



Penerapan Data Mining Klasterisasi Data Penumpang Pesawat pada Bandara Indonesia Menggunakan Algoritma *K-Means*

¹Winda Helena Ratoebandjoe, ²Made Kamisutara, ³Achmad Muchayan
^{1,2,3}Universitas Narotama

Alamat Surat

Email: windahelena16@gmail.com, made.kamisutara@narotama.ac.id,
achmad.muchayan@narotama.ac.id

Article History:

Diajukan: 25 Oktober 2023; Direvisi: 15 Maret 2024; Accepted: 22 April 2024

ABSTRAK

Saat ini, data mining semakin dikenal sebagai alat manajemen data yang penting seiring dengan bertambahnya jumlah data. Tujuan penelitian ini adalah untuk mengelompokkan atau mengklasifikasikan jumlah penumpang pesawat di bandara Juanda dan bandara Ngurah Rai selama lima tahun terakhir yaitu tahun 2018 sampai dengan tahun 2022. Dari banyaknya data jumlah penumpang pesawat yang mengalami kenaikan dan penurunan di lima tahun terakhir, maka dari itu jumlah data yang banyak ini peneliti ingin menambang data untuk mendapat pengetahuan atau informasi baru dari data tersebut. Data mining atau menambang data merupakan proses yang memanfaatkan teknik-teknik statistik, matematika dan kecerdasan buatan untuk mengekstrak dan mengidentifikasi informasi dan wawasan selanjut (atau pola) dari kumpulan data yang sangat besar. Salah satu teknik dari data mining yaitu *clustering*, yang digunakan untuk mendapatkan kelompok objek dengan sifat umum dalam data yang cukup besar. Dalam *clustering* terdapat metode algoritma *K-Means* dengan menentukan jumlah *cluster* yang diinginkan yaitu dua *cluster*, tiga *cluster* dan seterusnya. Setelah melakukan proses klasterisasi dilanjutkan dengan melakukan pengodingan menggunakan bahasa pemrograman *Python* untuk menampilkan visual agar mudah dipahami. Dimana penelitian ini menghasilkan 3 klaster, setiap klaster memiliki nilai atau postensi pesawat di bandara juanda dan bandara Ngurah Rai.

Kata kunci: *Data Mining, Clustarisasi, Algoritma K-Means, Python, Bandara Juanda, Bandara Ngurah Rai*

ABSTRACT

*For some people who do not have financial problems, the practical way is to bring their computer or as the amount of data increases. The purpose of this study is to group or classify the number of aircraft passengers at Juanda airport and Ngurah Rai airport for the last five years, namely 2018 to 2022. From the large amount of data on the number of airplane passengers that have increased and decreased in the last five years, therefore this large amount of data researchers want to mine data to get new knowledge or information from the data. Data mining is a process that utilizes statistical, mathematical and artificial intelligence techniques to extract and identify information and further insights (or patterns) from very large data sets. One technique of data mining is clustering, which is used to obtain groups of objects with common properties in large enough data. In clustering there is a *K-Means* algorithm method by determining the desired number of clusters, namely two clusters, three clusters and so on. After doing the clustering process, proceed with coding using the Python*

programming language to display visuals for easy understanding. Where this research produced 3 clusters, each cluster has an aircraft value or positency at Juanda Airport and Ngurah-Rai Airport.

Keywords: *Data Mining, Clustarisasi, Algorithm K-Means, Python, Juanda Airport, Ngurah Rai Airport*

1. LATAR BELAKANG

Transportasi atau pengangkutan adalah perpindahan manusia atau barang dari suatu tempat ke tempat yang lainnya dengan menggunakan sebuah alat transportasi atau sebuah kendaraan yang bisa digerakkan oleh manusia atau mesin (Desga, W., Putri, F. M., & Yulanda, 2017). Tujuan utama dari transportasi adalah untuk memungkinkan mobilisasi dan hubungan antar orang, barang dan tempat.

Salah satu jenis transportasi yang diminati oleh masyarakat saat ini yaitu transportasi jalur udara. Transportasi jalur udara sangat penting, karena dapat menjangkau wilayah yang lebih luas (Ikhsan et al., 2017). Jenis prasarana transportasi udara di Indonesia yaitu bandar udara yang biasa dikenal dengan sebutan bandara. Peranan bandara yaitu sebagai jaringan transportasi udara dengan menjadikan pertemuan beberapa rute penerbangan, tempat masuk dan keluarnya perekonomian dalam hal upaya peningkatan pertumbuhan maupun perkembangan sebuah negara, sebagai pendorong di berbagai industri, pariwisata dan perdagangan (Ikhsan et al., 2017).

Salah satu bandara di Indonesia yang memiliki peranan yang sangat penting dalam perekonomian Indonesia yaitu Bandara Juanda dan Bandara Ngurah Rai. Bandara Juanda juga memiliki peran vital bagi Indonesia. Bandar udara ini merupakan gerbang utama terjadinya investasi perindustrian maupun kepariwisataan. Sedangkan Bandara Ngurah Rai sendiri merupakan bandara yang memiliki konsentrasi aktivitas tertinggi ke tiga yang ada di Indonesia. Sebagai salah satunya faktor yang memiliki aktivitas tertinggi ke tiga Bali diakui sebagai tujuan pariwisata bagi turis dan pariwisataawan (Verdwiansyah Subandi et al., 2021).

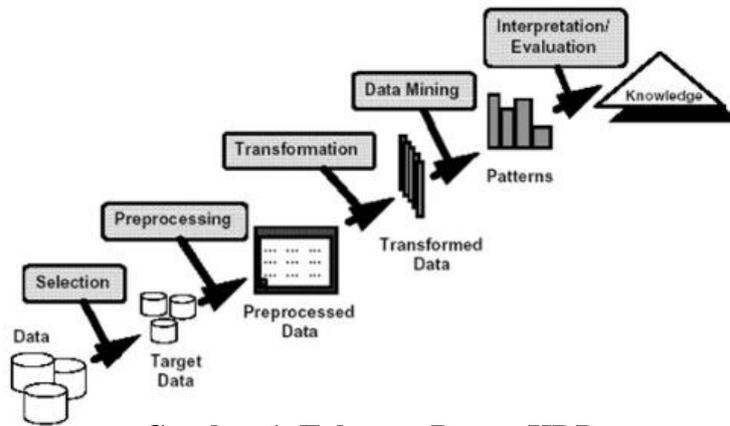
Jumlah data penumpang di kedua bandara tersebut mengalami kenaikan dan penurunan di lima tahun terakhir ini. Dengan adanya kenaikan dan penurunan data penumpang pesawat memiliki dampak atas pertumbuhan ekonomi di Indonesia. Atas kenaikan dan penurunannya data penumpang pesawat di kedua bandara tersebut peneliti ingin mengangkat sebuah topik penelitian penerapan *data mining* untuk mengklusterisasi atau mengelompokkan data penumpang pesawat pada bandara Indonesia terkhususnya bandara Juanda dan bandara Ngurah Rai menggunakan klusterisasi algoritma *K-Means*.

Penambangan data atau *Data Mining* merupakan proses dari menggunakan teknik statistik, matematika, *mechine learning* dan kecerdasan buatan untuk mengekstrak dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari *database* yang besar. Dari jumlah penumpang pesawat yang begitu banyak, maka dari data penumpang pesawat yang sangat banyak ini peneliti ingin mendalami data tersebut untuk menggali informasi yang baru dari data tersebut. Salah satu dari teknik *Data Mining* yaitu *Clustering*, gunanya untuk mendapat kelompok-kelompok dari objek dengan karakteristik umum dalam data yang cukup besar. Tujuan dari penelitian ini untuk mengetahui hasil dari penerapan *data mining* klusterisasi *K-Means* jumlah penumpang pesawat pada bandara Juanda dan bandara Ngurah Rai.

2. KAJIAN LITERATUR

2.1 *Knowledge Discovery in Database (KDD)*

Knowledge Discovery in Database (KDD) adalah proses menemukan pengetahuan yang berguna dari sebuah data yang bervolume besar, dan sering disebut dengan data mining atau pengembangan data.(Nur Khomarudin, 2003) Data mining adalah inti dari proses KDD untuk melibatkan dalam menyimpulkan algoritma yang mengeksplorasi data, mengembangkan pola dan menemukan pola-pola yang sebelumnya tidak diketahui. Ilustrasi dari *Knowledge Discovery in Database (KDD)* sebagai berikut:



Gambar 1. Tahapan Proses KDD

Dari gambar di atas terlihat bahwa data yang dikumpulkan dari berbagai sumber heterogen terintegrasi dalam satu gudang data yang disebut data target. Data yang relevan diputuskan untuk dianalisis dan diperoleh dari pengumpulan data. Kemudian itu adalah pra-diproses dan berubah atau mentransformasi data menjadi format standar yang sesuai (Verdwiansyah Subandi et al., 2021). *Data Mining* atau penambangan data merupakan langkah yang paling inti dalam algoritma atau teknik cerdas yang digunakan untuk mengekstraksi pola atau aturan yang bermakna. Pada akhirnya pola atau aturan yang dapat ditafsirkan menjadi informasi atau pengetahuan baru dan berguna.

2.2 Data Mining

Secara sederhana, *data mining* sebagai proses penggalian atau penggalian informasi dari suatu kumpulan data. Informasi yang diperoleh dapat digunakan di berbagai bidang seperti manajemen bisnis, pendidikan, perawatan kesehatan, perekonomian dan sebagainya. *Data mining* adalah proses menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari *database* yang besar (Handoko & Lesmana, 2018). Beberapa teknik yang sering di sebutkan dalam literatur data mining meliputi: *Classification*, *Clustering*, *Predication Association rule*, *Neural network*, *Decision trees*, *Nearest neighbor method*.

Tahap-tahap dari data mining adalah sebagai berikut: (1) *Data Cleaning* atau pembersihan data merupakan proses menghilangkan noise atau data yang tidak relevan, (2) *Data Integration* atau integrasi data merupakan penggabungan data dari *database* ke dalam satu *database* baru, (3) *Data Selection* atau seleksi data merupakan data yang ada pada *database* seringkali tidak semuanya dipakai, (4) *Data Transformation* atau transformasi data merupakan data gabungan ke dalam format yang sesuai untuk diproses dalam *data mining*, (5) Proses mining merupakan teknik diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data, (6) *Pattern Evaluation* atau evaluasi pola untuk mengidentifikasi pola-pola menarik, (7) *Knowledge Presentation* merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

2.3 Clustering

Clustering mengelompokkan data berdasarkan kemiripan antar objek, sehingga *clustering* tergolong metode pembelajaran tanpa pengawasan. Pada tahap ini dilakukan metode dimulai dengan menentukan jumlah *cluster* yang diinginkan yaitu dua *cluster*, tiga *cluster* dan seterusnya disebut dengan *non-hierarchical*, Metode ini umumnya dikenal dengan *K-Means Clusterin*. *K-Means Clustering* memiliki langkah-langkah sebagai berikut:

Langkah 1: Tentukan jumlah *cluster*

Langkah 2: Menentukan nilai untuk titik pusat awal

Langkah 3: Hitung pusat *cluster*

Langkah 4: Tentukan *centroid* baru dari masing-masing *cluster*

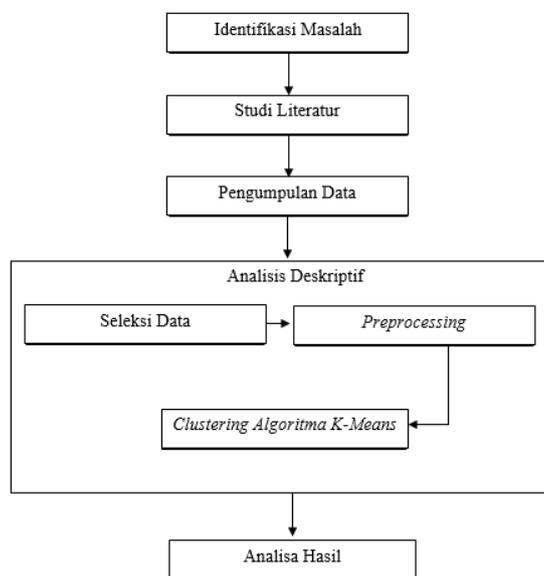
Langkah 5: kembali ke langkah 3 apabila masih ada data yang masih berpindah *cluster*, jika tidak ada perubahan maka hentikan proses *clustering* (Dan Pemetaan Jumlah Penumpang Kereta Api Di et al., 2019)

2.4 Terkait Penelitian Terdahulu yang dijadikan Acuan

Terkait penelitian terdahulu yang dijadikan acuan telah dilakukan oleh peneliti lainnya berkaitan pada penggunaan metode *K-Means* yang berjudul Penerapan data mining pada jumlah penumpang menggunakan metode *clustering* di susun oleh peneliti Koko Handoko pada tahun 2018 (Handoko & Lesmana, 2018). Letak perbedaan penelitian ini dengan penelitian yang terkait terletak pada objek yang diteliti dan *software* yang digunakan untuk perhitungan. Pada penelitian yang terkait, objek yang digunakan adalah data jumlah penumpang pesawat udara di bandar udara Hang Nadim Batam pada tahun 2015 hingga 2017 sedangkan penelitian ini, objek yang diteliti adalah data Penumpang pesawat di bandar udara Juanda dan bandar udara Ngurah Rai pada tahun 2018 sampai dengan tahun 2022. *Software* yang digunakan yaitu aplikasi *RapidMiner* sedangkan pada penelitian ini menggunakan pemrograman *Python di google colab*.

3. METODE

Terdapat 7 alur proses penelitian untuk melakukan penerapan data mining untuk *clusterisasi* jumlah data penumpang pesawat pada bandara di Indonesia yakni dimulai dengan identifikasi masalah, studi literatur, pengumpulan data, analisis deskriptif itu ada seleksi data, *preprocessing*, dan *clustering algoritma K-Means*, dan terakhir adalah analisa hasil penelitian.



Gambar 2. Kerangka Kerja Penelitian

Penjelasan dari setiap bagaian yang tercantum pada kerangka kerja penelitian dalam gambar diatas yaitu gambar 2 kerangka penelitian:

3.1 Identifikasi Masalah

Tahap ini, peneliti melakukan identifikasi masalah dan merumuskan permasalahan yang ada pada penelitian serta menawarkan solusi yang dapat digunakan untuk menyelesaikan masalah. Hal

ini bertujuan untuk mengumpulkan data atau objek berdasarkan tingkat kemiripan pada data penumpang pesawat menggunakan metode *clusterisasi K-Means*.

3.2 Studi Literatur

Tahap ini, peneliti melakukan *literature review* yaitu dengan mempelajari buku referensi, artikel dan hasil penelitian sejenis yang berkaitan dengan masalah yang sedang diteliti. Tujuan dari studi literatur ini adalah untuk mendapatkan landasan teori dari permasalahan yang diteliti, seperti teknik data mining, proses pengelompokan atau *clustering* menggunakan algoritma *K-Means*, *tools* yang bisa digunakan untuk data mining.

3.3 Pengumpulan Data

Pada tahapan ini peneliti mengumpulkan informasi atau data yang di perlukan yaitu data penumpang pesawat pada bandara di Indonesia di ambil dari Badan Pusat Statistik (BPS) dengan memasuki website BPS. Pengambilan data jumlah penumpang pesawat per bulan di Indonesia dari tahun 2018 sampai dengan tahun 2022 pada bandara Juanda dan Ngurah Rai. Data yang dihasilkan dalam format *.xls (Extensible Stylesheet Language)*.

3.4 Analisis Deskriptif

Proses analisis deskriptif berlangsung dalam beberapa tahapan, berikut ini penjelasan alurnya analisis deskriptif:

3.4.1 Seleksi Data

Data pemilihan (seleksi) digunakan dalam proses *data mining* disimpan dalam file terpisah yaitu file *Excel*. Kegunaan dari seleksi data adalah memilih data yang akan digunakan dalam penelitian ini. Data yang akan digunakan adalah bulan Januari hingga bulan Desember tahun 2018 sampai dengan tahun 2022. Dan Bandara Utama yang akan digunakan adalah bandara Juanda dan bandara Ngurah Rai.

3.4.2 Preprocessing

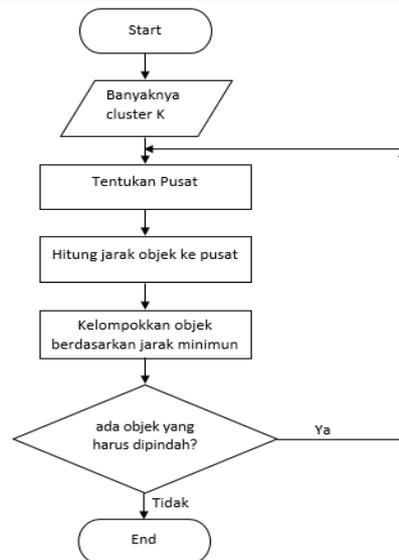
Data yang sudah dikumpulkan dari Badan Pusat Statistik (BPS) kemudian dilakukan Preprocessing data untuk mengikuti batasan masalah dari penelitian. Dengan membuang data Bandara Utama yaitu bandara Polonia, bandara Soekarno Hatta, dan bandara Hasanudin, dan Total di setiap tahun 2018 sampai dengan 2022.

3.4.3 Clustering K-Means

Dari tabel 2 dan tabel 3 merupakan dataset yang telah akan diteliti dalam penelitian ini, dimana masing-masing dari kedua bandara akan dilakukan pengelompokan menjadi beberapa kelompok dan setiap kelompok memiliki karakteristik yang mirip atau sama dengan yang lainnya tetapi dengan kelompok lainnya memiliki karakteristik yang berbeda. Pada tahapan *clusterisasi K-Means*, terdapat 3 *cluster* yang akan digunakan yang mana akan mempresentasikan data jumlah penumpang banyak, penumpang sedang, dan penumpang sedikit. Istilah dalam algoritma *K-Means clustering* yaitu:

1. *Cluster* : Kelompok atau grup
2. *Centroid* : titik pusat untuk menentukan *euclidean distance*
3. Iterasi : pengulangan proses, berhenti ketika hasil iterasi tidak mengalami perubahan

Proses klasterisasi menggunakan algoritma *K-Means* memiliki langkah-langkah berikut ini:



Gambar 3. Flowchart Proses Algoritma K-Means

1. Tentukan jumlah k sebagai *cluster* yang akan di bentuk.
2. Bangkitkan k *centroid* awal (titik pusat *cluster*) secara random. Untuk menentukan titik pusat dilakukan dalam bentuk acak dari beberapa objek yang tersedia sebanyak jumlah k *cluster* kemudian menghitung *centroid* ke- i berikutnya, dengan rumus:

$$v_{ij} = \frac{\sum_{i=1}^n x_{ij}}{n}; \quad (1)$$

Dimana:

v_{ij} = titik pusat pada *cluster*

x_{ij} = objek k - i

n = banyaknya jumlah objek yang menjadi *cluster*

3. Kemudian hitung antara setiap objek ke titik pusat *cluster* yang telah ditentukan secara random dengan menggunakan rumus *Euclidean Distance*.

$$D_e = \sqrt{(x_1 - s_i)^2 + (y_1 - t_i)^2} \quad (2)$$

Dimana

D_e = *Euclidean Distance*

I = banyaknya objek

(x,y) = koordinat objek dan

(s,t) = koordinat *centroid*

4. Hitung jarak dan alokasikan data ke dalam klaster menurut jarak minimum dari *centroid*
5. Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama.

Perhitungan dilakukan dengan menghitung manual dan bahasa pemrograman *Python* di google colab dengan menggunakan metode klasterisasi *K-Means*.

3.5 Identifikais Masalah

Pada tahap ini, setelah melakukan klasterisasi, dapat dilakukan visualisasi klaster menggunakan plot atau diagram yang memperlihatkan sebaran data disetiap *cluster*. Hal ini dapat membantu peneliti dalam memahami karakteristik setiap *cluster* dan menentukan interpretasi yang lebih jelas terhadap hasil klasterisasi.

4. HASIL DAN PEMBAHASAN

4.1 Perhitungan Manual *K-Means* Pada Bandara Juanda

Penulis melakukan perhitungan dengan persamaan untuk menghitung jarak antar data menggunakan proses klasterisasi jumlah data penumpang pesawat bandara juanda menggunakan *K-Means* menggunakan rumus *Euclidian Distance (ED)*.

1. Menentukan *cluster*

Seperti yang telah ditentukan pada bab sebelumnya bahwa telah ditetapkan oleh penulis menggunakan *cluster* sebanyak 3 *cluster* pada bandara Juanda untuk pengelompokan sehingga dapat ditentukan nilai k adalah k=3.

Tabel 1. Sampel Data Penumpang Pesawat Bandara Juanda

No	Bulan	Tahun				
		2018	2019	2020	2021	2022
1	Januari	689.756	560.242	553.747	189.560	301.152
2	Februari	597.290	473.070	481.881	152.366	220.164
3	Maret	660.707	481.147	408.725	198.457	319.093
4	April	671.763	443.457	97.748	219.430	285.101
5	Mei	592.835	378.933	5.397	210.700	501.021
6	Juni	706.372	607.024	64.137	346.782	404.070
7	Juli	856.010	620.237	121.240	74.127	447.785
8	Agustus	720.669	541.273	186.467	97.491	346.118
9	September	346.118	516.321	162.402	166.499	339.071
10	Oktober	688.559	543.876	182.889	231.536	367.656
11	November	634.151	549.010	238.036	278.750	371.289
12	Desember	642.080	574.001	244.579	305.669	413.988

2. Memilih *Centroid* Awal Secara Acak

Menentukan titik pusat *cluster* dilakukan secara acak dari sampel data yang tersedia sebanyak jumlah *cluster* k. Nilai titik *centroid* awal pada bandara juanda dalam penelitian ini sebagai berikut:

$$\begin{aligned}
 C1 &= 689.765; 560.242; 553.747; 189.560; 189.560 \\
 C2 &= 592.835; 378.933; 5.397; 210.700; 501.021 \\
 C3 &= 706.926; 516.321; 162.402; 166.499; 339.071
 \end{aligned}$$

3. Hitung jarak

Setelah menentukan titik *centroid*, kemudian akan dihitung jarak guna memisahkan data menjadi kelompok berdasarkan kesamaan atau kedekatan dari setiap data terhadap titik *centroid* awal. Hasil jarak ini akan menentukan data tersebut masuk kedalam *cluster* mana yang tersedia. Untuk menghitung jarak ke *centroid* masing-masing *cluster* menggunakan rumus *Euclidian Distance* adalah sebagai berikut:

$$D_{1,1} = \sqrt{
 \begin{aligned}
 &(689.756 - 689.756)^2 + (560.242 - 560.242)^2 \\
 &+ (553.747 - 553.747)^2 + (189.560 - 189.560)^2 \\
 &+ (301.152 - 301.152)^2
 \end{aligned}
 } \quad (3)$$

$$D_{1,2} = \sqrt{
 \begin{aligned}
 &(597.290 - 689.756)^2 + (473.070 - 560.242)^2 \\
 &+ (481.881 - 553.747)^2 + (152.366 - 189.560)^2 \\
 &+ (220.164 - 301.152)^2
 \end{aligned}
 } \quad (4)$$

$$\begin{aligned}
 D_{1,2} &= 171.044,1185 \\
 D_{1,3} &= \sqrt{\begin{aligned} &(660.707 - 689.756)^2 + (481.147 - 560.242)^2 \\ &+ (408.725 - 553.747)^2 + (198.457 - 189.560)^2 \\ &+ (319.093 - 301.152)^2 \end{aligned}} \quad (5) \\
 D_{1,3} &= 168.915,0082
 \end{aligned}$$

4. Kelompokkan setiap data

Setelah melakukan proses perhitungan jarak selanjutnya dilanjutkan dengan mengelompokkan atau meng-klasterisasi setiap data berdasarkan hasil perhitungan dan mengalokasikan data menurut jarak minimum sebagai berikut:

Tabel 2. Hasil Perhitungan Jarak Pada Iterasi 1

No.	C1	C2	C3	Jarak	Cluster
1	0	619.151,7366	396.666,6515	0	1
2	171.004,1185	546.094,0826	360.965,5791	171.004,1185	1
3	168.915,0082	459.320,6658	255.868,8352	168.915,0085	1
4	472.278,9682	256.162,7261	128.219,306	128.219,306	3
5	619.151,7366	0	291.074,423	0	2
6	526.776,4906	310.297,1952	233.688,2142	233.688,2142	3
7	503.118,3874	402.982,0811	234.670,4168	234.670,4168	3
8	383.025,6869	335.102,028	78.755,06873	78.755,0673	3
9	396.666,6515	291.074,423	0	0	3
10	379.459,7495	293.416,1714	81.013,9215	81.031,9215	3
11	340.243,3897	325.906,8831	160.385,6597	160.385,6597	3
12	352.505,9234	338.050,9988	198.156,2294	198.156,2294	3

Selesaiannya melakukan pengelompokkan data maka didapatkan *cluster* sebagai berikut:

- Nomor 1, 2, 3 adalah anggota dari C1;
- Nomor 5 adalah anggota dari C2;
- Nomor 4, 6, 7, 8, 9, 10, 11, 12 adalah anggota dari C3.

5. Menentukan posisi *centroid* baru

Menjumlahkan data kemudian dibagi dengan jumlah anggota *clusterisasi* yang dapat diperoleh dari hasil *cluster* diatas adalah:

a) Anggota C1

$$\begin{aligned}
 C_{1,1} &= \left(\frac{689.756 + 597.290 + 660.707}{3} \right) \\
 C_{1,1} &= 649.251 \quad (6)
 \end{aligned}$$

$$\begin{aligned}
 C_{1,2} &= \left(\frac{560.242 + 473.070 + 481.147}{3} \right) \\
 C_{1,2} &= 504.819,667 \quad (7)
 \end{aligned}$$

b) Anggota C2

$$\begin{aligned}
 C_{2,1} &= \left(\frac{592.835}{1} \right) \\
 C_{2,1} &= 592.835 \quad (8)
 \end{aligned}$$

$$\begin{aligned}
 C_{2,2} &= \left(\frac{378.933}{1} \right) \\
 C_{2,2} &= 378.933 \quad (9)
 \end{aligned}$$

Dengan menghitung titik pusat *cluster* yang baru maka nilai C1, C2, dan C3 yang telah dihitung adalah:

$$\begin{aligned} C1 &= 649.251; 504.819,6; 481.451; 180.127,6; 280.136,3 \\ C2 &= 592.835; 378.933; 5.397; 210.700; 501.021 \\ C3 &= 703.316,2; 549.399,8; 162.187,2; 215.035,5; 371.884,7 \end{aligned}$$

6. Menghitung *centroid* baru

Langkah berikut ini adalah mengulangi perhitungan di langkah ketiga dengan menghitung nilai *centroid* baru hingga titik *cluster* tidak mengalami perubahan. Karena untuk membuktikan hasil tidak mengalami perubahan, maka perlu dilakukan posisi *centroid* sama.

7. Kelompokkan setiap data dari iterasi ke-2

Seperti pada langkah sebelumnya dengan mengalokasikan masing-masing data ke dalam *cluster* menurut jarak minimum.

Tabel 3. Hasil Perhitungan Jarak Pada Iterasi Ke-2

No	C1	C2	C3	Jarak	Cluster
1	102.321,2042	619.151,7366	399.089,7053	102.321,2042	1
2	89.864,13443	564.094,0826	382.385,3573	89.864,13443	1
3	88.511,53878	459.320,6658	265.172,9263	88.511,53878	1
4	391.240,9495	256.162,7261	154.669,4805	154.669,4805	3
5	543.490,8761	0	287.303,3453	0	2
6	480.617,03	310.297,1952	177.021,9181	177.021,9181	3
7	474.515,1486	402.982,0811	235.851,6699	235.851,6699	3
8	323.461,49	335.102,028	124.246,9524	124.246,9524	3
9	330.015,1302	291.074,423	67.378,20423	67.378,20423	3
10	320.175,4092	293.416,1714	31.096.63812	31.096.63812	3
11	281.898,6184	325.906,8831	120.817.6689	120.817.6689	3
12	307.608,0616	338.050,9988	145.363.7132	145.363.7132	3

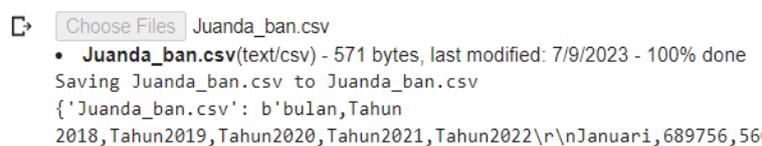
Terlihat bahwa pada iterasi ke-1 dan ke-2 tidak ada lagi perubahan pada titik *cluster* sehingga dapat di simpulkan bahwa iterasi dapat di hentikan pada iterasi yang ke-2 dengan hasil:

- a) *Cluster* satu (1) terdapat 3 item
- b) *Cluster* dua (2) terdapat 1 item
- c) *Cluster* tiga (3) terdapat 8 item

4.2 Pemograman *Python K-Means* Pada Bandara Juanda

Proses klasterisasi menggunakan metode *K-Means* akan dilakukan terhadap 60 data jumlah penumpang pesawat pada bandara Juanda berdasarkan per bulan dari tahun 2018 sampai dengan tahun 2022. Adapun langkah-langkahnya sebagai berikut:

- a. Proses awal adalah dengan menghubungkan google colab dengan *drive*
- b. Proses mengupload file csv, agar bisa dibaca sistem.



Gambar 4. Mengimport File .csv

- c. Selanjutnya memilih tipe data integer dan menentukan variabel untuk melakukan proses tahap perhitungan metode *K-Means* seperti pada gambar 5.

	Tahun 2018	Tahun2019	Tahun2020	Tahun2021	Tahun2022
0	689756	560242	553747	189560	301152
1	597290	473070	481881	152366	220164
2	660707	481147	408725	198457	319093
3	671763	443457	97748	219430	285101
4	592835	378933	5397	210700	501021

Gambar 5. Memilih Tipe Data

- d. Perhitungan menggunakan metode *clustering K-Means* pusat kluster awal yang digunakan untuk memulai proses klusterisasi metode *K-Means* yang sebagai pemandu jalan perhitungan diperoleh secara random dari bahasa pemrograman *Python*. Sebelumnya data jumlah penumpang pesawat bandara juanda 5 variabel tersebut diubah menjadi data array, seperti pada gambar 6 data array:

```
[[689756 560242 553747 189560 301152]
 [597290 473070 481881 152366 220164]
 [660707 481147 408725 198457 319093]
 [671763 443457 97748 219430 285101]
 [592835 378933 5397 210700 501021]
 [706372 607024 64137 346782 404070]
 [856010 620237 121240 74127 447785]
 [720669 541273 186467 97491 346118]
 [706926 516321 162402 166499 339071]
 [688559 543876 182889 231536 367656]
 [634151 549010 238036 278750 371289]
 [642080 574001 244579 305669 413988]]
```

Gambar 6. Perubahan Data Menjadi Array

Setelah data tersebut berubah menjadi array proses perhitungan menentukan kluster awal pada metode *K-Means* di *Python* ini yakni menstandarkan ukuran variabel, seperti pada gambar 7 dengan menstandarkan ukuran variabel:

```
array([[0.36827586, 0.75137171, 1.          , 0.42336653, 0.28836027],
 [0.0169279 , 0.39011786, 0.86894137, 0.28695238, 0.          ],
 [0.25789684, 0.42359016, 0.73553023, 0.45599751, 0.35223975],
 [0.29990691, 0.26739714, 0.16841616, 0.53291889, 0.23121019],
 [0.          , 0.          , 0.          , 0.50090041, 1.          ],
 [0.43141256, 0.94524334, 0.10712136, 1.          , 0.65480298],
 [1.          , 1.          , 0.21125741, 0.          , 0.81045158],
 [0.48573763, 0.67276133, 0.33020881, 0.08569071, 0.4484631 ],
 [0.43351762, 0.5693565 , 0.2863226 , 0.33878711, 0.42337204],
 [0.36372756, 0.68354855, 0.32368378, 0.57731932, 0.52514981],
 [0.1569906 , 0.70482462, 0.42425276, 0.75048321, 0.53808522],
 [0.18711884, 0.80839108, 0.43618492, 0.84921237, 0.69011632]])
```

Gambar 7. Menstandarkan Ukuran Variabel

Data yang sudah distandarkan ukuran variabelnya akan diambil 3 kluster pemandu jalan perhitungan metode *K-Means*, seperti gambar 8 penentuan titik *centroid*.

```
[[0.21436687 0.52169324 0.8681572 0.38877214 0.21353334]
 [0.74286881 0.83638067 0.27073311 0.04284535 0.62945734]
 [0.26752487 0.56839446 0.24942594 0.6499459 0.58039094]]
```

Gambar 8. Menentukan Titik Centroid

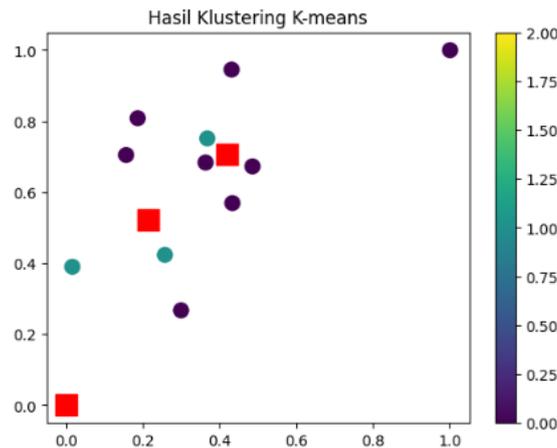
Dalam proses awal perhitungan menggunakan metode *K-Means* pada penelitian ini adalah tahapan menentukan kluster awal secara acak atau random menurut bahasa pemrograman *Python* yang dilakukan digoogle colab. Sebelumnya data jumlah penumpang pesawat bandara juanda di Indonesia yang memiliki 5 variabel tersebut diubah menjadi data array yang telah disebutkan dalam gambar 6 data array kemudian di standarkan ukuran variabelnya seperti pada gambar 7 menstandarkan ukuran variabel untuk menentukan 3 *cluster* pemandu jalan perhitungan metode *K-Means* yang ditentukan pada penelitian ini. Kemudian akan dihitung jarak dari setiap data yang ada terhadap ketentuan *cluster* awal. hasil jarak inilah yang akan menjadi penelitian data tersebut termasuk ke dalam *cluster* yang tersedia.

- e. Pada penelitian iterasi yang dilakukan dengan menggunakan standar 300 batas iterasi, sehingga menghasilkan *cluster* sebagai berikut:

[1 1 1 0 2 0 0 0 0 0 0 0]

Gambar 9. Cluster Bandara Juanda

- f. Setelah mendapat hasil dari pemrograman *Python* menggunakan metode *K-Means* akan ditampilkan visualisasi hasil *cluster* metode *K-Means* seperti gambar 10 visualisasi hasil *K-Means*.



Gambar 10. Visualisasi Hasil K-Means

- g. Hasil dari klasterisasi metode *K-Means* menggambarkan visualisasi seperti pada gambar 10 visualisasi hasil *K-Means* bandara juanda. Visualisasi tersebut memiliki arti warna ungu menggambarkan hasil *cluster* 0 (tinggi), dan warna hijau menggambarkan hasil *cluster* 1 (sedang). Pada salah satu titik merah terdapat *cluster* 2 (rendah). Arti titik warna merah pada gambar 10 merupakan titik tengah atau titik *centroid* tiap masing-masing *cluster*.

4.3 Perhitungan Manual *K-Means* Pada Bandara Ngurah-Rai

Sama seperti Sub-Bab sebelumnya, perhitungan claterisasi dilakukan dengan cara pengelompokkan menggunakan metode *K-Means* di mana bandara Juanda dan bandara Ngurah-Rai akan dilakukan pengelompokkan secara masing-masing. *Clusterisasi* bandara Ngurah-Rai dengan langkah-langkah yang sama seperti bandara Juanda yaitu sebagai berikut:

1. Menentukan kluster

Cluster telah ditentukan di bab sebelumnya yaitu 3 *cluster* untuk bandara Ngurah-Rai untuk pengelompokkan, sehingga dapat ditentukan nilai k adalah k=3.

Tabel 4. Sampel Data Penumpang Pesawat pada Bandara Ngurah-Rai

No	Bulan	Tahun				
		2018	2019	2020	2021	2022
1	Januari	449.389	428.629	453.130	118.965	317.662
2	Februari	401.856	357.076	346.962	71.105	191.034
3	Maret	427.958	369.930	253.517	116.888	290.652
4	April	467.257	364.220	44.122	142.344	224.687
5	Mei	433.394	289.400	2.423	121.441	395.423
6	Juni	479.185	460.273	11.396	225.671	362.418
7	Juli	546.118	459.284	41.375	49.999	408.296
8	Agustus	509.763	480.603	83.260	47.409	341.349
9	September	464.489	424.781	80.930	113.912	299.634
10	Oktober	481.774	435.86	99.469	224.887	341.780
11	November	440.613	435.288	169.655	274.179	317.266
12	Desember	475.739	450.458	189.289	318.264	388.008

2. Pilih *centroid* awal secara acak

Titik pusat *cluster* di pilih secara acak sebanyak jumlah *cluster* k. Nilai titik *centroid* awal pada bandara Ngurah-Rai dalam penelitian ini adalah:

C1 = 401.856; 357.076; 346.962; 71.105; 191.034

C2 = 433.394; 289.400; 2.423; 121.441; 39.5423

C3 = 509.763; 480.603; 83.260; 47.409; 341.349

3. Hitung jarak

Hasil jarak ini akan menentukan data tersebut masuk kedalam *cluster* mana yang tersedia. Untuk menghitung jarak ke *centroid* masing-masing *cluster* menggunakan rumus Euclidiance Distance adalah sebagai berikut:

$$D_{1,1} = \sqrt{\begin{matrix} (689.756 - 689.756)^2 + (560.242 - 560.242)^2 \\ + (553.747 - 553.747)^2 + (189.560 - 189.560)^2 \\ + (301.152 - 301.152)^2 \end{matrix}} \quad (12)$$

$$D_{1,1} = 0$$

$$D_{1,2} = \sqrt{\begin{matrix} (597.290 - 689.756)^2 + (473.070 - 560.242)^2 \\ + (481.881 - 553.747)^2 + (152.366 - 189.560)^2 \\ + (220.164 - 301.152)^2 \end{matrix}} \quad (13)$$

$$D_{1,2} = 171.044,1185$$

$$D_{1,3} = \sqrt{\begin{matrix} (660.707 - 689.756)^2 + (481.147 - 560.242)^2 \\ + (408.725 - 553.747)^2 + (198.457 - 189.560)^2 \\ + (319.093 - 301.152)^2 \end{matrix}} \quad (14)$$

$$D_{1,3} = 168.915,0082$$

4. Kelompokkan setiap data

Setelah melakukan proses perhitungan jarak selanjutnya dilanjutkan dengan mengelompokkan atau mengklusterisasi setiap data berdasarkan hasil perhitungan dan mengalokasikan data menurut jarak minimum, maka akan didapatkan data selengkapnya sebagai berikut:

Tabel 5. Hasil Perhitungan Jarak pada Iterasi 1

No.	C1	C2	C3	Jarak	Cluster
1	192.291,685	478.362,0575	385.786,8175	192.291,685	1
2	0	410.597,5952	345.828,982	0	1
3	146.963,726	283.831,8518	235.213,8054	146.963,726	1
4	319.762,2752	195.118,5764	198.761,7223	195.118,5764	2
5	410.597,5952	0	239.437,1868	0	2
6	427.196,7453	208.154,1147	196.809,5549	196.809,5549	3
7	415.077,8895	219.895,8968	895.496,9525	895.496,9525	3
8	345.828,982	239.437,1868	0	0	3
9	304.805,3215	186.253,8096	106.461,4949	106.461,4949	3
10	346.727,5703	216.305,3892	185.867,1262	185.867,1262	3
11	310.210,5117	280.605,377	257.495,4204	257.495,4204	3
12	372.727,7807	318.505,8177	298.073,8092	298.073,8092	3

Selesainya melakukan pengelompokkan data maka didapatkan *cluster* sebagai berikut:

- a) Nomor 1, 2, dan 3 adalah anggota dari C1,
- b) Nomor 4 dan 5 adalah anggota dari C2, dan
- c) Nomor 6, 7, 8, 9, 10, 11, dan 12 adalah anggota dari C3

5. Menentukan posisi *centroid* baru

Perhitungan pusat *cluster* baru dengan cara menjumlahkan data kemudian dibagi dengan jumlah anggota *clusterisasi* yang dapat diperoleh dari hasil *cluster* diatas adalah:

- a) Anggota C1

$$C1,1 = \left(\frac{449.389 + 401.856 + 427.958}{3} \right)$$

$$C1,1 = 426.401 \tag{15}$$

$$C1,2 = \left(\frac{428.629 + 357.076 + 369.930}{3} \right)$$

$$C1,2 = 385.211,6667 \tag{16}$$

- b) Anggota C2

$$C2,1 = \left(\frac{467.257 + 433.394}{2} \right)$$

$$C2,1 = 450.325,5 \tag{17}$$

$$C2,2 = \left(\frac{364.220 + 289400}{2} \right)$$

$$C2,2 = 326.801 \tag{18}$$

Dengan menghitung titik pusat *cluster* yang baru maka nilai C1, C2, dan C3 yang telah dihitung adalah:

C1 = 426.401; 385.211,6; 351.203; 102.319,3; 266.449,3

C2 = 450.325,5; 326.810; 23.272,5; 131.892,5; 310.055

C3 = 485.383; 449.506,8; 96.482; 179.188,7; 351.250,1

6. Menghitung *centroid* baru

Pada langkah ini yaitu perulangan dari langkah ketiga hingga posisi *centroid* lama dengan posisi *centroid* yang baru tidak mengalami perubahan lagi. Pehitungan *centroid* baru dilakukan dengan langkah ketiga yaitu menggunakan rumus *Euclidean Distance*.

7. Kelompokkan setiap data dari iterasi ke-2

Sesudah menghitung jarak data *clusterisasi* dapat dilakukan menurut jarak minimum setiap data sebagai berikut:

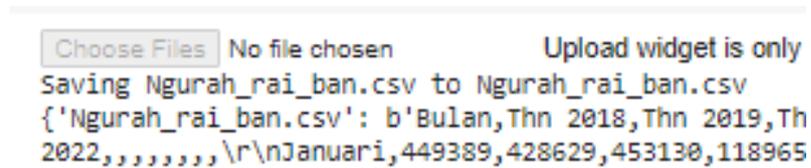
Terlihat bahwa pada iterasi ke-1 dan ke-2 tidak ada lagi perubahan pada titik *cluster* sehingga dapat di simpulkan bahwa iterasi dapat di hentikan pada iterasi yang ke-2 dengan hasil:

- a) *Cluster* satu (1) terdapat 3 item
- b) *Cluster* dua (2) terdapat 2 item
- c) *Cluster* tiga (3) terdapat 7 item

4.4 Pemograman Python K-Means Pada Bandara Ngurah-Rai

Setelah melakukan perhitungan manual akan dilanjutkan dengan melakukan perhitungan metode *K-Means* menggunakan bahasa pemrograman *Python*. Proses klasterisasi *K-Means* akan dilakukan terhadap 60 data jumlah penumpang pesawat pada bandara ngurah-rai berdasarkan per bulan dari tahun 2018 sampai dengan tahun 2022. Adapun langkah-langkahnya sebagai berikut:

- a) Mengupload file seperti pada gambar 11 untuk mengupload file csv, agar bisa terbaca oleh google colab.



Gambar 11. Mengupload File .csv

- b) Selanjutnya memilih tipe data integer dan menentukan variabel untuk melakukan proses tahap perhitungan metode *K-Means*, seperti pada gambar 12.

	Thn 2018	Thn 2019	Thn 2020	Thn 2021	Thn 2022
0	449389	428629	453130	118965	317662
1	401856	357076	346962	71105	191034
2	427958	369930	253517	116888	290652
3	467257	364220	44122	142344	224687
4	433394	289400	2423	121441	395423

Gambar 12. Memilih Tipe File

- c) Perhitungan menggunakan metode *clustering* K-means pusat klaster awal yang digunakan untuk memulai proses klasterisasi metode K-means yang sebagai pemandu jalan perhitungan diperoleh secara random dari bahasa pemrograman *Python*. Sebelumnya data jumlah penumpang pesawat bandara juanda 5 variabel tersebut diubah menjadi data array, seperti pada gambar 13 data array:

```
[[449389 428629 453130 118965 317662]
 [401856 357076 346962 71105 191034]
 [427958 369930 253517 116888 290652]
 [467257 364220 44122 142344 224687]
 [433394 289400 2423 121441 395423]
 [479185 460273 11396 225671 362418]
 [546118 459284 41375 49999 408296]
 [509763 480603 83260 47409 341349]
 [464489 424781 80930 113912 299634]
 [481774 435861 99469 224887 341780]
 [440613 435288 169655 274179 317266]
 [475739 450458 189289 318264 388008]]
```

Gambar 13. Tipe Data Array Bandara Ngurah-Rai

Setelah data tersebut berubah menjadi array proses perhitungan menentukan kluster awal pada metode K-means di *Python* ini yakni menstandarkan ukuran variabel, seperti pada gambar 14 yaitu menstandarkan ukuran variabel:

```
array([[0.32949079, 0.72817372, 1.          , 0.26418563, 0.58283547],
 [0.          , 0.35394842, 0.7644412 , 0.08748592, 0.          ],
 [0.18093469, 0.4211754 , 0.55711138, 0.25651732, 0.45851553],
 [1.          , 0.8885007 , 0.08642422, 0.00956231, 1.          ],
 [0.74799323, 1.          , 0.17935599, 0.          , 0.69186052],
 [0.53603166, 0.89367322, 0.01990872, 0.6581455 , 0.7888356 ],
 [0.45334877, 0.39131185, 0.09251909, 0.35050119, 0.15489593],
 [0.21861613, 0.          , 0.          , 0.27332706, 0.94074896],
 [0.43416146, 0.70804851, 0.17418633, 0.2455299 , 0.49985732],
 [0.55397818, 0.7659974 , 0.21531949, 0.65525096, 0.6938443 ],
 [0.26865703, 0.76300058, 0.37104372, 0.83723764, 0.58101279],
 [0.51214457, 0.84234034, 0.41460639, 1.          , 0.90661966]])
```

Gambar 14. Menstandarkan Ukuran Variabel

Data yang sudah distandarkan ukuran variabelnya akan diambil 3 kluster pemandu jalan perhitungan metode K-means, seperti gambar 15 penentuan titik centroid.

```
[[0.17014183 0.50109918 0.77385086 0.20272963 0.347117 ]
 [0.87399662 0.94425035 0.1328901 0.00478116 0.84593026]
 [0.42527683 0.6234817 0.18394053 0.57428461 0.65225922]]
```

Gambar 15. Menentukan Titik Centroid

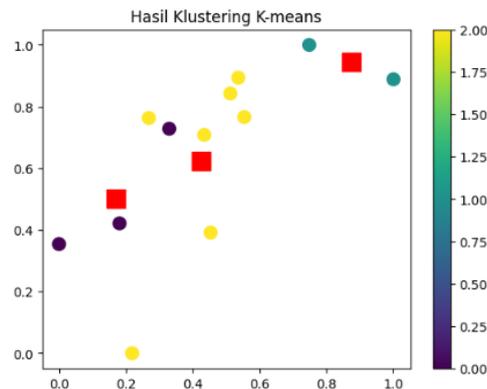
Dalam proses awal perhitungan menggunakan metode K-means pada penelitian ini adalah tahapan menentukan kluster awal secara acak atau random menurut bahasa pemrograman *Python* yang dilakukan di google colab. Sebelumnya data jumlah penumpang pesawat bandara juanda di Indonesia yang memiliki 5 variabel tersebut diubah menjadi data array yang telah disebutkan dalam gambar 13 data array kemudian di standarkan ukuran variabelnya seperti pada gambar 14 menstandarkan ukuran variabel untuk menentukan 3 *cluster* pemandu jalan perhitungan metode K-means yang ditentukan pada penelitian ini. Kemudian akan dihitung jarak dari setiap data yang ada terhadap ketentuan *cluster* awal hasil jarak inilah yang akan menjadi penelitian data tersebut termasuk ke dalam *cluster* yang tersedia.

- d) Penelitian iterasi yang dilakukan dengan menggunakan standar 300 batas iterasi, sehingga menghasilkan *cluster* sebagai berikut:

```
[0 0 0 1 1 2 2 2 2 2 2 2]
```

Gambar 16. Cluster Bandara Ngurah-Rai

- e) Setelah mendapat hasil dari pemrograman *Python* menggunakan metode k-means akan ditampilkan visualisasi hasil *cluster* metode k-means seperti gambar 17 visualisasi hasil K-Means.



Gambar 17. Visualisasi Hasil K-Means Bandara Ngurah Rai

Hasil dari klusterisasi metode K-means menggambarkan visualisasi seperti pada gambar 17 visualisasi hasil K-Means bandara Ngurah-Rai. Visualisasi tersebut memiliki arti warna ungu menggambarkan hasil *cluster* 0 (rendah), warna kuning menggambarkan hasil *cluster* 2 (sedang), dan warna hijau menggambarkan hasil *cluster* 1 (tinggi). Arti titik warna merah pada gambar 17 merupakan titik tengah atau titik centroid tiap masing-masing *cluster*.

5. SIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah di uraikan pada bab sebelumnya, maka penulis menarik kesimpulan adalah sebagai berikut:

1. Hasil algoritma K-Means Clustering menggunakan perhitungan manual dan pemograman *Python* di google colab adalah sama.
2. Hasil dari penerapan data mining yang dilakukan adalah untuk mengetahui potensi jumlah penumpang pesawat di Bandara Juanda dan Bandara Ngurah-Rai, dimana masing-masing bandara memiliki 60 sampel data yang didapat dari BPS menghasilkan 3 klaster menggunakan metode K-Means. Klaster tersebut mengacu pada jumlah penumpang pesawat di Bandara Juanda dan Ngurah Rai yang mana terdiri dari klaster rendah, klaster sedang dan klaster tinggi. Hasil klaster dari Bandara Juanda yaitu klaster tinggi terdapat 40 sampel data, klaster sedang terdapat 15 sampel data, dan klaster rendah terdapat 5 sampel data. Hasil klaster dari Bandara Ngurah-Rai yaitu klaster tinggi terdapat 15 sampel data, klaster sedang terdapat 35 sampel data, dan klaster rendah terdapat 10 sampel data.

5.2 Saran

Berdasarkan hasil penelitian dan kesimpulan, maka dapat diambil saran sebagai berikut:

1. Metode K-Means Klustering dapat dikembangkan lebih lanjut dengan menambahkan lebih banyak data dan klaster
2. Saran yang dapat dilakukan untuk penelitian selanjutnya yakni melakukan klusterisasi dengan menambahkan variabel data penumpang pesawat berdasarkan usia, jenis kelamin, atau preferensi penerbangan lainnya.

6. DAFTAR PUSTAKA

- Dan Pemetaan Jumlah Penumpang Kereta Api Di, A., Wijaya, B., Maulana Fahrudin, T., & Nugroho, A. (2019). Indonesia Menggunakan Metode Statistik Deskriptif Dan K-Means Clustering Articleinfo Abstract. In *JurnalMantik* (Vol. 3, Issue 2).
- Desga, W., Putri, F. M., & Yulanda, N. (2017). Pemodelan Bangkitan Perjalanan Di Nagari

- Siguntur, Nagari Barung-Barung Belantai Dan Nagari Nanggalo Kecamatan Koto Xi Tarusan Kabupaten Pesisir Selatan. *Jurnal Transportasi Multimoda*, 14(2), 77–82.
- Handoko, K., & Lesmana, L. S. (2018). Data Mining Pada Jumlah Penumpang Menggunakan Metode Clustering. *Snistek*, 1, 97–102.
- Ikhsan, M. T., Rusadi, D., & Ghalih, M. (2017). Analisis Pengaruh Jumlah Keberangkatan Penumpang di Bandara Pada Penerbangan Domestik dan Internasional di Indonesia. *Jurnal Riset Akuntansi Politala*, 2(1), 8–15.
- Nur Khomarudin, A. (2003). *Teknik Data Mining : Algoritma K-Means Clustering*.
- Verdwiansyah Subandi, A. R., Gunawan, H., & Lorenza Supriyanto, K. (2021). Penerapan Teknik Data Mining Untuk Memprediksi Jumlah Penumpang Pesawat Pada Bandara Indonesia. *Journal of Digital Ecosystem for Natural Sustainability (JoDENS)*, 1(2), 2798–6179.