

RANCANG BANGUN APLIKASI POLA KLAIM MENGGUNAKAN METODE K-MEANS CLUSTERING

STUDI KASUS : PESERTA PT ASABRI (PERSERO) KANCAB BATAM

Salsha Nanditya Abbas¹, Agus Fatulloh, S.T, M.T ²

^{1,2*} Teknik Informatika, Teknik Informatika, Politeknik Negeri Batam

Email: ¹slshabbs@gmail.com, ²agusf@polibatam.ac.id

Abstrak

PT ASABRI (Persero) Cabang Batam menghadapi sejumlah kendala dalam pengelolaan klaim peserta, terutama terkait hal efisiensi dalam pemrosesan klaim dan kepatuhan terhadap pelaporan. Kerumitan dalam verifikasi dokumen dan jumlah klaim yang tinggi sering kali menyebabkan keterlambatan dan peningkatan biaya operasional. Penelitian ini bertujuan untuk mengimplementasikan sistem identifikasi pola klaim peserta menggunakan machine learning dengan algoritma K-Means clustering.

Metode ini digunakan untuk mengelompokkan klaim berdasarkan berbagai variabel usia peserta, status pensiun, jenis pekerjaan dan mitra guna memahami pola klaim yang terbentuk. Proses analisis data klaim dilakukan melalui serangkaian tahapan data mining, meliputi pengolahan, standarisasi data, penentuan jumlah cluster, serta evaluasi hasil clustering dengan metode Silhouette Score.

Kata kunci: *PT ASABRI (PERSERO) Cabang Batam, K-Means Clustering, Machine Learning, Identifikasi Pola, Manajemen Klaim*

DESIGN OF CLAIM PATTERN APPLICATION USING K-MEANS CLUSTERING METHOD

CASE STUDY : PT ASABRI (PERSERO) PARTICIPANTS IN BATAM BRANCH

Abstract

PT ASABRI (Persero) Batam Branch faces a number of obstacles in managing participants' claims, particularly in terms of efficiency in claims processing and compliance with reporting. The complexity of document verification and the high number of claims often cause delays and increased operational costs. This study aims to implement a system for identifying participants' claim patterns using machine learning with K-Means clustering algorithm.

This method is used to cluster claims based on various variables of participant age, retirement status, type of work and partner to understand the pattern of claims formed. The process of analyzing claims data is carried out through a series of data mining stages, including processing, data standardization, determining the number of clusters, and evaluating clustering results with the Silhouette Score method.

Keywords: *PT ASABRI (PERSERO) Batam Branch, K-Means Clustering, Machine Learning, Pattern Identification, Claim Management*

I. PENDAHULUAN

PT ASABRI (Persero) Kancab Batam adalah Badan Usaha Milik Negara (BUMN) yang memiliki fungsi penting dalam manajemen manfaat asuransi bagi anggota TNI, Polri, serta ASN Kemhan/Polri. Layanan yang ditawarkan diantaranya program Jaminan Hari Tua, Jaminan Kematian, dan pembayaran pensiun. Layanan tersebut bertujuan untuk menjamin kesejahteraan para peserta dan keluarganya setelah masa dinas selesai atau saat menghadapi risiko tertentu [1]. Ketepatan dan efisiensi dalam pengelolaan data

menjadi landasan utama untuk memastikan bahwa hak-hak pensiunan diberikan secara tepat waktu dan untuk mencegah kemungkinan penyalahgunaan anggaran negara. Dalam menjalankan tugasnya, proses autentikasi yang dilakukan secara berkala menjadi langkah verifikasi yang krusial yang harus dilakukan oleh peserta pensiunan. Autentikasi merupakan verifikasi rutin yang memastikan bahwa penerima pensiun masih hidup dan berhak menerima tunjangan. Proses ini sangat penting untuk menjaga akuntabilitas dan efisiensi dalam pengelolaan dana pensiun negara.

Namun, pelaksanaan autentikasi sering menghadapi berbagai kendala, khususnya di kantor PT

ASABRI (Persero) Kancab Batam. Kendala utama adalah minimnya pemahaman yang mendalam mengenai variasi karakteristik pensiunan yang sangat beragam, baik dari segi usia, status pensiun, jenis pekerjaan, maupun tingkat literasi teknologi, maupun wilayah geografis. Pendekatan autentikasi atau penyuluhan yang seragam dan umum terbukti tidak berhasil karena tidak dapat memenuhi kebutuhan dan tantangan spesifik dari setiap kelompok pensiunan. Akibatnya, banyak pensiunan yang mengalami kesulitan atau gagal selama proses autentikasi sehingga menyebabkan inefisiensi operasional, antrean panjang, waktu tunggu yang terlalu lama, serta penggunaan sumber daya staf yang tidak optimal karena tidak diprioritaskan kepada kelompok yang paling membutuhkan [2].

Situasi ini menegaskan pentingnya pendekatan yang lebih sistematis dan berbasis data untuk mengenali pola perilaku dan karakteristik para pensiunan. Dengan pemahaman yang lebih mendalam mengenai pensiunan yang mengalami masalah autentikasi beserta penyebabnya, PT ASABRI (Persero) Kancab Batam dapat merancang strategi penanganan yang lebih tepat dan efisien. Selain itu, data pensiunan yang telah berhasil autentikasi juga merupakan aset yang sangat berharga. Analisis mendalam terhadap data ini dapat memberikan wawasan tentang profil pensiunan yang relevan, membantu dalam menemukan peluang untuk meningkatkan kualitas pelayanan, merancang program keterlibatan yang lebih efektif, atau bahkan mengidentifikasi anomali atau perubahan kecil dalam pola perilaku yang penting yang mungkin terlewat dari pemeriksaan rutin, sehingga mendukung deteksi risiko secara lebih proaktif [3].

Algoritma K-Means clustering adalah salah satu teknik dalam data mining yang berfungsi untuk mengelompokkan data berdasarkan kemiripan karakteristik. Dengan menerapkan K-Means pada data historis klaim dan autentikasi pensiunan, dimungkinkan untuk mengidentifikasi kelompok-kelompok pensiunan (klaster) yang memiliki profil serupa. Setiap klaster akan mewakili segmen pensiunan dengan kebutuhan serta tantangan autentikasi yang berbeda-beda [4]. Untuk data yang belum autentikasi, proses klasterisasi akan berperan dalam mendeteksi pensiunan yang berada dalam risiko tinggi yang membutuhkan pendampingan atau tindakan khusus dari staf, sehingga upaya untuk meningkatkan tingkat autentikasi dapat terlaksana dengan lebih efektif dan tepat sasaran. Sebaliknya, data yang sudah autentikasi, proses klasterisasi akan memberikan gambaran mengenai segmentasi perilaku dari pensiunan yang mematuhi aturan, yang memungkinkan PT ASABRI

(Persero) Kancab Batam untuk meningkatkan layanan mereka, merancang program keterlibatan yang sesuai, serta secara proaktif mengenali anomali atau perubahan pola yang mungkin memerlukan perhatian lebih, bahkan pada data yang sudah terverifikasi [5].

Berdasarkan permasalahan dan potensi solusi yang telah dijelaskan, penelitian ini bertujuan untuk merancang dan membangun sebuah sistem yang dapat mengidentifikasi pola klaim peserta PT ASABRI (Persero) Kancab Batam menggunakan algoritma K-Means clustering. Sistem ini tidak hanya akan melakukan pengelompokan data, tetapi juga menyajikan wawasan yang bermanfaat bagi pengambilan keputusan strategis guna meningkatkan efisiensi, efektivitas, dan kualitas layanan di dalam proses autentikasi, baik bagi pensiunan yang belum autentikasi maupun pensiunan yang sudah autentikasi.

II. METODE

1. Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah data primer yang merupakan catatan historis klaim peserta PT ASABRI (Persero) Kancab Batam. Data ini diperoleh dari database internal perusahaan dan mencakup beberapa variabel seperti usia peserta, status pensiun, jenis pekerjaan, dan mitra serta informasi penting lainnya yang menggambarkan karakteristik klaim tersebut. Data yang dikumpulkan meliputi periode April 2024 hingga Juli 2024.

Di samping itu, pengumpulan data ini memiliki keistimewaan karena peneliti terlibat langsung dalam proses entri data dan pengelolaan awal data klaim selama periode magang di PT ASABRI (Persero) Kancab Batam. Keterlibatan langsung dalam proses ini memberikan pemahaman mendalam mengenai struktur data, kemungkinan masalah data (seperti nilai yang hilang atau inkonsistensi), serta memberikan konteks operasional untuk setiap record klaim. Data kemudian diekspor dalam format terstruktur untuk analisis lebih lanjut.

Terakhir, pengumpulan data primer di lakukan dengan cara wawancara terstruktur bersama pihak-pihak kunci di PT ASABRI (Persero) Kancab Batam. Wawancara ini melibatkan Staf Pelayanan yaitu Ibu Lintang Chandra Dewi. Tujuan wawancara ini adalah:

- Melakukan verifikasi terhadap masalah yang terdeteksi dalam manajemen klaim.
- Memahami lebih jauh kebutuhan dan ekspektasi manajemen terhadap sistem identifikasi pola klaim.

- Mengonfirmasi interpretasi awal terhadap pola-pola klaim yang mungkin muncul dari data.
- Mendapatkan insight kualitatif mengenai faktor-faktor non-teknis yang memengaruhi efisiensi pengelolaan klaim [4].

2. Metode Pengembangan Sistem

Metode pengembangan sistem dalam penelitian ini menggunakan kerangka kerja CRISP-DM (Cross-Industry Standard Process for Data Mining), yang merupakan standar terstruktur untuk proyek data mining dan machine learning. Tahapan-tahapan yang dilakukan adalah sebagai berikut [6] [7] :

- Pemahaman Bisnis (Business Understanding)

Pada tahap ini, dilakukan identifikasi masalah utama dalam pengelolaan klaim di PT ASABRI (Persero) Kancab Batam, seperti inefisiensi serta kurangnya pemahaman terhadap pola klaim. Tujuan penelitian, yaitu mengembangkan sistem identifikasi pola klaim berbasis machine learning untuk meningkatkan efisiensi dan mendukung proses pengambilan keputusan, yang telah ditentukan secara jelas melalui tinjauan literatur serta wawancara dengan pihak terkait internal.

- Pemahaman Data (Data Understanding)

Data klaim historis yang telah dikumpulkan dianalisis secara mendalam. Tahap ini mencakup evaluasi awal terhadap kualitas data, penentuan format data, tipe variabel (numerik, kategorial), dan distribusi data. Analisis statistik deskriptif dilakukan untuk memahami karakteristik umum data serta mengidentifikasi potensi masalah seperti nilai hilang atau outliers.

- Persiapan Data (Data Preparation)

Data mentah yang sudah dianalisis kemudian menjalani serangkaian proses preprocessing untuk menjamin kualitas dan format yang sesuai untuk pemodelan machine learning. Langkah-langkah yang dilakukan meliputi:

- ✓ Data Cleaning : Penanganan nilai yang hilang (misalnya, imputasi atau penghapusan record) serta perbaikan inkonsistensi data.
- ✓ Data Transformation : mengubah variabel kategori ke dalam format numerik yang dapat di proses algoritma K-Means, seperti menggunakan One-Hot Encoding atau Label Encoding
- ✓ Feature Selection / Engineering : Memilih variabel yang paling signifikan untuk

pengelompokan atau menciptakan fitur baru dari data yang ada guna meningkatkan kinerja model.

- ✓ Standardisasi Data: Penerapan teknik normalisasi (contohnya : Min-Max Scaling atau Standardization-Z-Score normalization) terhadap variabel numerik untuk memastikan semua fitur memiliki skala yang konsisten, agar tidak ada variabel yang mendominasi perhitungan jarak pada algoritma K-Means.

- Pemodelan (Modeling)

Tahap ini berorientasi pada implementasi algoritma K-Means Clustering. Pemilihan jumlah cluster (K optimal) merupakan langkah penting dan dilakukan menggunakan satu pendekatan yaitu Metode Silhouette. Metode Silhouette merupakan metode menghitung Silhouette Score untuk setiap data point dan memilih nilai K yang menghasilkan nilai rata-rata Silhouette Score tertinggi, yang menunjukkan cluster yang lebih padat dan terpisah. Setelah K ditentukan, model K-Means dilatih menggunakan data klaim yang sudah disiapkan. Proses iterasi algoritma (penentuan cluster berdasarkan jarak Euclidean dan perhitungan ulang centroid sebagai rata-rata titik dalam cluster) dilakukan hingga konvergensi tercapai (posisi centroid tidak berubah atau pengelompokan tetap). Centroid diinisialisasi menggunakan pendekatan K-Means ++ untuk meningkatkan kualitas pengelompokan awal.

- Evaluasi (Evaluation)

Evaluasi sistem berfokus pada mutu pengelompokan yang dihasilkan oleh model K-Means dan pemahaman wawasan yang diperoleh. Metrik evaluasi intrinsik clustering yang digunakan adalah Silhouette Score untuk mengukur seberapa baik cluster terdefinisi dan terpisah. Selanjutnya, analisis mendalam terhadap ciri-ciri setiap kluster yang dilakukan dengan mengevaluasi nilai rata-rata dari fitur-fitur. Temuan ini selanjutnya dihubungkan lagi dengan tujuan bisnis untuk memberikan rekomendasi berbasis data yang dapat ditindaklanjuti untuk peningkatan efisiensi pengelolaan klaim. Validasi pemahaman dilakukan melalui diskusi dan tanggapan dari pihak manajemen atau staf PT ASABRI (Persero) Kancab Batam.

- Penyebaran (Deployment)

Hasil identifikasi pola klaim dan insight yang didapat ditampilkan dalam format yang mudah dipahami.

3. Algoritma K-Means Clustering

Clustering merupakan algoritma yang digunakan untuk mengelompokkan data ke dalam beberapa kluster berdasarkan adanya kemiripan satu sama lain, sementara K-Means clustering digunakan untuk membagi kumpulan data dalam jumlah besar dan tidak memiliki label menjadi beberapa cluster (K) yang berbeda [5].

Algoritma K-Means bekerja dengan prinsip iteratif untuk meminimalkan jumlah kuadrat jarak antara titik data dan centroid kluster mereka (juga dikenal sebagai Within-Cluster Sum of Square – WCSS atau inertia). Semakin kecil nilai WCSS, semakin padat (kohesif) kluster yang terbentuk.

Proses kerja algoritma K-Means secara umum memiliki beberapa tahap, yakni :

1. Menentukan jumlah kluster (K)
2. Inisialisasi centroid awal
Memilih K titik acak dari dataset sebagai centroid awal untuk setiap kluster
3. Penugasan kluster
Setiap titik data di assign ke kluster berdasarkan jarak terdekat ke centroid
4. Perhitungan ulang centroid
Hitung posisi centroid baru dengan mengambil rata-rata dari semua titik data dalam setiap kluster
5. Iterasi
Ulangi langkah 3 dan 4 hingga posisi centroid stabil atau tidak ada perubahan dalam penugasan kluster.

Salah satu tantangan dalam K-Means adalah menentukan nilai K (jumlah kluster) yang optimal, karena algoritma memerlukan nilai K sebagai input. Metode yang umum digunakan untuk membantu menentukan nilai K yaitu adalah metode Silhouette Score. Silhouette Score adalah metode yang mengukur seberapa mirip sebuah objek dengan kluster sendiri dibandingkan dengan kluster lain. Silhouette Score berkisar dari -1 hingga 1, dimana nilai yang lebih tinggi menunjukkan kluster yang lebih padat dan terpisah dengan baik. Nilai K yang menghasilkan Silhouette Score tertinggi dianggap optimal [8].

Kelebihan K-Means Clustering

- Algoritma yang sederhana dan mudah dimengerti
- Sangat efektif dalam mengelompokkan data besar dengan waktu komputasi yang cukup singkat

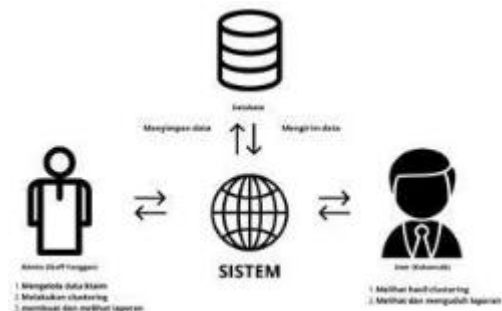
- Mampu menghasilkan pengelompokan yang konsisten dan stabil pada data yang cenderung stabil
- Dapat diterapkan pada data dengan banyak fitur tanpa adanya penurunan performa yang signifikan
- Saat data baru muncul, algoritma dapat dijalankan kembali untuk menyesuaikan data tersebut dalam kluster yang sudah ada [9].

Kekurangan K-Means Clustering

- Pemilihan jumlah kluster (K) yang tepat merupakan hal subjektif dan cukup menantang, karena nilai K yang tidak tepat bisa menghasilkan kluster yang tidak relevan
- Sensitif terhadap inisialisasi centroid yang mengakibatkan hasil pengelompokan yang kurang optimal dan bervariasi setiap kali algoritma dijalankan
- Kesulitan menangani outlier sehingga mempengaruhi hasil pengelompokan
- Kluster dengan ukuran dan kepadatan yang sangat berbeda dapat menghasilkan hasil pengelompokan yang tidak seimbang dan sulit untuk dipahami [9].

III. HASIL DAN PEMBAHASAN

1. Gambaran Umum Sistem



Gambar 1. Gambaran Umum Sistem

Aplikasi Sistem Clustering Data Klaim berbasis web merupakan sistem informasi yang digunakan untuk melakukan pengelompokan data klaim secara otomatis menggunakan metode machine learning (K-Means). Sistem ini mencakup fitur-fitur manajemen data klaim, proses clustering, visualisasi hasil serta pembuatan laporan analisis. Dengan adanya sistem ini, proses untuk menganalisis data klaim menjadi lebih efisien, akurat, dan terstruktur.

Sistem ini terdiri dari beberapa aktor utama, yaitu admin(staff yangangan) dan user(kancab) yang masing-masing memiliki peran dan hak akses yang

berbeda. Berikut penjelasan terkait aktor yang akan menggunakan sistem ini :

a. Admin (Staf Yanggan)

Admin sebagai staf yanggan berperan sebagai pengelola utama sistem. Admin memiliki akses penuh untuk melakukan pengelolaan data klaim, menjalankan proses clustering, mengelola user, dan membuat laporan analisis. Beberapa tugas yang dapat dilakukan admin antara lain :

- Mengelola data klaim (tambah, edit, hapus)
- Menjalankan proses klasterisasi data
- Melihat dan menganalisis hasil klasterisasi
- Membuat dan melihat laporan sistem
- Mengelola akun pengguna sistem

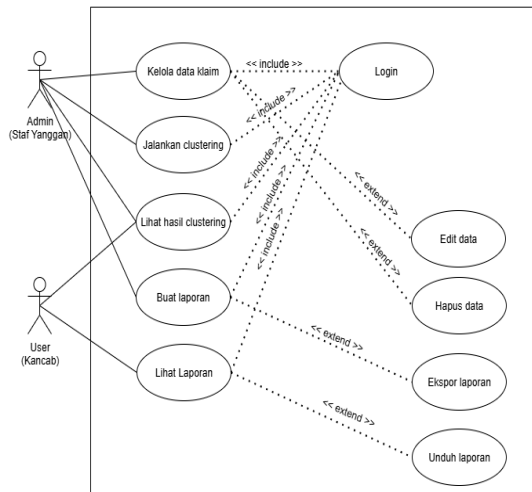
b. User (Kancab)

User sebagai kancab adalah pihak yang dapat mengakses sistem dengan hak terbatas. User dapat melihat hasil clustering dan laporan analisis, namun tidak dapat melakukan perubahan data atau menjalankan proses clustering. Peran user antara lain :

- Melihat hasil klasterisasi data
- Melihat dan mengunduh laporan analisis klaim

2. Perancangan Umum Sistem

a. Usecase



Gambar 2. Usecase

Use case adalah penjelasan terstruktur mengenai bagaimana pengguna (aktor) berinteraksi dengan sebuah sistem untuk mencapai tujuan tertentu dalam konteks yang spesifik. Dalam pengembangan perangkat lunak, use case berfungsi untuk menggambarkan skenario atau alur kerja yang menjelaskan langkah-langkah yang dilakukan

pengguna dan respons sistem terhadap setiap langkah tersebut [10].

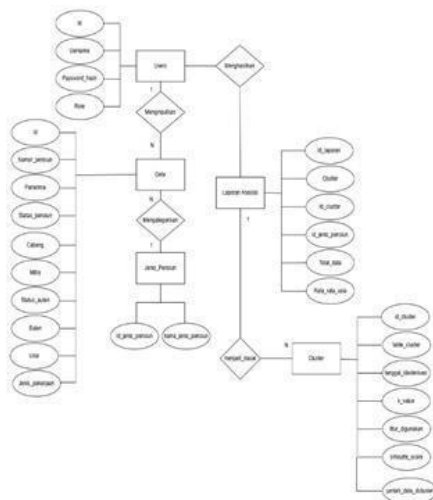
Pada gambar 2. Use case terdapat dua aktor yaitu admin (staf yanggan) dan user (kancab). Penjelasan terkait fungsionalitas diagram usecase dijelaskan dalam tabel 1. Definisi Use case Diagram

No	Use Case	Deskripsi
1	Login	Use case ini adalah tahapan awal pada sistem, yaitu user perlu melakukan login dengan memasukkan kredensial yang valid untuk mendapatkan akses ke sistem
2	Kelola data klaim	Use case ini mendefinisikan tentang pengelolaan data klaim, meliputi penambahan, pengubahan, dan penghapusan data klaim. (include login)
3	Jalankan clustering	Use case ini mendefinisikan tentang proses menjalankan algoritma clustering untuk mengelompokkan data. (Include login)
4	Lihat hasil clustering	Use case ini mendefinisikan tentang melihat dan menampilkan hasil dari proses clustering yang telah dijalankan. (include login)
5	Buat laporan	Use case ini mendefinisikan tentang pembuatan laporan berdasarkan data yang ada di sistem (include login)
6	Lihat laporan	Use case ini mendefinisikan tentang dan menampilkan laporan

		yang telah dibuat. (Include login)
7	Edit data	Use case ini mendefinisikan tentang proses pengubahan data yang sudah ada di sistem. (extend Kelola data klaim)
8	Hapus data	Use case ini mendefinisikan tentang proses penghapusan data dari sistem. (extend Kelola data klaim)
9	Ekspor laporan	Use case ini mendefinisikan tentang proses mengekspor laporan ke format tertentu (include login)
10	Unduh laporan	Use case ini mendefinisikan tentang proses mengunduh laporan ke perangkat pengguna (extend lihat laporan)

Tabel 1. Definisi Use case Diagram

b. Entity Relationship Diagram



Gambar 3. ERD

ERD adalah sebuah alat bantu visual yang digunakan dalam merancang database guna memodelkan struktur data. Ini adalah representasi grafis dari entitas (objek atau konsep penting dalam sistem) dan relasi (koneksi atau interaksi) di antara entitas-entitas tersebut [11].

Tabel users, tabel ini digunakan untuk menyimpan data pengguna yang terdaftar dalam sistem. Setiap pengguna memiliki peran (role) yang menentukan hak akses mereka dalam sistem, seperti mengunggah data atau melihat laporan. Atribut yang tersedia antara lain : id (PK), username, password_hash, dan role.

Tabel data, tabel ini berfungsi untuk menyimpan data mentah atau informasi terkait pensiun yang diunggah ke sistem. Data ini kemudian akan diproses, dikategorikan, dan dianalisis untuk tujuan yang relevan dengan pengelolaan pensiun. Atribut yang tersedia antara lain : id (PK), id_users (FK ke users), nomor_pensiun, nama_penerima, status_auten, bulan, mitra, usia, dan jenis_pekerjaan.

Tabel jenis_pensiun, tabel ini merupakan tabel yang menyimpan kategori jenis pensiun. Atribut yang tersedia antara lain : id_jenis_pensiun dan nama_jenis_pensiun.

Tabel laporan_analisis, tabel ini menyimpan hasil dari proses analisis data. Setiap entri pada tabel ini merupakan ringkasan atau detail dari laporan yang dihasilkan sistem. Atribut yang tersedia antara lain : id_laporan, cluster, id_cluster, id_jenis_pensiun, total_data, dan rata_rata_usia

Tabel cluster, tabel ini digunakan untuk menyimpan informasi mengenai pengelompokan (clustering) data yang dilakukan oleh sistem. Setiap kluster memiliki karakteristik unik yang mengidentifikasi pola atau kelompok data pensiun. Atribut yang tersedia : id_cluster, table_cluster, tanggal_klasterisasi, k_values, fitur_digunakan, silhouette_score, dan jumlah_data_dicluster.

IV. HASIL PENELITIAN

1. Implementasi Antarmuka

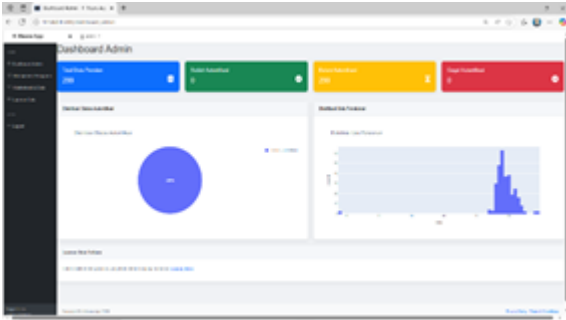
a. Halaman Login



Gambar 3. Halaman Login

Halaman login adalah gerbang utama bagi pengguna untuk mengakses sistem. Oleh karena itu, perancangan antarmuka harus intuitif, aman, dan efisien.

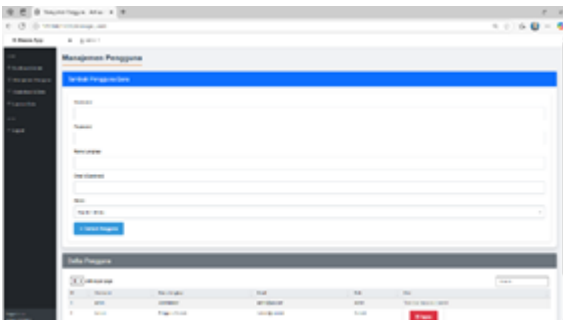
b. Halaman Dashboard (Admin)



Gambar 4. Halaman Dashboard Admin

Halaman dashboard admin adalah halaman utama yang akan terlihat oleh pengguna dengan peran admin setelah login berhasil. Halaman ini menampilkan summary card yang menunjukkan angka-angka penting seperti : Total Data Pensiun, Sudah Autentikasi, Belum Autentikasi, dan Gagal.

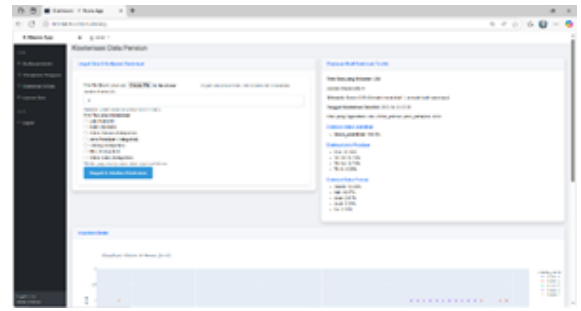
c. Halaman Manajemen Pengguna



Gambar 5. Halaman Manajemen Pengguna

Halaman ini memungkinkan admin untuk mengelola akun pengguna yang dapat mengakses aplikasi. Tampilan ini terdiri dari formulir untuk menambah pengguna baru, meliputi : username, password, nama lengkap, email, dan peran Serta tabel daftar pengguna yang sudah ada.

d. Halaman Klasterisasi dan Data (Pengelolaan Data dan Klasterisasi)

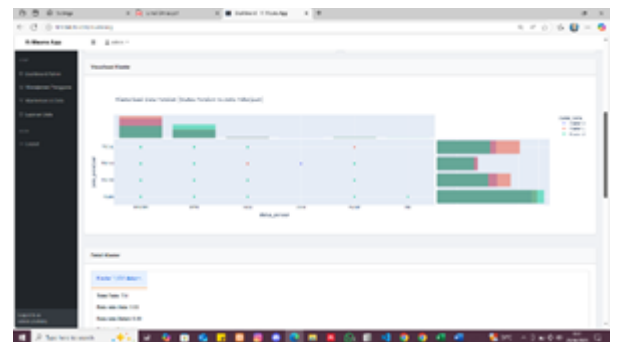


Gambar 6. Halaman Klasterisasi

Halaman ini merupakan pusat untuk mengunggah data baru dan menjalankan proses klasterisasi K-Means. Halaman ini terbagi atas dua bagian utama yaitu Unggah Data dan Konfigurasi Klasterisasi di sebelah kiri, dan Ringkasan Hasil Klasterisasi Terakhir di sebelah kanan.

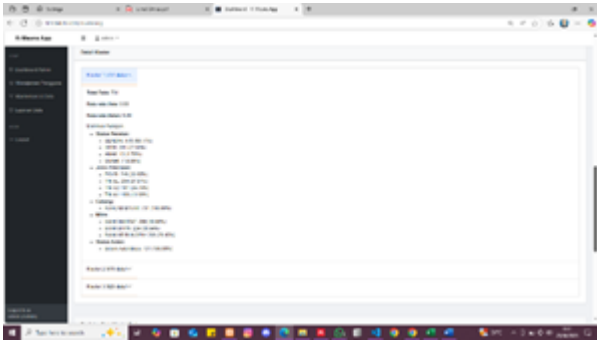
Unggah Data dan Konfigurasi Klasterisasi : terdapat form untuk memilih file excel, memasukkan “jumlah Kluster (K)”, dan memilih “Pilih Fitur untuk Klasterisasi”

Ringkasan Hasil Klasterisasi Terakhir : menampilkan ringkasan hasil klasterisasi terakhir yang dijalankan, seperti “Total Data yang di cluster”, Jumlah Kluster (K)”, “Silhouette Score”, “Tanggal Klasterisasi Terakhir”, dan “Fitur yang digunakan”



Gambar 6.1 Visualisasi Klasterisasi

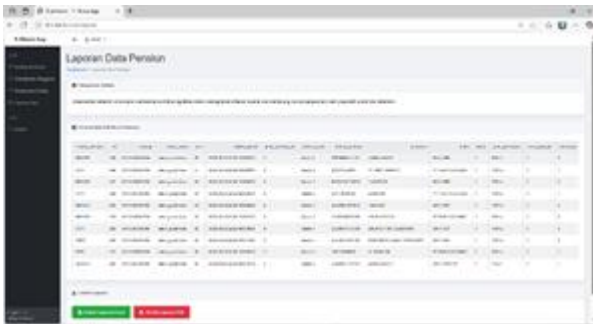
Pada bagian ini adalah komponen sentral untuk memahami hasil klasterisasi secara intuitif. Berdasarkan fitur yang dipilih, sistem akan menghasilkan plot interaktif menggunakan Plotly.



Gambar 6.2 Detail Kluster

Bagian ini memberikan analisis mendalam untuk setiap kluster yang terbentuk. Setiap kluster disajikan dalam format akordeon yang dapat diperluas, memungkinkan administrator untuk melihat karakteristik spesifik dari setiap kelompok.

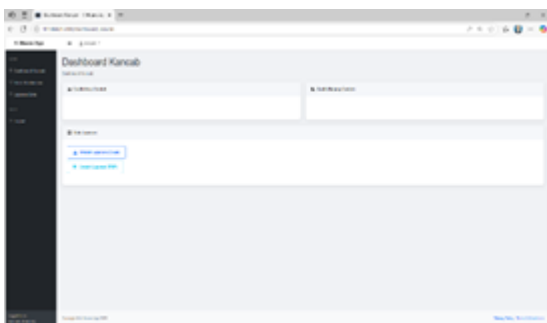
e. Halaman Laporan Data



Gambar 7. Halaman Laporan Data

Halaman laporan menyajikan detail data pensiun yang telah diunggah dan di klusterisasi dalam bentuk tabel, serta opsi untuk mengunduh laporan.

f. Halaman Dashboard Kancab

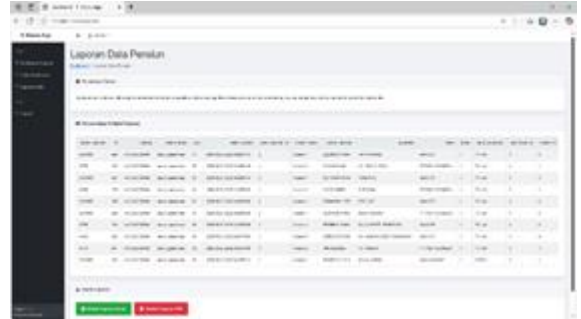


Gambar 8. Halaman Dashboard Kancab

Halaman ini adalah halaman khusus untuk Kancab yang dirancang lebih sederhana dibanding Dashboard Admin. Halaman ini menampilkan ringkasan data atau informasi yang relevan dengan

kebutuhan pengguna Kancab, seperti beberapa grafik ringkasan atau tautan langsung ke laporan.

g. Halaman Laporan Data



Gambar 9. Halaman Laporan Data

Halaman laporan menyajikan detail data pensiun yang telah diunggah dan di klusterisasi dalam bentuk tabel, serta opsi untuk mengunduh laporan.

2. Implementasi K-Means

Dalam penelitian ini, penggunaan algoritma K-Means clustering diterapkan sebagai elemen penting dari aplikasi web dengan menggunakan bahasa pemrograman Python dan pustaka-pustaka data science yang sesuai. Berikut adalah uraian mencakup struktur sistem, alur pemrosesan data, dan modul-modul penting, serta teknologi yang digunakan :

a. Implementasi Library

Sistem mengintegrasikan berbagai pustaka (Library) Python untuk mendukung operasional aplikasi web, pengelolaan data, proses klusterisasi, dan pembuatan laporan.

```
import os
from datetime import datetime
import pandas as pd
import numpy as np
from io import BytesIO
import json

from flask import Flask, render_template, request, redirect, url_for, flash, jsonify, send_file
from flask_sqlalchemy import SQLAlchemy
from flask_login import LoginManager, UserMixin, login_user, logout_user, login_required, current_user
from werkzeug.security import generate_password_hash, check_password_hash
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
import plotly.express as px
import plotly.graph_objects as go
from sqlalchemy import func

s = Import ReportLab for PDF generation
from reportlab.lib.pagesizes import A4
from reportlab.lib.units import inch
from reportlab.platypus import SimpleDocTemplate, Paragraph, Spacer, Table, TableStyle, PageBreak, Image
from reportlab.lib.styles import getSampleStyleSheet, ParagraphStyle
from reportlab.lib import colors
```

Gambar 10. Pustaka Python

- Flask adalah sebuah kerangka minimalis dan adaptif untuk pengembangan aplikasi web yang berfungsi sebagai dasar utama untuk menciptakan solusi berbasis web

- SQLAlchemy berperan sebagai Mapper Relasional Objek yang mempermudah interaksi antara aplikasi Python dan database SQLite, serta mendukung pengelolaan data efektif.
 - Flask-Login adalah sebuah tambahan pada flask yang menyajikan fasilitas untuk pengelolaan sesi pengguna, termasuk autentikasi dan otorisasi demi menjaga keamanan akses aplikasi
 - Pandas adalah sebuah Pustaka untuk visualisasi data interaktif yang digunakan dalam pembuatan grafik dan plot dinamis yang menyajikan hasil klasterisasi dalam antarmuka web
 - ReportLab adalah pustaka Python yang dirancang untuk membuat dokumen PDF secara otomatis, dimanfaatkan untuk menghasilkan laporan klasterisasi yang sistematis dan informatif.
- b. Mekanisme Kerja Algoritma K-Means
- Pada tahap ini menjelaskan secara rinci setiap tahapan dalam implementasi K-Means, mulai dari persiapan data hingga evaluasi dan penyimpanan hasil klasterisasi.
1. Pengumpulan dan Persiapan Data.
 - Implementasi : Data historis klaim peserta PT ASABRI (Persero) Kancab Batam dikumpulkan dalam format file excel (.xlsx). Dalam aplikasi, Admin sebagai staf yang bertugas mengunggah file ini melalui antarmuka web.
 - Kode : fungsi `pd.read_excel(file)` dari pustaka Pandas digunakan untuk mengambil data dari excel ke dalam objek DataFrame. Nama-nama kolom selanjutnya dinormalisasi (misalnya : `df_excel.columns.str.lower().str.replace(' ', '_')`).
 - Persiapan : Nilai yang hilang (missing values) dalam kolom-kolom tertentu (contohnya `status_auten`, `status_pensiun`, `jenis_pekerjaan`, `mitra`) diisi secara otomatis dengan nilai default seperti “Belum Autentikasi” atau “Tidak Diketahui” melalui metode `fillna()` pada DataFrame Pandas.
 2. Pemilihan Fitur (atribut).
 - Implementasi : Admin dapat memilih fitur-fitur tertentu yang akan digunakan untuk proses klasterisasi lewat antarmuka di halaman “Klasterisasi dan Data”. Fitur-fitur ini dipilih berdasarkan keterkaitannya dalam mengidentifikasi pola klaim, seperti “Usia”, “Status_pensiun”, “Jenis Pekerjaan”, dan “Mitra”.
 - Kode : Fitur yang dipilih dari form (`request.form.getlist('Features')`) digunakan untuk membuat tabel subset DataFrame bar (`df_cluster = df[available_features].copy()`) yang hanya berisi kolom-kolom yang akan diklasterisasi
 3. Standarisasi dan Encoding Data.
 - Implementasi : mengingat algoritma K-Means bergantung pada jarak, sangat penting untuk menjamin bahwa semua fitur memiliki skala yang seimbang. Selain itu, fitur kategori perlu diubah menjadi format angka.
 - Kode : langkah ini dilakukan dalam fungsi `get_data_for_clustering` :
 - Untuk fitur numerik seperti usia, `StandardScaler` dari `sklearn.preprocessing` diterapkan. Baris kode `scaler = StandardScaler()` diikuti dengan `df_cluster[feature] = scaler.fit_transform(df_cluster[[feature]])` akan mengubah data sehingga memiliki nilai rata-rata 0 dan deviasi standar 1.
 - Untuk fitur kategorikal seperti `Status_pensiun`, `Jenis_pekerjaan`, dan `Mitra` digunakan `LabelEncoder` dari `sklearn.preprocessing`. Baris kode `encoder = LabelEncoder()` diikuti dengan `df_cluster[feature] = encoder.fit_transform(df_cluster[feature])` akan mengonversi setiap kategori teks menjadi angka integer yang berbeda.
 4. Penentuan Jumlah Klaster (K).
 - Implementasi : Kuantitas klaster (K) ditetapkan oleh admin melalui input pada antarmuka web. Walaupun pemilihan K dapat rumit, aplikasi ini mendukung eksplorasi dengan memberi kesempatan kepada admin untuk menguji berbagai nilai K.
 - Kode : Nilai K diperoleh dari form (`num_cluster = request.form.get('num_clusters', 3, type=int)`). Selama proses evaluasi, metrik seperti `Silhouette Score` dihitung secara otomatis untuk menunjukkan kualitas klasterisasi pada K yang telah dipilih.
 5. Inisialisasi Centroid dan Iterasi (Proses K-Means).
 - Implementasi : setelah data siap, algoritma K-Means digunakan untuk mengelompokkan data.
 - Kode : kelas `KMeans` dari `sklearn.cluster` digunakan untuk menerapkan algoritma. Baris

kunci merupakan kmeans =
Kmeans(n_clusters=n_clusters,
random_state=42, n_init=10)

- n_clusters : mengatur jumlah kluster (K) yang ditentukan Admin
 - random_state=42 : memastikan hasil yang serupa setiap kali proses klusterisasi dijalankan dengan parameter yang sama
 - n_init=10 : mengarahkan algoritma untuk melaksanakan K-Means sebanyak 10 kali dengan inialisasi centroid yang berbeda (secara default menggunakan metode k-means ++), dan memilih hasil terbaik berdasarkan inersia (WCSS), untuk mengurangi ketergantungan pada inialisasi awal.
 - Metode kmeans.fit_predict(df_cluster) secara internal melakukan proses penugasan kluster dan pembaruan centroid secara iteratif, dan akhirnya mengembalikan label kluster untuk setiap titik data.
6. Evaluasi Hasil Klustering.
- Implementasi : Kualitas pengelompokan di nilai dengan menggunakan Silhouette Score yang ditampilkan secara langsung di antarmuka web dan juga dimasukkan dalam laporan.
 - Kode : fungsi silhouette_score(df_cluster, clusters) dari sklearn.metrics digunakan untuk menghitung nilai Silhouette Score. Hasil dari klusterisasi juga divisualisasikan dengan menggunakan pustaka Plotly (plotly.express dan plotly.graph_objects) yang menghasilkan plot interaktif (contoh : box plot atau scatter plot sesuai dengan fitur yang dipilih) yang disimpan dalam format JSON (fig.to_json()) dan ditampilkan di frontend.
7. Interpretasi Hasil dan Penyimpanan.
- Implementasi : setelah melakukan klusterisasi, karakteristik dari setiap kluster dihitung dan ditampilkan pada bagian “Detail Setiap Kluster” di halaman web. Hasil ini juga disimpan untuk laporan dan referensi historis dan pelaporan. Paling penting, karakteristik ini menjadi dasar untuk merumuskan strategi penanganan dan penyuluhan yang spesifik untuk setiap kluster.
 - Kode : karakteristik kluster (rata-rata usia, distribusi fitur kategorial) dihitung secara otomatis untuk setiap kluster yang terbentuk. Data ini bersama dengan Silhouette score, fitur yang digunakan, dan grafik visualisasi, disimpan ke dalam tabel Cluster di dalam database (db.session.add(new_cluster_run)).

Data asli yang sudah dilabel kluster disimpan ke tabel Data (db.session.add_all(data_to_save)).

c. Implementasi File Sistem

Bagian ini menjelaskan bagaimana interaksi dengan sistem berkas (file system) diimplementasikan dalam aplikasi, khususnya terkait dengan proses unggah dan penyimpanan data.

- Unggah file : aplikasi memungkinkan admin untuk mengunggah file data historis klaim dalam format .xlsx. Proses ini melibatkan penanganan request HTTP untuk unggahan file dan penyimpanan sementara file di server
- Proses file : setelah diunggah, file dibaca menggunakan Pandas, dan data diekstrak untuk proses klusterisasi
- Penyimpanan data : data yang telah diproses dan hasil klusterisasi disimpan ke dalam database SQLite melalui SQLAlchemy, memastikan persistensi data dan kemudahan akses untuk analisis lebih lanjut dan pelaporan.

3. Pengujian Aplikasi

Untuk memastikan fitur pada web rancang bangun K-Means Clustering studi kasus PT ASABRI (Persero) Kancab Batam berjalan dengan semestinya, maka dilakukan pengujian pada website dengan menggunakan metode Black Box Testing. Hasil Pengujiannya disajikan dalam Tabel 2. Pengujian Fitur Website Menggunakan Black Box Testing

No	Fitur	Pengujian	Status	Keterangan
1	Halaman Login	Login dengan username dan password	Berhasil	Berhasil login ke Dashboard sesuai role
2	Unggah data	Unggah file excel valid dari halaman “Klusterisasi & Data”	Berhasil	Data berhasil diunggah dan disimpan di DB
3	Jalankan klusterisasi	Pilih fitur dan K, lalu jalankan klusterisasi	Berhasil	Silhouette score dan visualisasi kluster muncul

4	Lihat detail klaster	Akses "Detail Seriap Klaster" dan perluas setiap klaster	Berhasil	Detail karakteristik klaster
5	Tambah pengguna	Isi form "Tambah Pengguna Baru" dengan data valid dan klik "Tambah Pengguna"	Berhasil	Pengguna baru berhasil ditambahkan ke daftar
6	Hapus pengguna	Klik "Hapus" pada pengguna (bukan akun sendiri) dan konfirmasi	Berhasil	Pengguna berhasil di hapus dari daftar.
7	Lihat laporan data (admin)	Akses halaman "Laporan Data" sebagai admin	Berhasil	Tabel data lengkap (termasuk hasil klasterisasi ditampilkan)
8	Lihat laporan data (kancab)	Akses halaman "Laporan Data" sebagai kancab.	Berhasil	Tabel data (termasuk hasil klasterisasi) ditampilkan (difilter jika diimplementasikan)
9	Unduh laporan Excel / PDF	Klik "Unduh data Excel" atau klik "Unduh PDF" dari halaman laporan data	Berhasil	File excel berisi data dan file PDF berisi laporan klasterisasi berhasil diunduh
10	Log out	Verifikasi pengguna dapat keluar dari sistem dengan aman	Berhasil	Pengguna diarahkan kembali ke halaman login dan sesi berakhir

Tabel 2. Pengujian Fitur Website

V. KESIMPULAN

Penelitian ini berhasil mengembangkan sebuah sistem yang mampu mengidentifikasi pola klaim yang memanfaatkan machine learning dengan menggunakan metode CRISP-DM (Cross-Industry Standard Process for Data Mining) di PT ASABRI (Persero) Kancab Batam. Melalui proses yang sistematis dari Business Understanding sampai dengan Deployment, sistem ini dikembangkan untuk mengatasi masalah ifisiensi dalam pengelolaan klaim sekaligus membantu proses pengambilan keputusan yang lebih efektif. Penerapan algoritma K-Means clustering terbukti berhasil dalam mengelompokkan data klaim yang sudah ada. Proses klasterisasi mencakup beberapa langkah penting, seperti pengumpulan serta persiapan data daei file Excel, pemilihan fitur yang relevan oleh admin melalui antarmuka web, serta standarisasi dan encoding data agar sesuai dengan algoritma berbasis jarak. Menetapkan jumlah klaster (K) yang fleksibel memungkinkan penelusuran pola yang bervariasi, dengan penilaian kualitas pengelompokan menggunakan Silhouette Score yang ditampilkan secara interaktif. Penentuan centroid menggunakan K-Means ++ dan pemilihan iterasi yang tepat menjamin pengelompokan yang robust.

Hasil dari klasterisasi memberikan pemahaman mendalam tentang karakteristik berbagai pola klaim yang beragam, termasuk segmentasi berdasarkan status autentikasi peserta, usia, jenis pekerjaan, dan mitra. Analisis hasil ini memungkinkan PT ASABRI (Persero) Kancab Batam untuk menyusun strategi penyuluhan dan edukasi yang lebih spesifik dan sesuai sasaran bagi setiap klaster. Sistem ini tidak hanya berfungsi untuk meningkatkan efisiensi operasional dalam pengelolaan klaim tetapi juga menawarkan dasar data yang kuat untuk pengambilan keputusan strategis.

Walaupun sistem ini sudah menunjukkan kemampuannya, masih terdapat beberapa keterbatasan yang dapat menjadi fokus dalam penelitian selanjutnya. Pengembangan lebih lanjut bisa melibatkan integrasi dengan sumber data real-time, eksplorasi algoritma klastering lainnya untuk perbandingan, atau penambahan modul prediksi untuk mengantisipasi pola klaim di masa mendatang.

VI. DAFTAR PUSTAKA

- [1] P. A. (Persero), "BUMN yang menaungi Asuransi Sosial Prajurit TNI, Anggota Polri, ASN Kemhan, dan ASN Polri," ASABRI, [Online]. Available: www.asabri.co.id.

- [2] L. N. E. W. A. S. Rusmini Icha Indryani, "Pengelolaan Dana Pensiun Yang Efektif dan Efisien dalam Menjamin Kesejahteraan Keuangan di Masa Tua," 2024.
- [3] R. A. S. I. S. H. R. S. Deanyta Ratnasari Permana, "Penggunaan Big Data untuk Peningkatan Efisiensi Pelayanan Publik di Pemerintahan Lokal di Kabupaten Subang".
- [4] W. I. R. Seta Permana, "Sistem Aplikasi Data Pegawai Pensiun Menggunakan Metode K-Means (Studi Kasus : PT. Kawasan Berikat Nusantara (Persero))," 2020.
- [5] Z. F. Moh. Baha Uddin, "Penerapan Data Mining Clustering K-Means Dalam Mengelompokkan Data Penduduk Penyandang Disabilitas," 2024.
- [6] D. A.-P. Learning, "Teknik Analisis Data CRISP-DM dalam Data Mining," DQLAB AI-Powered Learning, [Online]. Available: www.dqlab.id.
- [7] I. N. Nur Cholifah Sastya, "Penerapan Metode CRISP-DM dalam Menganalisis Data untuk," 2023.
- [8] H. P. E. A. Khoirulli Nurul Fatimah, "Penerapan Algoritma K-Means Clustering Dalam Pengelompokan Hasil Belajar Siswa di SMP Budi Luhur Samarinda".
- [9] D. A.-P. Learning, "K-Means Clustering, Salah Satu Contoh Teknik Analisis Data Populer," DQLab AI-Powered Learning. [Online].
- [10] W. H. N. P. D. P. Habib Rahman M Akbar Putra, "Estimasi Biaya Perangkat Lunak Menggunakan Metode Usecase Point (Studi aksus : PT PLN (Persero) Area Malang," 2019.
- [11] R. F. T. L. N. G. N. F. Syahidannur Mukminah Pulungan, "Analisis Teknik Entity Relationship Diagram Dalam Perancangan Database," 2023.