**PROPOSAL PROYEK AKHIR MATA KULIAH PENAMBANGAN DATA**



**PREDIKSI KELULUSAN MATA KULIAH MAHASISWA DENGAN ALGORITMA SVM**

**OLEH:**

|  |  |
| --- | --- |
| **12S16033** | **Noel Christoper B Butarbutar** |
| **12S17003** | **Lavenia Situmorang** |
| **12S17052** | **Ruth Angeli Sbarani** |
|  |  |

**PROGRAM STUDI SARJANA SISTEM INFORMASI**

**FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO**

**INSTITUT TEKNOLOGI DEL**

**NOVEMBER 2020**

# **DAFTAR ISI**

[**DAFTAR ISI**](#_gjdgxs) **1**

[**BAB 1 BUSINESS UNDERSTANDING**](#_2et92p0) **2**

[1.1 Determine Business Objectives](#_ssmpovlaxva0) 2

[1.2 Situation Assessment](#_91r1737bgc6r) 2

[1.3 Determine Data Mining Goal](#_256yjxskz19p) 2

[**BAB 2 DATA UNDERSTANDING**](#_745sn09wyk0m) **4**

[2.1 Collect the Initial Data](#_l8e6o3gxr9lc) 4

[2.2 Describe the Data](#_ns43tfye0uy6) 4

[2.3 Explore the Data](#_vemg6ajwbmae) 4

[2.4 Verify Data Quality](#_9ofo5kbbtll5) 4

# **BAB 1 BUSINESS UNDERSTANDING**

*Business Understanding* merupakan tahap pertama dalam proses CRISP-DM atau disebut juga sebagai tahap pemahaman penelitian [1]. Tahap ini dibutuhkan pemahaman mengenai substansi dari kegiatan *data mining* yang akan dilakukan serta kebutuhan dari perspektif bisnis. Pada tahap ini juga diperlukan pemahaman tentang latar belakang dan tujuan pada proses bisnis yang berhubungan dengan nilai mahasiswa.

## **1.1 Determine Business Objectives**

Pada tahap *Determine Business Objectives*, dijelaskan tujuan bisnis dan faktor-faktor penting yang terlibat dalam penelitian yang direncanakan dan untuk memastikan bahwa hasil akhir dari penelitian sesuai dengan yang diharapkan.

*Business objective* dari penelitian ini adalah melakukan prediksi kelulusan mahasiswa pada suatu mata kuliah dengan menambang data dari nilai suatu mata kuliah setiap mahasiswa. Setiap atribut yang mempengaruhi kelulusan mahasiswa seperti skor atau nilai dari setiap mahasiswa akan dianalisis dan dimodelkan dengan SVM menggunakan bahasa pemrograman python, sehingga didapatkan hasil prediksi tentang kelulusan mahasiswa melalui nilai yang diperoleh.

## **1.2 Situation Assessment**

*Situation Assessment* (menilai situasi) pada prediksi kelulusan mata kuliah mahasiswa ini dilihat dari sistem informasi akademik, yang artinya sistem informasi akademik ini adalah sistem yang berfungsi untuk mengelola data-data akademik dengan penerapan teknologi komputer baik hardware maupun software sehingga seluruh proses kegiatan akademik dapat dikelola menjadi informasi yang bermanfaat dalam pengelolaan manajemen perguruan tinggi dan pengambilan keputusan di perguruan tinggi.

Data akademik pada umumnya berisi atribut seperti nama mahasiswa, nilai matakuliah, dan rata-rata nilai secara menyeluruh. Setiap atribut ini dapat dijadikan target acuan bagi seorang mahasiswa untuk lulus atau tidak.

## **1.3 Determine Data Mining Goal**

*Determine data mining goal* adalah tahap mengubah pengetahuan pada domain bisnis menjadi sebuah definisi masalah data mining dan menentukan tujuan *data mining* (penelitian).

Tujuan *data mining* ini adalah untuk menggali pengetahuan (*discovering knowledge*) mengenai pola (*pattern*) nilai mahasiswa dalam kelulusan mata kuliah mahasiswa.

# 

# 

# 

# 

# 

# 

# 

# **BAB 2 DATA UNDERSTANDING**

*Data Understanding* atau pemahaman data adalah fase mengumpulkan data awal, mempelajari data untuk bisa mengenal data yang akan dipakai. Fase ini mencoba mengidentifikasikan masalah yang berkaitan dengan kualitas data, mendeteksi subset yang menarik dari data untuk membuat hipotesa awal. Pada penelitian yang berjudul “*Business and Data Understanding* dalam rangka pembentukan model tata letak dan tata ruang pasar tradisional dalam meningkatkan laba penjualan menggunakan metode *association rule dan decision tree* (studi kasus di kota Bandung)[3], dijelaskan bahwa “tahapan *data understanding* dimulai dengan pengumpulan data awal dan hasil kegiatan dalam rangka untuk membiasakan diri dengan data untuk mengidentifikasi masalah data, untuk menentukan wawasan pertama ke data atau mendeteksi subset menarik untuk membentuk hipotesis untuk informasi yang tersembunyi”.

## **2.1 *Collect the Initial Data***

*Collect initial data* adalah proses pengumpulan data untuk dapat digunakan untuk ditambang. Data dapat diperoleh dengan melakukan kuisioner maupun langsung mengambil sampel data nya dari internet. Penulis akan menggunakan data yang bersumber dari kaggle yaitu Students Performance.csv. Sesuai dengan topik penelitian yang akan dilakukan, maka data yang relevan untuk diolah adalah data mahasiswa yang berisi nilai studi dari mahasiswa.

## **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 karektiristik dari data.

## **2.3 *Explore the Data***

Pada tahap *explore the data,* dilakukan eksplorasi terkait data yang digunakan. Eksplorasi ini melibatkan eksplorasi terhadap atribut dan objek data secara menyeluruh. Eksplorasi data juga memperhatikan ekstensi dari data yang 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.

## 