BAB 2

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1. TINJAUAN PUSTAKA

Pada tinjauan pustaka akan berisi tentang referensi penelitian yang digunakan sesuai dengan judul atau metode dari skripsi yang dibuat sebagai referensi yang digunakan seperti berikut:

Olivia (2015) dalam penelitiannya membahas sistem informasi tentang pola belanja konsumen untuk penentuan *layout* barang di dalam rak sesuai keinginan konsumen dan proses pengambilan keputusan untuk strategi bisnis yang penting bagi perusahaan dengan mengunakan metode algoritma *apriori* dan mengunakan *tools Magic* dengan hasil pencapaian untuk *minimum support* 50% dan *minimum confidence* 60% didapat hasil pola peletakan yaitu roti_biskuit, snacks, dan soft_drink.

Jananto (2012) dalam penelitiannya membahas tentang penjajakan pemetaaan dari nilai-nilai matakuliah yang telah ditempuh oleh mahasiswa atau calon lulusan terhadap kompetensi dasar yang harus dimilki oleh seorang lulusan informatika yang disusun oleh APTIKOM sehingga dihasilkan peta kompetensi dari calon lulusan. Metode yang digunakan yaitu algoritma *apriori* dengan mengunakan *tools data mining* yaitu Tanagra dengan pencapaian kompetensi pada *level minimum support* 70% dan *minimum confidence* 75% didapat hasil pola kompetensi fakuktas teknologi informasi UNISBANK yaitu Sistem Informasi(IS), System Integration (SI), dan Network and Communication (NC).

Prisca (2015) dalam penelitiannya membahas implementasi algoritma *apriori* dalam sebuah aplikasi penjualan dimana algoritma *apriori* dapat menganalisa pola konsumsi pelanggan untuk melakukan promosi pembelian produk untuk meningkatkan penjualan. Pada penelitian ini hanya membahas sistem informsinya tanpa dibandingkan dengan data manual. Metode yang digunakan yaitu algoritma *apriori* dan mengunakan bahasa pemrograman java.

Rahmaniar (2015) pada penelitiannya penulis membuat aplikasi untuk mengelola data perceraian pada pengadilan agama agar dapat menjadi sebuah informasi yang dapat digunakan untuk menekan angka perceraian. Aplikasi ini mengunakan bahasa pemrogramana java dan metode *association rule* dengan algorotma *apriori*. Adapun hasil dari penelitian diperoleh 5 rule dengan nilai *minimum support* 5% dan *minimum confidence* 20% serta 2 rule dengan niali *minimum support* 10% dan *minimum confidence* 25%. Sehingga diketahui bahwa semakin tinggi nilai *minimum support* dan *minimum confidence*, semakin sedikit hasil analisa yang didapatkan dengan pola yang di dapat yaitu ketidak ada harmonisan, tidak ada tanggung jawab, dan latar belakang pekerjaan serta pendidikan.

Dari referensi yang telah diutarakan di atas maka akan dibangun sebuah sistem yang dapat mebantu kaprodi untuk melihat kompetensi yang di miliki lulusannya apakah sudah sesuai target yang di inginkan dengan mengunakan algoritma apriori. Sistem yang akan di bangun hampir sama dengan penelitian milik Jananto (2012) akan tetapi berbeda pada bagian prodi yang digunakan untuk penelitian, *curricula* APTIKOM , dan tools. Sistem yang akan dibangun mengacu

pada prodi Manajemen Informatika STMIK Akakom Yogyakarta, serta *curricula* APTIKOM yang digunakan adalah *curricula* 2013, dan untuk membangun sistem mengunakan bahasa pemrograman java.

Berikut ini adalah tabel perbandingan dari penelitian sebelumnya :

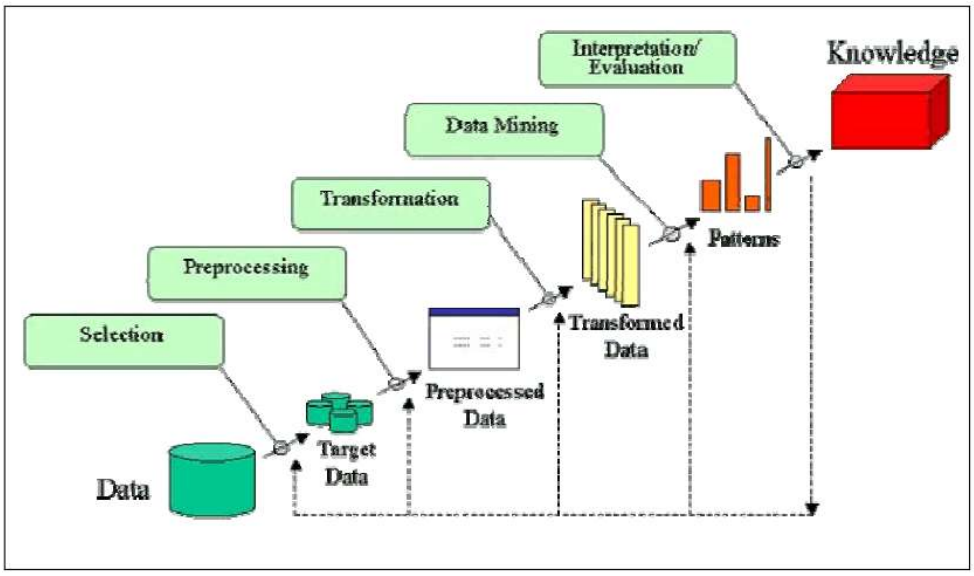
Tabel 2.1 Perbandingan Tinjauan Pustaka

No	Penulis	Judul	Objek Penelitian	Metode	Bahasa Pemrograman /*Tools*
1	Olivia(2015)	Perancangan Sistem Informasi Data Mining Dengan Algoritma Apriori Untuk Penentuan Layout Produk Pada PT. Metro Makmur Nusantara	PT. Metro Makmur Nusantara	Association Rule	Magic
2	Prisca,R. (2015)	Penerapan Association Rule Dengan Algoritma Apriori Pada Transaksi Penjualan Di Minimarket	Minimarket di kediri	Association Rule	Java
3	Jananto,A (2012)	Penggunaan Market Basket Analysis untuk Menentukan Pola Kompetensi Mahasiswa	Fakultas Teknologi Informasi UNISBANK	Association Rule	Tanagra
4	Rahmaniar,A. (2015)	Penerapan Metode Association Rule Mengunakan Algoritma Apriori Untuk Analisa Pola Data Perceraian	Pengadilan Agama Sleman	Association Rule	Java
5	Subekti,N. (2018)	Penerapan Algoritma Apriori Untuk Menentukan Pola Kompetensi Mahasiswa	Prodi D3 STMIK Akakom	Associaton Rule	Java

2.2. DASAR TEORI

2.2.1 *Data Mining*

Alasan utama mengapa *data mining* diperlukan adalah karena adanya sejumlah besar data yang dapat digunakan untuk menghasilkan informasi dan *knowledge* yang berguna. Informasi dan *knowledge* yang didapat tersebut dapat digunakan pada banyak bidang, mulai manajemen bisnis, control produksi, kesehatan, dan lain-lain. Secara sederhana, *data mining* dapat diartikan sebagai proses mengekstrak atau "menggali" *knowledge* yang ada pada sekumpulan data. Banyak orang yang setuju bahwa *data mining* adalah sinonim dari *Knowledge Discovery in Database*(KDD)(Han dan Kamber, 2001).



Gambar 2.1. Proses KDD.

Data mining merupakan salah satu langkah dalam proses penemuan pengetahuan yang terdiri dari *iterasi* urutan langkah-langkah seperti pada gambar 2.1 yaitu:

1. *Data Selection*

   Pemilihan (seleksi) data dari sekumpulan data operasi perlu dilakukaan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan di gunakan untuk *data mining*, disimpan dalam suatu berkas, terpisah dari baris data operasional.

2. *Pre-processing/ Cleaning*

   Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membangun *duplikasi* data, memeriksa data yang *inkonsisten* dan memperbaiki kesalahan pada data, seperti kesalahan cetak(*tipografi*). Juga dilakukan proses *enrichment*, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD seperti data atau informasi *eksternal*.

3. *Transformation*

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

4. *Data Mining*

   Data mining dalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Interpretation/Evaluation*

   Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau *hipotesis* yang ada sebelumnya(Han dan Kamber,2011).

2.2.2 Standar Kompetensi Aptikom

Menurut standar kurikulum yang baru, dimana APTIKOM akan kembali mengacu hasil studi IEEE dan ACM yang disampaikan melalui dokumen publikasi Computing Curricula 2013. Berdasarkan standar adopsi ini, ciri khas seorang lulusan informatika adalah apabila memiliki 7 (tujuh) kompetensi dasar yang berkaitan dengan aspek-aspek pengetahuan berupa 1). Pembentukan Karakter, 2).Matematika dan Statistika, 3). Algoritma Dan Pemrograman , 4).

Rekayasa Perangkat Lunak, 5). Arsitektur Komputer , 6). Kecakapan Hidup, 7). Manajemen Informasi (Aptikom, 2015).

2.2.3 Algoritma *Apriori*

Algoritma *apriori* adalah algoritma paling terkenal untuk menentukan pola frekuensi tinggi. Pola frekuensi tinggi adalah pola-pola *item* di dalam suatu *database* yang memiliki frekuensi di atas ambang batas tertentu yang disebut dengan istilah *minimum support* atau *threshold. Threshold* adalah batas *minimum* transaksi. Jika jumlah transksi kurang dari *threshold,* maka *item* atau kombinasi *item* tidak akan diikutkan dalam perhitungan selanjutnya. Penggunaan *threshold* dapat mempercepat perhitungan(Rahmaniar,2015).

Prinsip algoritma *apriori* terbagi dalam beberapa tahap:

1. Kumpulkan jumlah *item* tunggal, dapatkan *item* yang lebih besar dari nilai batas.
2. Dapatkan *candidate pairs*, hitung => *large pairs* dari *item-item*.
3. Dapatkan *candidate triplets*, hitung => *large triplets* dari *item-item* seterusnya.

2.2.4 *Association Rule*

Analisis asosiasi atau *associaton rule mining* adalah teknik *data mining* untuk menemukan aturan asositif antara suatu kombinasi *item*(Kusrini,dkk, 2009). Algoritma aturan asosiasi akan mengunakan

data latihan, sesuai dengan pengertian *data mining* untuk menghasilkan pengetahuan. Aturan asosiasi yang berbentuk "*if...then...*" atau "jika...maka..." merupakan pengetahuan yang dihasilkan dari fungsi aturan asosiasi. Contoh asosiasi yang biasa terjadi seperti seberapa besar kemungkinan pembeli membeli roti dengan selai secara bersamaan. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter yaitu *support* dan *confidence.*

1. *Support* (nilai penunjang) adalah presentase kombinasi *item* tersebut dalam *databases*.
2. *Confidence* (nilai kepastian) adalah kuatnya hubungan antar *item* dalam aturan asosiasi.

Analisis asosiasi didefinisikan sebagai suatu proses untuk menemukan semua aturan asosiaasi yang memenuhi syarat *minimum* untuk *support* (*minimum support*) dan syarat *minimum* untuk *confidence* (*minimum confidenc*). Jika support-nya >= minimum support dan confidence-nya >= minimum confidence, maka rule tersebut bisa dikatakn sebagai *iteresting rule*.

Metodologi dasar analisis asosiasi terbagi menjadi 2 tahap(Rahmaniar,2015), yaitu:

a. Analisis pola frekuensi tinggi

Tahap ini mencari kombinasi *item* yang memenuhi syarat *minimum* dari nilai *support* dalam *databases*. Nilai *support* sebuah *item* di peroleh dengan rumus 2.1:

$$\boldsymbol{Support(A)} = \frac{\sum Transksi\ mengandung\ A}{\sum\ Transaksi} \quad (2.1)$$

Sementara itu, nilai *support* dari 2 *item* diperoleh dari rumus 2.2:

$$\boldsymbol{Support(A,B)} = \frac{\sum Transksi\ mengandung\ A\ dan\ B}{\sum\ Transaksi} \quad (2.2)$$

b. Pembentukan aturan asosiasi

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat batas *minimum* untuk *confidence* dengan mengitung nilai *confidence* dari aturan A -> B diperoleh rumus 2.3:

$$\boldsymbol{Confidence\ P(B|A)} = \frac{\sum Transksi\ mengandung\ A\ dan\ B}{\sum Transaksi\ mengandung\ A} \quad (2.3)$$

Pada tabel 2.2 merupakan *dataset* sebanyak 15 data dari alumni:

Tabel 2.2 Nilai Mahasiswa

No	NIM	1	2	3	4	X	X	56	57
1	113110001	A	B	B	B	X	X	A	A
2	113110002	A	B	A	A	X	X	A	B
3	113110003	A	B	B	C	X	X	B	B
4	113110004	B	C	B	C	X	X	B	A
5	113110005	B	C	B	C	X	X	B	B
6	123110001	A	C	A	B	X	X	A	B
7	123110005	B	B	B	C	X	X	A	C
8	123110006	A	B	B	C	X	X	A	C
9	123110008	A	B	B	A	X	X	A	B
10	123110009	A	A	A	A	X	X	A	A
11	133110003	A	C	A	C	X	X	A	B
12	133110005	A	A	A	B	X	X	A	A
13	133110015	A	B	A	C	X	X	A	B
14	133110018	A	A	A	C	X	X	A	A
15	133110020	A	B	A	B	X	X	A	B

Keterangan:

1-57 adalah perumpamaan untuk matakuliah

Langkah mengolah data nilai mahasiswa pada tabel 2.2:

1. Dari data diatas kemudian kita ambil nilai bobot dari setiap nilai matakuliah matakuliah sehingga seperti tabel 2.3:

Tabel 2.3 Nilai Bobot

No	NIM	1	2	3	4	X	X	56	57
1	113110001	4	3	3	3	X	X	4	4
2	113110002	4	3	4	4	X	X	4	3
3	113110003	4	3	3	2	X	X	3	3
4	113110004	3	2	3	2	X	X	3	4
5	113110005	3	2	3	2	X	X	3	3
6	123110001	4	2	4	3	X	X	4	3
7	123110005	3	3	3	2	X	X	4	2
8	123110006	4	3	3	2	X	X	4	2
9	123110008	4	3	3	4	X	X	4	3
10	123110009	4	4	4	4	X	X	4	4
11	133110003	4	2	4	2	X	X	4	3
12	133110005	4	4	4	3	X	X	4	4
13	133110015	4	3	4	2	X	X	4	3
14	133110018	4	4	4	2	X	X	4	4
15	133110020	4	3	4	3	X	X	4	3

2. Distribusikan matakuliah ke kompetensi seperti pada tabel 2.4:

Tabel 2.4 Distribusi Matakuliah ke Kompetensi

No	Kode	Nama Mata Kuliah	PEMBENTUKN KARAKTER	MATEMATIKA DAN STATISTIKA	ALGORITMA DAN PEMROGRAMAN	REKAYASA PERANGKAT LUNAK	ARSITEKTUR KOMPUTER	KECAKAPAN HIDUP	MANAGEMEN INFORMASI
1	AK1201T	Agama Islam	1						
2	AK2011T	Algoritma dan pemrograman			1				
3	AK1208T	Bahasa Inggris 1	1						
4	AK2101T	Matematika Dasar		1					
5	MI1101T	Pengantar Bisnis							
6	MI1001T	Pengantar Teknologi Informasi				1			

3. Hitung rata-rata setiap kompetensi berdasarkan matakuliah yang telah di distribusikan ke kompetensi sehingga hasilnya seperti pada tabel 2.5:

Tabel 2.5 Rata-Rata Nilai Bobot Per Kompetensi

No	NIM	PEMBENTUKN KARAKTER	MATEMATIKA DAN STATISTIKA	ALGORITMA DAN PEMROGRAMAN	REKAYASA PERANGKAT LUNAK	ARSITEKTUR KOMPUTER	KECAKAPAN HIDUP	MANAGEMEN INFORMASI
1	113110001	3,20	3,25	3,17	3,17	3,25	3,33	3,50
2	113110002	3,80	3,75	3,50	3,67	3,50	3,33	3,50
3	113110003	3,60	2,50	3,50	3,00	3,50	3,00	3,00
4	113110004	2,80	2,50	2,83	2,67	3,00	3,33	2,00
5	113110005	2,80	2,00	3,00	2,83	2,75	3,00	2,00
6	123110001	3,20	2,25	2,40	2,50	3,25	3,00	2,33
7	123110005	2,80	2,00	2,53	2,83	3,50	2,67	2,67
8	123110006	2,80	2,25	2,80	2,63	3,25	3,00	2,67
9	123110008	3,00	3,50	3,13	3,17	3,75	3,00	2,50
10	123110009	3,40	3,50	3,44	3,43	3,75	3,67	3,50
11	133110003	3,67	2,50	2,90	3,44	3,00	3,33	3,00
12	133110005	3,83	3,25	3,67	3,56	3,75	4,00	3,00
13	133110015	3,83	2,50	3,10	3,13	3,33	3,33	2,67
14	133110018	3,83	2,25	3,25	3,11	3,75	3,33	2,33
15	133110020	3,50	3,00	3,19	3,44	3,50	3,33	3,00

4. Ambil kompetensi yang memiliki rata-rata lebih dari sama dengan 3,00 seperti pada tabel 2.6:

Tabel 2.6 Pembineran data Rata-Rata

No	NIM	K1	K2	K3	K4	K5	K6	K7
1	113110001	1	1	1	1	1	1	1
2	113110002	1	1	1	1	1	1	1
3	113110003	1	0	1	1	1	1	1
4	113110004	0	0	0	0	1	1	0
5	113110005	0	0	1	0	0	1	0
6	123110001	1	0	0	0	1	1	0
7	123110005	1	1	1	1	1	1	1

Tabel 2.6 Lanjutan

No	NIM	K1	K2	K3	K4	K5	K6	K7
8	123110006	0	0	0	0	1	1	0
9	123110008	1	1	1	1	1	1	0
10	123110009	1	1	1	1	1	1	1
11	133110003	1	0	0	1	1	1	1
12	133110005	1	1	1	1	1	1	1
13	133110015	1	0	1	1	1	1	0
14	133110018	1	0	1	1	1	1	0
15	133110020	1	1	1	1	1	1	1

Keterangan:

K1-K7 merupakan istilah untuk nama kompetensi yang ada.

5. *Data itemset* yang di gunakan untuk tahap *apriori* seperti pada tabel 2.7 dan jumlah mahasiswa per kompetensi seperti pada tabel 2.8:

Tabel 2.7 *Data Itemset*

No	*ITEM SET*
1	K1,K2,K3,K4,K5,K6,K7
2	K1,K2,K3,K4,K5,K6,K7
3	K1,K3,K4,K5,K6,K7
4	K5,K6
5	K3,K6
6	K1,K5,K6
7	K1,K2,K3,K4,K5,K6,K7
8	K5,K6
9	K1,K2,K3,K4,K5,K6
10	K1,K2,K3,K4,K5,K6,K7
11	K1,K4,K5,K6,K7
12	K1,K2,K3,K4,K5,K6,K7
13	K1,K3,K4,K5,K6
14	K1,K3,K4,K5,K6
15	K1,K2,K3,K4,K5,K6,K7

Tabel 2.8 Jumlah Mahasiswa per kompetensi

KOMPENTENSI	*count* (mahasiswa)
K1	12
K2	7
K3	11
K4	11
K5	14
K6	15
K7	8

6. Menghitung *support* untuk *1-itemset* dengan *minimum support* = 75 seperti tabel 2.9:

Tabel 2.9 Menghitung *support 1-Itemset*

1-ITEMSET	Rumus *support*	Hasil *Support*
K1	(12/15)*100	80
K2	(7/15)*100	47
K3	(11/15)*100	73
K4	(11/15)*100	73
K5	(14/15)*100	93
K6	(15/15)*100	100
K7	(18/15)*100	53

Keterangan:

Bagian yang diberi warna merah merupakan nilai *support* yang memenuhi *minimum support* dan dapat dikombinasikan ke tahap selajutnya.

7. Menghitung *support* untuk *2-itemset minimum support* = 75 seperti pada tabel 2.10:

Tabel 2.10 Menghitung *support 2-Itemset*

2-ITEMSET		Rumus *Support*	Hasil *Support*
Comb 2	*Count*		
K1,K5	12	(12/15)*100	80
K1,K6	12	(12/15)*100	80
K5,K6	14	(14/15)*100	93

8. Menghitung *support* untuk *3-itemset minimum support* = 75 seperti pada tabel 2.11:

Tabel 2.11 Menghitung *support 3-Itemset*

3- itemset				Rumus *Support*	Hasil *Support*
comb 3			*Count*		
K1	K5	K6	12	(12/15)*100	80

9. Aturan yang terjadi sebagai berikut dengan *minimum confidence* = 90 seperti pada tabel 2.12:

Tabel 2.12 Menghitung *confidence*

ATURAN (X → Y)	SUP(X U Y)	SUP X	*CONFIDENCE*
K1,K5 -> K6	80	80	100
K1,K6 -> K5	80	80	100
K5,K6 -> K1	80	93	86
K1 -> K5	80	80	100
K5 -> K1	80	93	86
K1 -> K6	80	80	100
K6-> K1	80	100	80
K5 ->K6	93	93	100
K6 ->K5	93	100	93

Dari hasil tabel asosiai, maka dapat digenerasi suatu aturan asosiasi sebagai berikut :

Jika K1,K5 true, maka K6

Jika K1,K6 true, maka K5

Dimana K1=Pembentukan Karakter, K5=Rekayasa Perangkat Lunak, dan K6=Kecakapan Hidup. Ketiga kompetensi tersebut lah yang lebih menonjol dibandingkan kompetensi yang lain.