengelompokkan dua grup terdekat bersama-sama setiap kali seluruh kumpulan data diulang atau dibagi menjadi beberapa cluster. Langkah untuk melakukan hirarki *clustering* terlebih dahulu mengidentifikasi elemen-elemen yang berdekatan, kemudian menggabungkan elemen-elemen tersebut menjadi satu cluster, selanjutnya menghitung jarak antar cluster, dan akhirnya semua terhubung, ulangi dari awal sampai selesai. Contoh teknik pengelompokan hierarki termasuk tautan tunggal, tautan penuh, tautan rata-rata, dan tautan kelompok rata-rata. Terdapat beberapa metode yang sering digunakan yaitu *single linkage*, *complete linkage*, *average linkage*, dan *ward's minimum variance*.

2.3. Algoritma K-Means

Algoritma *K-Means* adalah salah satu algoritma *unsupervised learning* yang termasuk dalam analisis kluster (*cluster analysis*) *non hirarki* yang digunakan untuk mengelompokkan data berdasarkan *variabel* atau *feature* dengan tujuan untuk mendapatkan kelompok data dengan memaksimalkan kesamaan karakteristik dalam kluster dan memaksimalkan perbedaan antar kluster. Algoritma *K-means* mengelompokkan data berdasarkan jarak antara data terhadap titik *centroid* kluster yang didapatkan melalui proses berulang. Analisis perlu menentukan jumlah K (banyaknya kluster) sebagai input algoritma.

2.4. Rapid Miner

RapidMiner adalah perangkat lunak yang berfungsi sebagai alat pembelajaran dalam ilmu data mining. Platform dikembangkan oleh perusahaan yang didedikasikan untuk semua langkah yang melibatkan sejumlah besar data dalam bisnis komersial, penelitian, pendidikan, pelatihan, dan pembelajaran. RapidMiner memiliki sekitar 100 solusi pembelajaran untuk pengelompokan, klasifikasi dan analisis regresi. RapidMiner juga mendukung sekitar 22 format file, seperti .xls, .csv, dan sebagainya.

RapidMiner membawa kecerdasan buatan kepada perusahaan melalui platform ilmu data yang terbuka dan dapat diskalakan. RapidMiner dibangun untuk tim analisis, mengintegrasikan seluruh siklus ilmu data, dari persiapan data hingga pembelajaran mesin hingga penyebaran model prediksi. Lebih dari 625.000 profesional analitis menggunakan produk RapidMiner untuk meningkatkan pendapatan, mengurangi biaya, dan menghindari risiko.

RapidMiner adalah perangkat lunak independen yang digunakan untuk menganalisa data dan mesin penambangan data, yang dapat diintegrasikan dengan berbagai bahasa pemrograman secara mudah. RapidMiner ditulis dengan Bahasa pemrograman Java, sehingga dapat berkerja pada banyak sistem operasi. RapidMiner menyediakan UI untuk mendesain pipa analisis, di mana akan

menghasilkan file XML yang dapat menjelaskan proses analisis yang ingin diterapkan oleh pengguna ke data. RapidMiner akan membaca file ini untuk menjalankan analisa secara otomatis. RapidMiner menyediakan tampilan (UI) yang ramah pengguna, sehingga memudahkan pengguna saat menggunakannya. Tampilan yang terdapat pada RapidMiner disebut *Perspective*. Terdapat 3 *Perspective*, yaitu *Welcome Perspective*, *Design Perspective* dan *Result Perspective*.

3. Methodologi Penelitian

3.1. Jenis Data

Jenis data dalam sebuah penelitian dapat dibedakan berdasarkan sifat dan sumbernya. Pada penelitian ini, data yang digunakan adalah data kuantitatif, yakni data yang dalam bentuk angka atau bilangan yang dapat dioperasikan dengan perhitungan matematika. Data tersebut seringkali dijadikan sebagai dasar untuk permasalahan dalam sebuah penelitian yang bersifat statistik. Data kuantitatif dibedakan menjadi dua, yaitu data diskrit dan data kontinu. Data diskrit juga dapat disebut sebagai data nominal yang diperoleh dari hasil perhitungan. Sedangkan data kontinu dapat dibedakan menjadi tiga, yaitu data ordinal, interval, dan rasio. Jenis data berdasarkan sumber pada penelitian ini merupakan data sekunder yang diperoleh dari situs BPS pada tahun 2021.

3.2. Variabel Penelitian

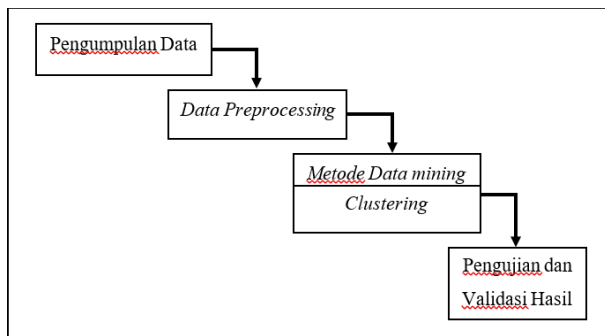
Variabel penelitian adalah suatu atribut atau sifat atau nilai dari orang, objek, organisasi, atau kegiatan yang mempunyai variasi tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian ditarik kesimpulannya [24]. Pada penelitian ini, variabel yang digunakan adalah variabel yang menjadi indikator pengangguran. Indikator-indikator tersebut meliputi:

- a. Penduduk Miskin (PM)
- b. Usia Muda 15-24 tahun (UM)
- c. Tingkat Pengangguran Terbuka (TPT)
- d. Tingkat Setengah Pengangguran (TSP)

3.3. Data Yang Digunakan

Data yang digunakan dalam penelitian ini merupakan data dengan variabel yang menjadi indikator dalam masalah pengangguran yang terjadi di Indonesia. Data tersebut merupakan data sekunder milik Badan Pusat Statistik (BPS) berupa persentase dalam bentuk tabel yang kemudian diolah dalam penelitian ini dengan penerapan metode *clustering* dengan rumus jarak terdekat atau *Euclidean distance* untuk menghasilkan beberapa kelompok data pengangguran terbuka di Indonesia.

3.4. Metode Penelitian



Gambar 1. Tahapan Penelitian

3.5. Metode Pengumpulan Data

Metode pengumpulan data dapat dikatakan sebagai dataset collection yang merupakan langkah pertama untuk menghasilkan data yang akan diteliti. Metode pengumpulan data dalam penelitian ini adalah:

a. Studi Dokumen

Studi dokumen merupakan sebuah teknik pengumpulan data yang sumbernya dari dokumen yang sudah ada. Dokumen tersebut pada umumnya dihasilkan oleh lembaga resmi seperti Badan Pusat Statistik (BPS) yang dipublikasikan sehingga data dapat dengan mudah diakses dan diolah sesuai kebutuhan dalam penelitian.

b. Studi Pustaka

Studi pustaka merupakan teknik pengumpulan data dengan cara mencari sumber-sumber yang berkaitan dengan judul penelitian, sumber yang dimaksud dalam teknik studi pustaka dapat berupa jurnal, buku, berita, ataupun artikel yang berhubungan dengan penelitian yang sedang dilakukan.

3.6. Data Preprocessing

Data preprocessing merupakan sebuah tahapan dalam pengolahan dataset yang bertujuan untuk meminimalisir atau menghilangkan *noise*, *missing value*, dan data yang tidak konsisten. Data yang telah diperoleh pada tahapan pengumpulan data akan dipahami dan diidentifikasi yang selanjutnya data tersebut akan diolah pada tahapan *data preprocessing*. Pada penelitian ini, *data preprocessing* yang diterapkan adalah transformasi data, yaitu perubahan yang dilakukan pada data provinsi yang diringkas menjadi bentuk ordinal dan persentase menjadi bentuk desimal agar data mudah untuk dikelola.

4. Pengolahan Data

Pengolahan data pada penelitian ini diawali oleh pengumpulan data yang akan digunakan sebagai dataset penelitian, selanjutnya data diolah dan diuji dengan algoritma *k-means* menggunakan rumus *Euclidean distance* atau rumus jarak terdekat dan *RapidMiner Studio*.

4.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari situs resmi Badan Pusat Statistik (BPS) mengenai indikator pengangguran berdasarkan beberapa variabel, yaitu Penduduk Miskin, Usia Muda >15 tahun, Tingkat Pengangguran Terbuka, dan Tingkat Setengah Pengangguran, data-data tersebut merupakan data dalam bentuk persentase pada tahun 2021.

4.2. Transformasi Data

Data yang telah diperoleh pada Tabel 4.1 adalah data awal yang akan ditransformasi atau dilakukan perubahan pada setiap variabelnya. Transformasi yang dilakukan adalah meringkas data atau perubahan data menjadi bentuk ordinal dan nilai persentase yang diubah menjadi nilai desimal agar memudahkan dalam proses kelola dataset. Di bawah ini terdapat Tabel 4.2 dan Tabel 4.3 yang merupakan tabel proses dan tabel hasil transformasi data pada setiap variabel yang akan digunakan sebagai dataset yang diolah.

4.3. Pembahasan Hasil Pengujian

Hasil pengujian diperoleh dari data awal yang telah melalui proses transformasi data yang kemudian menjadi sebuah dataset dan diolah menggunakan perhitungan manual dengan algoritma *k-means* dan rumus *Euclidean distance* atau rumus jarak terdekat dan diuji dengan menggunakan *tools RapidMiner Studio*.

1. Penutup

Penelitian yang telah dilakukan adalah pengelompokan pengangguran terbuka di Indonesia dengan menggunakan metode *clustering* dan algoritma *k-means* dengan rumus *euclidean distance* atau rumus jarak terdekat dapat menghasilkan informasi mengenai kelompok pengangguran terbuka di Indonesia dengan menghasilkan tiga *cluster*, yaitu *cluster 1* adalah kelompok wilayah dengan tingkat pengangguran rendah yang meliputi 3 wilayah, *cluster 2* adalah kelompok wilayah dengan tingkat pengangguran sedang yang meliputi 10 wilayah, dan *cluster 3* adalah kelompok wilayah dengan tingkat pengangguran tinggi yang meliputi 21 wilayah. Hasil analisa memperoleh persentase pada setiap *cluster*, yaitu *cluster 1* memiliki 9% wilayah, *cluster 2* memiliki 29% wilayah, dan *cluster 3* memiliki 62% wilayah.

Hasil tersebut diharapkan dapat memberikan informasi kepada masyarakat dan pemerintah setempat khususnya di wilayah yang memiliki status tingkat pengangguran tinggi untuk dapat melakukan gagasan-gagasan guna mengurangi tingkat persentase pengangguran di Indonesia.

Daftar Pustaka

- [1] A. Hari, Alat Evaluasi Keterampilan Bermain Bola Basket, Jakarta: Dipdiknas: Jurnal Nasional Pendidikan Jasmani dan Ilmu Keolahragaan, 2003.
- [2] A. P. Mangunegara, Evaluasi Kinerja SDM, Kota Bandung: Refika Aditama, 2010.
- [3] Badan Pusat Statistik (BPS), "Keadaan Ketenagakerjaan Indonesia Agustus 2021," Jakarta, 2021.
- [4] E. Alfariani, "Penentuan Cluster Optimum untuk mengelompokan Kabupatendi Jawa Timur Berdasarkan Tingkat Pengangguran dan Tingkat Kemiskinan menggunakan Metode K-means," 2019.

- [5] F. Sarasati, "Penerapan Clustering K-means untuk pemetaan tingkat pengangguran terbuka Provinsi Jawa Timur," 2019.
- [6] Kementerian Sekretariat Negara, "Peraturan Presiden (PERPRES) tentang Perubahan atas Peraturan Presiden Nomor 36 Tahun 2020 tentang Pengembangan Kompetensi Kerja Melalui Program Kartu Prakerja," Jakarta, 2020.
- [7] M. LJ, Management and Organisational Behaviour, Prentice hall, 2005
- [8] PDDikti - Pangkalan Data Pendidikan Tinggi, "Statistik Perguruan Tinggi," 2020.
- [9] Pusat Data dan Teknologi Informasi, "Statistik Pendidikan," 2021.
- [10] R. Damanik dkk, "Implementasi Metode K-means Dalam Perbandingan Tingkat Pengangguran dengan Tenaga Kerja berdasarkan Provinsi," 2020.
- [11] U. A. Fitriyadi and A. Kurniawati, "Analisis Algoritma K-means dan K-Medoids Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional," 2021.