

IMPLEMENTASI ALGORITMA K-MEANS DALAM MENENTUKAN CLUSTERING PADA PENILAIAN VERIFIKASI PPIU

Tuhfah Humaira Mufidah¹, Muhammad Risky², Zakaria Anshori³

¹Mahasiswa, Prodi Pendidikan Sistem dan Teknologi Informasi Universitas Pendidikan Indonesia ²Universitas Budi luhur dan Direktorat Jenderal Penyelenggaraan Haji dan Umrah Kementerian Agama, ³ Direktorat Jenderal Penyelenggaraan Haji dan Umrah Kementerian Agama

¹Alamat (Jl. Veteran No. 8 Purwakarta 41115 Telepon dan Fax: (0264)-200395; e-mail: tuhfahumaira@upi.edu), ²(Jl. Ciledug Raya No.99, RT.10/RW.2, Petukangan Utara, Kec. Pesanggrahan, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12260; email : 2011600703@student.budiluhur.ac.id), ³(Kementerian Agama Republik Indonesia Jl. Lapangan Banteng Barat Nomor 3-4 Jakarta 10710; email : Zack.anshori@gmail.com)

(Received: November 2022, Revised : Januari 2023, Accepied : April 2023)

Abstract— The Ministry of Religious Affairs becomes and serves as a medium for organizing government in the field of religion, thus making it a good service provider, one of which is in organizing Hajj and Umrah travel. This makes the Ministry of Religion a supervision and guidance on the granting of licenses to the Umrah Worship Travel Organization (PPIU). In KMA Number 540 of 2021, it states that PPIU passes when it gets a score in the range of 65-100, however, it turns out that PPIU is found with a score lower than the predetermined value range, this needs to be followed up carefully because it sees the importance of providing Umrah worship travel services for Muslims who can be trusted and the Ministry of Religion as a form of improvement in the social, political, religious, and economic fields of the country in granting PPIU operational licenses to be faster, more transparent, and accountable. In connection with that, researchers conducted clustering on the PPIU dataset using the PPIU Name variable with the Total Verification Value found in the entire data obtained as many as 230 using RapidMiner with the k-means algorithm so that the number of k was obtained using the Elbow Method so that it was obtained at k-3 with Avg. Distance 0.595. At k-3, the Davies Bouldin Index (DBI) is 0.027 with an interval range value of 61-100.

Keyword: Clustering, K-Means, Rapidminer, PPIU.

Intisari— Kementerian Agama menjadi dan bertugas sebagai media penyelenggaraan pemerintahan dalam bidang agama, sehingga menjadikannya sebagai penyedia layanan yang baik salah satunya dalam penyelenggaraan perjalanan haji dan umrah. Hal tersebut menjadikan Kementerian Agama sebagai pengawasan dan pembinaan atas pemberian izin kepada Penyelenggaraan Perjalanan Ibadah Umrah (PPIU). Pada KMA Nomor 540 Tahun 2021 menyatakan PPIU lulus ketika mendapatkan nilai dalam rentang 65-100 namun, ternyata ditemukan PPIU dengan nilai lebih rendah dari rentang nilai yang telah ditetapkan, hal tersebut perlu ditindak lanjuti dengan cermat karena melihat pentingnya penyediaan layanan perjalanan ibadah umrah bagi umat islam yang dapat di percaya dan Kementerian Agama sebagai bentuk perbaikan dalam bidang sosial, politik, agama, dan ekonomi negara pada pemberian perizinan operasioanl PPIU menjadi lebih cepat, transparan, dan akuntabel. Sehubungannya dengan itu, peneliti melakukan clustering

pada dataset PPIU dengan menggunakan variable Nama PPIU dengan Jumlah Nilai Verifikasi ditemukan keseluruhan data yang di dapat sebanyak 230 menggunakan RapidMiner dengan algoritma k-means sehingga di dapatkan banyaknya k dengan menggunakan Elbow Methode sehingga di dapat pada k-3 dengan Avg. Distance 0,595. Pada k-3 di peroleh Davies Bouldin Indeks (DBI) sebesar 0,027 dengan nilai rentang intervalnya dari 61-100.

Kata Kunci: Clustering, K-Means, Rapidminer, PPIU

I. PENDAHULUAN

Mayoritas penduduk di Indonesia merupakan penganut agama islam[1]. Salah satu kebutuhan ibadah mereka adalah menunaikan ibadah umrah, sehingga perjalanan ibadah umrah sangat diminati di Indonesia, seiring perkembangan waktu, semakin banyak yang berminat melaksanakan ibadah umrah semakin banyak pula yang menyediakan layanan penyelenggaraan perjalanan ibadah umrah. Layanan ini disediakan oleh perusahaan swasta yang merupakan Biro Perjalanan Wisata (BPW) yang telah mendapatkan izin usaha untuk melakukan perjalanan ibadah umrah. Dalam hal ini, Kemeterian Agama berperan untuk mengatur (regulator) dan melakukan pembinaan dan pengawasan terhadap Penyelenggara Perjalanan Ibadah Umrah (PPIU). Untuk mendapatkan izin usaha tersebut, BPW harus memenuhi standard kegiatan usaha yang pemenuhannya dilakukan dengan penilaian dan pengelompokkan verifikasi PPIU[2].

Sesuai dengan Undang – Undang Nomor 20 Tahun 2020 tentang Cipta Kerja dan Peraturan Pemerintah Nomor 5 Tahun 2021 mengenai Penyelenggaraan Perizinan Berusaha Berbasis Risiko, bahwa Kementerian Agama sudah melaksanakan penyesuaian Pelayanan Perizinan berupaya memperbaiki secara digital. Dengan ini, Kementerian Agama telah melaksanakan usaha membentuk inovasi dan membuat perubahan sebagai bentuk perbaikan dalam bidang sosial, politik, agama, dan ekonomi negara pada pemberian perizinan operasioanl Penyelenggara Perjalanan Ibadah Umrah (PPIU) menjadi lebih cepat, transparan, dan akuntabel. Hal ini diwujudkan dengan penerbitan Keputusan Menteri Agama Republik Indonesia Nomor 540 Tahun 2021 Tentang Pedoman Penerbitan Surat Rekomendasi Kantor Wilayah

Kementerian Agama Provinsi Bagi Perizinan Berusaha Penyelenggara Perjalanan Ibadah Umrah[2].

Untuk menguji upaya Kementerian Agama dalam meningkatkan kualitas pelayanan publik sebagai bentuk pengawasan dan pembinaan terhadap PPIU, pada artikel ini peneliti akan mengkaji dengan menganalisis hasil dari nilai verifikasi PPIU sehingga dapat mengelompokkan kedalam beberapa kelompok sesuai dengan kualitas penilaian yang di peroleh, menggunakan algoritma *clustering k-means* karena data yang diperoleh berjumlah banyak sehingga lebih tepat dan mudah menggunakan *k-means*[3]. *K-means* juga merupakan salah satu cara yang paling sederhana dalam pengelompokan suatu data[4]. Dengan ini, peneliti menggunakan algoritma clustering k-means sebagai metode penelitian. Peneliti melakukan pengelompokan berdasarkan 2 variabel yang berbeda yaitu diantaranya nama PPIU dengan akumulasi Jumlah Nilai Verifikasi. Sesuai dengan KMA Nomor 540 tahun 2021 penilaian verifikasi PPIU dinyatakan lulus jika nilai berada di rentang dari 65-100. Sedangkan jika nilai <65 maka dinyatakan tidak lulus. Dengan 7 (tujuh) kriteria seperti Tabel 1 berikut.

Tabel 1. Nilai Kriteria Penilaian Verifikasi PPIU

Kriteria	Bobot kriteria	Jumlah Bobot indikator	Nilai Elemen yang Dipilih	Total Nilai Indikator
Finansial (Aspek Pengelolaan)	25	75	100	25,00
SDM (Aspek Pengelolaan)	15	100	100	15,00
Memiliki rencana pengembangan SDM	10	100	100	10,00
Kementerian/Kerja sama dengan Pihak Kedua (Aspek Pengelolaan)	15	100	100	15,00
Operasional (Aspek Pengelolaan)	10	100	100	10,00
Sarana & Prasana (Aspek Pengelolaan)	15	100	100	15,00
Layanan / Service (Aspek Pengelolaan)	10	100	100	10,00

Pada tabel tersebut dapat diketahui atribut yang di pakai untuk penghitungan jumlah yang akan menentukan nilai akhir dari setiap PPIU yang mengajukan, dengan rumus penghitungan yang di pakai sebagai berikut[2],

$$(Bobot\ kriteria/Nilai\ elemen\ yang\ dipilih) \times (Bobot\ indikator/Nilai\ elemen\ yang\ dipilih) \times Nilai\ elemen$$

Sehingga menghasilkan jumlah yang bervariasi, maka dibutuhkan clustering dengan mengelompokkan sesuai dengan jumlah yang dihasilkan.

Dengan pengelompokan nilai verifikasi PPIU diharapkan Kementerian Agama dapat melakukan tindakan yang sesuai dengan ukuran tingkatan

Implementasi algoritma k-means dalam menentukan clustering pada penilaian verifikasi PPIU

pengelompokkan, dapat mempertimbangkan keputusan peraturan, meningkatkan pemeliharaan serta pemeliharaan terhadap proses kerja PPIU itu sendiri, sehingga terus melaksanakan dan mengembangkan misi dari Kementerian Agama yaitu sebagai administrasi layanan yang baik dengan meningkatkan layanan publik sehingga terhindari dari oknum penipuan khususnya dalam pelayanan perjalanan ibadah umrah. Dari pihak PPIU sendiri dengan adanya pengelompokan ini dapat menjadikan bahan evaluasi untuk kedepannya agar menjadi PPIU yang baik sesuai standar negara sehingga dapat melayani sehingga dapat dipercaya dan bertambahnya jumlah jamaah pada suatu PPIU karena masyarakat muslim sangat membutuhkan perjalanan ibadah umrah untuk menunaikan ibadahnya.

II. TINJAUAN PUSTAKA

A. Analisis Clustering

Menurut [5] Analisis Clustering merupakan runtutan dalam membagi objek dari kumpulan data menjadi beberapa cluster yang homogen. Tujuan utama dari metode clustering adalah untuk mengelompokkan sekumpulan data/objek ke dalam kelompok-kelompok (cluster) sehingga setiap cluster berisi sebanyak mungkin data yang mirip. Clustering bertujuan untuk menempatkan objek yang mirip atau bertetangga dalam sebuah cluster dan membuat jarak antar cluster seluas mungkin. Dengan ini maka objek dalam satu cluster sangat mirip satu sama lain dan berbeda dengan objek di cluster lain.

B. Algoritma K-Means

Menurut [6] Algoritma K-Means merupakan metode non-hierarchical clustering yang bertujuan untuk membagi objek yang ada menjadi satu atau lebih cluster atau kelompok objek berdasarkan karakteristiknya, seperti objek dengan karakteristik yang sama dikelompokkan dalam cluster yang sama dan objek dengan karakteristik yang berbeda dikelompokkan dalam cluster lain.

C. Pengawasan Penyelenggara Perjalanan Ibadah Umrah (PPIU)

Menurut [7] PPIU perlu diawasi agar mengikuti aturan yang ditetapkan Kementerian Agama. Pengawasan dilakukan untuk membangun kesadaran dan budaya profesional, membangun sistem dan prosedur, serta

mengatasi pelanggaran hal tersebut sesuai dengan Pasal 20 PMA No. 18 tahun 2015.

III. METODE PENELITIAN

Metode penelitian yang digunakan oleh peneliti berupa metode kuantitatif. Data yang didapatkan diperoleh dari pihak Direktorat Jenderal Penyelenggaraan Haji dan Umrah yaitu berupa excel dari hasil angket lewat google form yang berisikan indikator penilaian yang telah diisi oleh para PPIU yang mengajukan izin dan melewati tahap akumulasi sehingga didapatkan hasil nilai dari setiap PPIU.

A. Data Mining

Data mining merupakan langkah untuk menemukan pola yang unik dari sebaran data dalam jumlah banyak, sehingga dapat menentukan korelasi antara data lainnya dan membentuk sebuah informasi atau pengetahuan baru[8]. Data dapat disimpan dalam sistem digital pada komputer seperti database, datawarehouse, dan penyimpanan data lainnya[3]. Fungsi dari data mining yaitu sebagai fungsi deskripsi (description), fungsi estimasi (estimation), fungsi prediksi (prediction), fungsi klasifikasi (classification), fungsi pengelompokan (clustering), dan fungsi asosiasi (association)[9]. Karena data mining merupakan langkah dan terdiri dari tahapannya sehingga terdapat 7 (tujuh) susunan tahapan data mining, diantaranya [10], [11]:

1. Pembersih data (Data cleaning), pada tahap ini data dipilah, diperiksa dari duplikasi data dan dibersihkan dari data yang tidak konsisten dan tidak sesuai, kemudian diperbaiki kesalahan data yang ada seperti tipografi. Pada data yang diberikan pada pihak Direktorat Jenderal Penyelenggaraan Haji dan Umrah merupakan data yang dihasilkan dari pengisian PPIU pada google form, sehingga PPIU dapat mengisi google form lebih dari satu kali. Maka dengan ini, terdapat data yang tidak sesuai seperti nama PT yang tidak jelas dan duplikasi data sehingga dengan ini peneliti menghapus data yang tidak sesuai dan memperbaiki data tersebut.
2. Integrasi data (Data integration), pada tahap ini data dapat digabungkan dengan berbagai data dari database yang lain dan membentuk data yang baru. Sehingga tidak heran bahwa terkadang data yang satu dengan data yang lainnya membuat suatu relasi.
3. Seleksi Data (Data selection), pada tahap ini dari kumpulan database tidak semua atribut digunakan untuk bahan penelitian, maka dipilah atribut yang diperlukan untuk bahan analisis. Pada kasus ini peneliti membutuhkan data yang terdiri dari [2]:
 - a. Nama PPIU,
 - b. Laporan Keuangan audited dari KAP terigister di Kemenkeu dengan opini Wajar Tanpa Pengecualian (WTP),
 - c. Laporan ekuitas mencakup jumlah ekuitas dalam neraca laporan perusahaan pada bagian ikhtisar keuangan,
 - d. Kemampuan dalam membayar utang lancar yang direfleksikan dalam bentuk rasio kas (kas setara kas dibagi liabilitas jangka pendek(kewajiban lancar) * 100%
 - e. kemampuan memenuhi kewajiban jangka pendek dan jangka panjang (Debt to Asset Ratio) rumus: total hutang dibagi total Asset * 100% (Dominan).
 - f. Kecukupan kebutuhan SDM kuantitas dan kualitas yang selaras dengan struktur organisasi (ticketing, keuangan/akuntan, administrasi dan pemasaran).
 - g. Tenaga ahli (tour leader, tour guide, pembimbing ibadah) (Dominan).
 - h. Standar Kompetensi (tour leader, tour guide, pembimbing ibadah).
 - i. Memiliki rencana pengembangan SDM (Dominan).
 - j. BPW memiliki surat dukungan kerjasama/Kontrak kerja/SPK atau setara dengan perusahaan transportasi pada saat menjalankan paket wisata selama 1 tahun beroperasi.
 - k. BPW memiliki surat dukungan kerjasama/Kontrak kerja/SPK atau setara dengan perusahaan akomodasi selama 1 tahun beroperasi.
 - l. BPW memiliki surat dukungan kerjasama/Kontrak kerja/SPK atau setara perusahaan bidang jasa boga/catering selama 1 tahun beroperasi.
 - m. BPW Memiliki surat dukungan kerjasama/Kontrak kerja/SPK atau setara dengan perusahaan Asuransi selama 1 tahun beroperasi.
 - n. Paket wisata yang disediakan oleh BPW memuat informasi.
 - o. BPW menyediakan buku saku / panduan wisatawan.
 - p. BPW berpengalaman menyediakan paket rombongan wisatawan/Group.
 - q. BPW memiliki mekanisme untuk mengidentifikasi wisatawan selama perjalanan.
 - r. Ketersediaan kantor pelayanan yang dibuktikan dengan legalitas kepemilikan (Sertifikat SHM/HGB, IMB dan PBB) maupun surat perjanjian sewa. (Dominan).
 - s. Luas bangunan kantor pelayanan cukup memadai dalam menjalankan pelayanan dibuktikan dengan surat ukur/sertifikat/pajak bumi dan bangunan/pengukuran langsung (Dominan).
 - t. Tersedia sarana teknologi informasi dan komunikasi yang berfungsi dengan baik.
 - u. Tersedia sarana transportasi (Roda2 dan Roda4) yang dapat dioperasikan. Sarana transportasi haruslah merupakan aset perusahaan atau disewa dengan bukti perjanjian.
 - v. Fasilitas Bangunan Kantor (Dominan).

- w. Kejelasan Kantor Pelayanan Memiliki papan nama.
- x. Memiliki Data Base pelanggan.
- y. Pemberian informasi dilakukan sebelum keberangkatan, perjalanan dan di tempat tujuan. Dibuktikan dengan dokumentasi (SOP/pedoman kerja) dan foto/video kegiatan (Dominan).
- z. Pelayanan pemberangkatan ke dan dari tempat tujuan terkoordinir dengan baik dan dilakukan oleh petugas khusus. Dibuktikan dengan SOP dan Program Kegiatan atau foto/video kegiatan (Dominan).
- aa. Terdapat mekanisme pelayanan administrasi dan dokumen umrah yang meliputi :1. Pendaftaran jamaah 2. paspor dan visa dokumen perjalanan, 3. pembatalan jamaah dan 4. dokumen hilang selama perjalanan. Dibuktikan dengan SOP (Dominan).
- bb. Terdapat mekanisme pelayanan Penyelesaian dan tindak lanjut keluhan dibuktikan dengan hasil kuesioner kepuasan pelanggan atau data keluhan pelanggan.

Selain atribut tersebut tidak diperlukan untuk di analisis.

4. Transformasi data (Data transformation), pada tahap ini data digabungkan dan diubah ke dalam bentuk format yang diperlukan untuk dapat di proses karena beberapa metode data mining diperlukan format khusus untuk dapat di olah. Pada kasus ini menggunakan metode clustering k-means sehingga membutuhkan data yang bersifat integer, maka data yang berbentuk text diubah ke data integer untuk dapat diolah.
5. Proses mining, pada tahap ini merupakan tahap utama pada metode yang diterapkan untuk membuat informasi dan pengetahuan baru dari data.
6. Evaluasi pola (Pattern evaluation), pada tahap ini untuk menemukan pola-pola menarik kedalam knowledge based yang ditemukan. Sehingga untuk membuktikan apakah hipotesa yang didapat telah tercapai.
7. Presentasi pengetahuan (Knowledge presentation), pada tahap ini data yang telah diolah di visualisasikan dan dipresentasikan hasil dari pengetahuan yang di dapatkan.

B. K-Means

K-Means merupakan algoritma pengelompokan menjadi beberapa k kelompok berdasarkan kemiripan data yang sama dari beberapa data serta menggunakan titik centeroid secara acak sebagai bahan ukur[3], [10]. K-Means merupakan algoritma clustering yang bersifat Unsupervised learning dimana memerlukan semua variabel (atribut) untuk mencari pola tertentu. Algoritma K-Means ini digunakan untuk data yang bersifat numerik[4]. Sering digunakan karena mudah dan cepat dalam menghasilkan jawaban yang dibutuhkan. Proses

kerja algoritma K-Means sangat sederhana, meliputi[4], [12] :

1. Input data.
2. Menentukan jumlah cluster (random).
3. Menentukan titik pusat cluster.
4. Menghitung jarak data ke pusat cluster.
5. Mengelompokkan data sesuai dengan jarak ke pusat cluster.
6. Jika sudah sesuai, maka hasil sudah dapat di pakai. Jika belum atau ditemukannya data yang mengalami perpindahan posisi ataupun perubahan nilai centroid, maka mengulangi tahap ketiga.

C. Metode Elbow

Metode Elbow adalah metode untuk menentukan banyaknya cluster terbaik dengan melihat hasil dari presentase perbandingan diantara banyaknya cluster yang akan menghasilkan bentuk siku pada suatu titik[10]. Percobaan dilakukan dengan menginput avg distance, avg distance merupakan rata-rata jarak antar titik, pada tiap banyaknya K, dari avg distance ini kemudian dibuatkan grafiknya agar terlihat titik yang membentuk siku[13]. Semakin bertambahnya jumlah K maka akan semakin menurun jumlah avg distancenya. Sehingga dapat dilihat titiknya, semakin siku maka semakin tepat dalam penentuan banyaknya K pada clustering K-Means.

D. Aplikasi RapidMiner

RapidMiner adalah sebuah aplikasi penambangan data pembelajaran mesin, penambangan teks dan analisis prediktif. RapidMiner merupakan salah satu pengolahan data pertambangan. Pekerjaan RapidMiner meliputi penambangan teks adalah tentang analisis teks, untuk mengekstrak pola dari kumpulan data besar dan hubungan ke metode statistik, kecerdasan buatan, dan basis data. Tujuan dari analisis tekstual ini adalah mendapatkan informasi kualitas tertinggi dari teks ditempa RapidMiner menyediakan metode Penambangan data dan pembelajaran mesin termasuk: ETL (ekstraksi, transformasi, memuat), pemrosesan data, visualisasi, pemodelan dan evaluasi. Proses data mining terdiri dari operator bersarang dijelaskan dengan XML dan dengan antarmuka pengguna grafis. Presentasi direkam dengan pemrograman Java. Rapidminer digunakan untuk keperluan bisnis, penelitian, pendidikan, pelatihan, pengembangan aplikasi dan segala hal yang berhubungan tentang validasi data, pengoptimalisasi, dan visualisasi data[14], [15].

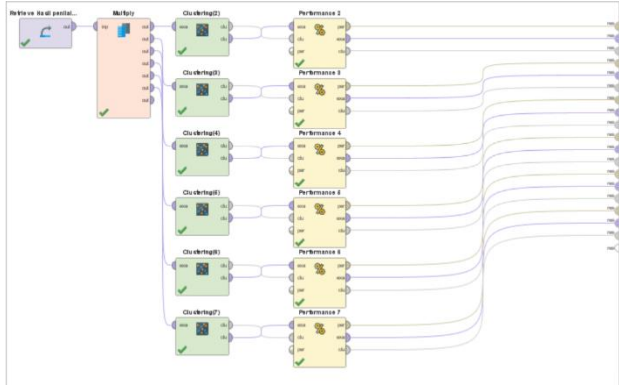
E. Davies Bouldin Indeks (DBI)

Davies Bouldin Indeks (DBI) merupakan suatu ukuran yang ditemukan di tahun 1979 oleh Donal W. Bouldin dan David L. Davies[16]. Dapat ditemukan ketika menggunakan aplikasi Rapidminer pada bagian cluster distance performance, DBI akan dipergunakan untuk mengevaluasi algoritma K-Means dimana semakin DBI mendekati nol (0) maka semakin optimal skema clustering yang dimiliki.

IV. HASIL DAN PEMBAHASAN

A. Pengolahan Data

Data didapatkan dari hasil mengumpulkan data melalui google form yang diberikan kepada Penyelenggara Perjalanan Ibadah Umrah (PPIU) yang telah terdaftar di Kementerian Agama dengan jumlah yang telah mengisi



google form sebanyak 235 PPIU, dengan pertanyaan yang disediakan meliputi pertanyaan seputar kriteria yang dibutuhkan dan sesuai dalam KMA Nomor 540 tahun 2021 untuk dapat dilihat nilai dari setiap point pertanyaan dalam bentuk numerik dari hasil pengkalian dan pembagian yang meliputi bobot dan nilai elemen kemudian dari nilai tersebut seluruhnya di jumlahkan, sehingga mendapatkan nilai verifikasi dari tiap PPIU. Namun ternyata ditemukan data yang tidak valid/cacat dan data yang berduplikat, sehingga perlu perbaikan dengan menghapus data cacat dan salah satu data yang berduplikat yang dapat dilihat pada tabel berikut.

Tabel 2. Jumlah Nilai Verifikasi PPIU

No.	PPIU	Laporan Keuangan	Laporan ekuitas ...	Kemampuan
1.	Aa	80	60	60	...
2.	PT. XYZ	100	80	100	...
3.	dd	100	80	80	...
4.	PT sukses internasional Tour Travel	80	80	80	...
5.	A	80	60	60	...

Ditemukan 5 data yang cacat dan berduplikat. Dimana data cacat dikarenakan nama yang tidak valid dan yang berduplikat menginput dua kali dengan membandingkan kedua data yang sama sehingga dapat disimpulkan data input kesatu tidak valid dibandingkan data input kedua dilihat dari point yang didapatkan. Maka data tersebut perlu dihilangkan karena akan mempengaruhi proses clustering jika tidak ditelaah. Dengan berkurangnya data tersebut total jumlah data yang akan di olah sebanyak 230. Berikut visual beberapa data valid yang didapatkan.

Tabel 3. Jumlah Nilai Verifikasi PPIU

No.	Nama PPIU	Jumlah
1.	PT. Salma Amanah Wisata	75,45

2.	PT. AMANAH SAFARI INTERNASIONAL	77,9
3.	PT. NAJAH HURRAHMAN	74,4
4.	PT. AL KAMILA TRAVEL & TOUR	81,4
.....
230.

Dapat dilihat jumlah data valid menjadi 230 dengan atribut yang dibutuhkan untuk digunakan pada clustering sebanyak 2 buah, meliputi Nama PPIU dan Jumlah. Didapatkan pula banyak jumlah yang bervariasi sehingga diperlukan adanya pengelompokan/clustering.

B. Menentukan Banyaknya K Dengan Metode Elbow

Data yang valid peneliti input menggunakan RapidMiner dengan K sebanyak n untuk percobaan, agar mempermudah dan mempercepat proses clustering. Peneliti melakukan percobaan sebanyak K7 dengan data yang sama. Dapat di lihat pada gambar berikut.

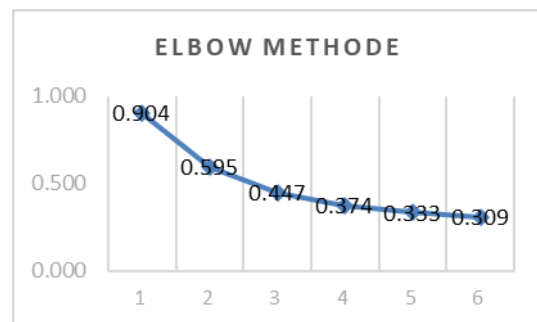
Gambar 1. Clustering K-Mean dengan Banyaknya K7

Peneliti mencoba membagi kelompok sebanyak 7 kali untuk menentukan banyaknya K terbaik sesuai dengan metode elbow dengan melakukan percobaan dengan kelompok 2, 3, 4, 5, 6, dan 7. Setelah dari semua K yang ada, pada setiap K nya diambil jumlah Avg Distance pada bagian Cluster Distance Performance dengan hasil yang didapatkan sebagai berikut.

Tabel 4. Tabel Menentukan Banyaknya K

K	Avg Distance
2	0,904
3	0,595
4	0,447
5	0,374
6	0,333
7	0,309

Pada Tabel 4 didapatkan K percobaan sebanyak K 7 dengan masing-masing Avg Distancenya, untuk dapat melihat posisi titik dari setiap nilai Avg Distance dibuatkan sebuah diagram linier yang dapat dilihat pada gambar berikut.



Gambar 2. Elbow Methode

Dapat dilihat pada Gambar 2. titik x (3) dengan titik y bernilai 0,595 memiliki titik yang paling mendekati siku-siku sedangkan pada titik x (4) sampai x (6) memiliki titik yang tumpul. Sesuai dengan metode elbow dimana K

terbaik ketika memiliki bentuk paling siku. Maka didapatkan titik yang paling siku ketika titik berada pada y (Avg Distance) dengan nilai y berjumlah 0,595 dengan nilai x (K) bernilai 3.

C. Eksekusi Clustering

Setelah mendapatkan banyaknya K terbaik dengan metode elbow, kemudian peneliti mengeksekusi data yang sama dengan melakukan clustering dengan K sebanyak 3 pada RapidMiner sehingga membuat kelompok 1, 2, dan 3 seperti gambar berikut.

Gambar 3. Clustering K-Means dengan K 3

Pada percobaan kedua ini peneliti telah membagi data kedalam 3 kelompok sesuai dengan jarak dengan titik senteroidnya. Namun, hasil Davies-Bouldin Index (DBI) masih bernilai negatif seperti tabel berikut.

Tabel 5. Hasil DBI Tiap Kelompok

No.	Pengelompokkan	DBI
1	2 Kelompok	-0.743
2	3 Kelompok	-0.827

Dilihat dari Tabel 5 bahwa DBI dengan nilai terbaik yaitu nilai yang paling mendekati nol, maka didapatkan DBI dengan nilai -0.743 namun nilai tersebut masih negatif.

Untuk menghasilkan nilai yang positif pada pemodelan menggunakan parameter maximize sehingga mengubah nilai negatif yang dihasilkan dari Davies Bouldin Index (DBI). Parameter ini menentukan apakah kinerja harus dimaksimalkan. Jika benar, hasilnya tidak dikalikan dengan minus satu. Pada kualitas parameter dimaksimalkan untuk setiap kualitas dari masing-masing kelompok, maka parameter maximize perlu sertakan. Dengan ini maximize dilakukan untuk mengoptimalkan hasil dari DBI, sehingga hasil yang didapatkan menjadi seperti pada tabel berikut.

Tabel 6. Hasil DBI Setelah Maximize

No.	Pengelompokkan	DBI
1	2 Kelompok	0.743
2	3 Kelompok	0.827

Setelah menggunakan parameter maximize dapat dilihat nilai negatif kini berubah menjadi nilai positif. Untuk lebih mengoptimalkan nilai DBI perlu dilakukan normalize karena parameter normalize digunakan untuk membuat skala agar sesuai dengan rentang tertentu[17]. Hal tersebut penting ketika terdapat data dengan unit dan skala yang berbeda. Misalnya, menggunakan jarak Euclidean, semua atribut harus berada pada skala yang sama untuk perbandingan yang adil. Normalisasi berguna untuk membandingkan atribut dengan ukuran berbeda. Operator ini menormalkan atribut yang dipilih. Tersedia empat metode pemulihan. Parameter ini menentukan apakah hasil harus dinormalisasi. Jika benar, kriteria akan dibagi dengan jumlah fitur. Maka hasil yang didapatkan ketika menggunakan parameter maximize dan normalize akan tampak pada tabel berikut.

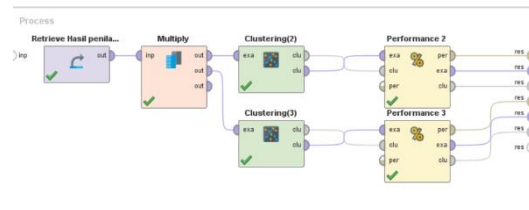
Tabel 7 Hasil DBI Setelah Maximize dan Normalize

No.	Pengelompokkan	DBI
1	2 Kelompok	0.027
2	3 Kelompok	0.030

Dapat dilihat setelah menggunakan parameter maximize dan normalize hasil nilai DBI semakin optimal.

D. Hasil Clustering

Setelah melakukan tahap penentuan banyaknya cluster dan mengelola hasil DBI agar semakin optimal, didapatkan jumlah items dari masing-masing kelompok dengan nilai



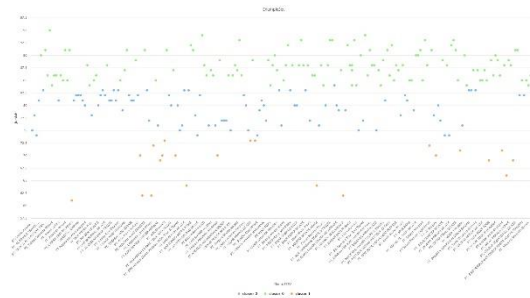
sesuai dengan cluster model pada gambar berikut.

Cluster Model

```
Cluster 0: 121 items
Cluster 1: 22 items
Cluster 2: 87 items
Total number of items: 230
```

Gambar 4. Cluster Model

Dapat dilihat pada Gambar 4 bahwa kelompok 1 memiliki sebanyak 121 items, kelompok 2 memiliki sebanyak 22 items, dan kelompok 3 memiliki sebanyak 87 items. Persebaran data dapat dilihat pada gambar grafik berikut.



Gambar 5. Grafik Persebaran Data Berdasarkan Clusternya

Pada grafik persebaran data berdasarkan clusternya menggunakan plot type scatter/bubble dimana X sebagai Nama PPIU dan Y sebagai Jumlah untuk memperjelas persebaran dan pembagian kelompok pada data tersebut.

Tahap selanjutnya memisahkan data sesuai dengan kelompoknya, digunakan untuk mempermudah tahap mencari rentang nilai dari hasil nilai maximum dan minimum dengan data tersebut. Pembagiannya pun dapat disajikan dengan tabel sebagai berikut.

Tabel 8. Cluster 1

No.	Cluster 1	
1	PT. ASIA PERMATA WISATA	90,15

2	PT. TELAGA AL-KAUTSAR TRAVEL	91,15
3	PT. RITAS RAUDAH	86,4
4	PT. GT TRAVEL INDONESIA	95,2
5	PT. TAIBAH SEMESTA WISATA	84,05
...
121	PT. SARMANIYA BINA UTAMA	85,85

Pada Tabel 8 merupakan atribut yang termasuk kedalam cluster 1 dengan jumlah seluruhnya terdapat 121 items.

Tabel 9. Cluster 2

No.	Cluster 2	
1	PT. CHARIS BERKAH ABADI	60,9
2	PT. ALHALID KUSUMA WARDANI	69,75
3	PT. ARABIA NUSANTAMA RAYA	61,95
4	PT. LABBAIKA CIPTA IMANI	61,75
5	PT.EMMA TOUR DAN TRAVEL	71,65
....
22	PT. RINDU BAITULLAH	69,25

Pada Tabel 9 merupakan atribut yang termasuk kedalam cluster 2 dengan jumlah seluruhnya terdapat 22 items.

Tabel 10 Cluster 3

No.	Cluster 3	
1	PT. SALMA AMANAH WISATA	75,45
2	PT. AMANAH SAFARI INTERNASIONAL	77,9
3	PT. NAJAH HURRAHMAN	74,4
4	PT. AL KAMILA TRAVEL & TOUR	81,4
5	PT. ALMARWAH FITRI WISATA	83,5
...
87	PT. LAMAHU TOUR & TRAVEL	82,65

Pada Tabel 10 merupakan atribut yang termasuk kedalam cluster 3 dengan jumlah seluruhnya terdapat 87 items. Setelah data sudah dipisahkan berdasarkan cluster kemudian dicari nilai maximum dan nilai minimum untuk mendapatkan nilai interval pada tiap cluster. Dengan hasil yang didapatkan sebagai berikut.

Tabel 11 Nilai Max, Min, dan Interval Tiap Cluster

	Cluster 1	Cluster 2	Cluster 3
Nilai Max	95,2	73,45	83,6
Nilai Min	83,5	60,9	73,85
Nilai Interval	95,2-83,5	73,45-60,9	83,6-73,85

Dari hasil Tabel 11 didapatkan nilai maximum dan minimum dari tiap cluster dengan menggunakan aplikasi excel untuk mempermudah proses dimana menggunakan rumus dengan sintaks $MIN(Number1; [Number2]; ...)$ Implementasi algoritma k-means dalam menentukan clustering pada penilaian verifikasi PPIU

untuk mencari nilai maximum dan sintaks $MIN(Number1; [Number2]; ...)$ untuk mencari nilai minimum. Sehingga dapat menentukan nilai interval atau rentang dari dua angka yang diberikan dengan rumusnya sebagai berikut

$$Nilai Interval = Nilai Maximum - Nilai Minimum$$

Sehingga rentang nilai untuk penilaian verifikasi PPIU dapat disimpulkan menjadi 3 cluster dengan tiap rentang clusternya disimpulkan pada tabel berikut.

Tabel 12 Akumulasi Pengelompokkan

Akumulasi Pengelompokkan	
Interval cluster 1	83,5-95,2
Interval cluster 2	60,9-73,45
Interval cluster 3	73,85-83,6

Dengan ini, untuk mengoptimalkan penilaian terhadap verifikasi PPIU melakukan pembaharuan pada KMA 540 tahun 2021 penilaian verifikasi PPIU dengan menetapkan rentang kelulusan PPIU dengan rentang nilai dari 100-73 dinyatakan lulus, < 73 dinyatakan tidak lulus. Hal tersebut juga menjadikan Kementerian Agama dapat mudah menjadi pengawas, mengambil dan melakukan tindakan, mempertimbangkan keputusan peraturan, serta meningkatkan pemeliharaan pada proses kerja PPIU sesuai dengan pengelompokkan dengan sifat data yang sama.

V. PENUTUP

A. Kesimpulan

Berdasarkan penelitian yang telah dilakukan didapatkan kesimpulan akan perlunya optimalisasi pelaksanaan verifikasi PPIU berdasarkan ketentuan yang ada dalam KMA 540 tahun 2021 Tentang Pedoman Penerbitan Surat Rekomendasi Kantor Wilayah Kementerian Agama Provinsi Bagi Perizinan Berusaha Penyelenggara Perjalanan Ibadah Umrah. Hal mana sebelumnya PPIU dinyatakan lulus jika memenuhi rentang nilai dengan akumulasi 100-65 dan tidak lulus apabila < 65 namun, setelah diteliti terdapat nilai PPIU di bawah rentang kelulusan. Hal tersebut perlu diperbaiki dan ditelusuri sehingga mendapatkan cara penanggulangannya dengan melakukan clustering K-Means menggunakan RapidMiner, membagi cluster menjadi 3 kemudian dapat dilihat pada setiap cluster rentang nilainya (nilai interval) sehingga disimpulkan Interval cluster 1 didapatkan 83,5-95,2 ; Interval cluster 2 didapatkan 60,9-73,45 ; Interval cluster 3 didapatkan 73,85-83,6 . Dengan ini KMA dapat diubah dengan menetapkan rentang kelulusan PPIU menjadi 100-73 dinyatakan lulus, < 73 dinyatakan tidak lulus.

B. Saran

Pada penelitian berikutnya dapat mengelompokkan dengan lebih terperinci pada aspek variable yang berbeda tidak hanya antara nama PPIU dengan Jumlah Nilai Verifikasi. Dapat pula menggunakan algoritma clustering

lainnya untuk dikomparasi dan diteliti perbedaannya dan ditemukan yang lebih optimal.

DAFTAR PUSTAKA

- [1] F. M. Yunus, "AGAMA DAN PLURALISME," *Jurnal Ilmiah Islam Futura*, vol. 13, no. 2, p. 213, Feb. 2014, doi: 10.22373/jiif.v13i2.72.
- [2] "KEPUTUSAN MENTERI AGAMA REPUBLIK INDONESIA."
- [3] Y. Prastyo, P. Eosina, and F. Fatimah, "PEMBAGIAN TINGKAT KECANDUAN GAME ONLINE MENGGUNAKAN K-MEANS CLUSTERING SERTA KORELASINYA TERHADAP PRESTASI AKADEMIK."
- [4] D. Marlina, N. Fauzer Putri, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," *Jurnal CoreIT*, vol. 4, no. 2, 2018.
- [5] "216-Article Text-1616-1692-10-20210122".
- [6] A. Nur Khomarudin, "Teknik Data Mining : Algoritma K-Means Clustering," 2003. [Online]. Available: <https://agusnkhom.wordpress.com>
- [7] "YEYET ROHILAH-FDK".
- [8] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS," *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [9] M. Pengetahuan and B. Data, "PENGANTAR DATA MINING."
- [10] E. Muningsih, "OPTIMASI JUMLAH CLUSTER K-MEANS DENGAN METODE ELBOW UNTUK PEMETAAN PELANGGAN."
- [11] "jurnal juna eska".
- [12] N. W. Rohmawati, S. Defiyanti, M. Jajuli, and T. H. Informatika Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang Jl Ronggowaluyo Telukjambe Timur Karawang, "IMPLEMENTASI ALGORITMA K-MEANS DALAM PENGKLASTERAN MAHASISWA PELAMAR BEASISWA," 2015.
- [13] A. Winarta and W. J. Kurniawan, "OPTIMASI CLUSTER K-MEANS MENGGUNAKAN METODE ELBOW PADA DATA PENGGUNA NARKOBA DENGAN PEMROGRAMAN PYTHON," *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 5, no. 1, 2021.
- [14] R. Nofitri and N. Irawati, "ANALISIS DATA HASIL KEUNTUNGAN MENGGUNAKAN SOFTWARE RAPIDMINER," *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 5, no. 2, pp. 199–204, Jul. 2019, doi: 10.33330/jurteksi.v5i2.365.
- [15] E. Fadhilah, "Implementasi Metode Profile Matching Terhadap Sistem Pendukung Keputusan Penerimaan Dana Zakat pada Badan Amil Zakat Pertamina (BAZMA)," *MATICS*, vol. 10, no. 2, p. 39, Mar. 2019, doi: 10.18860/mat.v10i2.5745.
- [16] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *Jurnal Sains dan Manajemen*, vol. 9, no. 1, 2021, [Online]. Available: www.bps.go.id
- [17] M. Risky, A. Wibowo, and Z. Anshori, "Komparasi Pengelompokan Peningkatan Sertifikasi Travel Umrah Berizin dengan Algoritma Klusterisasi K-Means dan K-Medoids," *InComTech: Jurnal Telekomunikasi dan Komputer*, vol. 12, no. 1, p. 47, Apr. 2022, doi: 10.22441/incomtech.v12i1.14528.