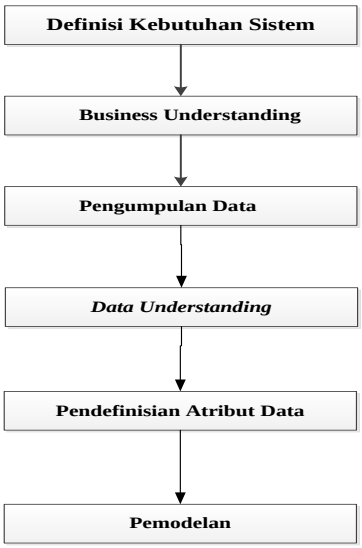
BAB III
METODOLOGI PENELITIAN

III.1 Tahapan Penelitian

Sebagai langkah awal, penelitian ini dilakukan guna menganalisis hasil pengolahan data menggunakan algoritma dalam *Association Rules Mining* khususnya algoritma *FP-Growth* guna menentukan keterhubungan antara atribut yang satu dengan yang lain dalam suatu basis data. Dalam hal ini langkah-langkah proses penelitian akan diuraikan dalam bentuk diagram alir.



Gambar III.1 Diagram Alir Metodelogi Penelitian

Adapun penjelasan dari metodologi yang digunakan dalam penelitian ini adalah:

a. Definisi Kebutuhan Sistem

Tahap ini merupakan pendefinisian kebutuhan sistem berdasarkan masalah atau kendala yang dihadapi pada tempat studi kasus dengan cara melakukan wawancara dan observasi langsung.

Pada tahap ini juga pembelajaran tentang topik-topik yang relevan terhadap penelitian yang akan dilakukan. Selain itu studi juga dilakukan untuk mempelajari mengenai data mining dan metode data mining yang akan digunakan.

b. Business understanding

Tahap ini berisi langkah-langkah yang bertujuan untuk memahami lingkup permasalahan pada institusi/jurusan tempat studi kasus serta memahami kondisi yang ada pada institusi tersebut. Tahapan ini akan berusaha untuk mengetahui secara rinci proses bisnis yang terjadi pada sistem berjalan khususnya berkaitan dengan data polling yang selama ini dijadikan acuan pada saat penentuan matakuliah pilihan di jurusan. Pada penelitian ini selain data *polling* juga akan digunakan data nilai sebagai acuan rekomendasi. Sehingga aturan dan kebijakan untuk histori nilai harus dipahami dari institusi/jurusan tempat studi kasus.

c. Pengumpulan data

Pengumpulan data meliputi tahap pengumpulan *training* data dan *testing* data pada institusi/jurusan tempat studi kasus yang akan digunakan untuk membentuk model keputusan. Data yang akan dikumpulkan berupa data *polling* yang disebarkan dan histori nilai dari nilai prasyarat matakuliah pilihan yang akan ditawarkan.

d. Data Understanding

Tahap persiapan data ini terbagi menjadi tiga tahap utama, yaitu [3]:

1) Describe data

Tahap ini merupakan tahap pendeskripsian data yang sudah didapat dari tempat studi kasus.Setiap atribut data yang sudah didapat didefinisikan supaya bisa dimengerti dan dipelajari dengan baik.

2) Explore data

Tahap ini berisi langkah-langkah untuk melihat keterhubungan atribut-atribut pada data yang dikumpulkan serta melihat atribut-atribut yang mungkin digunakan dalam menjawab pertanyaan *data mining*.

3) Verify data quality

Tahap ini memverifikasi hasil deskripsi dan eksplorasi data yang didapat. Dari tahap ini pula didapat gambaran kualitas data yang akan digunakan dalam pemodelan.

e. Pendefinisian atribut data

Tahap ini dilakukan untuk menyiapkan data untuk tahap pemodelan. Tahap ini terbagi menjadi lima tahap, yaitu [3]:

1) Select data

Tahap ini memilih data serta atribut yang akan digunakan dalam proses *data mining*.

2) Clean data

Tahap ini dilakukan untuk meningkatkan kualitas data dengan cara menangani nilai yang kosong serta memperhatikan *outliers* yang ada pada data.

3) Construct data

Tahap ini merupakan tahap dimana atribut-atribut turunan dibentuk sebagai atribut tambahan dalam penemuan pola.

4) Integrate data

Tahap ini menggabungkan data dari beberapa tabel sehingga menghasilkan nilai atau *record* baru.

5) Format data

Tahap ini mengubah kumpulan data yang sudah diproses menjadi suatu format file yang dimengerti oleh alat bantu pemodelan yang digunakan.

f. Pemodelan

Tahap pemodelan pada tesis ini digunakan untuk mencari model keputusan yang tepat untuk domain permasalahan tesis ini. Pada tahap ini akan dihasilkan model keputusan dengan menggunakan *Association Rule Mining* dengan algoritma *FP-Growth* dari data yang telah dihasilkan sebelumnya. Model keputusan yang dihasilkan merupakan hasil irisan dari *rule* yang didapatkan dari *polling* (kemauan) dan histori nilai (kemampuan).

III.2 Skenario Penelitian

Kegiatan yang dilakukan dalam penelitian ini diantaranya adalah pembentukan data transaksi yang berasal dari data hasil *polling* dan data nilai untuk memperoleh aturan-aturan yang dibutuhkan. Proses dimulai dengan mendefinisikan kebutuhan system pada tempat studi kasus dengan cara wawancara dan observasi langsung untuk mengetahui masalah atau kendala yang terjadi. Selan itu pula kegiatan ini dibarengi dengan studi pustaka untuk mengetahui cara atau metode yang tepat yang dapat digunakan sebagai penyelesaian masalah. Kegiatan berikutnya yaitu pemahaman proses bisnis dari system yang ada yaitu dengan mengetahui kebijakan dan aturan dalam pengisian *polling* untuk penentuan matakuliah pilihan. Penentuan matakuliah pilihan ini akan diuji untuk mencari irisan dengan data nilai bagi mahasiswa yang telah mengisi polling untuk menentukan matakuliah pilihan yang direkomendasikan bagi pihak jurusan. Aturan data nilai yang digunakan berdasarkan kebijakan dan aturan yang diberikan pihak jurusan.

Metodologi *data mining* didasarkan pada dua tahapan yang dilakukan untuk menemukan *Association Rules* pada dataset polling dan nilai dengan metode *FP-Tree* yang terdapat pada algoritma *FP-Growth*. Kedua tahapan tersebut adalah :

1. Seleksi atribut dataset dan *cleaning* data

2. *Association rule mining* dengan menggunakan algoritma *FP-Growth*

Pada tahap pertama, seleksi atribut dalam dataset untuk mendapatkan data record dengan atribut yang relevan dan konsisten terhadap hasil output /keluaran yang diinginkan serta dilakukan pemrosesan awal terhadap data transaksi untuk menghapus data atau record yang tidak konsisten, redundan dan *missing value* sehingga memperoleh *dataset polling* dan nilai yang siap untuk diolah ke tahap selannjutnya. Pada tahap kedua, mengekstrak data atau mentransformasikan *dataset* yang akan digunakan untuk menemukan pola/informasi sehingga dataset tersebut siap untuk di-*mining*. Dari masing-masing dataset yaitu polling dan nilai, *rule* yang dihasilkan akan dilakukan irisan untuk mendapatkan model keputusan akhir yang dapat merekomendasikan matakuliah pilihan yang akan ditawarkan pihak jurusan. Pengolahan data ini menggunakan metode *Association Rule* dengan algoritma *FP-Growth*.

III.3 Waktu dan Tempat Penelitian

Waktu pelaksanaan penelitian dilakukan mulai bulan Maret 2012 hingga bulan Juni 2012. Adapun tempat penelitiannya dilakukan di jurusan Teknik Informatika UNIKOM.