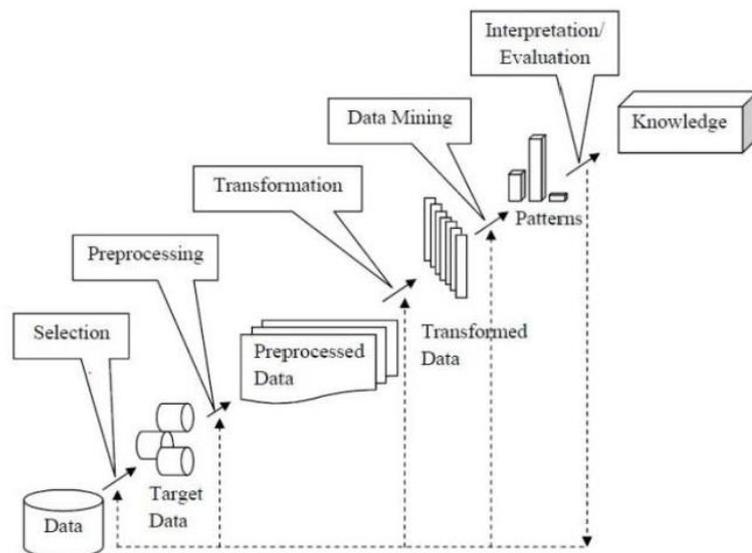


BAB II LANDASAN TEORI

2.1 *Knowledge Discovery in Database (KDD)*

Kegiatan yang meliputi beberapa proses mulai dari pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar, itulah yang disebut KDD (Yosepta dan Aprilianto, 2017). Proses dari KDD merupakan proses menggunakan metode *data mining* untuk mengekstrak pengetahuan yang dianggap sesuai dengan ukuran dan batas, menggunakan *database* dan melakukan *preprocessing*, pengambilan sampel dan mentransformasi *database* (Nuruliyani dan Warnars, 2019).



Gambar 2.1 Proses *Knowledge Discovery in Database* (Han and Kamber, 2006)

Proses KDD dapat dijelaskan sebagai berikut (Bradley, Fayyad and Mangasarian, 1999).

1. *Data Selection*

Pada proses *data selection* dilakukan pemilihan himpunan data, menghasilkan data target, atau memfokuskan pada *subset* variabel (sampel data) dimana *discovery* akan dilakukan. Hasil seleksi kemudian disimpan dalam suatu berkas yang terpisah dari baris data operasional.

2. *Pre-Processing* dan *Cleaning Data*

Pada proses ini dilakukan pembuangan data yang tidak konsisten dan *noise*, duplikasi data, memperbaiki kesalahan data, dan bisa diperkaya dengan eksternal yang relevan.

3. *Transformation*

Salah satu proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining* adalah *coding*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data Mining*

Proses ini merupakan proses utama saat metode diterapkan untuk menemukan pengetahuan yang tersembunyi dari data. *Data mining* merupakan proses mencari pola atau informasi menarik dalam data yang digunakan dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* yang bervariasi. Pemilihan metode atau algoritma tersebut tergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Interpretation* atau *Evaluation*

Pada proses ini dilakukan identifikasi pola-pola dari data kelulusan yang ditemukan. Pola informasi yang ditemukan dari proses *data mining* ditampilkan dalam bentuk yang mudah dimengerti. Tahap ini adalah bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan terhadap pola atau informasi yang ditemukan sesuai atau tidak dengan hipotesis yang ada sebelumnya.

2.2 Data Mining

Data mining merupakan studi pengumpulan, membersihkan, memproses, menganalisis, dan mendapatkan wawasan yang berguna dari data. Analisis *data mining* menggunakan beberapa proses dimana data kasus dikumpulkan, dibentuk, dan kemudian diubah menjadi format standar (Yosepta dan Aprilianto, 2017).

Data mining juga diartikan sebagai proses ekstraksi pola-pola yang implisit, tidak diketahui sebelumnya, dan berpotensi untuk dimanfaatkan dari data yang berukuran besar. *Data mining* merupakan penemuan atau penggalian informasi baru dengan mencari dan menemukan pola atau aturan tertentu dari sejumlah data yang sangat besar (Han and Kamber, 2006). *Data mining* merupakan sebuah proses menemukan pengetahuan dan pola yang menarik dari data yang berjumlah besar. Data-data tersebut dapat bersumber dari *database*, gudang data, web, *repository* informasi lainnya, ataupun data yang dialirkan ke sistem secara dinamis (Han, Kamber, and Pei, 2012).

Data mining juga didefinisikan sebagai suatu analisis dari peninjauan kumpulan data untuk menemukan keterkaitan yang tidak diduga dan meringkas data dengan cara yang berbeda yang dapat dipahami dan bermanfaat bagi pemilik data (Irfan, 2015). *Data mining* sendiri memiliki beberapa karakteristik diantaranya sebagai berikut (Davies, 2004):

1. *Data mining* berhubungan dengan penemuan pola data tertentu dan sesuatu yang tersembunyi yang tidak diketahui sebelumnya.
2. *Data mining* biasa digunakan pada data yang sangat besar. Penggunaan data yang sangat besar ini bertujuan untuk membuat hasil lebih dipercaya.
3. *Data mining* berguna dalam pengambilan keputusan yang kritis, terutama dalam strategi.

Berdasarkan beberapa pengertian di atas, dapat ditarik kesimpulan bahwa *data mining* merupakan suatu teknik penggalian informasi berharga yang tersembunyi pada kumpulan *database* yang sangat besar sehingga ditemukan pola yang menarik yang tidak diketahui sebelumnya (Susanto dan Dedy, 2010).

2.3 Teknik *Data Mining*

Data mining memiliki enam teknik yaitu sebagai berikut (Mustafa, Ramadhan dan Thenata, 2018) :

1. Deskripsi

Deskripsi merupakan fungsi untuk menggambarkan secara singkat atau ringkas terhadap pola dari data-data yang berskala besar dan memiliki

banyak jenis. Beberapa contoh metode *data mining* deskripsi yaitu, *Decision Tree*, *Exploratory Data Analysis* dan *Neural Network*.

2. Pengelompokkan

Pada fungsi pengelompokkan ini dilakukan proses identifikasi data dengan karakteristik tertentu. Pada fungsi *Clustering* ini metode yang diterapkan antara lain, *Hierarchical Clustering*, *K-Means*, dan *Self Organizing (SOM)*.

3. Klasifikasi

Pada metode ini merupakan proses menemukan model yang dapat membedakan kelas data, tujuannya untuk dapat memprediksi suatu kelas yang diperoleh dari objek dengan label yang tidak diketahui. Contoh metode yang menerapkan fungsi klasifikasi antara lain, *Neural Network*, *Decision Tree*, *Naive Bayes* dan *K-Nearest Neighbor*.

4. Asosiasi

Fungsi asosiasi disebut juga sebagai analisis keranjang pasar, fungsi ini sering kali digunakan untuk mengidentifikasi *item – item* produk yang mungkin dibeli oleh konsumen secara bersamaan dengan produk lainnya. Metode yang menerapkan fungsi asosiasi antara lain, *Apriori*, *Generalized Sequential Pattern (GSP)*, *FP-Growth* dan *GRI Algorithm*.

5. Prediksi

Pada Fungsi ini dilakukan perkiraan suatu nilai yang akan terjadi di masa yang akan datang. Metode yang diterapkan pada fungsi prediksi adalah *Neural Network*, *Decision Tree*, dan *K-Nearest Neighbor*.

6. Estimasi

Fungsi ini merupakan proses menerka suatu nilai yang belum diketahui kebenarannya. Metode yang menerapkan fungsi estimasi antara lain, *Point Estimation*, *Confidence Interval Estimation*, *Simple Linear Regression*, *Correlation*, dan *Multiple Regression*.

2.4 Klasifikasi

Metode yang berguna untuk menemukan model atau fungsi yang dapat membedakan kelas data yang memiliki tujuan memprediksi objek yang belum diketahui label dari kelas tersebut disebut klasifikasi. Klasifikasi merupakan bentuk analisis data dengan tujuan mengekstrak model yang mendeskripsikan kelas data (Han, Kamber, and Pei, 2012).

Tujuan dari model klasifikasi dalam *data mining* adalah sebagai berikut (Tan, Steinbach and Kumar, 2006).

1. Deskriptif

Pemodelan ini berperan sebagai alat yang bisa menjelaskan dengan tujuan membedakan antar objek dengan kelas yang berbeda.

2. Prediktif

Pemodelan prediktif ini berperan sebagai alat yang dapat memperkirakan label kelas yang *recordnya* belum diketahui.