**DOKUMEN PROYEK**

**12S4054 - PENAMBANGAN DATA**

***Binary Classification Using KNN***

**Disusun Oleh:**

| **12S18005** | **Lusiana Ros Romantika Siahaan** |
| --- | --- |
| **12S18042** | **Indah Oktavia M. Sibarani** |
| **12S18067** | **Grace Vidia Rosa Panjaitan** |



**PROGRAM STUDI SARJANA SISTEM INFORMASI**

**FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO (FITE)**

**INSTITUT TEKNOLOGI DEL**

**TAHUN 2020/2021**

# **BUSINESS UNDERSTANDING**

CRISP-DM or *Cross Industry Standard Process for Data Mining* merupakan serangkaian langkah/kerangka kerja yang jelas untuk melaksanakan proyek ilmu data/penambangan data apa pun. Penggunaan kerangka kerja ini akan memastikan kami memiliki proses yang efisien dalam mengerjakan proyek akhir *Data Mining*. *Business Understanding* adalah tahap pertama dalam CRISP-DM. Pada tahapan ini, untuk membangun model terbaik perlu untuk digali lebih dalam apa yang dibutuhkan dari proyek data mining. Untuk itu dibutuhkan pengetahuan mengenai objek bisnis, bagaimana membangun atau mendapatkan data, dan bagaimana menyesuaikan tujuan pemodelan terhadap tujuan bisnis untuk membangun pemodelan yang baik. Kegiatan yang dilakukan dalam tahapan ini antara lain: menetapkan tujuan dan persyaratan dengan jelas secara keseluruhan, menerjemahkan tujuan tersebut serta menentukan pembatasan dalam perumusan masalah data mining, dan mempersiapkan strategi awal untuk mencapai tujuan tersebut. [Indah]

## **1.1 Determine Business Objective[Lusi]**

Badan Penyelenggara Jaminan Sosial atau BPJS merupakan layanan kesehatan sosial yang berfungsi dalam memberikan jaminan kesehatan. BPJS memiliki lima program Sistem Jaminan Sosial Nasional (SJSN) yang membantu para masyarakat untuk

## **1.2 Situation Assessment [Grace]**

Dalam pengerjaan proyek ini, dibutuhkan sumber daya yang terdiri dari *hardware*, *data sources*, dan personel. *Hardware* yang digunakan selama pengerjaan proyek berupa Laptop Lenovo Ideapad dengan 8GB RAM. Dataset yang digunakan pengerjaan proyek ini adalah data BPJS yang digunakan dalam kegiatan BPJS Hackathon dengan format file adalah CSV. Dataset train tersebut terdiri dari 200217 observasi dan 53 variable. Jumlah personil yang dibutuhkan dalam pengerjaan proyek ini adalah 3 orang mahasiswa yang terlibat dalam setiap proses dalam proyek ini, baik itu dalam proses *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*.

## **1.3 Determine Data Mining Goal [Inda]**

Tujuan bisnis pada penelitian ini adalah

Tujuan pengerjaan proyek ini adalah membantu rumah sakit dalam memprediksi besar fraud yang mungkin terjadi pada saat klaim pelayanan rumah sakit dengan menggunakan algoritma KNN (*K-Nearest Neighbour*) dan menerapkan metodologi CRISP-DM (Standar Kompetensi Kerja Nasional: KepMen Ketenagakerjaan No 299 thn 2020). Pengerjaan proyek ini juga bertujuan memberikan tampilan hasil *classification* dalam bentuk visualisasi untuk memudahkan membaca hasil *classification.*

## **1.4 Produce Project Plan [Lusiana]**

Tahap perencanaan yang dilakukan untuk mencapai tujuan data mining dan mencapai tujuan bisnis pada penelitian “*Binary Classification Using KNN*” ini adalah sebagai berikut.