

LAPORAN PROYEK DATA MINING



Disusun Oleh :

12S20019	Kristina Margaret Sitorus
12S20042	Mastawila Febryanti Simanjuntak
12S20043	Sevia Rajagukguk
12S20052	Eka Rohani Gultom

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

2023

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I

Business Understanding

Pada bab ini akan dijelaskan mengenai pemahaman substansi dari aktivitas data mining yang akan dilaksanakan serta kebutuhan dari perspektif bisnis. Aktivitasnya antara lain menentukan objektif bisnis, menentukan tujuan bisnis, dan membuat rencana proyek

1.1 Menentukan Objektif Bisnis

Pada tahapan business understanding dilakukan pemahaman terhadap data yaitu data sampel BPJS Kesehatan tahun 2021. Menentukan objektif bisnis merupakan tahapan awal pemahaman dengan melihat objektif bisnis pada data BPJS tersebut. Objektif bisnis dilakukan untuk dapat menganalisa kebutuhan bisnis tersebut yaitu BPJS. BPJS (Badan Penyelenggara Jaminan Sosial) adalah lembaga publik yang bertanggung jawab pada jaminan terhadap perlindungan dan pelayanan kesehatan bagi para pengguna atau masyarakat. BPJS dibentuk untuk menyediakan perlindungan bagi masyarakat terutama dalam hal kesehatan, ketenagakerjaan, dan jaminan sosial. Lembaga ini berdiri dengan program jaminan sosial sesuai dengan undang - undang yang berlaku.

Pada data sampel BPJS kesehatan berisi berbagai informasi dalam memanfaatkan layanan kesehatan . Data tersebut memiliki 135 variabel yang berisikan informasi demografis, pemanfaatan pelayanan kesehatan dan biaya kesehatan. Variabel pada data tersebut memiliki potensi yang dapat dimanfaatkan pada berbagai bidang yaitu analisa kesehatan populasi, evaluasi kinerja penyedia layanan kesehatan, manajemen biaya dan efisiensi, prediksi kebutuhan layanan kesehatan, peningkatan kualitas kesehatan, evaluasi program pencegahan dan pengobatan, keamanan data dan kepatuhan regulasi, penyelidikan penyakit spesifik, dan pengembangan program kesehatan.

Objektif bisnis tersebut ditetapkan dengan baik untuk memastikan bahwa analisis terhadap data tersebut telah sesuai dengan proses bisnis untuk mencapai keberhasilan bisnis terhadap arah dan pengambilan keputusan terhadap proyek tersebut.

1.2 Menentukan Tujuan Bisnis

Tujuan pengerjaan proyek ini adalah membangun sebuah model dengan penggunaan *Data Mining* untuk

1.3 Membuat Rencana Proyek

Berikut ini adalah rencana proyek yang akan dikerjakan. Rencana proyek berisi tahapan, waktu pengerjaan dan kegiatan yang dilakukan setiap tahapan untuk mencapai tujuan pada proyek *data mining* dan tujuan bisnis dalam

Tahapan	Waktu Pengerjaan	Kegiatan
<i>Business Understanding</i>	4 hari	Menentukan objektif bisnis, menentukan tujuan proyek serta membuat rencana proyek
<i>Data Understanding</i>	3 hari	Menelaah data yang akan digunakan dan melakukan validasi pada data tersebut.
<i>Data Preparation</i>	4 hari	Memilih data yang akan digunakan, membersihkan data, konstruksi data, menentukan label data dan mengintegrasikan data yang akan digunakan.
<i>Modeling</i>	3 hari	Membangun skenario pengujian dan membangun model.
<i>Evaluation</i>	3 hari	Melakukan evaluasi hasil pemodelan dan melakukan review terhadap proses pemodelan.
<i>Deployment</i>	5 hari	Membuat rencana deployment model, monitoring and maintenance terhadap deployment model proyek.

Dalam proses pengerjaan tiap tahapan pada proyek ini, berikut ini adalah beberapa spesifikasi *development environment* yang digunakan adalah sebagai berikut

Tools	<ul style="list-style-type: none"> - Jupyter Notebook - Google Collab - Visual Studio Code
Bahasa Pemrograman	Python 3
Algoritma	
Web Interface	
Deployment Cloud	

BAB II

DATA UNDERSTANDING

Dalam tahapan data understanding yang merupakan tahapan pemahaman terhadap data yang akan digunakan, tahapan ini dimulai dari mengumpulkan data, mendeskripsikan data dan memahami data yang akan digunakan dalam penelitian.

2.1 Mengumpulkan Data

Pengumpulan data merupakan tahap awal untuk menemukan data yang akan digunakan dalam penelitian. Data yang digunakan adalah data sampel BPJS Kesehatan 2015-2021. Data Sampel BPJS Kesehatan 2015-2021 adalah kumpulan data yang berisi informasi mengenai peserta BPJS Kesehatan, termasuk data demografi, data pelayanan kesehatan, dan data pembiayaan kesehatan. Data ini dirilis oleh BPJS Kesehatan setiap tahun untuk mendukung penelitian dan pengembangan di bidang kesehatan masyarakat.

Data sampel ini terdiri dari 2.305.435 peserta BPJS Kesehatan, yang diambil dari populasi database BPJS Kesehatan dengan menggunakan metode sampel acak. Data ini mencakup

periode waktu 2015-2021, sehingga dapat digunakan untuk melihat tren kesehatan masyarakat dalam jangka waktu tersebut.

2.2 Menelaah Data

Menelaah data pada data understanding adalah proses untuk mendapatkan pemahaman awal mengenai data dengan melakukan analisis dan visualisasi data. Data sampel kepesertaan kontekstual Diabetes Mellitus (DM) adalah seluruh peserta tersampel yang ada dalam sistem informasi BPJS Kesehatan sebagai representative peserta JKN-KIS pernah didiagnosis DM. Data kepesertaan terkait karakteristik selain jenis kelamin bersifat dinamis, dan data kepesertaan berikut menunjukkan status peserta pada tanggal 31 Desember 2021. Hasil analisis pada tabel adalah jumlah dan persentase tertimbang terhadap populasi menggunakan variabel bobot.

Berikut tabel yang membahas terkait atribut pada dataset kepesertaan pada file DM 2019 kepesertaan.dta

No.	Variable	Nama variable	Tipe variable	
1.	PSTV01	Nomor peserta	Integer	Kategorik (Nominal)
2.	PSTV02	Nomor keluarga	Integer	Kategorik (Nominal)
3.	PSTV03	Tanggal lahir peserta	String	Numerik (Interval-scaled)
4.	PSTV04	Hubungan keluarga	String	Kategorik (Nominal)
5.	PSTV05	Jenis kelamin	String	Kategorik (Nominal)
6.	PSTV06	Status perkawinan	String	Kategorik (Nominal)
7.	PSTV07	Kelas rawat	String	Kategorik (Nominal)
8.	PSTV08	Segmentasi peserta	String	Kategorik (Nominal)
9.	PSTV09	Provinsi tempat tinggal peserta	String	Kategorik (Nominal)
10.	PSTV10	Kabupaten/Kota Tempat Tinggal Peserta	String	Kategorik (Nominal)
11.	PSTV11	Kepemilikan faskes	String	Kategorik (Nominal)
12.	PSTV12	Jenis faskes	String	Kategorik (Nominal)
13.	PSTV13	"Provinsi Fasilitas Kesehatan Peserta Terdaftar	String	Kategorik (Nominal)
14.	PSTV14	Kabupaten/Kota Fasilitas Kesehatan Peserta Terdaftar	String	Kategorik (Nominal)
15.	PSTV15	Bobot	Float	Numerik (Ratio)
16.	PSTV16	Tahun Sample	Integer	Numerik (Ratio)
17.	PSTV17	Status Kepesertaan	String	Numerik (Ordinal)
18.	PSTV18	Tahun Meninggal	Integer	Numerik (Ratio)

2.3 Memvalidasi Data

Pada tahap ini dilakukan validasi terhadap data yang akan digunakan dengan memeriksa kelengkapan data untuk menghindari terjadinya *error* ataupun masalah *input data* yang terjadi *missing value*. *Missing data* adalah row atau baris pada data yang bernilai *null/Nan/NaT*, beberapa *feature* pada row yang kosong (*missing*) sudah ditangani dengan

memberikan nilai “*missing*”. *Duplicate value* merupakan row atau baris pada data yang memiliki nilai (*value*) yang sama. Informasi mengenai *missing value* dan *duplicate value* untuk setiap data dapat dilihat pada tabel berikut.

No.	Data	<i>Missing Value</i>	<i>Duplicate Value</i>
1.	Kepesertaan	-	-
2.	FKTP-Kapitasi	-	2 rows
3.	FKTP Non-Kapitasi	42.503 rows	38 rows
4.	FKRTL	-	-
5.	FKRL Diagnosis Sekunder	-	-