**PROGRAM STUDI SARJANA SISTEM INFORMASI**

**LAPORAN AKHIR PROYEK AKHIR MATA KULIAH**

**12S4054 - DATA MINING**



***Fraud Detection Using SVM Algorithm***

**OLEH:**

|  |  |
| --- | --- |
| **12S17009** | **Prince Ephraim Prabowo Silaban** |
| **12S17043** | **Enjelin Ida Hutahaean** |
| **12S18004** | **Rosalia Pane** |
| **12S18017** | **Putri Yohana Panjaitan** |

**PROGRAM STUDI SARJANA SISTEM INFORMASI**

**FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO**

**INSTITUT TEKNOLOGI DEL**

**2021**

**DAFTAR ISI**

[**BAB 1 BUSINESS UNDERSTANDING** 3](#_Toc88687914)

[**1.1 Determine Business Objectives** 3](#_Toc88687915)

[**1.2 Access the Situation** 3](#_Toc88687916)

[**1.3 Determine Data Mining Goals** 3](#_Toc88687917)

[**1.4 Produce Project Plan** 4](#_Toc88687918)

[**BAB 2 DATA UNDERSTANDING** 6](#_Toc88687919)

[**2.1 Collect the Initial Data** 6](#_Toc88687920)

[**2.2 Describe the Data** 6](#_Toc88687921)

[**2.3 Explore the Data** 6](#_Toc88687922)

[**2.4 Verify Data Quality** 6](#_Toc88687923)

**DAFTAR TABEL**

[**Tabel 1 Perencanaan Proyek** 5](#_Toc88690460)

[**Tabel 2 Deskripsi dataset fraud\_detection\_train.csv** 7](#_Toc88690461)

# **BAB 1 BUSINESS UNDERSTANDING**

*Business Understanding* merupakan tahap awal atau tahap pemahaman dalam penelitian. Pada tahap ini dibutuhkan pemahaman mengenai substansi dari kegiatan *data mining* yang akan dilakukan serta kebutuhan dari sebuah perspektif bisnis. Pada tahap *Business Understanding* juga diperlukan pemahaman tentang latar belakang dan tujuan pada proses bisnis yang berhubungan dengan *Fraud Detection.*

## **1.1 Determine Business Objectives**

Pada tahap *Determine Business Objectives*, dijelaskan tujuan bisnis untuk menentukan faktor-faktor penting dalam penelitian yang direncanakan dan memastikan bahwa hasil akhir dari penelitian sesuai dengan yang diharapkan. Semakin berkembangnya teknologi maka semakin banyak informasi yang tersedia. Informasi dapat diakses dengan mudah melalui penggunaan teknologi yang dikaji agar lebih efisien dan optimal melalui internet.

*Business Objectives* dari penelitian ini adalah melakukan *Fraud Detection* dengan menggunakan data dari BPJS Hackathon*.* Setiap atribut yang terdapat pada data BPJS Hackathon akan dianalisis dan dilakukan pemodelan dengan menggunakan algoritma Support Vector Machine (SVM) menggunakan bahasa pemrograman Python.

## **1.2 Access the Situation**

Proyek ini akan melibatkan pencarian fakta yang lebih rinci untuk semua sumber daya (*sources)* seperti sumber daya perangkat keras, sumber daya data (*data sources)* dan sumber daya personal.

1. *Data sources* yang digunakan pada proyek ini adalah dataset *Fraud Detection train* pada studi kasus BPJS Hackathon.
2. Sumber daya perangkat keras yang digunakan pada proyek ini adalah laptop IdeaPad Lenovo 4GB RAM, Processor Intel Core i5-7200U Dual Core 2.5 GHZ Turbo Boost 3.1 GHZ, CD/DVD ROM Drive.
3. Sumber daya personal pada proyek ini terdiri dari 4 orang mahasiswa yang berperan pada pengerjaan proyek mulai dari tahap *business understanding, data understanding, data preparation, modeling, evaluation,* dan *deployment.*

## **1.3 Determine Data Mining Goals**

*Determine Data Mining Goals* adalah tahapan untuk mengubah pengetahuan pada domain bisnis menjadi sebuah *definisi problem data mining* serta untuk menetapkan tujuan *data mining.* Tujuan data mining dalam pengerjaan proyek ini adalah untuk menggali *Discovering Knowledge* mengenai pola (*pattern*) item mengenai *Fraud Detection* menggunakan BPJS Hackathon.

## **1.4 Produce Project Plan**

Tahapan yang dilakukan disini adalah memaparkan rancangan kerja yang ditujukan untuk mencapai tujuan dari data mining sehingga mampu untuk mencapai tujuan bisnis, kemudian menentukan teknik dan t*ools* yang selanjutnya akan dipergunakan.

*Project plan* proyek yaitu untuk dapat menyelesaikan tujuan data mining serta mencapai tujuan bisnis adalah sebagai berikut :

**Tabel 1 Perencanaan Proyek**

|  |  |  |
| --- | --- | --- |
| Tahap | Waktu | Kegiatan |
| *Business Understanding* | 4 Hari | Menetapkan tujuan bisnis,  menilai situasi, menetapkan sasaran *data mining*, dan menyusun beberapa rencana proyek. |
| *Data Understanding* | 7 Hari | Mengumpulkan data, memahami data, melakukan eksplorasi data, dan mengidentifikasi kualitas data. |
| *Data Preparation* | 12 Hari | Memilih *dataset* akhir dari  data mentah, melakukan *cleaning data*,  mengimplementasikan data, melakukan integrasi data, serta membuat format data |
| *Modeling* |  | Memilih metode pemodelan  dan sejumlah parameter yang disesuaikan untuk memperoleh nilai yang optimum. |
| *Evaluation* |  | Melakukan evaluasi hasil, *review* proses, dan menentukan langkah selanjutnya. |
| *Deployment* |  | Menyusun rencana  penerapan, menyusun  rencana pemantauan  (*monitoring*) dan pemeliharaan (*maintenance*), membuat laporan akhir |

Teknik yang digunakan pada pengerjaan proyek ini adalah menggunakan algoritma Support Vector Machine (SVM) yang termasuk dalam *Supervised Learning* pada penambangan data (*Data Mining*). Serta Tools yang digunakan pada proyek ini adalah bahasa pemrograman Python.

# **BAB 2 DATA UNDERSTANDING**

*Data Understanding* atau pemahaman data merupakan tahap pengumpulan data awal dan meneliti data yang bertujuan untuk mengidentifikasi dan mempelajari data untuk bisa mengenal data yang akan dipakai. Tahap ini mencoba mengidentifikasikan masalah yang berkaitan dengan kualitas data, mendeteksi subset yang menarik dari data untuk membuat hipotesa awal.

## **2.1 Collect the Initial Data**

*Collect Initial Data* adalah proses pengumpulan data untuk dapat digunakan, data dapat diperoleh dengan melakukan kuesioner, wawancara, mengambil langsung sampel data dari lapangan, maupun dari internet. Penulis akan menggunakan data *Fraud Detection train* pada studi kasus BPJS Hackathon.

## **2.2 Describe the Data**

Pada tahap *describe the data,* penulis akan memahami karakteristik dari data dengan menganalisa setiap atribut yang berada di dalam data, termasuk juga melakukan analisis apakah ada objek data yang bersifat *noisy*. Hal ini dilakukan untuk memperoleh informasi terkait data yang akan digunakan. Pada tahap ini, penulis akan melakukan *exploratory data analysis* (EDA) untuk memahami karakteristik dari data.

Adapun deskripsi dari data yang akan digunakan dapat dilihat pada tabel berikut.

**Tabel 2 Deskripsi dataset fraud\_detection\_train.csv**

|  |  |
| --- | --- |
| ***Variable name*** | ***Variable description*** |
| visit\_id | id kunjungan |
| kdkc | Kode wilayah kantor cabang BPJS Kesehatan |
| dati2 | Kode kabupaten/kota |
| typeppk | Kode tipe dari rumah sakit |
| jkpst | Jenis kelamin peserta JKN-KIS |
| umur | Umur peserta saat mendapatkan pelayanan rumah sakit |
| jnspelsep | Tingkat pelayanan: 1.rawat inap; 2.rawat jalan; |
| los | Lama peserta dirawat di rumah sakit |
| cmg | Klasifikasi CMG (Case Mix Group) |
| severitylevel | Tingkat urgensi |
| diagprimer | Diagnosa primer |
| dx2\_a00\_b99 - dx2\_z00\_z99 | Diagnosa sekunder |
| proc00\_13 – procv00\_v89 | Kode kelompok procedure |
| label | Flag fraud: 1:fraud; 0:tidak fraud |

## **2.3 Explore the Data**

Pada tahap *Explore the Data* akan dilakukan eksplorasi terkait data yang digunakan. Eksplorasi dilakukan pada atribut dan objek data secara menyeluruh. Eksplorasi data juga memperhatikan ekstensi dari data yang akan digunakan. Oleh karena itu eksplorasi data pada penelitian ini akan dilakukan dengan melakukan analisis terhadap dimensi dari data yang digunakan, termasuk mengelompokkan data berdasarkan variabel target.

## **2.4 Verify Data Quality**

Pada tahap *verify data quality,* dilakukan verifikasi terhadap pengerjaan eksplorasi data untuk memastikan tidak ada data yang bersifat *noisy*. Hal ini dilakukan untuk menghindari kesalahan pada tahap pemodelan. Tahap mengevaluasi kualitas data dan kelengkapan data atau nilai-nilai yang hilang sering terjadi, terutama jika data yang dikumpulkan di jangka waktu yang lama. Memeriksa atribut yang hilang atau kosong. Menilai apakah semua nilai masuk akal, ejaan nilai-nilai, dan apakah atribut dengan nilai yang berbeda memiliki arti yang sama.