

BAB II

LANDASAN TEORI

2.1 *Knowledge Discovery In Database (KDD)*

Menurut (Fajhrin, 2018), *Knowledge Discovery In Database (KDD)* diartikan sebagai ekstraksi informasi potensial, implisit dan tidak dikenal dari sekumpulan data. Proses KDD melibatkan hasil proses dari *data mining* (Proses mengekstrak kecendrungan pola suatu data), kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami. KDD sendiri diartikan sebagai keseluruhan proses *non-trivial* untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dimengerti.

Menurut (Kennedi Tampubolon, hoga saragih, 2013), ada enam elemen yang paling esensial dalam teknik pencarian informasi atau pengetahuan dalam KDD yaitu :

1. Mengerjakan sejumlah data besar.
2. Diperlukan efisiensi berkaitan dengan volume data.
3. Mengutamakan ketetapan/keakuratan.
4. Membutuhkan pemakaian Bahasa tingkat tinggi.
5. Menggunakan beberapa bentuk dari pembelajaran otomatis.
6. Menghasilkan hasil yang menarik.

2.2 *Data Mining*

Menurut (Kennedi Tampubolon, hoga saragih, 2013) *data mining* merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai *database* besar/ *data warehouse*.

Menurut (Fajhrin, 2018), *data mining* bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan *data mining* adalah kenyataan bahwa *data mining* mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terebih dahulu. Berawal dari beberapa disiplin ilmu, *data mining* bertujuan untuk memperbaiki teknik tradisional sehingga bisa menangani :

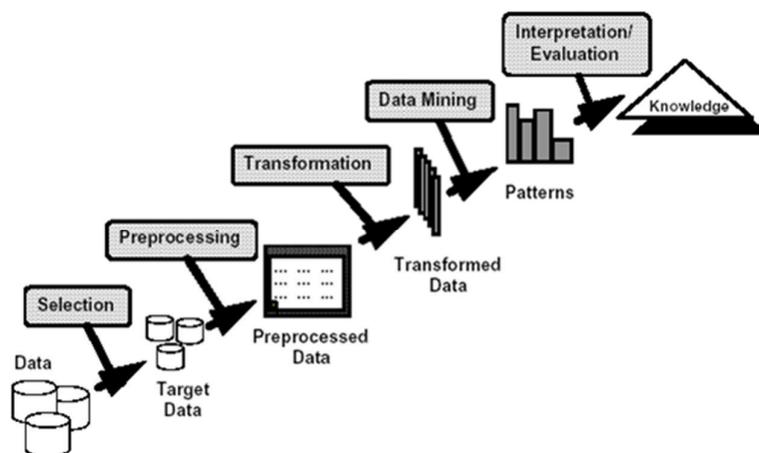
1. Jumlah data yang sangat besar.
2. Dimensi data yang tinggi.
3. Data yang heterogen dan berbeda sifat.

Menurut (Koko Handoko, 2019) *Data Mining* merupakan sebuah aktivitas yang dilakukan dengan menggabungkan berbagai ilmu statistika, matematika, *artificial intelligence* dan *machine learning* dengan tujuan untuk melakukan ekstraksi dan identifikasi dari berbagai informasi yang ada sehingga menghasilkan suatu pengetahuan baru yang bermanfaat pada hasil kedepannya pada sebuah *database* yang besar.

2.2.1 Proses Tahapan *Data Mining*

Menurut (Vulandari, 2017), *Data mining* merupakan salah satu dari rangkaian KDD. KDD berhubungan dengan teknik *integrasi* dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah data. Serangkaian proses tersebut memiliki tahapan sebagai berikut :

1. Pembersihan data (untuk membuang data yang tidak konsisten dan *noise*).
2. Integrasi data (penggabungan data dari beberapa sumber).
3. Transformasi data (data diubah menjadi bentuk yang sesuai untuk di-*mining*).
4. Aplikasi teknik *Data Mining*, proses ekstraksi pola dari data yang ada).
5. Evaluasi pola yang ditemukan (proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan).
6. Presentasi pengetahuan (dengan teknik visualisasi).



Gambar 2. 1 Tahapan *Knowledge Discovery In Database (KDD)*

Sumber: Data Mining Teori dan Aplikasi Rapidminer

2.2.2 Pengelompokan *Data Mining*

Menurut (Fajhrin, 2018), pengelompokan *data mining* dibagi menjadi beberapa kelompok yaitu:

1. Deskripsi

Deskripsi merupakan cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data yang dimiliki.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik dari pada kearah kategori. Model yang dibangun menggunakan record lengkap yang menyediakan nilai variabel target sebagai nilai prediksi.

3. Prediksi

Prediksi menerka sebuah nilai yang belum diketahui dan juga memperkirakan nilai untuk masa mendatang.

4. Klasifikasi

Dalam klasifikasi terdapat target variabel kategori, misal penggolongan pendapatan dapat dipisahkan dalam tiga kategori yaitu tinggi, sedang, dan rendah.

5. Pengklusteran

Merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.

6. Asosiasi

Asosiasi bertugas menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

2.2.3 Fungsi *Data Mining*

Menurut (Vulandari, 2017), fungsi - fungsi yang umum diterapkan dalam *data mining* (haskett, 2000) yaitu :

1. *Assosiation*, adalah proses untuk menemukan aturan asosiasi antara suatu kombinasi item dalam suatu waktu.
2. *Sequence*, proses untuk menemukan aturan asosiasi suatu kobinasi item dalam suatu waktu dan diterapkan lebih dari satu periode.
3. *Clustering*, adalah proses pengelompokan sejumlah data/objek kedalam kelompok data sehingga setiap kelompok berisi data yang mirip.
4. *Classification*, proses penemuan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan dari suatu objek yang labelnya tidak diketahui.
5. *Regression*, adalah proses pemetaan data dalam suatu nilai prediksi.
6. *Forecasting*, adalah proses pengestimasian nilai prediksi berdasarkan pola-pola didalam sekumpulan data.
7. *Solution*, adalah proses penemuan akar masalah dan *problem solving* dari persoalan bisnis yang dihadapi atau paling tidak sebagai informasi dalam pengambilan keputusan.

2.2.4 Kategori *Data Mining*

Menurut (Vulandari, 2017), *data mining* dibagi menjadi dua kategori utama (hand dan Kamber, 2006) yaitu:

1) Prediktif

Tujuan dari tugas prediktif adalah untuk memprediksi nilai dari atribut tertentu berdasarkan pada nilai atribut lain. Atribut yang diprediksi biasanya dikenal sebagai target atau variabel tak bebas, sedangkan atribut-atribut yang digunakan untuk membuat prediksi dikenal sebagai *explanatory* atau variabel bebas.

2) Deskriptif

Tujuan dari tugas deskriptif adalah untuk menurunkan pola-pola (korelasi, *trend*, *cluster*, teori, dan *anomaly*) yang meringkas hubungan kecocokan dalam suatu data. Tugas *data mining* deskriptif merupakan penyelidikan dan sering kali memerlukan teknik *post-processing* untuk validasi dan penjelasan hasil.

2.2.5 Manfaat *Data Mining*

Menurut (Vulandari, 2017), Pemanfaatan *data mining* dilihat dari dua sudut pandang, yaitu sudut pandang komersial dan sudut pandang keilmuan.

- a) Sudut pandang komersial, pemanfaatan *data mining* dapat digunakan untuk menangani meledaknya *volume* data, dengan menggunakan teknik komputasi dapat digunakan untuk menghasilkan informasi-informasi yang dibutuhkan yang merupakan *asset* yang dapat meningkatkan daya saing suatu institusi.

- b) Sudut pandang keilmuan, *data mining* dapat digunakan untuk *capture*, menganalisis serta menyimpan data yang bersifat *real time* dan sangat besar, misalnya :
1. *Remote sensor* yang ditempatkan pada suatu satelit.
 2. *Telescope* yang digunakan untuk memindai langit.
 3. Simulasi saintifik yang membangkitkan data dalam ukuran *terabytes*.

2.2.6 Penerapan *Data Mining*

Menurut (Vulandari, 2017), Penerapan *data mining* dibagi menjadi tiga bagian yaitu sebagai berikut :

1. Analisa pasar dan kumpulan sumber data yang digunakan seperti transaksi kartu kredit, kartu anggota *club* tertentu, kupon diskon, keluhan pembeli, ditambah dengan *study* dengan gaya hidup *public*. Beberapa solusi yang dapat diselesaikan dengan *data mining* antara lain (menebak target pasar, melihat pola beli pemakai dari waktu ke waktu, *cross market* analisis, *profil customer*, identifikasi kebutuhan *customer*, menilai loyalitas *customer*, informasi *summary*)
2. Analisa perusahaan dan *management* resiko
 - a) Perencanaan keuangan dan evaluasi *asset*, *data mining* dapat membantu melakukan, analisis dan prediksi *cash flow* serta dapat melakukan *contigent claim analysis* untuk mengevaluasi *asset*.
 - b) Perencanaan sumber daya, dengan melihat ringkasan informasi serta pola pembelanjaan dan pemasukan dari masing-masing *resource*.

- c) Persaingan, *data mining* dapat membantu untuk memonitor pesaing-pesaing dengan melihat *market direction* mereka.
3. Telekomunikasi, *data mining* melihat jutaan transaksi yang masuk dan melihat transaksi mana saja kah yang masih harus ditangani secara manual.

2.3 Metode *Data Mining*

Menurut (Robi yanto, 2015) Analisis asosiasi atau *association rule mining* adalah teknik *data mining* untuk menemukan aturan asosiasi antara suatu kombinasi *item interestingness measure* yang dapat digunakan dalam *data mining* adalah:

1. *Support*, adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu *item* atau *itemset* dari keseluruhan transaksi.
2. *Confidence*, adalah suatu ukuran yang menunjukkan hubungan antar dua *item* secara *conditional* (berdasarkan suatu kondisi tertentu).

2.3.1 *Algoritma Apriori*

Menurut (Robi yanto, 2015) *Algoritma apriori* adalah *algoritma* pengambilan data dengan aturan asosiatif (*Association rule*) untuk menentukan hubungan asosiatif suatu kombinasi item.

Menurut (Irsyad Djamaludin, 2017) *Algoritma Apriori* adalah salah satu *algoritma* yang melakukan pencarian *frequent itemset* dengan menggunakan teknik *association rule*. *Algoritma Apriori* menggunakan pengetahuan sebelumnya dari suatu *itemset* dengan frekuensi kemunculan yang sering atau juga yang bisa disebut frekuensi *itemset*.

2.3.2 Analisis Pola Frekuensi Tinggi dengan *Algoritma Apriori*

Mencari suatu nilai frekuensi *item* yang memenuhi syarat minimum dari nilai *support* dalam basis data. Nilai *support* sebuah *item* diperoleh dengan menggunakan rumus berikut:

$$\text{Support (A)} = \frac{\text{Jumlah transaksi mengandung A}}{\text{Total Transaksi}} \times 100\%$$

Rumus 2. 1 Frekuensi tinggi 1 *Itemset*

Nilai *support* dari kombinasi 2 *item* diperoleh dengan menggunakan rumus:

$$\text{Support (A,B)} = P(A \cap B)$$

$$\text{Support (A,B)} = \frac{\Sigma \text{transaksi mengandung A dan B}}{\Sigma \text{Transaksi}} \times 100\%$$

Rumus 2. 2 Menghitung kombinasi 2 *Itemset*

2.3.3 Pembentukan Aturan Asosiasi

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiatif A U B. Nilai *confidence* dari aturan A U B diperoleh dengan rumus berikut.

$$\text{Confidence} = P(B|A) = \frac{\Sigma \text{transaksi mengandung A dan B}}{\Sigma \text{Transaksi mengandung A}} \times 100\%$$

Rumus 2. 3 Menghitung nilai *Confidence*

Untuk menentukan aturan asosiasi yang akan dipilih maka harus diurutkan berdasarkan *Support* dan *Confidence*. Aturan diambil sebanyak jumlah aturan yang memiliki hasil terbesar.

2.4 *Software Pendukung*

2.4.1 *Software Tanagra*

Tanagra adalah perangkat lunak pengolahan data, sumber terbuka, ramah penggunaan dan mudah untuk pembacaan hasil keputusan yang dibuat bagi siswa dan peneliti hingga pembisnis untuk menambang jumlah data yang sangat banyak bagi mereka. *Software* ini menyediakan beberapa metode penambangan dari data analisis, data eksplorasi, pembelajaran statistik, pembelajaran mesin dan bidang basis data. *Software* ini diprogram dalam Bahasa *pascal*. *Software* ini memiliki berbagai komponen seperti visualisasi data, *statistics*, *clustering*, *association*, *analysis factorial*, *regression* dan banyak lagi. Dengan menggunakan *Tanagra*, peneliti dapat memvisualisasikan suatu data menjadi lebih baik. Visualisasi data yang ada disajikan dalam bentuk *dataset*, memplot nilai pada grafik dan sebar plot. Ini digunakan untuk menunjukkan hubungan antara atribut dalam sumbu 2D. *Tanagra* mencakup *algoritma* pengelompokan dasar seperti *K-Means*, *EM – clustering* dll. Dalam basis data (Jain Vishal, 2013).



Gambar 2. 2 *Tanagra Software*

Sumber: *Tanagra software logo*

2.4.2 *Microsoft Excel*

Aplikasi Microsoft excel 2013 adalah aplikasi pengolahan data yang menyajikan informasi dengan baik dan digunakan oleh berbagai kalangan

contohnya, dosen, mahasiswa, siswa, serta perusahaan-perusahaan besar untuk mempresentasikan hasil produknya. Dengan *Microsoft excel 2013* penyampaian informasi apapun akan mudah untuk dipahami. (Auliya Rahman, Fitrah Yuridka, 2015).

2.5 Penelitian Terdahulu

Penelitian terdahulu merupakan acuan dalam sebuah penelitian, ada beberapa penelitian terdahulu yang digunakan dalam penelitian ini sebagai berikut.

1. Menurut Alfannisa Annurullah fajrin dan Algifanri Maulana tahun 2018 yang berjudul **“PENERAPAN *DATA MINING* UNTUK ANALISIS POLA PEMBELIAN KONSUMEN DENGAN *ALGORITMA FP-GROWTH* PADA DATA TRANSAKSI PENJUALAN SPARE PART MOTOR”** Pada perusahaan yang memiliki banyak cabang atau dealer seperti CV.TJAJAJA BARU ini, membutuhkan lokasi yang harus diperhatikan dalam mendirikan sebuah cabang baru, harus dapat mempengaruhi pola pembelian konsumen, karena pola pembelian setiap konsumen berbeda-beda. Hal ini perlu dianalisis lebih mendalam sehingga dapat menghasilkan informasi yang bermanfaat, serta memaksimalkan keuntungan yang bisa diperoleh. *Data Mining* dapat digunakan oleh perusahaan besar untuk menggali data serta menemukan informasi yang dapat menunjang dan meningkatkan suatu proses keberlangsungan bisnis perusahaan tersebut. Maka dalam penelitian ini dilakukan pengujian menggunakan *algoritma FP-*

Growth untuk membantu perusahaan menemukan pola pembelian konsumen dan transaksi penjualan *spare part*.

2. Menurut Irsyad Djamaludin dan Agus Nursikuwagus tahun 2017 yang berjudul **“ANALISIS POLA PEMBELIAN KONSUMEN PADA TRANSAKSI PENJUALAN MENGGUNAKAN ALGORITMA APRIORI”** Penambangan data perlu dilakukan jika informasi disimpan dalam suatu penyimpanan. Aktivitas ini menjadi dasar utama untuk menentukan suatu keputusan. Aktivitas yang sudah berjalan menemukan informasi transaksi penjualan dan pembelian yang hanya tersimpan pada *repository* sebagai suatu laporan saja. Sehingga informasi hanya menjadi data histori yang tidak digunakan. Berdasarkan kejadian ini, maka diperlukan suatu analisis untuk mendukung suatu keputusan. Analisis ini adalah teknik *mining* dengan menggunakan *algoritma Apriori*. Tujuan dari penelitian ini adalah mendapatkan pola penjualan dan pembelian suatu produk. Untuk menjawab tujuan maka digunakan *algoritma Apriori* dengan aturan asosiasi. Penentuan *itemset* yang digunakan adalah berjumlah 1, 2, dan 3 *itemset* produk. Sedangkan untuk *Minimum Support* yang digunakan adalah 30% dari jumlah transaksi yang ada. Untuk asosiasi antar produk ditentukan dengan *confidence* sebesar 70%. Penelitian ini diselesaikan dengan menggunakan *UML* yang terdiri dari *Usecase Diagram*, *Class Diagram*, dan *Activity Diagram*. Hasil dari penelitian ini adalah perangkat lunak analisis pola penjualan dan pembelian konsumen dengan *Algoritma Apriori*. Sedangkan keputusan yang diperoleh adalah asosiasi antar produk dengan menggunakan nilai *confidence*.

Untuk *Final Rule Association* adalah produk bolu *cake meses* dan bolu *cake keju* dengan nilai *confidence* sebesar 84,62%.

3. Menurut Dewi Kartika Pane tahun 2013 yang berjudul **“IMPLEMENTASI DATA MINING PADA PENJUALAN PRODUK ELEKTRONIK DENGAN ALGORITMA APRIORI (STUDI KASUS : KREDITPLUS)”** Penjualan produk elektronik, khususnya laptop mengalami peningkatan setiap bulannya, produk yang ditawarkan bermacam merek, merek mempengaruhi masyarakat untuk membeli produk tersebut, untuk mengetahui merek dengan penjualan terbanyak diperlukan *algoritma apriori* untuk dapat mengetahuinya, dan dengan bantuan tools *Weka*, produk dengan penjualan terbanyak dapat diketahui. *Algoritma apriori* termasuk jenis aturan asosiasi pada *data mining*. Salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien adalah analisis pola frekuensi tinggi (*frequent pattern mining*). Penting tidaknya suatu asosiasi dapat diketahui dengan dua tolak ukur, yaitu : *support* dan *confidence*. *Support* (nilai penunjang) adalah persentase kombinasi *item* tersebut dalam database, sedangkan *confidence* (nilai kepastian) adalah kuatnya hubungan antar-*item* dalam aturan asosiasi. *Algoritma apriori* dapat membantu untuk pengembangan strategi pemasaran.
4. Menurut Robi Yanto dan Riri Khoiriah tahun 2015 yang berjudul **“IMPLEMENTASI DATA MINING DENGAN METODE ALGORITMA APRIORI DALAM MENENTUKAN POLA PEMBELIAN OBAT”** *Data mining* merupakan proses untuk mendapatkan informasi yang berguna dari

gudang basis data yang berupa ilmu pengetahuan. penelitian ini melakukan analisa data dengan menggunakan *data mining* dan metode *algoritma apriori*. Sistem yang dibangun ditujukan untuk pemenuhan dalam penentuan pola pembelian obat dengan menggunakan bahasa pemrograman *Visual Basic 6.0* dan *database Mysql* pada studi kasus di sektor kesehatan. Sistem ini dibangun berdasarkan kebutuhan pengguna yang diperoleh melalui metode wawancara dan studi lapangan. Metodologi pengembangan sistem yang digunakan yaitu metode *waterfall* yang terdiri analisis, desain, pengkodean dan pengujian. Hasil pengujian dengan *algoritma apriori* dan sistem yang dibangun menunjukkan hasil yang telah memenuhi kebutuhan dalam penentuan pola pembelian obat berdasarkan kecenderungan pembelian obat oleh pelanggan. Dibandingkan dengan sistem yang sedang berjalan kinerja tersebut ditunjukkan pada efektifitas informasi dari sistem tentang penentuan pola pembelian obat untuk ketersediaan obat dan tata letak obat untuk memudahkan dalam mengetahui keberadaan obat yang dilihat dari 2 *itemset* obat.

5. Menurut Goldie Gunadi Dana Indra Sensuse tahun 2012 yang berjudul **“PENERAPAN METODE *DATA MINING* MARKET BASKET ANALYSIS TERHADAP DATA PENJUALAN PRODUK BUKU DENGAN MENGGUNAKAN *ALGORITMA APRIORI* DAN *FREQUENT PATTERN GROWTH (FP-GROWTH)* : STUDI KASUS PERCETAKAN *PT. GRAMEDIA*”** *Data mining* merupakan proses analisa data untuk menemukan suatu pola dari kumpulan data tersebut. *Data mining*

mampu menganalisa data yang besar menjadi informasi berupa pola yang mempunyai arti bagi pendukung keputusan. Salah satu teknik *data mining* yang dapat digunakan adalah *association data mining* atau yang biasa disebut dengan istilah *market basket analysis*. *Market basket* didefinisikan sebagai suatu *itemset* yang dibeli secara bersamaan oleh pelanggan dalam suatu transaksi. *Market basket analysis* adalah suatu alat yang ampuh untuk pelaksanaan strategi *cross-selling*. Metode ini dimulai dengan mencari sejumlah *frequent itemset* dan dilanjutkan dengan pembentukan aturan-aturan asosiasi (*association rules*). *Algoritma Apriori* dan *frequent pattern growth (FP-growth)* adalah dua algoritma yang sangat populer untuk menemukan sejumlah *frequent itemset* dari data-data transaksi yang tersimpan dalam basis data. Dalam penelitian ini *algoritma Apriori* dan *frequent pattern growth (FP-growth)* digunakan untuk membantu menemukan sejumlah aturan asosiasi dari basis data transaksi penjualan produk buku di Percetakan *PT. Gramedia*, sehingga untuk selanjutnya dapat digunakan sebagai pertimbangan dalam membuat strategi pemasaran dan penjualan yang efektif.

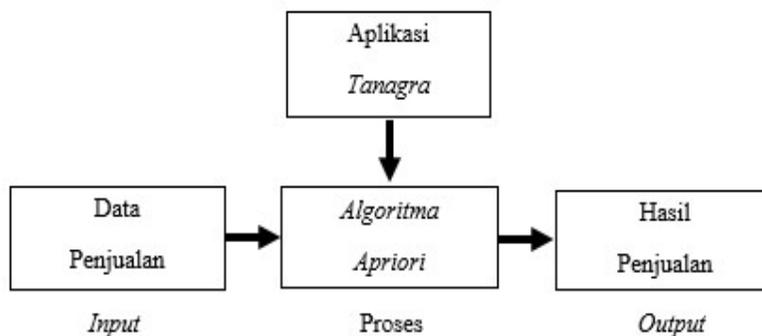
6. Menurut Jiao Yabing tahun 2013 yang berjudul ***“Research of an Improved Apriori Algorithm in Data Mining Association Rules”*** *Algoritma Apriori* adalah algoritma dari aturan asosiasi, yang menyebutkan semua set *item* yang sering muncul. Saat ini *algoritma apriori* mengalami data yang padat karena besar sejumlah pola panjang muncul, kinerja *algoritma* ini menurun drastis. Untuk menemukan aturan yang lebih baik, penulis mengusulkan adanya peningkatan *algoritma* aturan asosiasi, kepada *algoritma Apriori* klasik yang

pada akhirnya, algoritma yang ditingkatkan akan diverifikasi, kemudian hasilnya akan menunjukkan bahwa peningkatan *algoritma apriori* lebih efektif, dapat mengekstraksi lebih banyak informasi nilai maupun data.

7. Menurut Wiwit Agus Triyanto tahun 2014 yang berjudul “**ASSOCIATION RULE MINING UNTUK PENENTUAN REKOMENDASI PROMOSI PRODUK**” Rekomendasi Promosi Produk merupakan model aplikasi dari hasil observasi terhadap keadaan dan keinginan pelanggan untuk membeli suatu produk. *Association Rule Mining* dapat memberikan dukungan keputusan antara barang yang dibeli secara bersama-sama oleh pelanggan. Penemuan pola barang yang dibeli oleh pelanggan sangat penting dikarenakan dapat membantu rekomendasi promosi produk sehingga strategi pemasaran menjadi lebih tepat sasaran. Penelitian ini menggunakan *Algoritma FP-Growth* untuk pendekatan asosiasi. Hasil dari penelitian ini adalah *rule* yang dapat digunakan untuk membantu menentukan rekomendasi promosi suatu produk secara lebih tepat, yaitu dengan menghasilkan 3 *rule* ketika menggunakan *minimum support* 40% dan *minimum confidence* 80%.

2.6 Kerangka Pemikiran

Konsep yang dikembangkan dalam penelitian ini sebagai berikut :



Gambar 2. 3 Kerangka Pemikiran

Sumber: Data Penelitian (2019)

Data penjualan yang bersumber dari data transaksi penjualan rumah yang digunakan sebagai *input*, kemudian data dikelola dengan menggunakan metode *algoritma apriori* sehingga mendapat hasil atau *output* yang digunakan sebagai pengembangan dan peningkatan pembangunan rumah. Data penjualan yang bersumber data transaksi mencakup tiga variabel yaitu tipe rumah, luas tanah, harga jual, dan hasil akhir penjualan yang dikumpulkan dalam waktu tiga tahun terakhir oktober 2016 sampai dengan september 2019 dengan data yang diambil dari data perbulannya selama tiga puluh enam bulan. Algoritma apriori yang digunakan akan membantu kita menambang suatu data yang banyak sehingga menghasilkan aturan dan pengetahuan baru.