

I. PENDAHULUAN

1.1 Latar Belakang

PT TASPEN Tabungan dan Asuransi Pensiun merupakan lembaga yang menyediakan program jaminan sosial dan pensiun di Indonesia dan mengelola berbagai manfaat seperti pensiun hari tua, pensiun cacat, dan pensiun janda/duda (Taspen, 2018). Berdasarkan Undang-Undang No. 11 Tahun 1969, Pensiunan di Indonesia yang dikelola PT Taspen adalah pegawai negeri sipil (PNS), anggota TNI, Polri, serta beberapa kategori profesi lain yang terdaftar dan sudah memasuki masa pensiun serta menerima manfaat dari negara sesuai dengan ketentuan.

Pada tanggal 28 November 2022 diluncurkan Aplikasi Taspen Otentikasi oleh PT Taspen dalam proses meningkatkan keamanan dan efisiensi pengelolaan layanan pensiun dan jaminan sosial yang disediakan oleh PT Taspen. Aplikasi taspen otentikasi merupakan syarat utama peserta taspen dalam mencairkan dana pensiun setiap bulan. Aplikasi taspen otentikasi sangat diperlukan dalam perlindungan dan pemeliharaan dana pemerintah agar dana pensiunan dapat disalurkan kepada orang yang tepat dalam menjaga keamanan dan pengelolaan dana pensiun negara (Taspen, 2022).

Berdasarkan data Aplikasi Taspen Otentikasi Tahun 2023, setelah satu tahun peluncuran aplikasi oleh Pemerintah Indonesia dan PT Taspen, terdapat ketidaksesuaian antara implementasi di lapangan dengan ketentuan yang berlaku. Secara spesifik, peserta tidak melakukan otentikasi secara online sesuai dengan prosedur yang ditetapkan dan peserta masih meminta pelayanan langsung kepada petugas PT Taspen, meskipun aplikasi otentikasi dimaksudkan untuk menggantikan proses manual tersebut. Ketidaksesuaian ini menunjukkan bahwa penggunaan aplikasi belum mencapai tujuan yang diharapkan, yaitu mempermudah dan mempercepat proses otentikasi secara digital.

Menurut data Aplikasi Taspen Otentikasi Provinsi Jambi, efektifitas Provinsi Jambi dalam menggunakan aplikasi taspen otentikasi 60-70% dengan tingkat ketidakefektifan 30-40% yang sesuai dengan pernyataan Kepala Pelayanan PT Taspen Jambi Solichah, S (2023) mengatakan bahwa "Penggunaan aplikasi taspen otentikasi belum sepenuhnya efektif yang sesuai dengan data status pemblokiran dari penerimaan pembayaran pensiunan terus terjadi dan para peserta yang terus melakukan otentikasi secara offline yang tidak sesuai dengan arahan dan petunjuk dari aplikasi".

Menurut data Taspen Otentikasi Tahun 2023, terdapat perbedaan ketidakefektifan penggunaan aplikasi taspen otentikasi di setiap Provinsi; Provinsi DKI Jakarta, Jawa barat, Jawa Timur dan Bali adalah daerah dengan efektivitas penggunaan aplikasi yang tinggi sebaliknya Provinsi; Papua, Nusa Tenggara Timur, Kalimantan Tengah dan Maluku Utara memiliki efektifitas penggunaan aplikasi yang rendah. Sehingga, menurut Laporan Taspen (2023) dalam rangka peningkatan efektifitas penggunaan aplikasi diperlukan sosialisasi berdasarkan karakteristik setiap provinsi. Sosialisasi berdasarkan karakteristik tersebut dilakukan karena masing-masing provinsi di Indonesia memiliki perbedaan dalam infrastuktur teknologi dan penggunaan aplikasi. Permasalahan karakteristik efektifitas penggunaan aplikasi dapat dilakukan dengan cara mengelompokkan provinsi-provinsi yang memiliki kesamaan dan perbedaan karakteristik (Widodo et al., 2021). Salah satu metode pengelompokan permasalahan tersebut pada bidang matematika yaitu analisis *Cluster* (Hair et al., 2019).

Analisis *Cluster* adalah metode mengelompokkan objek ke dalam beberapa kelompok yang memiliki sifat sama dalam satu kelompok dan sifat berbeda antar kelompok (Backhaus et al., 2023). Pengelompokan *Clustering* tersebut dapat menunjukkan provinsi yang memiliki karakteristik sama antar kelompok dan yang memiliki karakteristik sangat berbeda antar kelompok *Cluster*. Salah satu cara yang dapat dilakukan dalam pengelompokan *Clustering* ini yaitu menggunakan analisis *Cluster k-means* dan *k-medoids*.

Pengelompokan *Clustering k-means* dan *k-medoids* memiliki cara yang berbeda dalam mempresentasikan pusat *Cluster* yang dapat mempengaruhi hasil *Clustering* (Muller et al., 2016). Dalam konteks pengelompokan provinsi berdasarkan efektivitas penggunaan aplikasi, diketahui bahwa terdapat dataset yang beragam dan terdapat data dengan nilai yang jauh berbeda antara data lainnya. Berdasarkan dataset yang beragam penggunaan *k-means* dapat digunakan untuk memfilter data sebelum menerapkan metode *k-medoids* yang lebih kompleks (Christopher, 2006). Hal ini dikarenakan *k-means* memilih pusat *Cluster* berdasarkan rata-rata (*centroid*) dari semua titik data dalam *Cluster* dan secara matematis, yang berarti *k-means* meminimalkan jarak *Euclidean kuadrat* dari setiap titik data ke *centroid*.

Salah satu tantangan utama dalam metode *Clustering*, seperti *k-means* dan *k-medoids*, adalah menentukan jumlah *Cluster* yang optimal (*k*) dan mengevaluasi kualitas pembagian data ke dalam *Cluster* tersebut. Penentuan jumlah *Cluster* yang tepat sering kali bersifat subjektif, dan tanpa pendekatan sistematis, bisa mengarah pada pemilihan yang tidak konsisten

(Muller et al, 2016). Metode *elbow* menawarkan solusi dengan memplot nilai fungsi objektif terhadap berbagai nilai k . Pada grafik *elbow*, titik di mana penurunan nilai objektif mulai melambat secara signifikan menunjukkan jumlah *Cluster* yang optimal. Pendekatan ini memberikan metode berbasis data yang lebih objektif untuk menentukan k , menghindari keputusan subjektif, dan memberikan indikator kuantitatif yang jelas tentang kualitas *Clustering*. Secara matematis, titik *elbow* dapat diidentifikasi dengan analisis perubahan gradien pada grafik, yang mencerminkan titik keseimbangan antara kompleksitas model dan kualitas pengelompokan (Muller et al., 2016).

Berdasarkan kedua teknik yang digunakan, perlu dilakukan validasi untuk menilai hasil pengelompokan dari kedua metode tersebut. Validasi *Cluster* adalah prosedur yang digunakan untuk mengevaluasi kualitas hasil *Clustering*. Validasi yang dilakukan yaitu *Davies Bouldien Indeks* yang merupakan mengukur tingkat kesamaan antarcluster dengan mempertimbangkan rasio antara jarak dalam *cluster* dan jarak antarcluster dan Koefisien *silhoutte* untuk menilai kedekatan setiap objek terhadap *cluster*-nya sendiri dibandingkan dengan *cluster* lain (Mustika, et al., 2021).

Beberapa faktor dan indikator dalam menentukan efektifitas suatu aplikasi yaitu aplikasi mobile taspen otentikasi diantaranya adalah (faktor *User* (pengguna), *Session* dan *Session Interval*, *Retention*, *Efektivitas Login*, *Churn Rate*, *New Users*, *Error Count*, *Conversion Rate* dan *Response Time*) yang saat ini masih belum merata dan berdasarkan karakteristik tersebut dapat menjadi acuan dalam pengelompokan *Cluster* tingkat efektifitas penggunaan aplikasi provinsi-provinsi di Indonesia (Zufwari Fadli, 2016). Berdasarkan faktor tersebut, dilakukan pengelompokan provinsi berdasarkan efektivitas penggunaan aplikasi taspen otentikasi yang dapat menjadi langkah untuk meningkatkan penggunaan dan keberhasilan aplikasi taspen otentikasi di seluruh Indonesia dalam menentukan strategi yang tepat untuk masing-masing kelompok provinsi dan meningkatkan efektivitas penggunaan aplikasi dan mendukung pengguna dengan lebih baik (Ali Muhidin, 2009).

Penelitian ini *relevan* terhadap penelitian sebelumnya yang telah dilakukan oleh beberapa peneliti lain. Diantaranya penelitian yang dilakukan oleh Pratiwi (2016), yang menerapkan algoritma *K-Means Clustering* untuk memprediksi minat nasabah pada PT Asuransi Jiwa Bersama 1912 Bumiputera Prabumulih, penelitian ini menggunakan data nasabah yang diolah menggunakan aplikasi Weka dengan metode *K-Means*. Penelitian menghasilkan tiga *Cluster* yang optimal berdasarkan jumlah minat

nasabah terhadap aplikasi Weka. Kategori pengelompokan terdiri dari *Cluster* satu dengan kategori sedikit, *Cluster* dua dengan kategori sedang dan *Cluster* tiga dengan kategori banyak. Penelitian lainnya oleh Astria (2019), yang membahas tentang penerapan *k-medoids* pada kelompok rumah tangga yang memiliki sumber penerangan Perusahaan Listrik Negara (PLN) berdasarkan Provinsi menunjukkan bahwa dengan penggunaan 2 *Cluster* menghasilkan kelompok data dengan potensi yang dimiliki yaitu *Cluster* 1 yaitu *Cluster* tingkat perilaku memilah sampah tinggi diperoleh 11 provinsi yaitu Aceh, Kep. Bangka Belitung, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, dan Sulawesi Utara serta 23 provinsi lainnya termasuk dalam *Cluster* tingkat rendah.

Berdasarkan penelitian yang dilakukan oleh Agustin dan Sirait (2021), diperoleh hasil bahwa pada perbandingan analisis *Cluster K-Means* dan *K-Medoids* pada data penyakit di Indonesia, diperoleh metode *K-Means* dan menghasilkan nilai variansi yang lebih kecil dibandingkan metode *K-Medoids*, sehingga diperoleh bahwa analisis *Cluster K-Means* lebih optimal dibandingkan metode *K-Medoids*. Namun, penelitian lainnya yang dilakukan oleh Marlina, et al. (2018), menyampaikan bahwa dari kedua algoritma dalam metode non hierarki yakni *K-Medoids* dan *K-Means* dengan menggunakan diperoleh hasil bahwa pada metode *K-Medoids* mampu melakukan pengelompokan pada data sebaran anak cacat dengan lebih baik dibandingkan metode *K-Means*, karena pada metode *K-Medoids* mampu menghasilkan nilai *Silhouette Coefficient* yang lebih kecil dibandingkan metode *K-Means*.

Berdasarkan permasalahan yang telah dipaparkan di atas, dapat terlihat bahwa tingkat penggunaan aplikasi di Indonesia masih menjadi permasalahan yang harus diatasi. Dalam penelitian ini peneliti menggunakan kedua analisis *Cluster* dan membandingkan kedua metode *K-Means* dan *K-Medoids* dengan tujuan untuk melihat perbandingan hasil dari kedua metode tersebut dan melihat algoritma metode mana yang lebih efisien untuk digunakan dalam data penelitian ini. Oleh karena itu, peneliti ingin mengelompokkan provinsi di Indonesia menjadi beberapa kelompok berdasarkan tingkat indikator penggunaan aplikasi di provinsi Indonesia, sehingga peneliti mengangkat penelitian dengan judul **“Perbandingan Algoritma *K-Means* dan *K-Medoids* Clustering Dalam Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Efektivitas Aplikasi Taspen Otentikasi dengan *DBI* dan Koefisien *Silhouette*”**.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, maka rumusan masalah yang akan dikaji dalam penelitian ini adalah sebagai berikut:

1. Bagaimana pengelompokkan setiap Provinsi di Indonesia berdasarkan indikator efektivitas penggunaan aplikasi taspen otentikasi dengan analisis *cluster* menggunakan *k-means* dan *-medoids clustering* dengan *DBI* dan Koefisien *Silhouette*?
2. Bagaimana perbandingan hasil *cluster* dari penggunaan aplikasi taspen otentikasi analisis *cluster* menggunakan *k-means* dan *k-medoids clustering* dengan *DBI* dan Koefisien *Silhouette*?

1.3 Tujuan

Berdasarkan rumusan masalah, maka tujuan yang akan dicapai dalam penelitian ini adalah sebagai berikut:

1. Mengetahui hasil pengelompokkan setiap Provinsi di Indonesia berdasarkan efektivitas penggunaan aplikasi taspen otentikasi dengan analisis *cluster* menggunakan *k-means* dan *k-medoids clustering* dengan *DBI* dan Koefisien *Silhouette*
2. Mengetahui metode dan algoritma *cluster* yang optimal dalam indikator penggunaan aplikasi taspen otentikasi dengan analisis *cluster* menggunakan *k-means* dan *k-medoids clustering* dengan *DBI* dan Koefisien *Silhouette*

1.4 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah sebagai berikut:

1. Bagi penulis adalah untuk menerapkan teori matematika dalam bidang statistika khususnya pengelompokkan suatu objek.
2. Bagi pembaca adalah untuk menambah wawasan dan pengetahuan sebagai bahan referensi untuk penelitian selanjutnya.
3. Bagi perusahaan adalah sebagai bahan informasi mengenai pengelompokkan Provinsi di Indonesia berdasarkan efektivitas penggunaan aplikasi taspen otentikasi.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Menggunakan data aplikasi taspen otentikasi Provinsi Indonesia.
2. Terdapat faktor atau indikator efektivitas penggunaan aplikasi yang digunakan sebagai variabel dalam penelitian ini. Indikator tersebut terdiri atas *user* (pengguna), *session* dan *session interval*, *retention*, efektivitas *login*, *churn rate*, *new users*, *error count*, *conversion rate* (rasio konversi) dan *response time*.