**LAPORAN PROYEK DATA MINING**

***BPJS participants clustering with K-Means Algorithm***



**Disusun oleh:**

|  |  |
| --- | --- |
| 12S20014 | Lidia Ginting |
| 12S20036 | Winda Sari Butarbutar |
| 12S20045 | Christine Hutagaol |

**PROGRAM STUDI SARJANA SISTEM INFORMASI FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO INSTITUT TEKNOLOGI DEL**

**2023**

**DAFTAR ISI**

[DAFTAR TABEL 2](#_Toc134256995)

[DAFTAR GAMBAR 3](#_Toc616260751)

[BAB 1 3](#_Toc1308278696)

[BUSINESS UNDERSTANDING 4](#_Toc1640865319)

[1.1 Determine Business Objective 4](#_Toc1450185610)

[1.2 Technical Goal 4](#_Toc1051593734)

[1.3 Produce Project Plan 4](#_Toc1759203387)

[BAB 2 4](#_Toc999929518)

[DATA UNDERSTANDING 5](#_Toc1605383407)

[2.1 Collect Initial Data 5](#_Toc1619131135)

[2.2 Describe Data 5](#_Toc927072717)

[2.3 Validation Data 5](#_Toc1652434713)

[BAB 3 5](#_Toc1290289743)

[DATA PREPARATION 6](#_Toc1409833420)

[3.1 Data Cleaning 6](#_Toc1568985850)

[3.2 Data Construction 6](#_Toc1105226047)

[3.3 Labeling Data 6](#_Toc654116604)

[3.4 Data Integration 6](#_Toc2046658799)

[BAB 4 6](#_Toc1660975365)

[MODELING 7](#_Toc629481627)

[4.1 Testing Scenario 7](#_Toc251478117)

[4.2 Model Building 7](#_Toc1901323719)

[BAB 5 7](#_Toc1423846232)

[MODEL EVALUATION 8](#_Toc1122207661)

[5.1 Result Evaluation 8](#_Toc884601401)

[BAB 6 8](#_Toc1104617983)

[DEPLOYMENT 9](#_Toc1401174958)

[6.1 Deployment 9](#_Toc1828944540)

# DAFTAR TABEL

[DAFTAR TABEL 2](#_Toc1229892751)

[DAFTAR GAMBAR 3](#_Toc157112765)

[BAB 1 3](#_Toc1616071534)

[BUSINESS UNDERSTANDING 4](#_Toc555439692)

[1.1 Determine Business Objective 4](#_Toc617909283)

[1.2 Technical Goal 4](#_Toc1191037623)

[1.3 Produce Project Plan 4](#_Toc1219010304)

[BAB 2 4](#_Toc899943288)

[DATA UNDERSTANDING 5](#_Toc120532659)

[2.1 Collect Initial Data 5](#_Toc2100759902)

[2.2 Describe Data 5](#_Toc289042422)

[2.3 Validation Data 5](#_Toc1134148841)

[BAB 3 5](#_Toc1065096261)

[DATA PREPARATION 6](#_Toc87823445)

[3.1 Data Cleaning 6](#_Toc1158865744)

[3.2 Data Construction 6](#_Toc65745192)

[3.3 Labeling Data 6](#_Toc962226885)

[3.4 Data Integration 6](#_Toc727146602)

[BAB 4 6](#_Toc1246513157)

[MODELING 7](#_Toc293081793)

[4.1 Testing Scenario 7](#_Toc407087803)

[4.2 Model Building 7](#_Toc1068202180)

[BAB 5 7](#_Toc76097558)

[MODEL EVALUATION 8](#_Toc752281642)

[5.1 Result Evaluation 8](#_Toc775439393)

[BAB 6 8](#_Toc618242460)

[DEPLOYMENT 9](#_Toc1166479813)

[6.1 Deployment 9](#_Toc1649059903)

# DAFTAR GAMBAR

# BAB 1

# BUSINESS UNDERSTANDING

## Determine Business Objective (lidia)

## Business objective atau tujuan bisnis dari penggunaan Clustering (seperti K-Means atau Hierarchical Clustering) untuk mengelompokkan peserta BPJS berdasarkan pola penggunaan layanan kesehatan sebagai berikut:

## Segmentasi Peserta BPJS: Mengidentifikasi kelompok-kelompok yang memiliki pola penggunaan layanan kesehatan serupa atau karakteristik yang mirip dalam hal pemanfaatan fasilitas kesehatan. Ini memungkinkan BPJS untuk memahami variasi dalam kebutuhan layanan kesehatan antar kelompok.

## Personalisasi Layanan Kesehatan: Dengan memahami pola penggunaan layanan kesehatan, BPJS dapat menyesuaikan pendekatan mereka dalam memberikan layanan kepada masing-masing kelompok. Ini dapat melibatkan penyediaan informasi yang lebih spesifik, rekomendasi layanan tambahan yang sesuai, atau pengembangan program-program yang lebih fokus sesuai dengan kebutuhan masing-masing kelompok.

## Optimisasi Sumber Daya: Dengan mengelompokkan peserta berdasarkan pola penggunaan layanan, BPJS dapat mengoptimalkan alokasi sumber daya mereka, termasuk tenaga medis, fasilitas, dan anggaran, dengan cara yang lebih efisien dan efektif. Hal ini memungkinkan pengalokasian yang lebih baik sesuai dengan kebutuhan setiap kelompok.

## Perencanaan dan Pengambilan Keputusan yang Lebih Baik: Analisis pola penggunaan layanan kesehatan membantu BPJS dalam merencanakan langkah-langkah ke depan, mengidentifikasi tren-tren dalam permintaan layanan kesehatan, dan membuat keputusan strategis yang lebih tepat. Dengan pemahaman yang lebih dalam tentang profil setiap kelompok, BPJS dapat mengambil tindakan yang lebih terarah.

## Peningkatan Efisiensi dan Kualitas Layanan: Dengan memahami pola penggunaan layanan kesehatan peserta BPJS, mereka dapat mengidentifikasi area-area di mana efisiensi dapat ditingkatkan dan kualitas layanan dapat diperbaiki. Ini dapat melibatkan penyediaan layanan tambahan, mengurangi waktu tunggu, atau meningkatkan aksesibilitas layanan kesehatan.

## Penerapan teknik Clustering pada data peserta BPJS untuk mengelompokkan mereka berdasarkan pola penggunaan layanan kesehatan dapat memberikan wawasan yang berharga bagi BPJS untuk meningkatkan pengalaman peserta, mengoptimalkan sumber daya, dan membuat keputusan yang lebih terinformasi.

## Technical Goal (christine)

Tujuan dari proyek ini adalah mengembangkan sebuah model data mining utuk melakukan klastering pada sample BPJS data 2015-2021 dengan menggunakan algorithn K-Means. Algoritma K-Means pada proyek ini digunakan untuk membagi atau mengelompokkan sekumpulan data menjadi beberapa kelompok berdasarkan pola penggunaan layanan kesehatan, dengan memahami bagaimana peserta memanfaatkan layanan, BPJS dapat meningkatkan pengalaman mereka, mengoptimalkan pemanfaatan sumber daya, dan membuat keputusan yang lebih terinformasi untuk memberikan pelayanan yang lebih baik.

## Produce Project Plan (winda)

## Dalam proyek ini, Python digunakan sebagai bahasa pemograman. Python adalah bahasa pemograman yang populer untuk Data Science, Machine Learning, dan Internet of Things(IoT). Python adalah bahasa pemograman yang digunakan untuk eksekusi sejumlah instruksi multiguna secara langsung dengan metode orientasi objek(OOP). Python juga menggunakan semantik dinamis meningkatkan keterbacaan kode.

## Adapun algoritma yang digunakan adalah KNN, algoritma ini bertujuan untuk malakukan klasifikasi dan pengelompokan data berdasarkan kategorinya. Berikut adalah rencana proyek yang akan digunakan pada proyek ini.

|  |  |  |
| --- | --- | --- |
| Aktivitas | Detail | Durasi |
| Pemilihan Kasus | Pemilihan Kasus | 1 Hari |
| Pemilihan Algoritam | 1 Minggu |
| Business Understanding | Menentukan Objektif Bisnis | 2 Hari |
| Menentukan Tujuan Bisnis | 3 Hari |
| Membuat Rencana Proyek | 1 Hari |
| Data Understanding | Mengumpulkan data | 3 Hari |
| Menelaah Data | 4 Hari |
| Memvalidasi Data | 2 Hari |
| Data Preparation | Memilah Data | 2 Hari |
| Membersihkan Data | 2 Hari |
| Mengkontruksi Data | 2 Hari |
| Menentukan Label Data | 2 Hari |
| Mengintegrasikan Data | 2 Hari |
| Modeling | Membangun Skenario Pengujian | 5 Hari |
| Membangun Model | 3 Hari |
| Model Evaluation | Mengevaluasi Hasil Pemodelan | 4 Hari |
|  | Melakukan Review Proses Pemodelan | 4 Hari |
| Deployment | Melakukan Deployment Model | 5 Hari |
|  | Membuat Laporan akhir Proyek | 4 Hari |

# BAB 2

# DATA UNDERSTANDING

## Collect Initial Data

## Pengumpulan data merupakan langkah awal untuk menentukan data yang digunakan pada proyek, adapun dataset yang digunakan untuk mengelompokkan dan mengklasifikasikan data bpjs adalah 2020 Metadata Data Sampel BPJS Kesehatan.

## Describe Data (christine)

## Dataset *2020 Metadata Data Sampel BPJS Kesehatan* digunakan untuk mengelompokkan data peserta BPJS. Dataset ini terdiri dari 2 (dua) variabel dimana kedua variabel tersebut adalah variabel informasi dan variabel nilai yang memperlihatkan data. Variabel informasi terdiri dari data variable, position, dan label. Variabel nilai terdiri dari data value dan label.

## **Validation Data (lidia)**

## Tahap ini akan mencakup validasi data untuk memeriksa keakuratan, kelengkapan, dan kualitas data yang akan digunakan. Validasi bertujuan untuk mencegah terjadinya kesalahan atau masalah dalam data, seperti keberadaan nilai yang hilang. Proses validasi melibatkan pemeriksaan terhadap potensi gangguan atau ketidakakuratan dalam sumber data. Jika terdeteksi adanya gangguan, langkah selanjutnya adalah membersihkan data guna memastikan bahwa data yang digunakan konsisten, lengkap, dan akurat.

# BAB 3

# DATA PREPARATION

## 3.1 Data Cleaning

## 3.2 Data Construction

## 3.3 Labeling Data

## 3.4 Data Integration

# BAB 4

# MODELING

Pada bab ini akan menjelaskan Testing scenario dan modeling building, berikut merupakan penjelasannya.

## 4.1 Testing Scenario

## 4.2 Model Building

# BAB 5

# MODEL EVALUATION

## 5.1 Result Evaluation

# BAB 6

# DEPLOYMENT

## 6.1 Deployment