

PENERAPAN ALGORITMA *K-MEANS CLUSTERING* UNTUK MENGELOMPOKKAN PENERIMA BANTUAN SOSIAL TUNAI (BST) DI JAWA BARAT

Indri Dinda Anjani¹, Agus Bahtiar²

¹ Komputerisasi Akuntansi, ² Sistem Informasi STMIK IKMI Cirebon
Jalan Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon
indridinda413@gmail.com

ABSTRAK

Bantuan Sosial Tunai (BST) adalah program bantuan dari pemerintah yang diberikan kepada masyarakat yang kurang mampu dengan tujuan untuk membantu meringankan beban keluarga penerima BST tersebut. Namun bantuan yang telah diterima tidak sesuai dengan apa yang diharapkan, masih ada beberapa keluarga yang tidak mendapatkan manfaat dari adanya BST ini. Karena data yang didapatkan tidak akurat atau tidak lengkap, ditemukan beberapa kasus penerima BST yang tidak tepat sasaran. Analisis pengelompokan penerima Bantuan Sosial Tunai (BST) berdasarkan lokasi dengan metode *K-Means Clustering* memiliki manfaat besar dalam meningkatkan efisiensi alokasi sumber daya dan menentukan penerima yang tepat sasaran di Jawa Barat. Dengan memahami pola distribusi penerima BST secara geografis, program bantuan sosial dapat dijalankan dengan lebih efektif untuk meningkatkan kesejahteraan sosial dan mengurangi tingkat kemiskinan di daerah tersebut. Data penelitian ini diambil dari opendata.jabarprov.go.id. Data ini mencakup lokasi penerima, jumlah penerima dan data pendukung lainnya. Kemudian pengelompokan akan dilakukan dengan *K-Means Clustering* menggunakan RapidMiner. Dengan adanya penelitian ini bisa menambah pengetahuan dalam menganalisis data dengan menggunakan metode *K-Means Clustering* dalam pengambilan keputusan yang lebih bijak. Diharapkan dengan adanya penelitian ini bisa membantu pemerintah dalam mengalokasikan BST tersebut agar lebih efisien dan tepat sasaran kepada masyarakat yang benar-benar membutuhkan. Hasil analisis clustering dapat mengidentifikasi pola distribusi penerima BST berdasarkan dan lokasi geografis keluarga penerima BST. Penelitian ini bisa memberikan dampak positif bagi pemerintah atau lembaga terkait agar bisa lebih efisien dalam mendistribusikan BST agar lebih tepat sasaran.

Kata Kunci : *K-Means Clustering, Data Mining, Penerima BST*

1. PENDAHULUAN

Bantuan Sosial Tunai (BST) adalah program yang paling penting dalam upaya pemerintah untuk meningkatkan kesejahteraan sosial dan mengurangi kemiskinan. Untuk memastikan efektifitasnya, perlu pemahaman yang mendalam tentang bagaimana bantuan ini didistribusikan. Ditengah keterbatasan sumber daya efisiensi dalam alokasi menjadi sangat penting. Menganalisis distribusi penerima BST dengan pendekatan data mining seperti *K-Means Clustering* dapat membantu dalam merancang program yang lebih efisien. *K-Means Clustering* adalah salah satu alat yang lebih efektif dalam analisis data ini. Analisis data menjadi semakin relevan dalam pemahaman masalah sosial.

Penerapan metode *K-Means Clustering* analisis data pengelompokan penerima BST untuk memahami pola distribusi penerima BST lokasi geografis keluarga tersebut. Teknik analisis data ini bisa digunakan untuk peningkatan efisiensi dalam alokasi sumber daya penerima BST lebih tepat sasaran dan untuk memastikan bahwa bantuan sosial didistribusikan dengan cara yang lebih efisien dan sesuai dengan kebutuhan penerima.

Penerapan algoritma *K-Means Clustering* banyak digunakan untuk beberapa penelitian. Salah satunya yang dikutip dari KOPERTIP: Jurnal Ilmiah Manajemen Informatika dan Komputer pada tahun 2020. Yang disusun oleh Anwar Pauji, Siti Aisyah,

Agus Surip, Riko Saputra, dan Irfan Ali dengan judul "Implementasi Algoritma *K-Nearest* Dalam Menentukan Penerima Bantuan Langsung Tunai". Analisis mendalam terhadap pola karakteristik penerima yang tidak tepat sasaran seperti identifikasi faktor-faktor yang memengaruhi ketidaksesuaian tersebut, kesalahan dalam pengumpulan data, perubahan status ekonomi, atau perubahan dalam komposisi rumah tangga. Dengan memahami pola ini, dapat dilakukan penyesuaian yang lebih baik dalam penentuan penerima [1].

Masalah yang terdapat pada penelitian tersebut yaitu pada pelaksanaan proses penentuan data dan pemberian bantuan kepada penerima bantuan tersebut tidak sesuai dengan apa yang diharapkan. Kemudian dilakukan percobaan klasifikasi penerima manfaat menggunakan metode *K-Nearest Neighbor* sehingga bantuan tersebut benar-benar tepat sasaran. Berdasarkan hasil penelitian tersebut dapat disimpulkan bahwa algoritma *K-Nearest Neighbor* bisa digunakan untuk membantu pemerintah dalam pengambilan keputusan yang lebih baik.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Menurut R. Harun, K. C. Pelangi, and Y. Lasena, *Data mining* merupakan suatu proses menggali data dari pengetahuan yang berbeda dan menghasilkan sebuah informasi penting yang dapat dipakai dalam

meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya [2]. *Data mining* adalah proses pencarian informasi yang berguna atau pola yang menarik dari sekumpulan data yang besar. *Data mining* digunakan untuk melakukan analisis dan memberikan informasi berguna bagi keperluan bisnis.

Data mining yang biasa disebut sebagai *knowledge discovery in database (KDD)* merupakan kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola hubungan dalam himpunan data yang berukuran besar [3].

2.2. Clustering

Menurut Budi Santosa dan A. Umam, *Clustering* adalah proses membagi-bagi obyek dari suatu set data menjadi beberapa klaster yang homogen. Tujuan utama dari metode klaster adalah pengelompokkan sejumlah data/obyek kedalam klaster (grup) sehingga dalam setiap klaster akan berisi data yang semirip mungkin. Clustering berusaha untuk menenpatkan obyek yang mirip atau yang jaraknya berdekatan dalam satu klaster dan membuat jarak antar klaster sejauh mungkin. Ini berarti obyek dalam satu klaster sangat mirip satu sama lain dan berbeda dengan obyek dalam klaster-klaster yang lain [2].

2.3. Metode K-Means Clustering

Algoritma K-means Cluster Analysis merupakan bidang penelitian dalam analisis dan data mining. Pada algoritma ini teknik pengelompokannya berdasarkan kemiripan data yang tidak memiliki acuan apapun (unsupervised). Tetapi, akan membagi keseluruhan data yang akan menjadi kelompok atau mempunyai kemiripan yang sama. Pada dasarnya algoritma ini menghitung jarak pada setiap data dengan pusat data (centroid) untuk mengukur kemiripan data. Metode ini digunakan bertujuan untuk meminimalisir fungsi objektif yang diatur pada proses clustering atau pengklasteran dengan meminimalkan variasi antar data yang terdapat dalam suatu cluster dan memaksimalkan variasi dengan data yang terdapat pada cluster lain [4].

2.4. RapidMiner

Menurut H. Rizqifalufhi & M. A. Yaqin *RapidMiner* adalah platform perangkat lunak data ilmu pengetahuan yang dikembangkan oleh perusahaan dengan nama yang sama, yang menyediakan lingkungan terpadu untuk pembelajaran mesin (*machine learning*), pembelajaran mendalam (*deep learning*), penambangan teks (*text mining*), dan analisis prediktif (*predictive analytics*). Aplikasi ini digunakan untuk aplikasi bisnis dan komersial serta untuk penelitian, pendidikan, pelatihan, pembuatan prototype dengan cepat, dan pengembangan aplikasi serta mendukung semua langkah proses pembelajaran mesin termasuk persiapan data, visualisasi hasil, validasi dan pengoptimalan. *RapidMiner* dikembangkan dengan model open core [5].

2.5. Bantuan Sosial Tunai (BST)

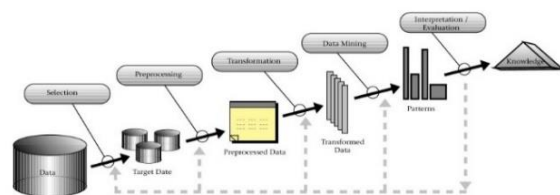
BST adalah sebuah program yang dijalankan oleh pemerintah yang diberikan kepada masyarakat berupa uang tunai untuk membantu perekonomian masyarakat yang memiliki tingkat ekonomi rendah. Salah satu kebijakan yang diambil oleh Pemerintah untuk menjaga daya beli masyarakat dan mencegah peningkatan angka kemiskinan adalah melalui Kebijakan Bantuan Sosial Tunai (BST) yang dikeluarkan oleh oleh Kementerian Sosial (Kemensos) Republik Indonesia bagi masyarakat miskin dan terdampak dari adanya pandemi Covid-19.

Melalui Keputusan Menteri Sosial (Kepmensos) Republik Indonesia Nomor 54/HUK/2020 tentang Pelaksanaan Bantuan Sosial Tunai dalam Penanganan Dampak analiCovid-19. Berdasarkan Kepmen tersebut, Bantuan Sosial Tunai diberikan berupa uang tunai kepada 9 (sembilan) juta keluarga miskin, tidak mampu, dan keluarga rentan yang terkena dampak dari adanya pandemi Covid-19 yang tersebar di 33 provinsi di Indonesia. Adapun nilai bantuan yaitu sebesar Rp. 600.000,- /bulan. Bantuan ini diberikan selama 3 (tiga) bulan berturut-turut mulai April hingga Juni 2020 [6].

2.6. Analisis Cluster

Analisis cluster atau analisis kelompok adalah teknik analisis data untuk mengelompokkan individu atau objek kedalam kelompok yang memiliki karakteristik yang berbeda antara setiap kelompoknya. Analisis cluster termasuk dalam analisis statistik multivariat metode interdependen, dan oleh karena itu tujuan analisis cluster tidak untuk menghubungkan ataupun membedakan dengan sampel/variabel lain. Analisis cluster berguna untuk meringkas data dengan jalan mengelompokkan objek-objek berdasarkan kesamaan karakteristik tertentu diantara objek-objek yang akan diteliti [7].

3. METODE PENELITIAN



Gambar 1. Rancangan Penelitian

Metodologi penelitian adalah proses untuk menganalisis data dan mengekstrak informasi yang berguna. Hasil analisis dapat digunakan untuk membantu dalam mengambil keputusan. Metode yang digunakan dalam penelitian ini yaitu *Knowledge Discovery in Databases (KDD)* yang terdiri dari

5 (lima) tahapan yaitu :

a. Data Selection

Pada proses ini data yang digunakan diambil dari Open Data Jabar, sebuah website pemerintah yang sudah resmi. Data tersebut sudah pasti dan bisa

diandalkan dalam penelitian ini. Dataset yang digunakan yaitu data jumlah alokasi dan realisasi keluarga penerima manfaat Bantuan Sosial Tunai (BST) berdasarkan tahap penyaluran dan kabupaten/kota di Jawa Barat dari tahun 2020 s.d 2021.

b. *Pre-Processing*

Setelah pemilihan data, selanjutnya dilakukan tahapan untuk menghilangkan beberapa data bermasalah yang dapat menghambat pemrosesan data. Proses *preprocessing* meliputi pembersihan duplikat data, data yang tidak konsisten dan perbaikan kesalahan pada data.

c. *Transformation Data*

Tranformasi data adalah proses perubahan data agar data dapat digunakan dalam proses *Data Mining*. Data disusun kedalam format yang dapat dianalisis untuk mendukung proses pengambilan keputusan. Kemudian ketiga data tersebut digabungkan menjadi satu sehingga dapat dilakukan proses *Data Mining*.

d. *Data Mining*

Data mining adalah proses pencarian informasi yang berguna atau pola yang menarik dari sekumpulan data yang besar. *Data mining* digunakan untuk melakukan analisis dan memberikan informasi berguna bagi keperluan bisnis. Teknik data mining ini meliputi klasterisasi, klasifikasi, asosiasi dan regresi.

e. *Data Evaluation*

Data evaluation adalah proses penilaian terhadap suatu program atau aktivitas yang melibatkan pengumpulan informasi dan hasil dari program tertentu. Tujuannya untuk membuat penilaian tentang sebuah program, meningkatkan efektivitas pengelolaan data dan memastikan data yang digunakan dalam proses analisis adalah data yang tepat dan relevan.

Implementasi algoritma *K-Means Clustering* pada Alplikasi RapidMiner :

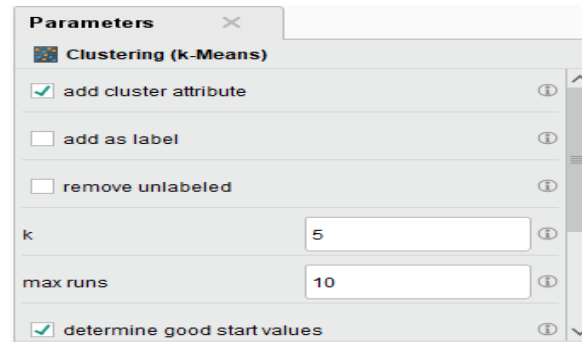
3.1. Data Sebelum Preprocessing

Gambar 2. Data Sebelum Preprocessing

Pada gambar 2 adalah data sebelum preprocessing, data yang diperoleh dari data penduduk penerima BST dari tahun 2020 sampai 2021. Pada

atribut Nama Kabupaten/Kota di dijadikan id agar memudahkan dalam proses *Clustering*

3.2. Menentukan Jumlah Cluster



Gambar 3. Menentukan Jumlah Cluster

Pada gambar 3 adalah cara menentukan nilai k, nilai k ditentukan dengan mengambil nilai *Davies Bouldin Index* (DBI) yang paling kecil. Dalam dataset ini nilai yang paling kecil adalah k = 5, dengan *max run* = 10. Perbandingan nilai DBI dapat dilihat pada tabel dibawah ini :

Tabel 1. Tabel Perbandingan Nilai DBI

Tabel Perbandingan BDI			
Nilai K	Max Run	Max Optimization Steps	Davies Bouldin Index
2	10	100	-0.509
3	10	100	-0.571
4	10	100	-0.439
5	10	100	-0.433
6	10	100	-0.455
7	10	100	-0.486
8	10	100	-0.464
9	10	100	-0.457
10	10	100	-0.475

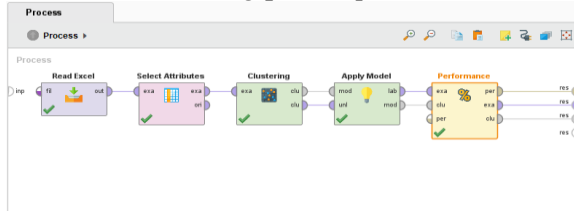
Hasil dari penentuan nilai DBI yaitu *Cluster 2* -0.509, *Cluster 3* memiliki nilai sebesar -0.571, *cluster 4* memiliki nilai -0.439, *cluster 5* memiliki nilai sebesar -0.433. Nilai k yang paling kecil adalah cluster 5.

3.3. Data Preprocessing

Gambar 4. Data PreProcessing

Pada gambar 4. adalah data yang telah dilakukan proses pemilihan atribut yang diperlukan dalam proses pengolahan data dalam mengelompokkan penerima BST berdasarkan lokasi peerima.

3.4. Proses Clustering pada Rapidminer



Gambar 5. Proses Clustering pada RapidMiner

Pada gambar 5. adalah proses pengolahan data menggunakan algoritma *K-Means Clustering* pada aplikasi *RapidMiner* ini menggunakan lima (5) operator yaitu *Read Excel* untuk membaca dataset jumlah penerima BST di Jawa Barat. *Select Attribute* untuk memilih atribut apa saja yang diperlukan untuk proses *clustering*, jumlah atribut yang digunakan pada proses pengelompokkan ini terdiri dari Id, Jumlah Penerima, Kode Kabupaten Kota, Kode Provinsi, Nama Kabupaten Kota dan Tahun. *K-Means Clustering* untuk mengelompokkan dataset dengan jumlah $K=5$ dan $Max\ Run=10$. *Apply Model* digunakan untuk membaca kumpulan data dalam cluster dari data yang telah dianalisis. *Cluster Distance Performance* digunakan untuk mengukur kinerja algoritma *K-Means Clustering* pada dataset yang telah dibaca.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Clustering

Row No.	nama_kabu...	cluster	id	kode_provinsi	kode_kabup...	jumlah_pen...	tahun
1	KABUPATEN ...	cluster_2	1	32	3201	107065	2020
2	KABUPATEN ...	cluster_2	2	32	3201	105230	2020
3	KABUPATEN ...	cluster_2	3	32	3201	107170	2020
4	KABUPATEN ...	cluster_2	4	32	3201	105441	2020
5	KABUPATEN ...	cluster_2	5	32	3201	107510	2020
6	KABUPATEN ...	cluster_2	6	32	3201	105335	2020
7	KABUPATEN ...	cluster_2	7	32	3201	95935	2020
8	KABUPATEN ...	cluster_2	8	32	3201	93995	2020
9	KABUPATEN ...	cluster_2	9	32	3201	95935	2020
10	KABUPATEN ...	cluster_2	10	32	3201	93994	2020
11	KABUPATEN ...	cluster_2	11	32	3201	97044	2020
12	KABUPATEN ...	cluster_2	12	32	3201	94914	2020
13	KABUPATEN ...	cluster_2	13	32	3201	97642	2020
14	KABUPATEN ...	cluster_2	14	32	3201	95522	2020

Gambar 6. Data Hasil Clustering Pada RapidMiner

Pada gambar 6 adalah data hasil *Clustering* pada *RapidMiner* yang dilihat pada bagian *ExampleSet* (*Apply Model*).

4.2. Cluster Model

Pada gambar 7 hasil dari penelitian ini adalah menghasilkan lima (5) *cluster* dengan jumlah *item* yang berbeda-beda. *Cluster 0* memiliki 170 *items*, *cluster 1* memiliki 44 *items*, *cluster 2* memiliki 153 *items*,

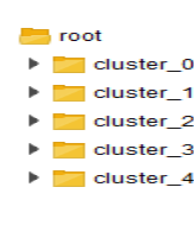
cluster 3 memiliki 281 *items*, *cluster 4* memiliki 36 *items*. Total *items* secara keseluruhan yaitu 684 *items*.

Cluster Model

```
Cluster 0: 170 items
Cluster 1: 44 items
Cluster 2: 153 items
Cluster 3: 281 items
Cluster 4: 36 items
Total number of items: 684
```

Gambar 7. Cluster Model

4.3. Folder View



Gambar 8. Folder View

Pada gambar 8 adalah tampilan untuk melihat atau mengelola file pada suatu sistem atau aplikasi, didalamnya berisi beberapa *cluster*.

5. KESIMPULAN DAN SARAN

Kesimpulan dari hasil penelitian ini yaitu : Analisis pengelompokkan menggunakan metode *K-Means Clustering* memberikan manfaat yang besar dalam meningkatkan efisiensi alokasi sumber daya dan menemukan penerima BST yang lebih tepat sasaran. Program Bantuan Sosial Tunai (BST) dapat dijalankan dengan lebih efektif dan dapat meningkatkan kesejahteraan sosial dan mengurangi tingkat kemiskinan.

Adapun saran bagi peneliti selanjutnya harus lebih dipahami dulu tentang data apa yang akan diteliti dan memahami lebih mendalam tentang metode yang akan digunakan dalam penelitian, dan sebaiknya menggunakan metode yang lain selain *K-Means Clustering*. Melakukan eksplorasi terhadap berbagai jenis metode analisis yang lainnya yang lebih canggih untuk menambah pengetahuan dan pemahaman yang lebih mendalam tentang metode analisis.

DAFTAR PUSTAKA

[1] T. Ramayanti, E. Haerani, J. Jasril, and L. Oktavia, "Penerapan Algoritma K-Medoids Pada Clustering Penerima Bantuan Pangan Non Tunai (BPNT)," *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1287–1296, 2023, doi: 10.30865/mib.v7i3.6475.

[2] A. Bahauddin, A. Fatmawati, and F. Permata Sari, "Analisis Clustering Provinsi Di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means," *J. Manaj. Inform. dan Sist. Inf.*, vol. 4, no. 1, pp. 1–8, 2021, doi: 10.36595/misi.v4i1.216.

- [3] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- [4] A. Rohmah, F. Sembiring, and A. Erfina, "Implementasi Algoritma K-Means Clustering Analysis untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: SMK Yaspim Gegerbitung)," *Semin. Nas. Sist. Inf. dan Manaj. Inform.*, pp. 290–298, 2021, [Online]. Available: <https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/32>
- [5] R. Nofitri and N. Irawati, "Analisis Data Hasil Keuntungan Menggunakan Software Rapidminer," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 5, no. 2, pp. 199–204, 2019, doi: 10.33330/jurteks.v5i2.365.
- [6] I. Gemiharto and E. R. Juningsih, "Komunikasi pemerintahan dalam implementasi kebijakan Bantuan Sosial Tunai di kabupaten Bandung," *J. Manaj. Komun.*, vol. 6, no. 1, p. 1, 2021, doi: 10.24198/jmk.v6i1.35510.
- [7] R. Sitepu and B. Gultom, "Clustering Analysis for Air Pollution Level on Industrial Sector in South Sumatera," *J. Penelit. Sains*, vol. 14, no. 3, pp. 11–17, 2011.