# Analisis Penerapan Knowledge Discovery in Database (KDD) pada Perpustakaan

****

**Disusun :**

**Oleh : Wahyuni Cahnia Sari**

**NIM : 09031181621122**

**Kelas : SI REG 5B**

**SISTEM INFORMASI**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS SRIWIJAYA**

**2018**

Dalam studi kasus Penerapan Knowledge Discovery in Database (KDD) pada perpustakaan menggunakan 7 sumber Paper atau Jurnal yang berjudul sebagai berikut :

1. Penggunaan metode Market Basket analysis untuk meningkatkan kemampuan search engine online buku tugas akhir pada Perpustakaan UK Petra (Paper)
2. Aplikasi Data Mining menggunakan metode association rule dan algoritma apriori berbasis web pada Perpustakaan di SMA Negeri 1 Bungursari (Jurnal)
3. Pemanfaatan Data Mining untuk penempatan buku di Perpustakaan menggunakan metode association rule (Jurnal)
4. Analisa Pola Peminjaman buku di Perpustakaan Universitas Klabat menggunakan Algoritma Apriori (Jurnal)
5. Teknik Data Mining menggunakan metode bayes classifier untuk optimalisasi pencarian pada aplikasi Perpustakaan (Jurnal)
6. Pengenalan Pola transaksi sirkulasi buku pada Database Perpustakaan menggunakan algoritma generalized sequential pattern
7. Penentuan pola sekuensial pada data transaksi Perpustakaan IPB menggunakan Algoritma graph search techniques

Berdasarkan dari sumber ketujuh jurnal tersebut menunjukkan bahwa Knowledge Discovery in Database (KDD) merupakan proses ekstraksi data dari informasi meliputi pengumpulan, pemakaian data historis untuk menemukan pola atau hubungan dalam data set dalam berukuran besar sesuai dengan algoritma yang digunakan dalam membuat sistem atau aplikasi tertentu. Dalam hal ini Perpustakaan merupakan salah satu studi kasus yang memiliki data set yang berukuran besar dan menggunakan beberapa metode algoritma dalam proses pembentukan Sistem yang terjadi di Perpustakaan. Seperti data judul buku dan data transaksi buku yang terdapat pada jurnal yang berjudul *. “Penggunaan metode Market Basket analysis untuk meningkatkan kemampuan search engine online buku tugas akhir pada Perpustakaan UK Petra”*, data buku yang digunakan untuk membuat aplikasi pencarian data buku yang terdapat dalam jurnal yang berjudul *“Teknik Data Mining menggunakan metode bayes classifier untuk optimalisasi pencarian pada aplikasi Perpustakaan”*, data peminjaman buku yang terdapat dalam jurnal yang berjudul *“Analisa Pola Peminjaman buku di Perpustakaan Universitas Klabat menggunakan Algoritma Apriori”* dan *“Aplikasi Data Mining menggunakan metode association rule dan algoritma apriori berbasis web pada Perpustakaan di SMA Negeri 1 Bungursari”*, data kunjungan perpustakaan, data peminjaman, data buku dan data tata letak buku yang terdapat dalam jurnal yang *berjudul “Pemanfaatan Data Mining untuk penempatan buku di Perpustakaan menggunakan metode association rule”* serta data transaksi yang terjadi di Perpustakaan yang terdapat dalam jurnal yang *berjudul “Pengenalan Pola transaksi sirkulasi buku pada Database Perpustakaan menggunakan algoritma generalized sequential pattern”* dan *“Penentuan pola sekuensial pada data transaksi Perpustakaan IPB menggunakan Algoritma graph search techniques”* maupun proses pembentukan sistem Perpustakaan lainnya yang menggunakan Knowledge Discovery in Database (KDD) atau Data Mining.

Tahapan dalam KDD yaitu Pengumpulan data, Pembersihan, Memilih data yang diperlukan sesuai kebutuhan, Mengekstrak Data, dan Mengevaluasi output sehingga data tersebut dapat digunakan untuk proses pembentukan sistem. Proses Knowledge Discovery in Database diantaranya yaitu Pembersihan Data (Menghilangkan data yang tidak konsisten dan mengandung gangguan yang menyebabkan eror atau variabel yang bervariasi atau acak), Integrasi Data (Menggabungkan data dari berbagai sumber), Seleksi Data (Mengambil data yang relevan untuk proses analisis), Transformasi Data (Mentransformasikan atau menggabungkan data ke dalam bentuk yang tepat dengan cara meringkas atau agregasi), Data Mining (Proses dengan metode yang akan diaplikasikan untuk menemukan pola-pola dalam data dengan cara diekstrak), Evaluasi Pola (Mengidentifikasi bebrapa pola-pola yang menarik dalam mempresentasikan knowledge yang dihasilkan) serta Presentasi Pengetahuan atau Knowledge (Penggunaan visualisasi atau teknik representasi untuk menunjukkan pengetahuan hasil penggalian data kepada user).



Dalam Studi kasus yang terjadi di Perpustakaan, menurut jurnal yang berjudul *“Penentuan pola sekuensial pada data transaksi Perpustakaan IPB menggunakan Algoritma graph search techniques”* dalam hal ini data set yang akan dilakukan dalam Proses KDD yaitu data Peminjaman (Sirkulasi) buku di Perpustakaan IPB. Tahapan Proses dalam KDD dimulai dari sebagai berikut:

1. Tahapan Transformasi data yang menghasilkan sebuah database dalam formast Paradox yang terdiri dari satu tabel, yaitu tebel C.
2. Pada tahapan Seleksi Data, koleksi data transaksi peminjaman buku yang tersedia kemudian diseleksi untuk mendapatkan data yang akan di proses melalui data mining dengan 4 prosedur yang dilakukan yaitu memilih tiga field dari 14 field yang tersedia, memilih data transaksi yang terjadi pada periode 1 september 2005 sampai 29 maret 2008, memilih data transaksi yang dilakukan oleh mahasiswa S1 tahun masuk 2005 dan memilih kelas buku M untuk dijadikan item sehingga menghasilkan data transaksi dengan 3 field yaitu field kelas buku M (Item), field nomor identitas peminjam (CID), dan field tanggal peminjaman (Time).
3. Pada Proses Pembersihan Data, data transaksi yang telah diseleksi kemudian dibersihkan untuk membuang record yang mempunyai nilai invalid.
4. Pada Proses Data Mining yang menggunakan metode Sekuensial untuk membentuk semua kemungkinan large sequence dengan menggunakan algoritma GST dengan melakukan berbagai percobaan.
* Pembentukan large sequence dengan time constraint 6 bulan



Grafik jumlah large sequence yang terbentuk dengan time constraint 6 bulan.

* Tabel 1 Daftar nilai minimum support di mana tidak terbentuk large sequence pada time constraint 6 bulan



* Pembentukan large sequence dengan time constraint 4 bulan



Grafik jumlah large sequence yang terbentuk dengan time constraint 4 bulan.

* Tabel 2 Daftar nilai minimum support di mana tidak terbentuk large sequence pada time constraint 4 bulan



* Pembentukan large sequence dengan time constraint 2 bulan



Grafik jumlah large sequence yang terbentuk dengan time constraint 2 bulan

* Tabel 3 Daftar nilai minimum support di mana tidak terbentuk large sequence pada time constraint 2 bulan



* Pembentukan large sequence dengan time constraint 20 hari



Grafik jumlah large sequence yang terbentuk dengan time constraint 20 hari.

* Daftar nilai minimum support di mana tidak terbentuk large sequence pada time constraint 20 hari



Berdasarkan Tabel dan Grafik diatas, maka menghasilkan proses data maining yaitu :



Grafik jumlah large sequence yang terbentuk untuk seluruh time constraint. 

Grafik perbandingan waktu yang dibutuhkan untuk membentuk large sequence.

1. Pada Proses Evaluasi Pola, seluruh dari proses data mining yang telah dilakukan, kemudian dievaluasi untuk mendapatkan pola sekuensial (pola yang telah ditentukan dalam jurnal). Pada Proses Presentasi Pengetahuan, pola sekuensial yang telah diperoleh sebelumnya sangat diperlukan agar pola yang telah ditentukan mudah untuk dipahami dan dapat diinterpretasikan sehingga menghasilkan dalam bentuk aturan logika, dan diharapkan dapat diperoleh pengetahuan yang beharga dari koleksi data yang telah melalui proses data mining.

Sedangkan menurut jurnal yang *berjudul “Aplikasi Data Mining menggunakan metode association rule dan algoritma apriori berbasis web pada Perpustakaan di SMA Negeri 1 Bungursari”* dalam hal ini menggunakan Data Set yaitu data buku, data transaksi dan data detail transaksi. Data Set tersebut telah didapatkan dari proses tahapan KDD sebelumnya yaitu proses pembersihan data karena tidak semua tabel digunakan maka perlu dilakukan pembersihan data agar data yang akan diolah menjadi relevan sesuai dengan yang dibutuhkan. Pembersihan data penting guna meningkatkan performa dalam proses data mining. Cara pembersihan data ini dilakukan dengan cara dengan menghapus atribut yang tidak terpakai dan menghapus data-data yang isinya tidak lengkap . Atribut yang digunakan terdiri dari atribut pada data koleksi buku, detail transaksi dan data peminjaman. Kemudian semua data tersebut diseleksi berdasarkan minimum threshold yang ditetapkan.

Di dalam jurnal yang *berjudul “Analisa Pola Peminjaman buku di Perpustakaan Universitas Klabat menggunakan Algoritma Apriori”* proses tahapan dalam KDD meliputi :



1. Pengumpulan Data

Pengumpulan data dilakukan pada transaksi peminjaman buku selama 5 tahun terakhir yang diambil dari database server Perpustakaan Universitas Klabat.

1. Pembersihan Data

Setelah data sudah terkumpul, selanjutnya dilakukan seleksi data. Gunanya adalah untuk mengetahui data mana yang valid atau sesuai dengan atribut yang ditentukan dan sesuai dengan metode analisa.

1. Pengolahan Data

Pengolahan data dilakukan untuk melihat permasalahan yang ada yaitu analisa pola peminjaman buku dan untuk mengolah datadata yang telah diseleksi, peneliti menggunkan tools WEKA dengan algoritma apriori sebagai metode pencarian minimum support dan confidence

1. Pengujian Metode

Pengujian metode dilakukan untuk menguji tingkat support dan confidence sehingga hubungan antar itemset diketahui sebelumnya.

1. Evaluasi Hasil

Evaluasi dilakukan untuk hasil yang didapat berupa pola peminjaman dari eksperimen yang telah dilakukan dengan menggunakan metode Algoritma Aproiori.

Sedangkan Jurnal atau Paper yang lainnya dalam tahapan Proses KDD didukung dengan berbagai macam metode dan algoritma yang digunakan. Diantaranya yaitu paper yang berjudul “Penggunaan metode Market Basket analysis untuk meningkatkan kemampuan search engine online buku tugas akhir pada Perpustakaan UK Petra” dan jurnal yang berjudul “Pemanfaatan Data Mining untuk penempatan buku di Perpustakaan menggunakan metode association rule” menggunakan algoritma apriori dan metode Generate Association Rules . sedangkan jurnal yang berjudul “Teknik Data Mining menggunakan metode bayes classifier untuk optimalisasi pencarian pada aplikasi Perpustakaan” menggunakan metode bayes classifier dalam aplikasi pencarian (search engine) dan jurnal yang berjudul “Pengenalan Pola transaksi sirkulasi buku pada Database Perpustakaan menggunakan algoritma generalized sequential pattern” menggunakan metode sequential generalized pattern. Dari keempat jurnal atau paper yang telah disebutkan sebelumnya, menunjukkan bahwa Knowledge Discovery merupakan gabungan dari banyak disiplin ilmu, sehingga proses KDD menggunakan metode dan algoritma yang berbeda walaupun studi kasusnya sama.