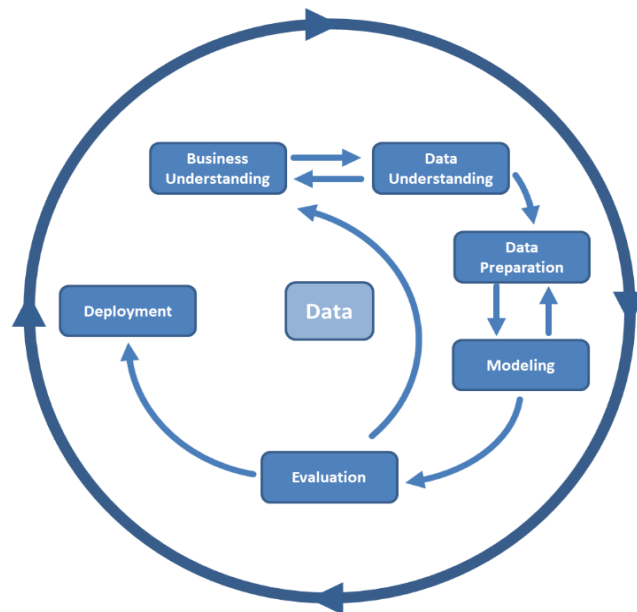


BAB III

METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Tahapan penelitian ini menggunakan Standar Pemrosesan *Data Mining* CRISP-DM :



Gambar 3. 1 Tahapan Penelitian CRISP-DM (M. North, *Data Mining for the Masses*. Global Text Project, 2012)

1. *Business Understanding*

Pemahaman bisnis merupakan awal dari proses CRISP-DM dalam pemahaman tujuan bisnis, penelitian situasi dan menerjemahkan tujuan bisnis ke dalam tujuan *data* mining. Penelitian ini membutuhkan pengetahuan untuk memperoleh data yang mendukung proses klasifikasi kelulusan siswa.

2. *Data Understanding*

Tahapan selanjutnya yaitu pemahaman data dilakukan dengan proses pengumpulan data nilai hasil pencapaian siswa melalui laporan nilai (Rapor) siswa. Setelah itu, dilakukan analisis dan evaluasi kualitas data untuk mengidentifikasi serta mempelajari data lebih lanjut. Sumber data yang digunakan dalam penelitian ini adalah data Rapor siswa kelas 12 semester 1 sampai 6 mulai tahun ajaran 2019/2020 sampai 2021/2022.

3. *Data Preparation*

Proses persiapan data dilakukan dalam penyusunan dataset akhir atau dataset yang nantinya akan digunakan sebagai input (Data Training dan Data Testing) menjadi pemodelan data mining. Tahapan dalam persiapan data diantaranya pemilihan data, transformasi data, dan pembersihan data.

a. *Data Cleaning*

Tahapan pertama yang dilakukan adalah Data Cleaning Han, Kamber & Pei menjelaskan Data Cleaning merupakan proses untuk dapat mengatasi nilai yang hilang, noise dan data yang tidak konsisten. (Han, Kamber & Pei, 2012). Proses *data cleaning* dilakukan pada data yang berjumlah 429 data. Data yang digunakan dalam penelitian ini yaitu data nilai akademik (Raport) semester 1 sampai 6 tahun lulus 2022 yang berasal dari SMA Negeri 1 Singaparna. Adapun atribut yang digunakan dalam penentuan klasifikasi kesiapan ujian ini adalah Nama Lengkap, Nomer Induk, NISN, jenis kelamin, nilai seluruh mata pelajaran dari semester 1 sampai semester 6.

b. *Integration Data*

Tahap ini dilakukan untuk pengelompokan nilai akademik dari seluruh kelas yang ada. Data yang ditampilkan dari proses ini adalah seluruh siswa yang sudah dilakukan proses *Clening Data*.

c. *Data Selection*

Data yang ada telah tersedia dan telah menjalani proses sebelumnya yaitu *Clening Data* dan *Integration Data* selanjutnya data masuk kedalam proses *Data Selection* yaitu mengambil data yang sesuai dengan apa yang akan dianalis.

d. *Data Transformation*

Pada tahapan ini hasil dari pengelompokan *Data Preprocessing* kemudian digunakan untuk *Data Training*. Proses pembentukan *Data Training* ini berdasarkan data yang ada harus diseleksi terlebih dahulu untuk menentukan atribut mana yang dapat mempengaruhi prediksi kelulusan siswa yang berpotensi lulus dan tidak lulus. Atribut yang digunakan pada hasil tahapan ini berbeda dengan tahapan sebelumnya yaitu karena atribut yang dihapus terdapat banyak data yang hilang dan tidak lengkap serta menghapus atribut yang tidak akan berpengaruh terhadap hasil. Selain itu jumlah data yang terdapat dari hasil ini dapat berubah.

4. *Modeling*

Tahapan ini dilakukan sesuai dengan teknik pemodelan yang akan diterapkan pada dataset yang dipilih untuk melakukan klasifikasi. Teknik

pemodelan dalam penelitian ini menggunakan algoritma *C4.5* dan Naïve Bayes. *Tools* yang digunakan dalam pemodelan ini adalah RapidMiner.

5. Evaluation

Tahapan ini akan menguji model yang telah dibuat dan mengevaluasi akurasi serta generalitasnya. Tahapan ini dilakukan untuk mengukur sejauh mana model yang dipilih memenuhi tujuan bisnis dan jika demikian apakah perlu lebih banyak model untuk dibuat lagi.

6. Deployment

Deployment adalah penyusunan laporan atau presentasi dari pemodelan dan evaluasi proses *data mining*. Hasil penelitian ini dapat digunakan untuk informasi tingkat kelulusan siswa sebagai bahan kajian untuk sekolah. Mengetahui informasi tingkat kelulusan dapat memudahkan staf dalam mengelola siswa serta memberikan informasi kepada siswa supaya dapat lulus.