

Analisa Penumpang Bus Transjakarta dengan Metode K-Means

Mohammad Farras Daffauzan

Program Studi Teknik Informatika,
Fakultas Ilmu Komputer,
Universitas Mercu Buana

*Corresponding author email:
arrasaffauzan2gmail.com

Manuscript history:

Diterima
Direvisi
Diterima untuk terbit

ABSTRAK

Data penumpang bus transjakarta merupakan data yang penting untuk menganalisa perilaku pengguna costumer dan pelayanan. Penelitian ini bertujuan untuk mengklasifikasikan data penumpang bus transjakarta dengan metode algoritma K-Means. Metodologi yang digunakan dalam penelitian ini adalah pengumpulan data, analisis data, implementasi algoritma K-Means, pengujian, dan validasi. Dataset yang digunakan dalam penelitian ini adalah dataset sebuah bus transjakarta yang mencakup trayek dan jumlah penumpang dibulan Desember. Hasil penelitian menunjukkan bahwa algoritma K-Means Weka dapat menghasilkan kelompok – kelompok costumer menggunakan jenis angkutan mikotrans. Hasil ini dapat memberikan bagi pihak transjakarta untuk lebih banyak menyediakan kendaraan bus agar tidak terjadinya penumpukan penumpang.

Kata Kunci: Analisa Data, Data Mining, Clustering, Algoritma K-Means, Transjakarta.

I. Pendahuluan

Data Penumpang Bus Transjakarta merupakan daya yang penting pihak transjakarta untuk menganalisa pengguna para costumer. Salah satu sektor kendaraan umum dengan banyaknya peminat salah satunya menggunakan bus transjakarta. Analisa data Bus Transjakarta dapat membantu bisnis dalam meningkatkan strategi pelayanan mereka, mengidentifikasi banyak pengguna, dan memprediksi penumpukan di halte. Salah satu algoritma pembelajaran mesin yang dapat digunakan untuk mengklasifikasikan data bus transjakarta adalah algoritma K-Means (Larose & Larose, 2014).. Algoritma K-Means adalah salah satu algoritma yang dapat digunakan untuk menganalisa data yang membagi data ke dalam kelompok atau kluster berdasarkan jenis kendaraan yang sering digunakan (Rokach, 2010). Tujuan dari penelitian ini adalah membagi data jenis kendaraan yang digunakan ke dalam beberapa kelompok berdasarkan pengguna sehari – hari.

II. Landasan Teori

A. Data Mining

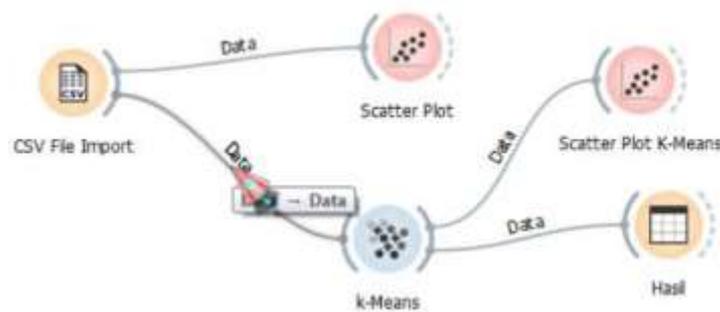
Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data disimpan dalam database, data warehouse, atau penyimpanan informasi lainnya (Abdillah et al., 2016). Clustering adalah metode yang mengelompokkan data kedalam cluster, dimana objek dengan kesamaan tinggi berada pada cluster yang sama, tetapi objek yang tidak sama berada pada cluster berbeda. Clustering adalah metode pengelompokan objek data kedalam kelompok yang berbeda (Rahmani et al., 2014). K-Means clustering merupakan salah satu metode data clustering non hierarki yang berusaha mempartisi data yang ada kedalam bentuk satu atau lebih cluster atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dengan data yang mempunyai karakteristik yan berbeda dikelompokkan ke dalam kelompok yang lainnya (Metisen & Sari, 2015).



Gambar 1 Flowchart K-Means Clustering

III. Metode Penelitian

Penelitian ini bertujuan untuk mengelompokkan jumlah data kategori jenis kendaraan di transjakarta menggunakan teknik data mining. Adapun tahap pertama mengumpulkan data dari bus transjakarta bulan desember tahun 2021 dari berbagai sumber terpercaya, lalu analisis cluster metode K-Means menggunakan software WEKA.



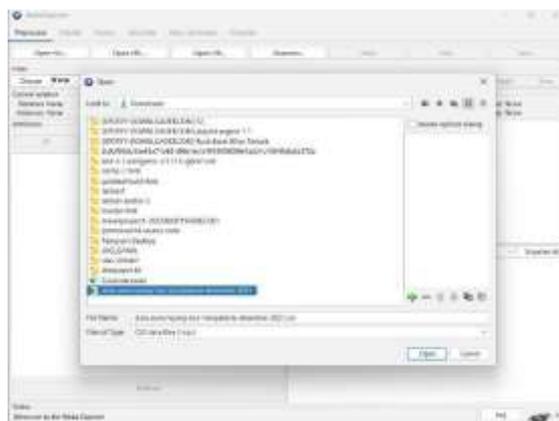
Gambar 2 Workflow yang digunakan untuk analisis cluster metode K-Means

Setelah semua data dipilih dan dianalisis, kemudian data tersebut telah dikelompokkan dengan menggunakan algoritma K-Means. Data yang sudah dikelompokkan dipilih berdasarkan kemiripan informasi yang ditampilkan data.

IV. Hasil Dan Pembahasan

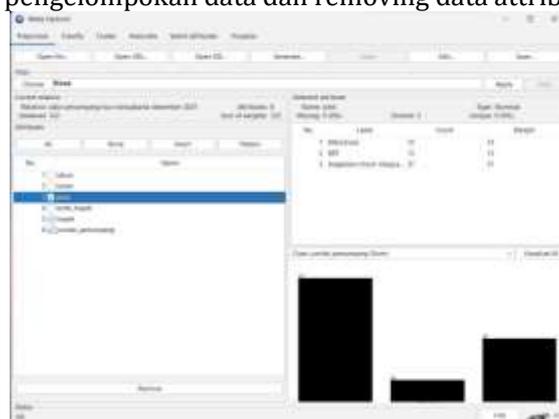
A. Hasil

Pre-Process data Bus Transjakarta Desember 2021 dilakukan dengan mengimport dataset yang sudah dikumpulkan kedalam software WEKA.



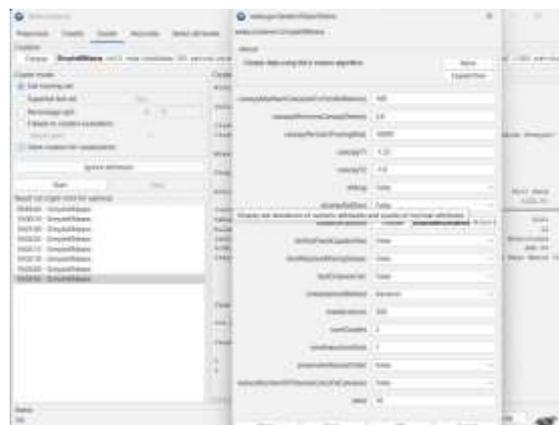
Gambar 3 Import File Dataset Ke Weka

Kemudian lakukan sebuah pengelompokan data dan removing data attribute yang tidak digunakan.



Gambar 4 Pengelompokan Dan Removing Data Attribute

Langkah selanjutnya konversikan data atribut yang dipilih, klik menu Cluster lalu button Choose dan pilih SimpleKMeans.



Gambar 5 Select SimpleKMeans

Hasil pengolahan pada aplikasi, seperti pada gambar dibawah ini

```
Cluster output
=====
Number of iterations: 2
Initial cluster mean of assigned objects: 100.00000000000000
Initial starting points (centroids):
=====
Cluster 0: 100.00000000000000, 0.00000000000000
Cluster 1: 100.00000000000000, 0.00000000000000
Missing values globally ignored with maximum:
=====
Final cluster centroids:
=====
Attribute          Full Size          Cluster#          %
-----
Value              1000.00             0                 1
Value              1000.00             1                 1
Value              1000.00             2                 1
Value              1000.00             3                 1
Value              1000.00             4                 1
Value              1000.00             5                 1
Value              1000.00             6                 1
Value              1000.00             7                 1
Value              1000.00             8                 1
Value              1000.00             9                 1
Value              1000.00             10                1
Value              1000.00             11                1
Value              1000.00             12                1
Value              1000.00             13                1
Value              1000.00             14                1
Value              1000.00             15                1
Value              1000.00             16                1
Value              1000.00             17                1
Value              1000.00             18                1
Value              1000.00             19                1
Value              1000.00             20                1
Value              1000.00             21                1
Value              1000.00             22                1
Value              1000.00             23                1
Value              1000.00             24                1
Value              1000.00             25                1
Value              1000.00             26                1
Value              1000.00             27                1
Value              1000.00             28                1
Value              1000.00             29                1
Value              1000.00             30                1
Value              1000.00             31                1
Value              1000.00             32                1
Value              1000.00             33                1
Value              1000.00             34                1
Value              1000.00             35                1
Value              1000.00             36                1
Value              1000.00             37                1
Value              1000.00             38                1
Value              1000.00             39                1
Value              1000.00             40                1
Value              1000.00             41                1
Value              1000.00             42                1
Value              1000.00             43                1
Value              1000.00             44                1
Value              1000.00             45                1
Value              1000.00             46                1
Value              1000.00             47                1
Value              1000.00             48                1
Value              1000.00             49                1
Value              1000.00             50                1
Value              1000.00             51                1
Value              1000.00             52                1
Value              1000.00             53                1
Value              1000.00             54                1
Value              1000.00             55                1
Value              1000.00             56                1
Value              1000.00             57                1
Value              1000.00             58                1
Value              1000.00             59                1
Value              1000.00             60                1
Value              1000.00             61                1
Value              1000.00             62                1
Value              1000.00             63                1
Value              1000.00             64                1
Value              1000.00             65                1
Value              1000.00             66                1
Value              1000.00             67                1
Value              1000.00             68                1
Value              1000.00             69                1
Value              1000.00             70                1
Value              1000.00             71                1
Value              1000.00             72                1
Value              1000.00             73                1
Value              1000.00             74                1
Value              1000.00             75                1
Value              1000.00             76                1
Value              1000.00             77                1
Value              1000.00             78                1
Value              1000.00             79                1
Value              1000.00             80                1
Value              1000.00             81                1
Value              1000.00             82                1
Value              1000.00             83                1
Value              1000.00             84                1
Value              1000.00             85                1
Value              1000.00             86                1
Value              1000.00             87                1
Value              1000.00             88                1
Value              1000.00             89                1
Value              1000.00             90                1
Value              1000.00             91                1
Value              1000.00             92                1
Value              1000.00             93                1
Value              1000.00             94                1
Value              1000.00             95                1
Value              1000.00             96                1
Value              1000.00             97                1
Value              1000.00             98                1
Value              1000.00             99                1
Value              1000.00             100               1
=====
Time taken to build model (with training data): 0 seconds
*** Model and evaluation on training set ***
=====
Clustered Instances
n = 100 (100%)
n = 100 (100%)
```

Gambar 6 Hasil Algoritma K-Means

Hasil penelitian menunjukkan bahwa algoritma K-Means memiliki tingkat akurasi sebesar 70% dalam mengklasifikasikan data jenis kendaraan.

V. Kesimpulan

Berdasarkan dari hasil penelitian, dapat disimpulkan bahwa algoritma K-Means dapat digunakan untuk menganalisa data transaksi dari sebuah bus transjakarta dan mengelompokkan customer berdasarkan jenis kendaraan yang sering digunakan berdasarkan data yang mereka miliki. Hasil dari analisis oengelompokan ini akan dipertimbangkan, Satu – satunya variable yang dipertimbangkan dalam situasi ini adalah jenis kendaraan dan jumlah kendaraan yang digunakan.

Daftar Rujukan

Abdillah, G., Putra, F. A., Renaldi, F., Informatika, P. S., Jenderal, U., Yani, A., Barat, J., & Cimahi, K. (2016). Penerapan data mining pemakaian air pelanggan untuk menentukan klasifikasi potensi pemakaian air pelanggan baru di PDAM Tirta Raharja menggunakan algoritma K-means. *Seminar Nasional Teknologi Informasi Dan Komunikasi, 2016*, 18–19.

Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining* (Vol. 4). John Wiley & Sons.

Metisen, B. M., & Sari, H. L. (2015). Analisis clustering menggunakan metode K-Means dalam pengelompokan penjualan produk pada Swalayan Fadhila. *Jurnal Media Infotama, 11*(2).

Rahmani, M. K. I., Pal, N., & Arora, K. (2014). Clustering of image data using K-means and fuzzy K-means. *International Journal of Advanced Computer Science and Applications, 5*(7).

Rokach, L. (2010). *Pattern classification using ensemble methods* (Vol. 75). World Scientific.