

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Penelitian Terdahulu

Tabel 2.1 Penelitian terdahulu

No	Nama Peneliti	Variabel Penelitian	Hasil
1	Fina Nasari (2016)	Diare	Hasil yang diperoleh dari penelitian mengenai kasus ini yaitu cluster 1 dan cluster 2 berada di Kecamatan Hinai, Sei Bingai, dan Sirapit. Cluster 1 jumlah penderita tingkat menengah atau bukan pusat penyebaran diare. Cluster 2 merupakan daerah-daerah penyebaran diare jadi perhatian pemerintah untuk penanganan diare.
2	Ilham Akbar (2016)	DBD	Hasil dari penelitian ini merupakan persebaran DBD di Kabupaten Jember yang menunjukkan dengan jumlah cluster <ol style="list-style-type: none">C1 terdapat 7 Kecamatan berwarna hijauC2 terdapat 15 Kecamatan berwarna kuningC3 terdapat 9 Kecamatan berwarna merah
3	Novita Meisida (2014)	Penyakit Karies Gigi	Dari hasil penelitian ini yaitu perbedaan hasil <i>cluster</i> pada perbandingan <i>cluster</i> kelas Black Poli Gigi Medical Center dan <i>cluster</i> kelas Black K-means karena pada data Poli Gigi yang melakukan pemeriksaan adalah dokter ahli secara langsung pada pasien, sedangkan <i>cluster</i> kelas K-means menggunakan metode

perhitungan berdasarkan nilai variabel yang dikelompokkan menjadi 6 *cluster*.

Berdasarkan penelitian diatas dapat disimpulkan bahwa penelitian yang diambil berbeda dengan penelitian terdahulu. Penelitian terdahulu untuk penelitiannya masih berbasis desktop dan tidak terdapat angka kematian, sedangkan pada penelitian ini terdapat data yang melibatkan angka kematian dan data dianalisa menggunakan aplikasi *RapidMiner*. Hasil perhitungannya memudahkan petugas untuk memetakan sebaran Tuberculosis.

2.2 Dasar Teori

2.2.1 Definisi Tuberkulosis

Paru-paru (Bahasa Inggris: Lung, dari kata Latin *pulmones* untuk paru-paru). Paru-paru merupakan organ tubuh yang penting didalam tubuh manusia. Tanpa adanya paru-paru, manusia tidak akan mungkin hidup di muka bumi ini. Struktur yang akan dilalui oleh udara yang kita hirup yaitu rongga hidung, faring, tenggorokan, dan cabang-cabang tenggorokan (Kurniawan, 2014)

Tuberkulosis merupakan penyakit menular yang disebabkan oleh bakteri *Mycobacterium tuberculosis*. Yang terdapat beberapa spesies *Mycobacterium*, antara lain: *M.Tuberculosis*, *M.Bovis*, *M.Leprae* dsb. Yang dikenal juga sebagai Bakteri Tahan Asam (BTA). Kelompok bakteri *Mycobacterium* selain *Mycobacterium tuberculosis* yang bisa menimbulkan gangguan pada saluran nafas dikenal sebagai MOTT (*Mycobacterium Other Than Tuberculosis*) yang terkadang bisa mengganggu penegakan diagnosis dan pengobatan TBC (Indah, 2018).

2.2.2 Klasifikasi Tuberculosis

Menurut Prabu (2008) klasifikasi penyakit *Tuberculosis* terbagi sebagai berikut :

1. TB Paru BTA Positif (+)

TB Paru BTA (+) apabila sekurang-kurangnya 2 dari 3 spesimen dahak SPS (Sewaktu Pagi Sewaktu) hasilnya positif, atau 1 spesimen dahak SPS positif disertai pemeriksaan radiologi paru menunjukkan gambaran TB aktif.

2. TB Paru BTA Negatif (-)

Apabila dalam 3 pemeriksaan spesimen dahak SPS BTA (-) dan pemeriksaan radiologi dada menunjukkan gambaran TB aktif. TB Paru dengan BTA (-) dan gambaran radiologi positif dibagi berdasarkan tingkat keparahan, apabila menunjukkan keparahan yakni kerusakan luas dianggap berat.

3. TB Ekstra Paru

TB yang menyerang organ tubuh di luar paru, termasuk *pleura* yaitu yang menyelimuti paru, serta organ lain seperti selaput otak, selaput jantung pericaditis, kelenjar limpa, kulit, persendian ginjal, saluran kencing, dan lain-lain.

2.2.3 Gejala-Gejala Tuberculosis

Beberapa tanda dan gejala TB aktif yang menyerang paru-paru (Perhimpunan Dokter Paru Indonesia, 2003):

1. Batuk tiga minggu secara terus-menerus. Mengalami batuk secara terus-menerus selama lebih dari 21 hari. Batuk darah merupakan kondisi yang paling parah dari munculnya penyakit ini
2. Nyeri pada dada

3. Penurunan berat badan. Pada umumnya penderita TB paru mengalami penurunan berat badan karena sakit yang tidak tertahankan. Hal tersebut membuat nafsu makan berkurang secara drastis. Hasilnya tubuh penderita semakin kurus.
4. Cepat merasa lelah
5. Demam mendadak
6. Berkeringat di malam hari. Ketika cuaca dingin akan merasa kedinginan hingga menggigil di malam hari. Karena infeksi dari bakteri TB paru akan menyebabkan tubuh mengalami kedinginan dan menggigil di malam hari.
7. Gejala TB Paru yakni urine berubah. Penderita akan mengalami perubahan warna urine pada tahap selanjutnya, jika TB tidak segera diobati. Tuberkulosis juga dapat mempengaruhi kesehatan bagian lain dari tubuh, termasuk ginjal, tulang belakang, hingga bagian otak. Ketika TB terjadi diluar paru-paru, tanda dan gejala bervariasi sesuai dengan organ yang terlibat.

2.3 Aplikasi RapidMiner

RapidMiner merupakan sebuah perangkat lunak untuk melakukan sebuah analisis terhadap data mining, text mining, dan analisis prediksi. Dalam aplikasi ini menggunakan teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat hasil keputusan yang paling baik. Pada aplikasi ini memiliki kurang lebih operator 500 data mining, yang termasuk operator untuk input, processing data, visualisasi, dan output. (Oktaria, 2018)

RapidMiner adalah salah satu software untuk pengolahan data mining. Pekerjaan yang dilakukan oleh RapidMiner text mining adalah berkisar dengan analisis teks, mengekstraksi pola-pola dari data yang besar dan mengombinasikannya dengan metode statistika, kecerdasan buatan, dan

database. Tujuan dari analisis ini untuk mendapatkan informasi bermutu tertinggi dari teks yang diolah. (Bertus, 2015)

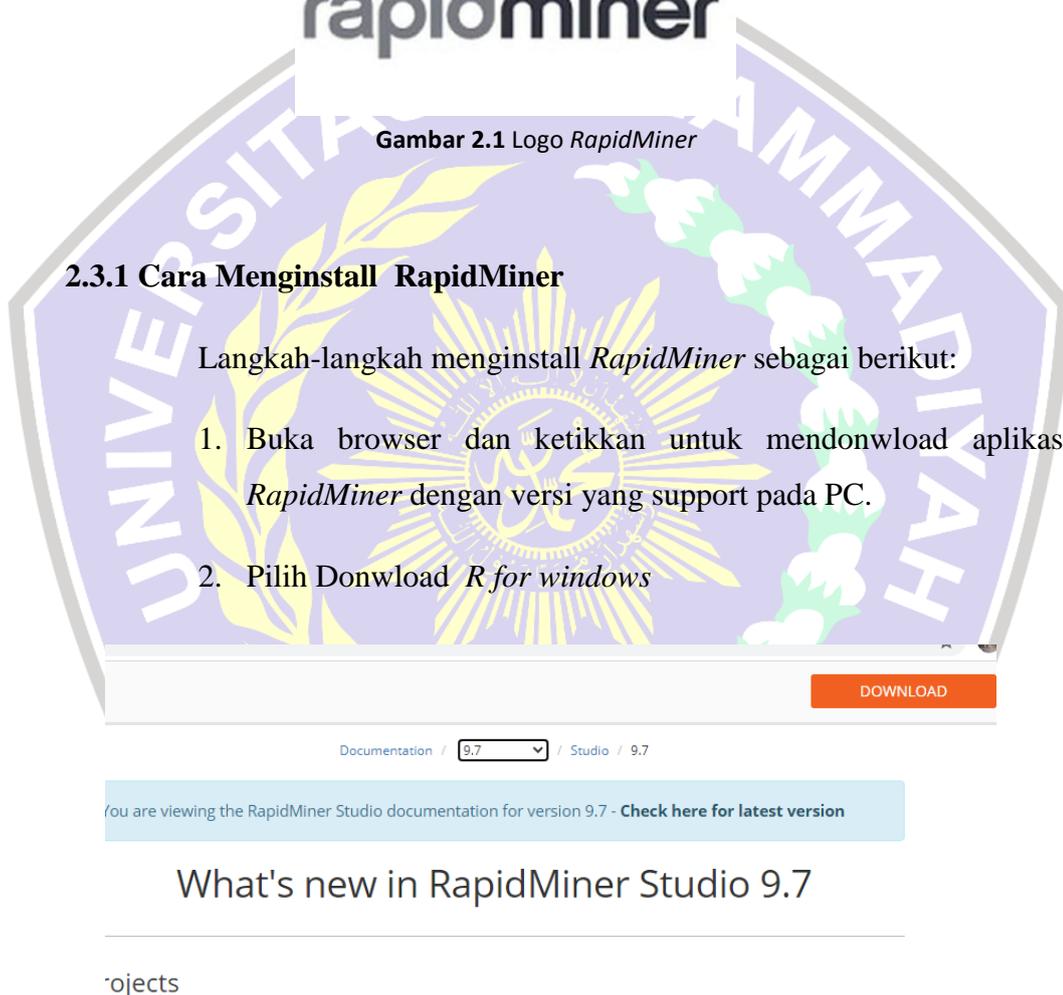


Gambar 2.1 Logo *RapidMiner*

2.3.1 Cara Menginstall *RapidMiner*

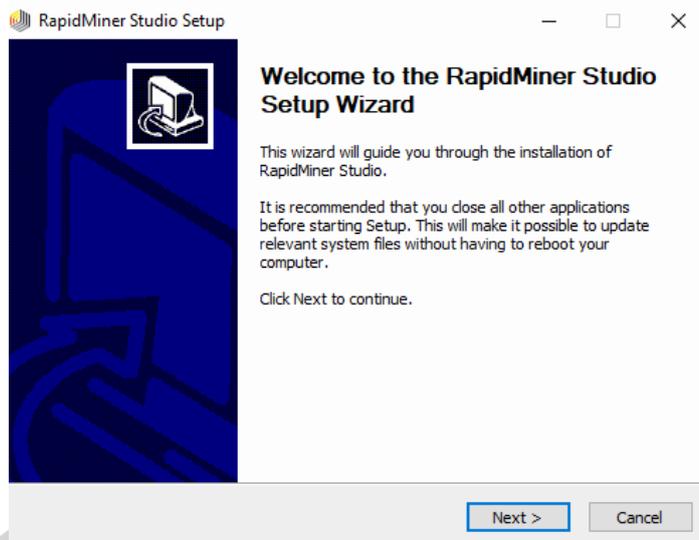
Langkah-langkah menginstall *RapidMiner* sebagai berikut:

1. Buka browser dan ketikkan untuk mendownload aplikasi *RapidMiner* dengan versi yang support pada PC.
2. Pilih Donwload *R for windows*



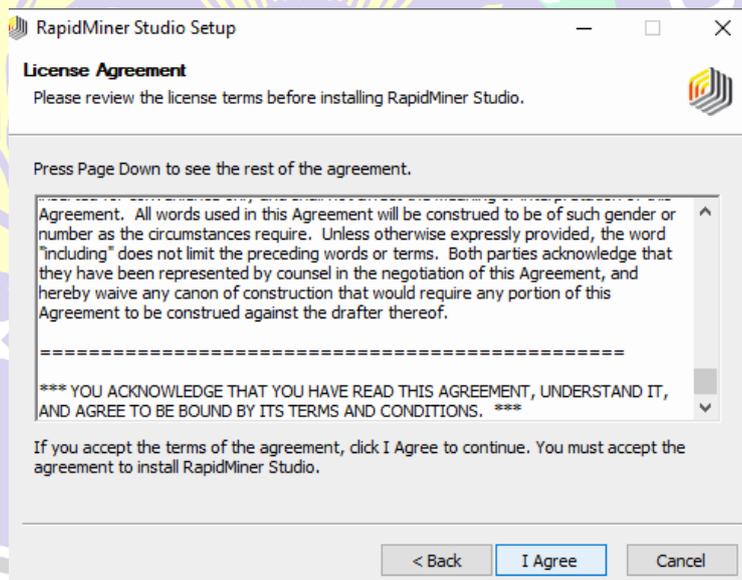
Gambar 2.2 Download *RapidMiner for Windows*

3. Tunggu hingga proses donwload tersebut hingga selesai
4. Jika sudah selesai selanjutnya menginstall, akan muncul seperti **Gambar 2.3** selanjutnya pilih *Next*



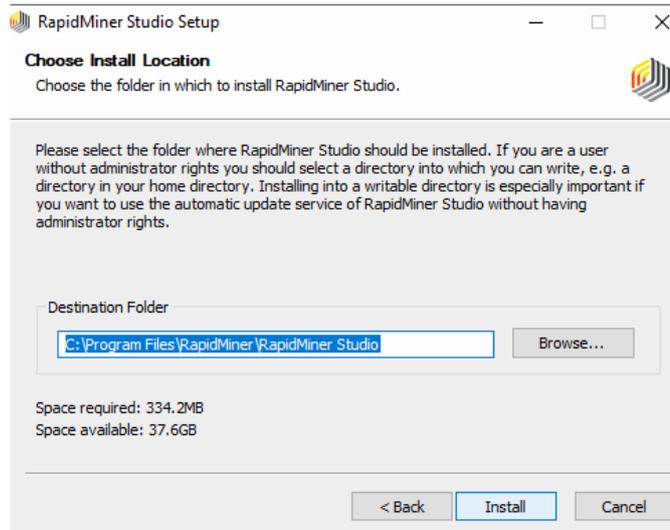
Gambar 2.3 Tahap install *RapidMiner*

5. Selanjutnya menyetujui untuk menginstall, ketika sudah muncul “*I Agree*” seperti Gambar 2.4



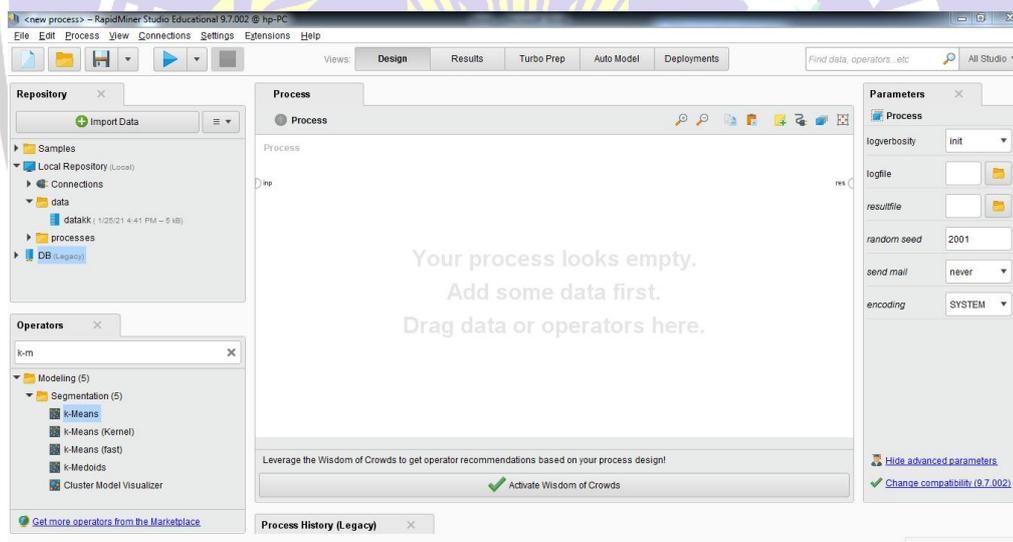
Gambar 2.4 persetujuan install *RapidMiner*

6. Memilih lokasi dimana kita akan menempatkan file instalasi RapidMiner. Setelah itu pilih tombol install, dan tunggu hingga selesai sampai muncul **Finish**.



Gambar 2.5 lokasi penyimpanan *RapidMiner*

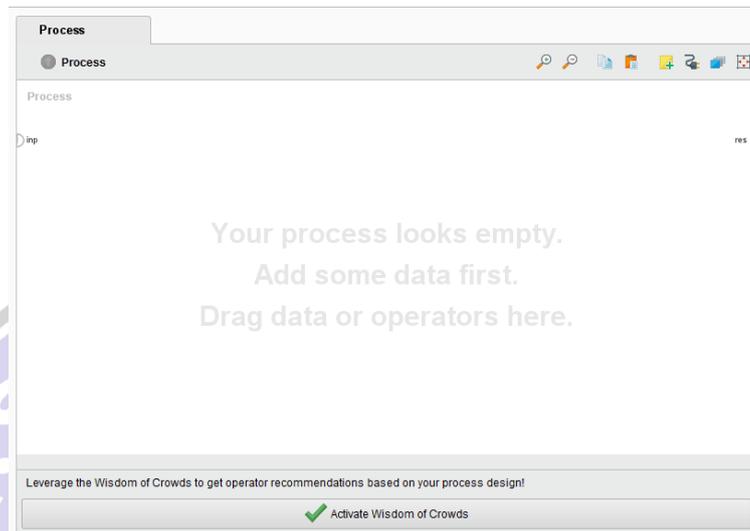
7. Setelah selesai, maka tampilan terpadu (IDE) seperti berikut



Gambar 2.6 Tampilan awal *RapidMiner*

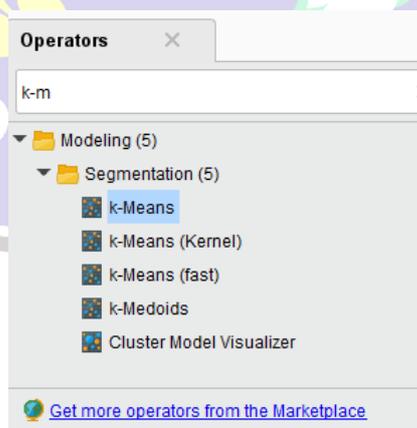
2.3.2 Istilah-istilah dalam RapidMiner

1. Proses adalah unit yang dapat dieksekusi yang dimilikinya. User menciptakan proses menggunakan operator dan menggabungkan secara bersama dengan cara yang diperlukan



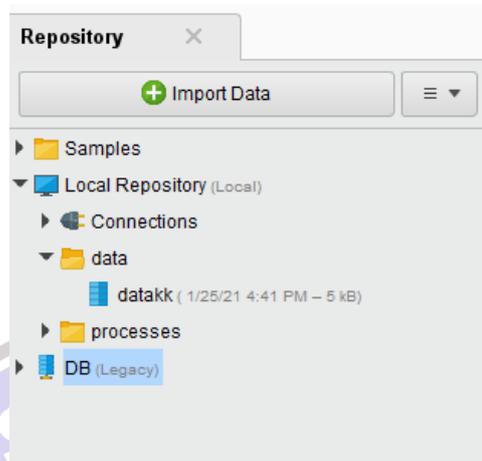
Gambar 2.7 proses pada RapidMiner

2. Operator yaitu fungsi yang disediakan oleh aplikasi RapidMiner yang digunakan untuk menjalankan perintah dengan menggunakan metode-metode yang dibutuhkan user untuk menuju suatu proses dan terhubung menuju proses lainnya.



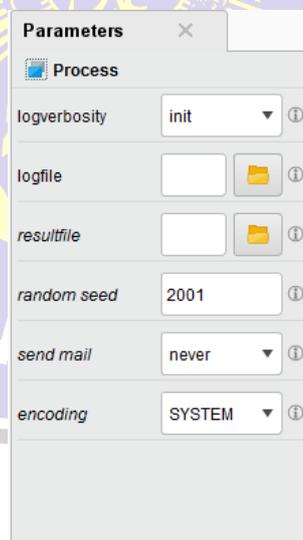
Gambar 2.8 operator pada RapidMiner

3. Repository merupakan dimana penyimpanan data,model, proses, dan file yang dapat dibaca baik RapidMiner atau dari sebuah proses.



Gambar 2.9 Repository pada RapidMiner

4. Parameters merupakan yang mendefinisikan bagaimana operator menggunakan atribut tersebut. jenis atribut ini antara lain id, cluster, prediction, dan outlier. RParameters juga dapat mengatur sebagaimana operator.



Gambar 2.10 Parameters pada RapidMiner

2.4 Data Mining

2.4.1 Definisi Data Mining

Data mining adalah proses yang menggunakan teknik statistika, matematika, kecerdasan buatan, *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Turban, 2005).

Fungsi dari data mining adalah membantu untuk mendapatkan informasi yang berguna serta meningkatkan bagi pengguna. Berikut empat fungsi data mining yaitu:

1. Prediksi (*Prediction*)

Proses ini untuk menentukan pola dari data dengan menggunakan variabel untuk memprediksi variabel lain yang tidak diketahui jenis atau nilainya.

2. Deskripsi (*Description*)

Proses untuk menentukan sebuah karakteristik yang penting dari data dalam suatu basis data. Jika pola tersebut berulang dan bernilai itulah karakteristik sebuah data yang diketahui.

3. Klasifikasi (*Classification*)

Suatu proses untuk menemukan fungsi atau model untuk menggambarkan class dari suatu konsep data. Proses ini untuk mendeskripsikan data yang penting penting serta dapat meramalkan kecenderungan data di masa depan.

4. Asosiasi (*Association*)

Pada proses ini digunakan untuk menemukan suatu hubungan yang terdapat pada nilai atribut dari sekumpulan data.

2.4.2 Tujuan Data Mining

1. *Explanatory* merupakan menjelaskan beberapa kondisi dari penelitian atau menjelaskan suatu kondisi, seperti mengapa penjualan buku di toko sumber makmur meningkat.
2. *Confirmatory* sarana untuk mengklarifikasi atau untuk mempertegas adanya hipotesa.
3. *Exploratory* digunakan untuk mencari pola baru yang sebelumnya tidak terdeteksi.

2.4.3 Metode- Metode Data Mining

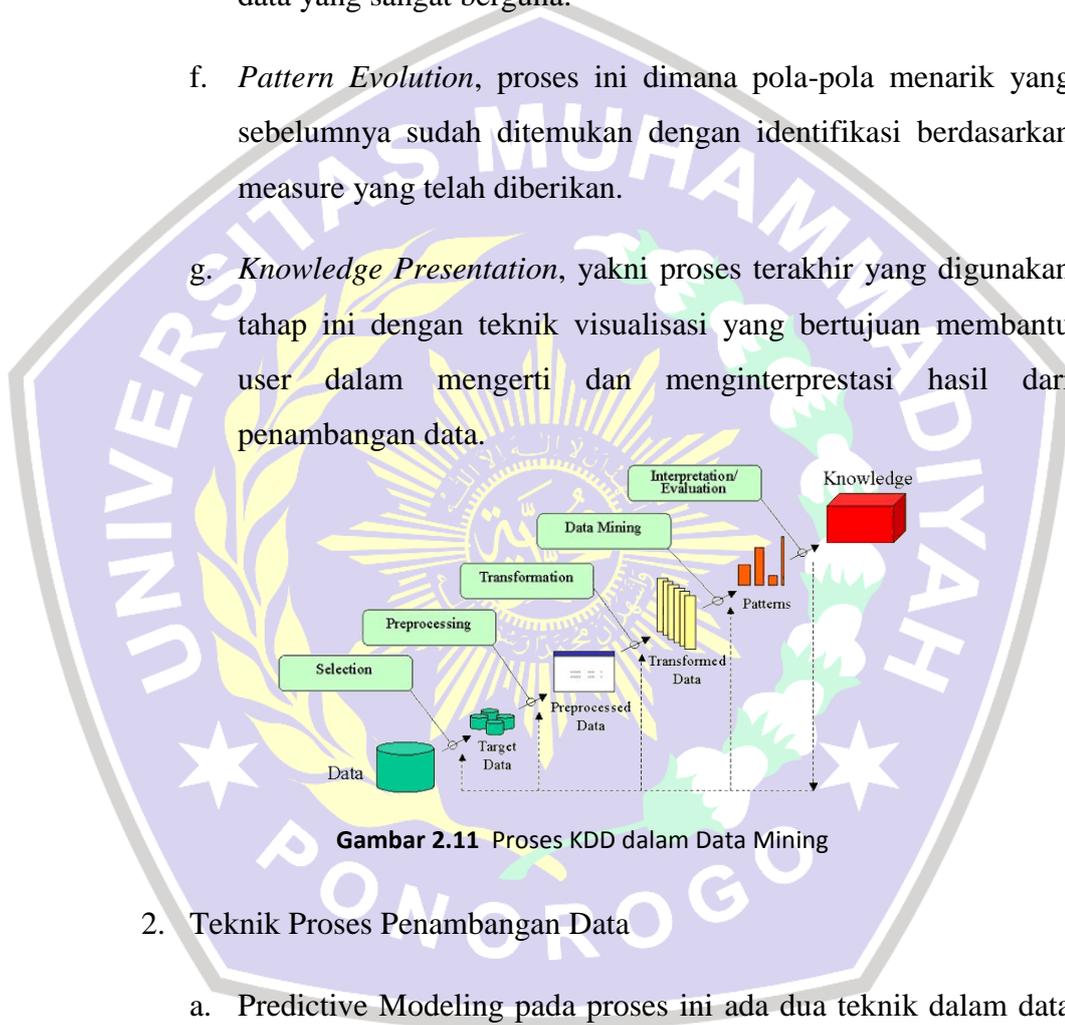
Untuk melakukan suatu pengumpulan data informasi menggunakan beberapa metode dalam data mining untuk membantu menemukan proses data tersebut. Berikut perencanaan dari data mining:

1. Proses Pengambilan Data

Dengan cara KDD tersebut kita dapat melakukan proses pengambilan data. Tahapan-tahapan dari data mentah sampai akhir dengan pengetahuan informasi yang sudah diolah. Berikut prosesnya:

- a. Data *Cleansing* yaitu dimana proses data-data yang tidak lengkap, mengandung eror dan tidak konsisten dibuang dari koleksi data.
- b. Data *Integration* yaitu proses integrasi data dimana yang berulang akan dikombinasikan
- c. *Selection* yaitu dimana proses ini melakukan tahap seleksi atau pemilihan data relevan terhadap analisis untuk diterima dari koleksi data yang ada.

- d. *Data Transformation*, yakni transformasi data yang sudah dipilih ke dalam bentuk *mining procedur* dengan cara dan gresi data.
- e. *Data Mining* yaitu proses yang sangat penting yang akan dilakukan berbagai teknik yang diaplikasikannya untuk mengekstrak berbagai pola-pola potensial untuk mendapatkan data yang sangat berguna.
- f. *Pattern Evolution*, proses ini dimana pola-pola menarik yang sebelumnya sudah ditemukan dengan identifikasi berdasarkan measure yang telah diberikan.
- g. *Knowledge Presentation*, yakni proses terakhir yang digunakan tahap ini dengan teknik visualisasi yang bertujuan membantu user dalam mengerti dan menginterpretasi hasil dari penambangan data.



Gambar 2.11 Proses KDD dalam Data Mining

2. Teknik Proses Penambangan Data

- a. Predictive Modeling pada proses ini ada dua teknik dalam data mining, teknik tersebut yakni :
 1. *Classification* untuk membuat perawalan tentang class yang spesifik pada setiap record dalam database
 2. *Value Prediction* untuk memperkirakan nilai numeric yang kontinu yang terasosiasi dengan record database.

- b. *Database Segmentation* pada proses ini untuk mempartisi database menjadi sebuah segmen, cluster yang sama.
- c. *Link Analysis* digunakan untuk membuat antara record individu atau sekumpulan record dalam database.
- d. *Deviation Detection* untuk mengidentifikasi outlier yang mengekspresikan deviasi dari ekspektasi yang diketahui sebelumnya. Pada operasi ini ditampilkan menggunakan teknik statistik dan visualisasi.
- e. *Nearest Neighbour* pada teknik ini hampir sama dengan pengelompokan untuk memperkirakan nilai prediksi ada dalam satu record, mencari kesamaan nilai prediktor dalam basis data historis.
- f. *Clustering* metode yang mengklasifikasikan data dalam kelompok-kelompok berdasarkan kriteria masing-masing data.
- g. *Decision Tree* yaitu teknik terakhir dalam sebuah model prediktif yang dapat digambarkan seperti pohon.

2.5 Clustering

Clustering adalah metode penganalisaan data, yang sering dimasukkan sebagai salah satu metode *Data Mining*, yang bertujuan untuk mengelompokkan data dengan karakteristik yang sama ke suatu wilayah yang sama dan data dengan karakteristik yang berbeda ke wilayah yang lain. Ada beberapa pendekatan yang digunakan dalam mengembangkan metode clustering dua pendekatan utama adalah clustering dengan pendekatan partisi dan clustering dengan pendekatan hirarki (Ohorella, 2019).

Tujuan *clustering* untuk mengelompokkan seluruh data ke dalam sub kelompok-kelompok atau yang relatif homogen, dimana kemiripan anggota *cluster* dapat ditingkatkan, dan kesamaan dengan anggota di luar kelompok

ini diminimalkan. Langkah awal yang dilakukan dalam proses data mining yaitu clustering. (Larose, 2005)

2.6 Algoritma K-Means

Algoritma K-Means merupakan pengklasteran nilai K rata-rata untuk pengelompokan sedemikian hingga setiap untuk menganalisa termasuk kedalam kelompok dengan rata-rata (titik tengah kelompok) terdekat. K-Means merupakan salah satu algoritma clustering yang bertujuan untuk membagi data menjadi beberapa kelompok. Algoritma K-Means mengelompokkan data untuk menentukan nilai K optimal kedalam kelompok data tersebut. pada setiap cluster terdapat titik pusat (*centroid*) yang akan mempresentasikan cluster tersebut. (Wikipedia, 2021)

Terdapat dua jenis data clustering yang sering digunakan dalam pengelompokan data yakni *Hierarchical* dan *Non-Hierarchical*, dan *K-Means* yang merupakan metode data clustering non-hierarchical atau partition clustering. Metode *Non Hirarki K-Means* ini menentukan terlebih dahulu jumlah *cluster* yang diinginkan (dua *cluster* atau lebih). Setelah jumlah *cluster* diketahui, baru proses cluster dilakukan tanpa mengikuti proses hirarki. (Larasati, 2019).

Hirarki merupakan analisis yang pengklasteran datanya dilakukan dengan cara mengukur jarak kedekatan pada setiap objek yang kemudian membentuk sebuah dendogram. Metode ini dimulai dengan pengelompokkan dua atau lebih objek yang mempunyai kesamaan paling dekat. Selanjutnya pada objek yang lain dan seterusnya hingga *cluster* akan membentuk semacam pohon dimana terdapat tingkatan (*hierarki*) yang jelas antara objek, dari yang paling mirip hingga yang paling tidak mirip (Richard, 2014)