

**PENERAPAN DATAMINING MENENTUKAN MINAT BACA MAHASISWA  
DI PERPUSTAKAAN UNIVERSITAS BINA DARMA PALEMBANG  
MENGUNAKAN METODE CLUSTERING**

***APPLICATION OF DATA MINING DETERMINES INTEREST READING IN  
STUDENTS AT UNIVERSITAS BINA DARMA PALEMBANG USING  
CLUSTERING METHOD***

**Lemi Iryani**

Universitas Sjakhyakirti  
[lemiiryani@unisti.ac.id](mailto:lemiiryani@unisti.ac.id)

**ABSTRACT**

*This study was conducted to determine student interest in reading in the library Bina Darma with clustering method. The data used in this study is Bina Darma library database that consists of three tables are tables collection, patron tables and tables of visitors who will be selected to be integrated and combined as a data warehouse. Application of data mining to determine the titles of the most demanding of students and courses are most borrowed books and read a book in the library building dharma. The information obtained is that the course is more interested in information systems to borrow and read books compared to the industrial engineering courses. Titles that most interest is the title of the book basics of programming, while the lowest book is the title of students interested in a practical handbook using Microsoft word and cluster categories of books that most interest students are computer science book category books category Indonesian whereas lower.*

**Keywords:** Clustering, Data Mining, Data Warehouse, Visitor, Collection, Patron

**ABSTRAK**

Penelitian ini dilakukan untuk menentukan minat baca mahasiswa Bina Darma dengan metode clustering. Data yang digunakan dalam penelitian ini adalah database perpustakaan Bina Darma yang terdiri dari tiga tabel yaitu tabel collection, tabel patron dan tabel visitor yang akan diintegrasikan dan diseleksi untuk digabungkan sebagai data warehouse. Penerapan data mining untuk menentukan judul buku yang paling banyak diminati mahasiswa dan program studi apa yang paling banyak meminjam buku dan membaca buku di perpustakaan bina darma. Informasi yang didapat adalah bahwa program studi sistem informasi lebih berminat meminjam dan membaca buku dibandingkan dengan program studi teknik industri. Judul buku yang paling banyak diminati adalah judul buku dasar-dasar pemrograman sedangkan buku yang paling rendah diminati mahasiswa adalah judul buku panduan praktis menggunakan Microsoft word dan cluster kategori buku yang paling banyak diminati mahasiswa adalah kategori buku ilmu komputer sedangkan kategori buku bahasa Indonesia lebih rendah.

**Kata Kunci:** Clustering, Data Mining, Data Warehouse, Visitor, Collection, Patron

**PENDAHULUAN**

Teknologi informasi telah menjadi fasilitas utama bagi kegiatan berbagai sektor kehidupan dimana memberikan andil besar terhadap perubahan-perubahan yang mendasar pada struktur operasi dan manajemen organisasi, pendidikan, transportasi, kesehatan dan penelitian. Perkembangan yang cepat

dalam teknologi pengumpulan dan penyimpanan data telah memudahkan organisasi untuk mengumpulkan sejumlah data berukuran besar.

Data mining adalah kegiatan untuk menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam database, data warehouse, atau penyimpanan informasi lainnya (Huda,

2010). Teknik data mining digunakan untuk memeriksa basis data berukuran besar sebagai cara untuk menemukan pola yang baru dan berguna. Salah satu teknik analisa yang dapat digambarkan dalam data mining yaitu clustering(Kurniawan & Fransiska, 2014). Clustering adalah proses pengelompokan data ke dalam cluster berdasarkan parameter tertentu sehingga obyek-obyek dalam sebuah cluster memiliki tingkat kemiripan yang tinggi satu sama lain dan sangat tidak mirip dengan obyek yang lain pada cluster yang berbeda(Metisen & Sari, 2015). Dalam clustering ini data yang dikelompokkan berupa judul buku dan data pengunjung yang paling diminati oleh mahasiswa untuk dibaca pada Perpustakaan Universitas Bina Darma Palembang.

Perpustakaan Universitas Bina Darma Palembang didirikan dengan tujuan untuk membantu dan mendukung dalam memenuhi kebutuhan informasi ilmiah sivitas akademika. Perpustakaan banyak digunakan mahasiswa untuk membaca dan meminjam buku. Dengan adanya perpustakaan kegiatan belajar dan mengajar akan lebih lancar karena didukung oleh berbagai fasilitas yang mendukung seperti buku bahan ajar, buku teks, dan lain-lain yang disediakan oleh perpustakaan. Minat adalah kecenderungan hati, gairah dan keinginan yang tinggi terhadap sesuatu(Agustin, 2018).

Dalam hal ini minat biasa muncul dari dalam diri mahasiswa sendiri. Akan tetapi lingkungan keluarga dan pergaulan juga sangat berpengaruh dalam menumbuhkan minat baca mahasiswa. Hal ini diperlukan karena pengembangan minat baca merupakan salah satu faktor pendukung pendidikan. Kurangnya minat baca mahasiswa mungkin disebabkan beberapa faktor kurangnya buku yang diminati, koleksi

buku yang sudah lama, kurangnya koleksi buku terbaru, serta kurangnya dosen dalam memberikan tugas dan lain-lain. Kondisi minat baca perpustakaan dapat dilihat dari persentase jumlah pengunjung, jumlah peminjam dan jumlah buku yang dipinjam(Riyandwyana, 2012).

Data peminjaman buku dijadikan pihak perpustakaan sebagai laporan perpustakaan. Dari laporan data peminjaman buku dapat kita lihat minat mahasiswa untuk membaca diperpustakaan setiap bulan dan tahunnya. Apakah data peminjaman buku hanya dijadikan pihak perpustakaan sebagai laporan bulanan dan tahunan saja? Sehingga dalam setiap tahunnya data akan membesar dan terjadi penumpukan data. Jika demikian maka sangat disayangkan bila data diperpustakaan hanya dijadikan laporan dan tumpukan saja oleh pihak perpustakaan.

## METODE

Dalam penelitian ini, penulis menggunakan metode deskriptif karena permasalahan yang sedang diteliti saat ini berdasarkan data-data yang bersifat fakta yang ada mengenai data anggota, data peminjam dan koleksi buku yang paling diminati oleh mahasiswa di perpustakaan Universitas Bina Darma Palembang.

Adapun untuk menganalisis data dalam penerapan data mining ini menggunakan proses tahapan knowledge discovery in databases (KDD) yang terdiri dari data cleaning, data integration, Data Selection, Data Transformation, Proses mining, pattern evaluation, knowledge presentation(Saputra, dkk., 2016). Tahap-tahap data mining ada 6 yaitu :

1. Pembersihan data (*data cleaning*)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari *database* suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa *data mining* yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya(Mardi, 2017).

## 2. Integrasi data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Tidak jarang data yang diperlukan untuk *data mining* tidak hanya berasal dari satu *database* tetapi juga berasal dari beberapa *database* atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

## 3. Seleksi Data (*data selection*)

Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan

orang membeli dalam kasus *market basket analysis*, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

## 4. Transformasi data (*data transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan *clustering* hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

## 5. Proses *mining*,

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

## 6. Evaluasi pola (*pattern evaluation*),

Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses *data mining*, mencoba metode *data mining* lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

## 7. Presentasi pengetahuan (*knowledge presentation*),

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari

proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami *data mining*. Karenanya presentasi hasil *data mining* dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses *data mining*. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil *data mining*.

## HASIL DAN PEMBAHASAN

### Analisis Data Mining

Dalam penelitian ini yang akan dicari adalah tingkat minat baca mahasiswa berdasarkan judul buku yang paling banyak dipinjam di perpustakaan Universitas Bina Darma Palembang. Untuk mengetahui tingkat minat baca mahasiswa data akan di *cluster* berdasarkan judul buku yang paling banyak dipinjam di perpustakaan.

Data yang digunakan berasal dari database perpustakaan yang terdiri dari tabel koleksi buku, tabel anggota dan tabel peminjaman untuk mencari keterkaitan antar tabel. Tidak semua atribut yang kira-kira berguna dan sebarannya tidak terlalu acak, karena data yang terlalu acak akan membuat proses mining memakan waktu lama dan tingkat hubungannya pun rendah. Data peminjaman yang akan dicari hubungannya meliputi *proses id\_visitor*, *date*, *time\_in*, *id\_patron*, *id\_collection*.

### Sumber Data

Data yang digunakan berasal dari perpustakaan Bina Darma Palembang berupa database perpustakaan yang terdiri dari tiga tabel, yaitu data koleksi, data peminjaman dan data anggota.

#### 1. Tabel *Collection* (koleksi buku)

Data koleksi adalah data koleksi buku yang di data ketika masuknya

buku-buku koleksi di perpustakaan dan di susun oleh pihak pustakawan pada perpustakaan dan dapat dipinjam ataupun dibaca mahasiswa yang berkunjung ke perpustakaan.

**Tabel 1. Atribut *collection***

Atribut	Keterangan
<i>Id_collection</i>	<i>Id_collection</i> adalah primary key untuk tabel koleksi buku
<i>Title</i>	<i>Title</i> adalah judul dari buku yang dihungunkan dengan id buku
<i>Publisher_city</i>	<i>Publisher_city</i> adalah kota terbit untuk identitas dari mana buku diterbitkan.
<i>Publisher</i>	<i>Publisher</i> adalah nama penerbit dari buku
<i>Published_year</i>	<i>Publisher_year</i> adalah tahun terbitnya buku

#### 2. Tabel *Patron* (anggota)

Data anggota adalah data mahasiswa yang telah mendaftar sebagai anggota perpustakaan agar dapat melakukan peminjaman buku.

**Tabel 2. Atribut *patron***

Atribut	Keterangan
<i>Id_patron</i>	<i>Id</i> Merupakan primary key
<i>Name</i>	<i>First name</i> adalah nama depan mahasiswa
<i>Sex</i>	<i>Sex</i> adalah jenis kelamin dari anggota
<i>Birth_place</i>	<i>Birth_place</i> adalah tempat lahir anggota
<i>Birth_date</i>	<i>Birth_date</i> adalah tanggal lahir anggota
<i>Blood_type</i>	<i>Blood_type</i> adlah golongan darah anggota
<i>Address</i>	<i>Address</i> adalah alamat mahasiswa
<i>Program studi</i>	<i>Program studi</i> adalah jurusan anggota

#### 3. Tabel *Visitirlog* (peminjaman)

Data peminjaman adalah data mahasiswa yang telah melakukan peminjaman buku di perpustakaan dan disimpan dalam serkulasi perpustakaan.

**Tabel 3. Atribut *visitirlog***

Atribut	Keterangan
<i>Id_visitorlog</i>	<i>Id</i> adalah kode identitas peminjam buku

<i>Date</i>	<i>Date</i> adalah tanggal peminjaman
<i>Time_in</i>	<i>Time_in</i> adalah waktu menjadi anggota
<i>Id_patron</i>	<i>Id</i> Merupakan primary key
<i>Id_collection</i>	<i>Id_collection</i> adalah primary key untuk tabel koleksi buku

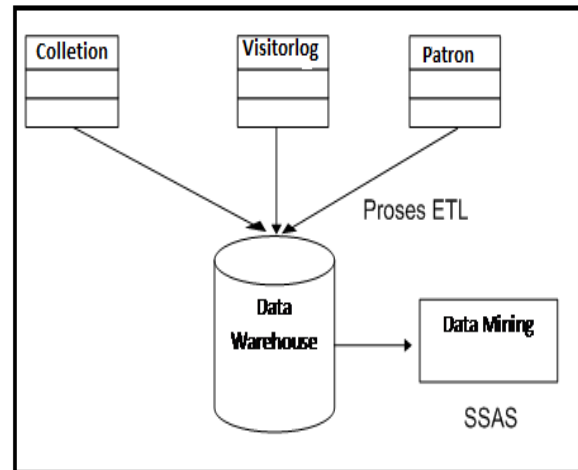
Data yang diambil dalam sampel adalah database dari perpustakaan. Hal ini didasarkan pada kebutuhan data yang akan di hubungkan dengan data koleksi, data anggota dan data peminjaman, dengan asumsi bahwa minat mahasiswa dalam meminjam buku diperpustakaan diperoleh berdasarkan judul buku yang paling banyak dipinjam oleh mahasiswa.

### Data Warehouse

Dalam proses ini dibangun *data warehouse* yang digunakan untuk menampung data dari *database* perpustakaan yang terdiri dari beberapa tabel yaitu tabel koleksi buku, tabel anggota dan tabel peminjaman buku. Tujuan utama pembangunan *data warehouse* adalah agar *database* tidak terganggu bila terjadi *error*, selain itu *data warehouse* memudahkan dalam menyatukan data dari tiga tabel. Proses ETL (*Ekstrak Transform Load*) dalam pembangunan *data warehouse* merupakan proses yang penting karena menentukan pembangunan *data warehouse* selanjutnya.

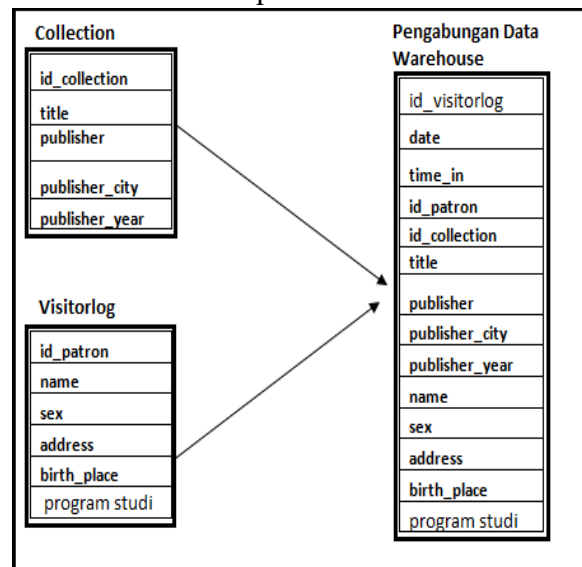
Dari database perpustakaan yang akan digunakan sebagai sumber data akan *dimport* kedalam *SQLSERVER2008R2* sebagai tempat proses untuk menyimpan database perpustakaan. Proses ETL (*Ekstrak Transform Load*) data yang diambil adalah data koleksi buku, data anggota dan data peminjaman akan diproses untuk dijadikan *data warehouse* sebagai penggabungan data. Data ditampilkan menjadi *data warehouse* baru diposes

lagi kedalam proses data mining yang sering disebut proses *SSAS*.



Gambar 1. Konsep proses ETL (*ekstrak transform load*).

Pembangunan *data warehouse* ini, terdiri dari tiga tabel yaitu tabel koleksi buku, tabel anggota dan tabel peminjaman buku yang akan digabungkan dalam satu tabel yang akan dijadikan sebagai *data warehouse*, data yang akan menjadi *data warehouse* adalah data peminjaman buku yang atributnya telah dipilih untuk dianalisis untuk mempermudah dalam pengambilan keputusan. Proses pemuatan data dari *database* ke *data warehouse* melalui proses ETL.

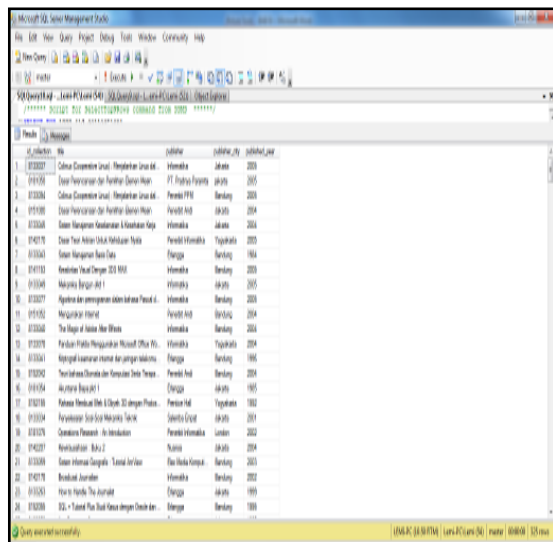


Gambar 2. Pengabungan data (*data warehouse*)

Semua data yang ada pada *tabel patron* dan *collection* diimport ke dalam *data warehouse*, kecuali jika ada data yang sama dari kedua tabel seperti, *id\_patron*, *id\_collection* dan yang lain, maka yang di *load* hanya salah satu dari *tabel* dengan asumsi memiliki nilai yang sama. Hal ini dilakukan agar tidak ada *redudancy* data atau data rangkap.

### Data Transformation (Transformasi Data)

Untuk dapat di *mining* sumber data yang digunakan dalam format *Mysql* akan di konversi kedalam format *Microsoft Office Excel*. Alasan penggunaan *Microsoft Office Excel* adalah untuk mempermudah imfort data kedalam *Sql Server Management Studio* agar dapat mengolah database sesuai dengan yang dibutuhkan. Tahap seleksi data dan pembersihan data juga dilakukan bersama pada tahap *transformasi* data. Hasil yang diperoleh dari tahap ini adalah sebuah basis data dalam format *Microsoft Office Excel* dapat dilihat pada gambar 4. data tranformasi dibawah ini.



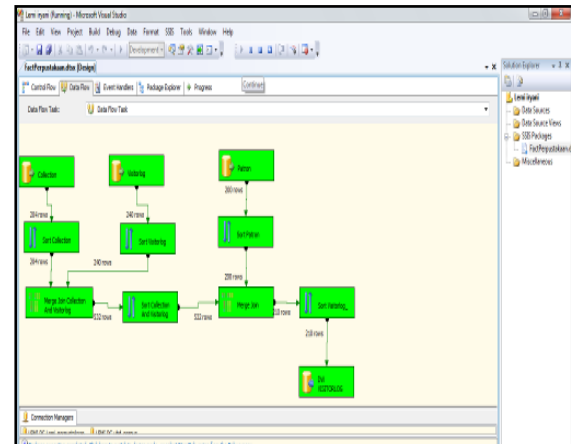
id_transaksi	isbn	judul	penerbit	tahun	jumlah_jual
1	9780000000000	Calculus	Yogyakarta	2005	1
2	9780000000000	Calculus	Yogyakarta	2005	1
3	9780000000000	Calculus	Yogyakarta	2005	1
4	9780000000000	Calculus	Yogyakarta	2005	1
5	9780000000000	Calculus	Yogyakarta	2005	1
6	9780000000000	Calculus	Yogyakarta	2005	1
7	9780000000000	Calculus	Yogyakarta	2005	1
8	9780000000000	Calculus	Yogyakarta	2005	1
9	9780000000000	Calculus	Yogyakarta	2005	1
10	9780000000000	Calculus	Yogyakarta	2005	1
11	9780000000000	Calculus	Yogyakarta	2005	1
12	9780000000000	Calculus	Yogyakarta	2005	1
13	9780000000000	Calculus	Yogyakarta	2005	1
14	9780000000000	Calculus	Yogyakarta	2005	1
15	9780000000000	Calculus	Yogyakarta	2005	1
16	9780000000000	Calculus	Yogyakarta	2005	1
17	9780000000000	Calculus	Yogyakarta	2005	1
18	9780000000000	Calculus	Yogyakarta	2005	1
19	9780000000000	Calculus	Yogyakarta	2005	1
20	9780000000000	Calculus	Yogyakarta	2005	1
21	9780000000000	Calculus	Yogyakarta	2005	1
22	9780000000000	Calculus	Yogyakarta	2005	1
23	9780000000000	Calculus	Yogyakarta	2005	1
24	9780000000000	Calculus	Yogyakarta	2005	1
25	9780000000000	Calculus	Yogyakarta	2005	1
26	9780000000000	Calculus	Yogyakarta	2005	1
27	9780000000000	Calculus	Yogyakarta	2005	1
28	9780000000000	Calculus	Yogyakarta	2005	1

Gambar 4. Data transformasi

### Data Warehouse

Proses *data warehouse* ini dibangun *data warehouse* yang digunakan untuk menampilkan data dari *database*

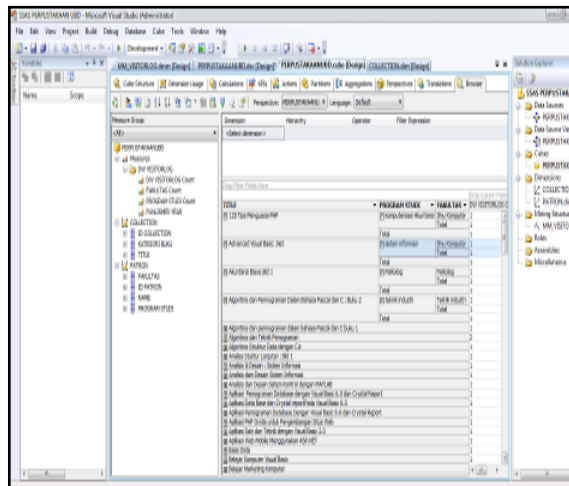
perpustakaan yang terdiri dari *tabel collection*, *table patron* dan *tabel visitor*. Dari *tabel-tabel* ini akan dibangun *tabel visitorlog* baru sebagai *data warehouse*. Cara membuat *data warehouse* Tambahkan *OLE DB Destination* untuk menggabungkan semua data sebagai *data warehouse* dapat dilihat pada gambar 3. *data warehouse* dibawah ini.



Gambar 3. Data warehouse

### Proses Data Mining

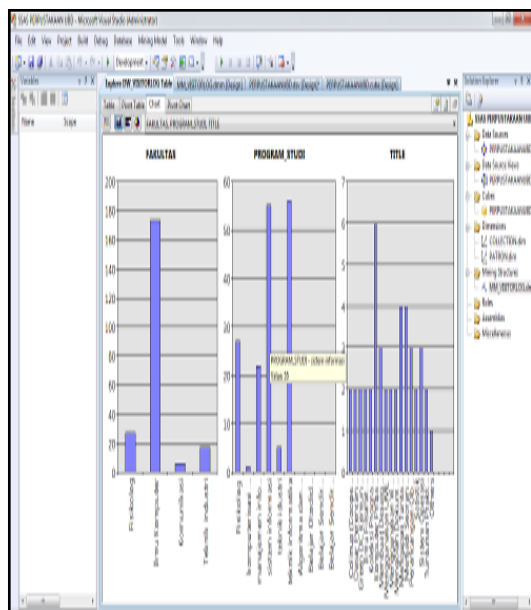
Proses *data mining* ini adalah untuk menampilkan semua data yang akan dijadikan *clustering* untuk mengetahui tingkat minat mahasiswa dalam meminjam buku di perpustakaan. Proses yang dilalui adalah proses setelah menentukan *data warehouse* yang dibangun sesuai dengan kebutuhan. Proses *mining* menggunakan *SSAS (sql server Analytical services)* untuk membuat *data mining* dan menampilkan minat baca mahasiswa berdasarkan judul buku di perpustakaan menggunakan metode *clustering*.



Gambar 5. Proses data mining

### Clustering

Proses *clustering* adalah sekumpulan data objek yang sama yang akan dikelompokkan menurut kemiripannya. *Clustering* ini diproses setelah melalui proses *mining* untuk menampilkan data yang dibutuhkan dan di gunakan dalam pengolahan data ini. Data yang akan *dicluster* adalah judul buku yang paling banyak diminati mahasiswa dapat dilihat pada gambar 6. *Clustering* dibawah ini.



Gambar 6. Clustering

### SIMPULAN

Kesimpulan yang dapat diambil dalam penelitian ini adalah:

1. Penerapan data *mining* untuk menampilkan informasi minat baca mahasiswa di Perpustakaan Bina Darma Palembang menggunakan metode *clustering*.
2. Informasi yang ditampilkan berupa buku yang paling banyak diminati mahasiswa adalah buku algoritma dan pemrograman.

Hasil yang didapat ada perbedaan antara program studi yang meminjam buku dapat dilihat bahwa program studi Sistem Informasi lebih berminat berkunjung dan meminjam buku di bandingkan dengan program studi Teknik Industri.

### DAFTAR PUSTAKA

- Agustin, W. (2018). Penerapan Metode Clustering K-Means dan Metode Six Sigma untuk Mengetahui Tingkat Minat Baca Mahasiswa (Studi Kasus: STMIK-AMIK Riau). *SATIN-Sains dan Teknologi Informasi*, 3(1), 22-31.
- Huda, N. M. (2010). *Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa (Studi Kasus di Fakultas MIPA Universitas Diponegoro)* (Doctoral dissertation, FACULTY OF MATHEMATICS AND NATURAL SCIENCES).
- Kurniawan, E., & Fransiska, M. (2014). Penerapan Algoritma K-Means untuk Clustering Dokumen E-Jurnal STMIK GI MDP.
- Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4. *5. Edik Informatika*, 2(2), 213-219.
- Metisen, B. M., & Sari, H. L. (2015). Analisis clustering menggunakan metode K-Means dalam pengelompokkan penjualan produk



- pada Swalayan Fadhila. *Jurnal media infotama*, 11(2).
- Riyandwyana, A., Suryani, E., & Mukhlason, A. (2012). Pengembangan Sistem Rekomendasi Peminjaman Buku Berbasis Web Menggunakan Metode Self Organizing Map Clustering Pada Badan Perpustakaan Dan Kearsipan (BAPERSIP) Provinsi Jawa Timur. *Jurnal Teknik ITS*, 1(1), A374-A378.
- Saputra, N. E., Tania, K. D., & Heroza, R. I. (2016). Penerapan Knowledge Management System (Kms) Menggunakan Teknik Knowledge Data Discovery (Kdd) Pada Pt Pln (Persero) Ws2jb Rayon Kayu Agung. *Sriwijaya Journal of Information Systems*, 8(2), 131900.