



Analisis Dan Pemetaan Jumlah Penumpang Kereta Api Di Indonesia Menggunakan Metode Statistik Deskriptif Dan K-Means Clustering

Benny Wijaya, Tresna Maulana Fahrudin , Aryo Nugroho

¹Teknik Informatika, Fakultas Ilmu Komputer, Universitas Narotama, Jl. Arief Rahman Hakim No. 51, Surabaya, 60117, Indonesia

E-mail: Bennywijaya2710@gmail.com, Tresna.maulana@narotama.ac.id, Aryo.nugroho@narotama.ac.id

ARTICLE INFO

Article history:

Received: 09 July 2019

Revised: 14 July 2019

Accepted: 01 Agustus 2019

Keywords:

PT.KAI, Analysis, K-means Clustering, Statistik Deskriptif, dan Mapping

ABSTRACT

The development of the population in Indonesia continues to increase, which will require more transportation facilities. PT. Kereta Api Indonesia (KAI) is one of the means of transportation in Indonesia. At present the railroad transportation facilities in Indonesia are still not comprehensive, the regions that have railroad transportation facilities are Java (Jabodetabek and outside Jabodetabek), and Sumatra. By taking data on the number of train passengers from the Central Statistics Agency (BPS), the analysis and mapping of the number of train passengers using descriptive statistics and K-means clustering was carried out in this study. This study produced 3 clusters in which each cluster has a measuring value. Cluster 0 is medium, cluster 1 is high, and cluster 2 is low. Calculated using k-means clustering produces a cluster of 0 there are 63, cluster 1 there is 47, and cluster 2 there are 46 with an accuracy of about 97.9%, and calculated using descriptive statistics to produce cluster 0 there are 108, cluster 1 there is 34, and cluster 2 exists 14 with an accuracy of about 93.6%

Copyright © 2019 Jurnal Mantik.
All rights reserved.

1. Pendahuluan

Perkembangan jumlah penduduk di Indonesia terus meningkat, dimana akan membutuhkan sarana transportasi yang lebih banyak agar dapat memwadhahi penduduk dalam menempuh perjalanan melakukan aktifitas. Kementerian transportasi Indonesia, sangat mendukung dalam pembangunan sarana dan prasarana perkembangan transportasi umum.

PT. Kereta Api Indonesia (KAI) merupakan salah satu sarana transportasi yang ada di Indonesia. Saat ini sarana transportasi kereta api di Indonesia masih terbatas hanya ada di beberapa wilayah bagian Indonesia, yakni di Jawa (Jabodetabek dan Luar Jabodetabek), dan Sumatera. Kondisi angkutan penumpang PT. KAI telah dibenahi secara besar-besaran[1]. Tidak ada lagi desakan, himpitan, dan jenis penderitaan lain yang menghiasi wajah angkutan penumpang kereta api. Kini, angkutan penumpang sudah mulai tertib, tidak ada lagi antrian yang mengular dan berdesakan, dan tidak ada lagi penumpang yang tidak memperoleh tempat duduk, bahkan penumpang kelas ekonomi sekalipun tidak lagi merasakan suasana kereta yang panas dan pengap karena PT. KAI secara bertahap telah menyediakan fasilitas AC di setiap kereta.

Perbaikan sejumlah fasilitas stasiun seperti peron dan toilet umum juga tidak luput dari perhatian PT. KAI. Serta telah terbenah meningkatkan pelayanan angkutan penumpang dan mempermudah masyarakat untuk membeli tiket kereta api (KA). Perluasan channel pembelian tiket kereta melalui kerjasama dengan jaringan *minimarket*, biro perjalanan, lewat telfon via *contactcenter* 121, melalui aplikasi yang ada di android, ios maupun blackberry, dan internet reservation, sangat memudahkan masyarakat dalam memperoleh informasi dan memesan tiket kereta api dengan sistem online.





Beberapa keunggulan yang dimiliki oleh angkutan kereta api yakni kemudahan dalam membeli tiket, layanan terjadwal, bebas macet, waktu tempuh lebih cepat, kepastian rencana perjalanan, kenyamanan fasilitas di stasiun maupun didalam kereta, daya angkut yang besar, tarif yang kompetitif. Berdasarkan UU Nomor 23 Tahun 2007 tentang perkeretaapian, setiap pengembangan sarana dan prasarana kereta api di Indonesia perlu terwujudnya kenyamanan dalam sarana dan prasarana. Sarana siap operasi dari PT. KAI semakin tahun semakin meningkat, agar terwadahnya jumlah penumpang yang ada. Dan setiap tahun ke tahun jumlah penumpang kereta api di Indonesia mengalami kenaikan yang cukup signifikan.

Dengan terjadinya lonjakan jumlah penumpang kereta api di hari-hari besar, PT. KAI sudah dapat mengantisipasi untuk menambah armada yang ada. Berdasarkan uraian di atas penulis mengangkat sebuah topik penelitian untuk menganalisis dan pemetaan jumlah penumpang kereta api di Indonesia menggunakan metode statistik deskriptif dan k-means *clustering*.

Penulis terlebih dahulu menyiapkan data jumlah penumpang kereta api di Indonesia dari Badan Pusat Statistik dalam rentan waktu 2006 sampai 2018. Dilakukan *preprocessing* data menjadi unsupervised agar dapat dilakukan perhitungan secara statistik deskriptif dan secara penerapan metode K-means. Sebagai acuan melakukan perhitungan menggunakan metode K-means dengan mengambil data random yang akan di hitung beberapa iterasi sampai nilai acuan data random tersebut tidak berubah[2].

Sehingga dapat menghasilkan nilai klasterisasi keterdekatan jarak antar data yang lain[3]. Dan dalam melakukan perhitungan menggunakan statistik deskriptif, data akan diproses menurut rumus yang ada. Sehingga dapat menghasilkan klasterisasi yang mana dapat di lakukan perhitungan akurasi baik menggunakan metode k-means maupun statistik deskriptif.

2. Tinjauan Pustaka

A. Statistik Deskriptif

Statistik Deskriptif ini adalah bagian dari ilmu pengetahuan statistik, yang mana ada dua golongan yakni statistik deskriptif dan statistik inferensial. Statistik deskriptif yang mana juga disebut statistik deduktif, statistik sederhana adalah metode yang berkaitan dengan cara pengumpulan, menyusun, mengolah, dan menyajikan suatu data sehingga dapat memberikan hasil yang jelas.

Statistik deskriptif bertujuan untuk mengorganisasi dan menganalisis data, agar memberikan hasil yang jelas[4][5]. Statistik deskriptif mempunyai 3 macam ciri pokok ;

1. Bekerja dengan angka
Angka ini dapat menjadi jumlah (frekuensi), dan angka dapat juga menjadi sebagai nilai.
2. Bersifat objektif
Yang mana sebagai alat penilaian hasil nyata yang dilihat dari objektif (data), bukan dilihat dari sisi subjektif.
3. Bersifat universal
Dapat digunakan untuk segala bidang penelitian yang mana berhubungan dengan data.

B. K-means Clustering

Klasterisasi adalah metode penganalisaan data, dimana bertujuan untuk mengelompokkan data dengan mengklasifikasi karakteristik data yang nilainya berdekatan untuk menjadi suatu klaster atau kelompok-kelompok tertentu yang telah di tentukan[6][7][8].

Klasterisasi terdapat 2 cara pendekatan, yakni pendekatan pertama adalah pendekatan partisi (*partition based clustering*) yaitu mengelompokkan data kedalam klaster-klaster yang telah ditentukan[9]. Pendekatan yang kedua yakni pendekatan hirarki (*hierarchical clustering*) yaitu mengelompokkan data dengan membentuk suatu hirarki.

Dalam penelitian ini menggunakan cara pendekatan partisi (*partition based clustering*) yang mana K-means menjadi metode yang digunakan didalam penelitian ini. Metode K-means merupakan salah satu metode klasterisasi di dalam data mining. Metode K-means populer dalam penelitian klasterisasi data, yang mana metode ini mengklasterisasi menggunakan data unsupervised, mudah di gunakan, dan membuat klaster data dengan sistem partisi[7]. Lloyd (1957, 1982), Forgy (1965), Friedman dan Rubin (1967), dan McQueen (1967) berikut beberapa orang yang telah menemukan metode K-means. Pada dasarnya algoritma K-means ini, metode yang mengelompokkan sendiri data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelas atau klasternya[10].

Langkah-langkah algoritma pada K-means sebagai berikut;

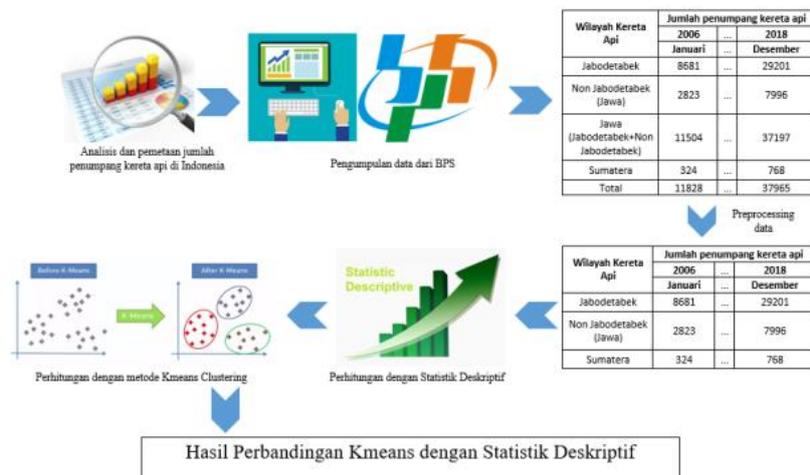




- 1) Menyiapkan data yang akan di klusterisasi terlebih dahulu
- 2) Menentukan berapa titik sentroid yang di perlukan
- 3) Kemudian titik sentroid tersebut memilih secara acak (*random*) data awal yang mana sebagai pengarah hasil klusterisasi
- 4) Kelompokkan data sesuai klaster yang telah di tentukan sebelumnya
- 5) Hitung kembali titik sentroid, untuk mendapatkan perbaharuan nilai titik centroid
- 6) Ulangi langkah 3 sampai 5 hingga sesuai dengan batas iterasi yang telah ditentukan atau hingga nilai titik centroid sudah tidak berubah.

3. Metodologi Penelitian

Dalam penelitian ini terdapat 5 alur proses penelitian untuk melakukan analisis dan pemetaan jumlah penumpang kereta api di Indonesia, yakni dimulai dengan pengumpulan data, preprocessing data, perhitungan menggunakan statistik deskriptif, perhitungan menggunakan metode K-means, dan hasil perbandingan statistik deskriptif dengan metode K-means. Seperti ditunjukkan pada Gambar 1 berikut.



Gambar 1. Desain analisis dan pemetaan jumlah penumpang kereta api di Indonesia.

A. Pengumpulan data

Pada penelitian ini data jumlah penumpang kereta api di Indonesia diambil dari Badan Pusat Statistik (BPS) dengan memasuki website BPS. Tahapan awal dalam pengumpulan data ini yakni memasuki website BPS kemudian masuk kehalaman publikasi lalu unduh data jumlah penumpang kereta api di Indonesia. Dalam penelitian ini penulis mengambil data jumlah penumpang kereta api di Indonesia dalam range waktu 2006 sampai 2018 per bulan. Data jumlah penumpang kereta api di Indonesia yakni dari wilayah Jabodetabek, luar Jabodetabek, dan Sumatera. Jumlah data yang diambil dan digunakan dari Badan Pusat Statistik yakni sebanyak 159 record dengan 3 variabel pada setiap wilayah. Data yang didapat berupa format .xls (*Extensible Stylesheet Language*). Data yang digunakan dalam penelitian ini memiliki satu tipe yaitu integer untuk tampilan data dan sampelnnya ditunjukkan pada tabel berikut ini.

Tabel 1. Keterangan tipe data

Variabel	Keterangan	Jenis data
Jabodetabek	Jumlah Penumpang	Integer
Luar Jabodetabek	Jumlah Penumpang	Integer
Sulawesi	Jumlah Penumpang	Integer



Tabel 2.Data jumlah penumpang kereta api di Indonesia dari BPS

Wilayah Kereta Api	Jumlah Penumpang kereta api				
	tahun 2006		...	tahun 2018	
	januari	februari	...	november	desember
Jabodetabek	8681	8144	...	28049	29201
Non Jabodetabek (Jawa)	2823	3561	...	6588	7996
Jawa (Jabodetabek + Luar Jabodetabek)	11504	11705	...	34637	37197
Sumatera	324	226	...	661	768
Total	11828	11931	...	35298	37965

B. Preprocessing data

Data yang sudah dikumpulkan dari Badan Pusat Statistik (BPS) kemudian dilakukan *preprocessing* data untuk menghindari perhitungan yang tidak sesuai dan tidak maksimal. *Preprocessing* data ini yakni membuang 2 variabel yang tidak di perlukan yaitu variabel Jawa (Jabodetabek dan luar Jabodetabek) dan total. Jadi ada 3 variabel yang digunakan dalam penelitian ini yakni Jabodetabek, luar Jabodetabek, dan Sumatera.

Tabel 3.Data Jumlah penumpang kereta api di Indonesia setelah *preprocessing* data

Wilayah Kereta Api	Jumlah Penumpang kereta api				
	tahun 2006		...	tahun 2018	
	januari	februari	...	november	desember
Jabodetabek	8681	8144	...	28049	29201
Non Jabodetabek (Jawa)	2823	3561	...	6588	7996
Sumatera	324	226	...	661	768

C. Perhitungan menggunakan statistik deskriptif

Pada tahapan perhitungan statistik deskriptif ini menggunakan microsoft excel. Berdasarkan variabel yang dimiliki, semua data yang ada akan dilakukan pengelompokkan dengan statistik deskriptif. Proses klusterisasi ini dimulai dengan menentukan 3 kluster yakni rendah, sedang dan tinggi. Adapun langkah-langkah pada tahapan perhitungan statistik deskriptif ini yakni pertama hitung SD (Standar Deviasi) pada persamaan 3, yaitu ;

$$SD = \sqrt{\sum \frac{(x_1 - Mean)^2 + (x_2 - Mean)^2 + \dots + (x_n - Mean)^2}{N-1}} \quad (3)$$

Dimana :

X = data ke n

SD = Standar Deviasi





Mean = nilai rata-rata data

N = banyaknya data

Kemudian hitung rata-rata data menggunakan rumus Mean seperti pada persamaan 4, yaitu :

$$\text{Mean} = \frac{X_1 + X_2 + X_3 + \dots + X_N}{N} \quad (4)$$

Dimana :

X = data ke n

Mean = nilai rata-rata data

N = banyaknya data

Lalu penentuan kluster sesuai penetapan dari statistik deskriptif yang telah dibuat, yaitu pada tabel IV

Tabel 4. Rumus dan peneteapan statistik deskriptif

FORMULA	Kategori	Kondisi
$X < M - 1\text{STD}$	RENDAH	batas rendah
$M - 1\text{SD} < X < M + 1\text{STD}$	SEDANG	batas rendah < X < batas tinggi
$X > M + 1\text{STD}$	TINGGI	Batas tinggi

D. Perhitungan menggunakan Kmeans

Pada tahapan perhitungan Kmeans ini dilakukan perhitungan menggunakan bahasa pemrograman python yang mana dilakukan perhitungan di googlecolab. Berdasarkan variabel yang dimiliki, semua data yang ada akan dilakukan pengelompokkan dengan metode Kmeans. Proses klusterisasi dimulai dengan menentukan jumlah kluster yang ditentukan, kemudian dilakukan mengidentifikasi data yang akan di kluster X_{ij} ($i=1, \dots, n$; $j=1, \dots, m$) dengan n adalah jumlah data yang akan di kluster dan m adalah jumlah variabel.

Pada awal iterasi, kluster yang ditentukan ditetapkan secara acak atau random. Dari hasil tersebut dihitung jarak antara setiap data dengan kluster yang telah ditentukan secara acak diawal. Berikut rumus Eucliden Distance, seperti pada persamaan 1, yaitu ;

$$(X, Y), (A, B) = \sqrt{(X - A)^2 + (Y - B)^2} \quad (1)$$

Suatu data akan menjadi anggota dari kluster ke-k apabila data tersebut kepusat kluster k bernilai paling kecil jika dibandingkan dengan jarak ke kluster lainnya. Nilai cluster yang baru dapat dihitung dengan cara mencari nilai rata-rata dari data yang menjadi anggota pada kluster tersebut, dengan menggunakan rumus pada persamaan (2) :

$$C_{ij} = \frac{\sum_i^p X_{ij}}{p} \quad (2)$$

Dimana X_{ij} cluster ke -k

P = banyaknya anggota kluster ke-k

Proses klusterisasi menggunakan metode algoritma Kmeans memiliki langkah-langkah sebagai berikut:

- Inisialisasi : tentukan berapa K sebagai jumlah kluster yang diinginkan.
- Pilih K buah titik centroid secara acak atau *random*.
- Lakukan perhitungan data yang tersedia dengan K yang telah dipilih, sehingga menghasilkan nilai keterdekatan masing-masing data terhadap K.
- Kelompokkan data sehingga terbentuk K buah kluster dengan titik centroid dari setiap kluster merupakan titik centroid yang telah dipilih sebelumnya.





- e. Perbaharui nilai titik centroid
- f. Kemudian hitung kembali centroid berdasarkan data yang mengikuti kluster masing-masing

Ulangi langkah c, d, e hingga kondisi konvergen tercapai yaitu perubahan fungsi objektif sudah dibawah ambang batas yang diinginkan atau tidak ada yang berpindah kluster atau posisi centroid yang sudah diambang batas yang ditetapkan.

4. Hasil dan Pembahasan

A. Hasil Statistik Deskriptif

Proses klusterisasi menggunakan statistik deskriptif ini dilakukan terhadap 156 data dari jumlah penumpang kereta api di Indonesia. Pada penelitian ini statistik deskriptif menghasilkan 3 kluster yakni kluster rendah, sedang, dan tinggi. Adapun langkah-langkah proses klusterisasi sebagai berikut:

- 1) Awal mula melakukan perhitungan rata-rata dari semua data, perhitungan tersebut menggunakan rumus Mean

$$\text{Mean} = \frac{X_1 + X_2 + X_3 + \dots + X_{156}}{156} = 21.307.000$$

Memasukkan data pertama ditambahkan sampai data terakhir ke-156 kemudian dibagi banyaknya data yakni 156 menghasilkan Mean 21.307.000

- 2) Perhitungan Standar Deviasi (SD) menggunakan rumus SD yang telah dijelaskan di awal :

$$\sqrt{\frac{(11.828.000 - 21.307.000)^2 + \dots + (X_{156} - \text{Mean})^2}{156 - 1}}$$

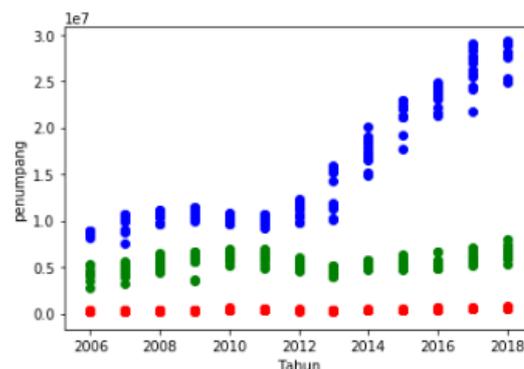
Standar Deviasi tersebut dilakukan dengan melakukan penambahan antara data pertama dikurangi Mean dalam kurung kwadrat hingga data terakhir yaitu data ke-156 kemudian dibagi data tersebut dikurangi 1 dimana hasilnya 7258,013371

- 3) Kemudian masukkan setiap data pada ketentuan yang dijelaskan di tabel 4. rumus dan penetapan statistik deskriptif, sehingga mendapatkan hasil kluster rendah ada 14 data, kluster sedang ada 108 data, dan kluster tinggi ada 34 data.

B. Hasil Metode Kmeans

Data pada penelitian ini yaitu jumlah penumpang kereta api di Indonesia yang diambil dari Badan Pusat Statistik (BPS) dengan mengunduh data berupa format excel .xls (*Extensible Stylesheet Language*). Kurang waktu pengumpulan data dilakukan maret – mei 2019, kemudian data dilakukan *preprocessing* data dengan melakukan sorting variabel yang tidak digunakan.

Kemudian data tersebut dilakukan perhitungan menggunakan metode Kmeans dan statistik deskriptif. Adapun hasil yang didapat dalam perhitungan menggunakan metode Kmeans yakni :



Gambar 2. Diagram jumlah penumpang kereta api di Indonesia



Setelah proses memunculkan diagram jumlah penumpang kereta api, kemudian proses klustersasi jumlah penumpang kereta api menggunakan metode Kmeans sebagai berikut :

1) Menentukan kluster

Dalam proses awal perhitungan menggunakan metode K-means pada penelitian ini ada tahapan menentukan kluster awal secara acak atau *random*. Sebelumnya data jumlah penumpang kereta api di Indonesia yang memiliki 3 variabel tersebut dirubah menjadi data array kemudian di standarkan ukuran variabelnya untuk menentukan 3 kluster pemandu jalan perhitungan metode K-means yang ditentukan pada penelitian ini. Kluster pertama adalah Jabodetabek, kluster kedua adalah luar Jabodetabek, dan kluster ketiga adalah Sumatera :

$$C_0 = (0.2124916, 0.57626136, 0.35344484)$$

$$C_1 = (0.80051785, 0.57441873, 0.60215054)$$

$$C_2 = (0.1258203, 0.31788803, 0.14161602)$$

2) Hitung Jarak

Kemudian akan di hitung jarak dari setiap data yang ada terhadap ketentuan kluster awal. Hasil jarak inilah yang akan menjadi penentu data tersebut termasuk kedalam kluster mana yang tersedia. Misalkan untuk menghitung jarak data jumlah penumpang kereta api di Indonesia dari data yang telah di standarkan variabelnya, data pertama (Jabodetabek 0.05126201 , luar Jabodetabek 0 , dan Sumatera 0.20430108) dengan pusat kluster pertama yakni:

$$ED_{0,0} = \sqrt{(0.05126201 - 0.2124916)^2 + (0 - 0.57626136)^2 + (0.20430108 - 0.35344484)^2}$$
$$= 0.616697654$$

$$ED_{0,1} = \sqrt{(0.05126201 - 0.80051785)^2 + (0 - 0.57441873)^2 + (0.20430108 - 0.60215054)^2}$$
$$= 1.024512266$$

$$ED_{0,2} = \sqrt{(0.05126201 - 0.1258203)^2 + (0 - 0.31788803)^2 + (0.20430108 - 0.14161602)^2}$$
$$= 0.3324773$$

3) Hitung kluster baru

Hitung pusat kluster yang baru dengan menambah nilai disetiap data yang klasternya sama kemudian dibagi dengan jumlah data yang klasternya sama. Misalkan C2 ada 2 data :

$$C_{2,0} \text{ baru} = \frac{\text{jumlah nilai kriteria 1 kluster 3}}{\text{jumlah data kluster 3}} = \frac{0.0778355}{2}$$
$$= 0.038918$$

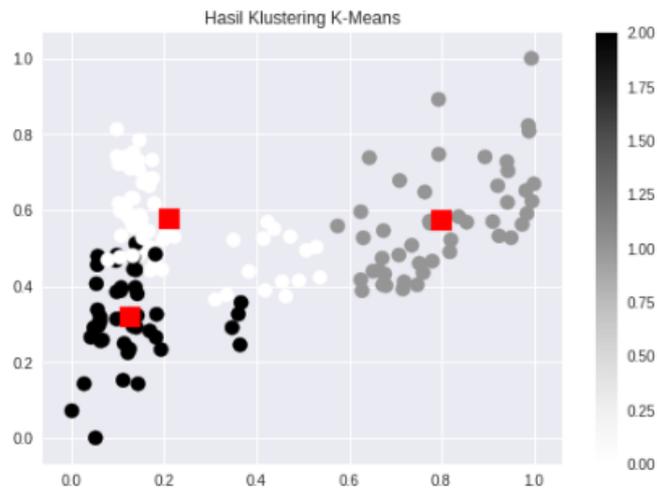
$$C_{2,1} \text{ baru} = \frac{\text{jumlah nilai kriteria 2 kluster 3}}{\text{jumlah data kluster 3}} = \frac{0.14266383}{2}$$
$$= 0.071331915$$

$$C_{2,2} \text{ baru} = \frac{\text{jumlah nilai kriteria 3 kluster 3}}{\text{jumlah data kluster 3}} = \frac{0.23297492}{2}$$
$$= 0.11648746$$

Setelah mendapatkan kluster baru, ulangi langkah perhitungan jarak terdekat atau Euclidian Distance (ED) lagi dan hitung kluster barunya sampai pada iterasi tertentu atau posisi data sudah tidak mengalami perubahan.

Dari langkah-langkah perhitungan mencapai hasil dari klustersasi jumlah penumpang kereta api menggunakan metode Kmeans yang mana divisualisasikan sebagai berikut ;





Gambar 3. Visualisasi klasterisasi jumlah penumpang kereta api di Indonesia

Hasil dari klasterisasi metode K-means menggambarkan visualisasi seperti pada gambar 3. Visualisasi klasterisasi jumlah penumpang kereta api di Indonesia. Visualisasi tersebut memiliki arti titik warna putih menggambarkan hasil klaster 0 (sedang), titik warna hitam menggambarkan hasil klaster 2 (rendah), dan titik warna abu-abu adalah hasil dari klaster 1 (tinggi). Arti titik warna merah adalah titik tengah tiap masing-masing klaster. Sehingga mendapatkan hasil klaster rendah ada 46 data, klaster sedang ada 63 data, dan klaster tinggi ada 47 data.

C. Akurasi Statistik Deskriptif

```
knn.score(X_test, y_test)
```

```
0.9361702127659575
```

Gambar4. Hasil Akurasi Statistik Deskriptif

Hasil akurasi data jumlah penumpang kereta api yang telah diklasterisasi menggunakan statistik deskriptif dalam perhitungan dengan bahasa pemrograman python di googlecolab menghasilkan 0.9361702127659575 yang mana jika di ubah persen menjadi sekitar 93.6%.

D. Akurasi Metode K-means

```
knn.score(X_test, y_test)
```

```
0.9787234042553191
```

Gambar 5. Hasil Akurasi K-means

Hasil akurasi data jumlah penumpang kereta api yang telah diklasterisasi menggunakan metode K-means dalam perhitungan dengan bahasa pemrograman python di googlecolab menghasilkan 0.9787234042553191 yang mana jika di ubah persen menjadi sekitar 97.9%.



5. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka diperoleh bahwa klusterisasi jumlah penumpang kereta api di Indonesia menggunakan metode K-means berhasil dilakukan dengan baik. Potensi jumlah penumpang kereta api dilihat dari tahun ke tahun slalu ada kenaikan jumlah penumpang kereta api. Dilihat potensi tinggi jumlah penumpang kereta api yang stabil dari hasil klusterisasi menggunakan statistik deskriptif dan metode K-means terjadi pada tahun 2016 sampai 2018.

Dari 156 data jumlah penumpang kereta api di Indonesia yang didapat dari BPS menghasilkan 3 kluster menggunakan metode K-means dan statistik deskriptif. Kluster tersebut mengenai potensi jumlah penumpang kereta api di Indonesia yang mana terdiri dari kluster rendah, kluster sedang, dan kluster tinggi.

Pengujian akurasi ketepatan perhitungan, metode K-means lebih unggul menghasilkan akurasi 97.9% dibanding klusterisasi menggunakan statistik deskriptif hanya menghasilkan akurasi 93.6%.

6. Referensi

- [1] M. Ruland, "Peningkatan Kinerja Pt. Kereta Api Indonesia Pada Pelayanan Keamanan Dan Keselamatan Publik Dalam Rangka K1 etahanan Nasional.pdf." 2009.
- [2] I. P. A. Pratama and A. Harjoko, "Penerapan Algoritma Invasive Weed Optimnization untuk Penentuan Titik Pusat Klaster pada K-Means," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 9, no. 1, p. 65, 2017.
- [3] Widiarina, "Klastering Data Menggunakan Algoritma Dynamic K-Means," *Klastering Data Menggunakan Algorith. Dyn. K-Means*, vol. 1, no. 2, pp. 260–265, 2015.
- [4] A. Sholikhah, "Statistik Deskriptif Dalam Penelitian Kualitatif," *KOMUNIKA J. Dakwah dan Komun.*, vol. 10, no. 2, pp. 342–362, 1970.
- [5] R. A. Vinarti and I. D. M. A. Baskara Joni, "Analisis Statistika Deskriptif pada Kepuasan Pengunjung Terminal Bus Purabaya," *S@Cies*, vol. 5, no. 1, pp. 1–8, 2018.
- [6] R. AKBAR, "Penerapan Data Mining dengan Menggunakan Metode Clustering K-Mean Untuk Mengukur Tingkat Ketepatan Kelulusan Mahasiswa Program Teknik Informatika S1 Fakultas Ilmu Komputer Universitas Dian Nuswantoro," *Dok. Karya Ilm.*, 2015.
- [7] R. Lynda, S. S. Widya, and S. Esti, "ANALISA CLUSTERING MENGGUNAKAN METODE K-MEANS DAN HIERARCHICAL CLUSTERING (STUDI KASUS : DOKUMEN SKRIPSI JURUSAN KIMIA , FMIPA , 2 . 3 Term Weighting dengan Term Frequency," vol. Volume 3 N, 2014.
- [8] Windha Mega Pradnya Dhuhita, "Clustering Menggunakan Metode K-Means untuk Menentukan Status Gizi Balita," *J. Inform.*, vol. 15, no. 2, pp. 160–174, 2016.
- [9] M. A. Wahyu, "Penerapan metode k-means clustering untuk mengelompokan potensi produksi buah – buahan di provinsi daerah istimewa yogyakarta," 2017.
- [10] L. Zahrotun, "Analisis Pengelompokan Jumlah Penumpang Bus Trans Jogja Menggunakan Metode Clustering K-Means Dan Agglomerative Hierarchical Clustering (Ahc)," *J. Inform.*, vol. 9, no. 1, pp. 1039–1047, 2015.