

## **BAB II**

### **TINJAUAN PUSTAKA DAN DASAR TEORI**

#### **2.1. Tinjauan Pustaka**

Pada tinjauan pustaka ini, penulis mendeskripsikan perkembangan penelitian yang penulis jadikan referensi. Tentang masalah yang diangkat dan yang belum terpecahkan dalam penelitian sebelumnya serta meyakinkan kelayakan penelitian ini. Beberapa penelitian menerapkan *association rules mining* dalam menganalisis pola pembelian konsumen menggunakan algoritma Apriori adalah sebagai berikut.

Survei yang berkaitan tentang *association rules mining* dalam *market basket analysis data mining* mengacu pada penggalian pengetahuan dari sejumlah besar data. *Association rules mining* mengidentifikasi hubungan antara sejumlah besar item data. Diamati dari analisis, bahwa *data mining tool* dapat secara efektif digunakan untuk mengoptimalkan pola yang terkait dengan perilaku dinamis dari transaksi yang dilakukan oleh pelanggan dalam membeli beberapa produk tertentu dan dapat membantu dalam desain katalog, *cross-marketing* dan berbagai proses pengambilan keputusan bisnis. Penulis telah menggunakan algoritma analisis keranjang belanja, suatu algoritma yang digunakan secara luas dan lebih dominan dari *association rules mining*. Dengan menggunakan algoritma ini, transaksi yang sering dilakukan oleh pelanggan telah dianalisis menggunakan *support* dan *confidence* pelanggan dalam membeli barang terkait. Dengan menggunakan metodologi ini terlihat bahwa ada hubungan tertentu antara produk pada saat pembelian produk oleh pelanggan (Gupta dan Mamtora, 2014).

Analisis penyebab kecelakaan lalu lintas berdasarkan menggunakan *Analytic Hierarchy Process (AHP)-Apriori*. Sifat multivariat, berlapis analisis lalu lintas kecelakaan menentukan bahwa metode analisis lalu lintas kecelakaan harus dikategorikan menurut faktor-faktor penyebab dan mencakup semua empat elemen lalu lintas dasar, yaitu pengemudi, kendaraan, jalan, dan lingkungan. AHP digunakan untuk mengurutkan faktor-faktor penyebab kecelakaan sesuai dengan kepentingan relatifnya, dengan tujuan memilih faktor-faktor yang paling

berpengaruh. Algoritma Apriori kemudian diterapkan untuk menganalisis tingkat kecelakaan atau tingkat pengaruh. Melalui aplikasi, metode ini membuktikan dirinya mampu menentukan jenis dan tingkat keparahan kecelakaan yang disebabkan oleh beberapa faktor. Metode ini juga menunjukkan peningkatan dalam akurasi dan efisiensi perhitungan dalam menangani sampel data berukuran besar (Xi dkk., 2016).

Evaluasi pengenalan kondisi kerja flotasi berbasis algoritma Apriori. Berdasarkan visi mesin dan teknologi pemrosesan gambar, sejumlah besar metode telah diusulkan untuk pengenalan kondisi kerja flotasi di industri flotasi buih. Namun, dalam metode pengenalan kondisi ini, tidak ada kriteria evaluasi internal atau eksternal, sehingga keandalan hasil pengakuan yang diperoleh oleh model pengakuan saat ini tidak tinggi untuk operator. Untuk mengatasi masalah ini, dengan mengambil deret waktu dari data kondisi kerja sebagai objek penelitian, algoritma Apriori yang ditingkatkan berdasarkan deret waktu diusulkan untuk evaluasi pengenalan kondisi kerja flotasi. Kemudian algoritma Apriori yang ditingkatkan digunakan untuk menemukan aturan asosiasi yang berisi hubungan antara rangkaian waktu dari database transaksi kondisi kerja untuk membentuk basis aturan untuk kondisi kerja, yang menyediakan evaluasi keandalan hasil pengakuan kondisi kerja *online* berdasarkan aturan asosiasi (Wang dkk., 2018).

Analisis kegagalan pola proses produksi asam sulfat dengan algoritma aturan asosiasi Apriori. PT. Asam Sulfat adalah industri asam sulfat yang mampu memenuhi kebutuhan sebagian besar produsen baterai di Indonesia. Masalah dalam operasi produksi sebagian besar disebabkan oleh peralatan produksi, seperti mesin berhenti tiba-tiba, lama penyetulan dan waktu penyesuaian dan penurunan kecepatan produksi mesin. Hasil ini menyebabkan penurunan tingkat efisiensi dan efektivitas mesin. Penelitian ini bertujuan untuk mengidentifikasi dan menganalisis mesin dan komponen yang mendominasi kerusakan proses produksi asam sulfat. Metode analisis yang digunakan adalah asosiasi aturan algoritma Apriori. Berdasarkan hasil identifikasi awal diperoleh 23 mesin dan 9 komponen sering rusak. Variabel yang dianalisis meliputi unit, tipe mesin, komponen, waktu perbaikan dan kekeruhan. Menggunakan nilai minimum *Support* 0,1 dan

*Confidence* 0,9 menghasilkan 17 aturan yang didominasi oleh kekeruhan rendah, terutama pada mesin AT Pump dan komponen Gasket (Septiani dkk., 2019).

Berdasarkan penelitian sebelumnya, tujuan dari penelitian ini adalah menerapkan *association rules mining* dalam menganalisis pola pembelian konsumen menggunakan algoritma Apriori, sehingga hasil efisiensi dari pola pembelian dapat digunakan untuk memilih pola pembelian konsumen yang dapat diimplementasi. Selain itu sistem juga diharapkan dapat membantu manajemen untuk memahami kebutuhan pembeli sesuai dengan kebutuhan dan pola pembelian pelanggan.

## **2.1. Dasar Teori**

### **2.1.1. Data Mining**

*Data mining* adalah proses bisnis intelijen yang dapat digunakan untuk mencari informasi yang berguna dan untuk mendukung proses pengambilan keputusan di perusahaan (Lam dkk., 2014). Namun pada kenyataannya kegiatan *data mining* tidak melaporkan pembuatan dan permintaan sama sekali. *Data mining* dilakukan dengan alat khusus, yang menjalankan operasi data yang telah ditentukan berdasarkan model analisis. *Data mining* adalah ekstraksi informasi atau pola yang penting atau menarik dari data yang berada di basis data besar yang sebelumnya tidak diketahui dan menghasilkan informasi yang berguna (Setiawan dkk., 2017).

#### **1. Arsitektur Data Mining**

Arsitektur *data mining* terdiri dari 4 bagian yaitu:

##### **a. Knowledge base**

*Knowledge base* berfungsi untuk melakukan evaluasi terhadap pola yang dihasilkan dari proses *data mining*. *Knowledge base* ini meliputi hirarki konsep yang digunakan untuk mengatur atribut atau nilai atribut ke dalam level abstraksi yang berbeda. *Knowledge base* ini juga dapat berupa pengetahuan pengguna, yang dapat digunakan untuk mengambil hal-hal penting dari pola yang diperoleh. Selain itu *knowledge base* juga dapat membantu menentukan nilai ambang yang tepat untuk proses *data mining*.

**b. Data mining engine**

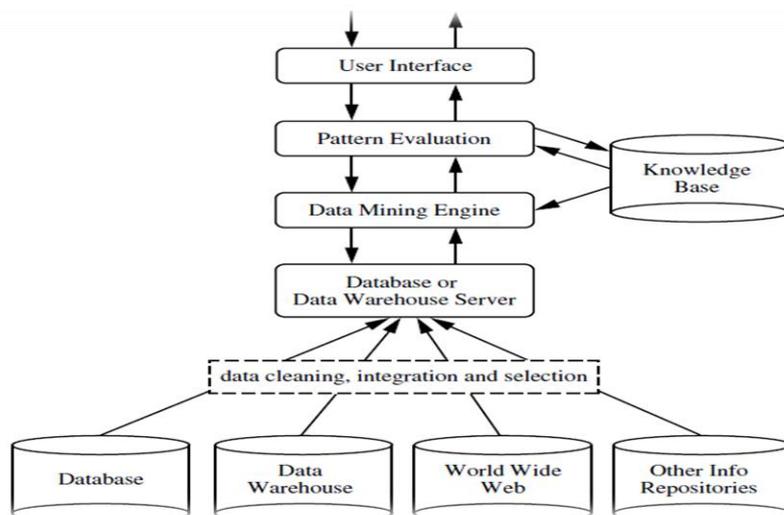
Bagian ini merupakan komponen penting dalam arsitektur sistem *data mining*. Komponen ini terdiri modul-modul fungsional *data mining* seperti karakterisasi, asosiasi, klasifikasi, dan analisis kelas.

**c. Pattern evaluation module**

*Pattern evaluation module* menggunakan ukuran-ukuran kemenarikan dan berinteraksi dengan modul *data mining* dalam pencarian pola-pola menarik. *Pattern evaluation module* dapat menggunakan nilai ambang untuk menyaring pola-pola yang didapat. Untuk melakukan *data mining* yang efisien sangat direkomendasikan untuk melakukan evaluasi pola dengan sangat mendalam ke dalam proses *data mining* sehingga membatasi pencarian hanya pada pola yang sangat menarik saja.

**d. User interface**

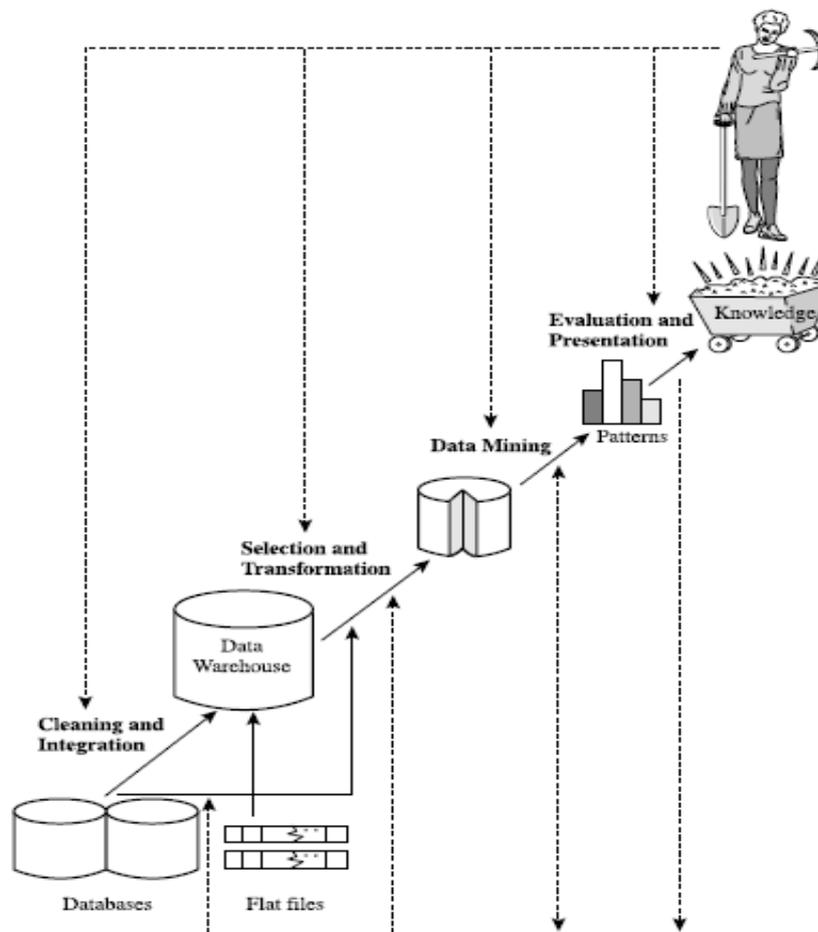
*User interface* berinteraksi dengan pengguna dan sistem *data mining*. Melalui *user interface*, pengguna berinteraksi dengan sistem dengan menentukan *query* atau *task data mining*. *User interface* juga menyediakan informasi untuk memfokuskan pencarian dan melakukan eksplorasi *data mining*. Selain itu *user interface* memungkinkan pengguna mencari basis data dan skema *data warehouse* atau struktur data, evaluasi pola yang diperoleh, dan visualisasi pola dalam beragam bentuk (Han dan Kamber, 2006). Arsitektur *data mining* dapat dilihat pada Gambar 2.1.



Gambar 2.1 Arsitektur *data mining* (Han dan Kamber, 2006)

## 2. Tahap-tahap *Data Mining*

Tahap-tahap pembuatan aplikasi *data mining* bersifat interaktif di mana pemakai terlibat langsung atau dengan perantara *knowledge base*. Tahap-tahap pembuatan aplikasi *data mining* ditunjukkan pada Gambar 2.2. Tahap-tahapnya adalah sebagai berikut:



Gambar 2.2 Proses *data mining* (Han dan Kamber, 2006)

### a. Pembersihan data dan integrasi data

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data yang tidak relevan. Integrasi data merupakan penggabungan data dari berbagai basis data ke dalam satu basis data baru.

Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik.

**b. Seleksi data dan transformasi data**

Data yang ada di dalam basis data sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari basis data. Transformasi data merupakan proses mengubah atau menggabungkan data ke dalam format yang sesuai untuk diproses dalam *data mining*. Beberapa teknik *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan.

**c. Aplikasi teknik *data mining***

Aplikasi teknik *data mining* merupakan suatu proses utama di mana metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Proses *mining* dapat menggunakan *metode predictive modelling, association analysis, clustering, dan anomaly detection*.

**d. Evaluasi pola dan presentasi pengetahuan**

Evaluasi pola mengidentifikasi pola-pola menarik untuk direpresentasikan ke dalam *knowledge base* yang ditemukan. Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesis yang ada memang tercapai. Presentasi pengetahuan merupakan visualisasi dan penyajian pengetahuan mengenai teknik 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 (Han dan Kamber, 2006).

**2.1.2. Market Basket Analysis**

*Market basket analysis* adalah asosiasi dalam *data mining* untuk menemukan atribut yang muncul dalam satu waktu transaksi. Proses ini dapat menentukan pola pembelian pelanggan dengan menemukan hubungan antara berbagai item dalam transaksi penjualan. Hasil yang telah diperoleh dapat digunakan oleh perusahaan ritel seperti minimarket dalam mengembangkan strategi pemasaran untuk melihat

barang-barang yang dapat dibeli secara bersamaan oleh konsumen (Setiawan dkk., 2017).

*Market basket analysis* berfokus untuk menemukan pola pembelian bersama dengan ribuan atau jutaan transaksi. *Association rules* memainkan peran penting untuk menemukan *itemset* yang sering dan hubungan dengan pembelian antara produk yang berbeda, di mana deklarasi eksplisit dari tipe tersebut dibuat adalah jika item X dibeli, maka item Y juga dibeli (Valle dkk., 2018).

*Market basket analysis* memiliki tujuan yaitu untuk produk-produk, yang mungkin akan dibeli secara bersamaan dalam transaksi pembelian. Pengetahuan yang diperoleh dari *market basket analysis* bisa sangat berharga karena, dapat digunakan oleh minimarket untuk mengatur ulang tata letaknya, mengambil produk yang sering dijual bersama dan menempatkannya dalam jarak yang dekat. Tetapi juga dapat digunakan untuk meningkatkan efisiensi kampanye promosi (Trnka, 2010). Produk yang terkait tidak boleh dipromosikan pada saat yang bersamaan. Dengan hanya mempromosikan salah satu produk terkait, memungkinkan minimarket untuk meningkatkan penjualan produk tersebut.

Dan sebagai tambahannya konsumen juga akan merasakan kenyamanan karena barang yang akan dibeli bersamaan ditempatkan berdekatan, yang akan meningkatkan nilai kepuasan pembeli. Agar *association rules* yang dihasilkan diakui sebagai valid, nilai *Support* dan *Confidence* harus sama atau lebih besar dari minimum *Support* dan minimum *Confidence* masing-masing (Lam dkk., 2014).

Meskipun *market basket analysis* memiliki banyak manfaat, namun ada beberapa masalah yang terkait dengan *market basket analysis* yang perlu ditangani. Akibatnya, pola pembelian konsumen akan berubah sesuai dengan waktu. Dengan demikian muncul kebutuhan untuk *market basket analysis* untuk dilakukan berulang kali pada seluruh basis data yang berisi transaksi pelanggan yang berbeda. Jika basis data berukuran besar, maka itu akan membutuhkan lebih banyak waktu yang lebih lama (Bhargav dkk., 2014).

### 2.1.3. Association Rules

*Association rules* adalah teknik *data mining* yang berguna untuk menemukan korelasi atau pola paling penting atau menarik dari kumpulan data besar (Septiani dkk., 2019). *Association rules mining* adalah teknik yang populer untuk analisis keranjang pasar, yang biasanya bertujuan menemukan pola pembelian untuk supermarket, pesanan melalui pos, dan pelanggan lainnya. Dengan *association rules mining*, analisis pemasaran mencoba menemukan *set* produk yang sering dibeli bersama, sehingga barang-barang tertentu lainnya dapat disimpulkan dari keranjang belanja yang berisi barang-barang tertentu. *Association rules* sering kali dapat digunakan untuk merancang promosi pemasaran misalnya, dengan mengatur produk di rak minimarket secara tepat dan dengan secara langsung menyarankan kepada pelanggan item-item yang mungkin menarik. Secara umum, *Support* dan *Confidence* adalah 2 langkah untuk mengevaluasi ketertarikan *association rules mining*. *Association rules mining* dianggap menarik jika *Support* dan *Confidence* masing-masing lebih besar daripada minimum *Support* dan minimum *Confidence* yang ditentukan pengguna. Metodologi dasar *association rules mining* terbagi menjadi 2 tahap yaitu sebagai berikut (Putra dan Suryani, 2018):

#### 1. Analisis Pola Frekuensi Tertinggi

Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai *Support* sebuah item. *Support* merupakan presentase dari *record-record* yang mengandung kombinasi dari item dibandingkan dengan jumlah total dari *record-record*. Dengan kata lain *Support* merupakan probabilitas sebuah transaksi yang mengandung item A dan B dibagi dengan keseluruhan transaksi. Persentase transaksi yang mengandung semua item dalam set item (misalkan pensil dan penghapus). Semakin tinggi *Support*, semakin sering terjadi *itemset*. Aturan dengan *Support* tinggi lebih disukai karena transaksi cenderung berlaku untuk sejumlah besar transaksi di masa depan (Gangurde dkk., 2017). Jadi perhitungan *Support* dapat dilakukan dengan persamaan 2.1 dan 2.2 sebagai berikut:

$$Support (A) = \frac{\sum \text{Transaksi yang mengandung A}}{\sum \text{Transaksi}} \quad (2.1)$$

Sedangkan nilai *support* dari 2 *itemset* diperoleh dari rumus berikut:

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

## 2. Pembentukan aturan asosiasi

Setelah pola frekuensi tertinggi ditemukan, selanjutnya mencari aturan asosiasi yang memenuhi syarat dari minimum *Confidence*. *Support* merupakan proses pengukuran kualitas *association rules* yang belum komplit. Akan timbul pertanyaan apakah nilai *support* 20% dari sebuah kombinasi potato chips dan susu merupakan *rules* yang baik?. Hal ini mungkin mempunyai arti bahwa 20% dari semua pelanggan membeli kedua *potato chips* dan susu, dan tidak ada satupun yang membeli susu tanpa membeli potato chips. Semakin tinggi *Confidence*, semakin besar kemungkinan bahwa barang yang berada pada sisi kanan akan dibeli (Gangurde dkk., 2017). *Confidence* dari sebuah *association rules* adalah keseluruhan transaksi yang mengandung A dan B dibagi dengan *support* atau keseluruhan transaksi yang mengandung A. *Confidence* ukuran keakuratan *association rules*, yang merupakan penyajian transaksi dalam *database* yang mengandung A dan mengandung B. Perhitungan *Confidence* dapat dilakukan dengan persamaan berikut 2.3.

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

## 3. Lift Ratio

Selain 2 parameter pada persamaan 2.1 dan 2.3 yang merupakan nilai *Support* dan nilai *Confidence*, ada parameter lain yang dikenal sebagai *Lift Ratio*. *Lift Ratio* menunjukkan kekuatan yang dimiliki oleh aturan atas aturan asosiasi yang dihasilkan (Septiani dkk., 2019). Secara keseluruhan, *Lift Ratio* merangkum kekuatan asosiasi antara produk. Semakin besar nilai *Lift Ratio*, semakin besar hubungan antara kedua produk (Gangurde dkk., 2017). *Lift Ratio* dapat dilakukan perhitungan di mana hasil dari *Confidence* dibagi dengan hasil dari *Support* transaksi yang mengandung B. Rumus dari *Lift Ratio* dapat dilihat pada persamaan 2.4.

$$Lift\ Ratio = \frac{Confidence\ (A\ dan\ B)}{Support\ (A)} \quad (2.4)$$

### 2.1.5. Algoritma Apriori

Algoritma Apriori adalah algoritma *data mining* mapan yang diperkenalkan oleh Agrawal & Srikant pada tahun 1994 untuk menambang *itemset* yang sering dalam basis data transaksional (Khalili, 2015). Selain Apriori, yang termasuk dalam kelompok ini adalah metode dan algoritma *hash berbasis induksi generalized rule*. Analisis asosiasi adalah teknik *data mining* untuk menemukan aturan asosiasi antara kombinasi item. Contoh aturan asosiasi dari analisis pembelian di supermarket mampu mengetahui seberapa besar kemungkinan seseorang untuk membeli roti bersama dengan susu. Berdasarkan pengetahuan ini, pemilik dapat mengatur sendiri penempatan barang atau merancang kampanye pemasaran dengan menggunakan kombinasi kupon diskon untuk barang-barang tertentu (Setiawan dkk., 2017).

Apriori adalah algoritma klasik untuk menambang *frequent itemset*, yang menggunakan metode berulang dengan pengetahuan sebelumnya tentang *frequent itemset* untuk mencari kandidat *itemset* lapis demi lapis. Langkah pertama algoritma apriori yaitu *Support* dari setiap item dihitung dengan melakukan *scan database*. Setelah *Support* dari setiap item didapat, item yang memiliki *Support* lebih besar dari *minimum Support* dipilih sebagai pola frekuensi tinggi dengan panjang 1 atau sering disingkat *1-itemset*. Singkatan *k-itemset* berarti 1 *set* yang terdiri dari k item. Iterasi kedua menghasilkan *2-itemset* yang tiap set-nya memiliki dua item. Pertama dibuat kandidat *2-itemset* dari kombinasi semua *1-itemset*. Lalu untuk tiap kandidat *2-itemset* ini dihitung *Support*-nya dengan men-*scan database*. *Support* artinya jumlah transaksi dalam *database* yang mengandung kedua item dalam kandidat *2-itemset*. Setelah *Support* dari semua kandidat *2-itemset* didapatkan, kandidat *2-itemset* yang memenuhi syarat *minimum Support* dapat ditetapkan sebagai *2-itemset* yang juga merupakan pola frekuensi tinggi dengan panjang 2 (Guo dkk., 2017). Untuk selanjutnya iterasi ke-k dapat dibagi lagi menjadi beberapa bagian:

#### **A. Pembentukan kandidat *itemset***

Kandidat *k-itemset* dibentuk dari kombinasi (*k-1*) *itemset* yang didapat dari iterasi sebelumnya. Satu ciri dari algoritma Apriori adalah adanya pemangkasan kandidat *k-itemset* yang *subset*-nya yang berisi *k-1* item tidak termasuk dalam pola frekuensi tinggi dengan panjang *k-1*.

#### **B. Penghitungan *support* dari tiap kandidat *k-itemset***

*Support* dari tiap kandidat *k-itemset* didapat dengan men-*scan database* untuk menghitung jumlah transaksi yang memuat semua item di dalam kandidat *k-itemset* tersebut. Ini adalah juga ciri dari algoritma apriori yaitu diperlukan penghitungan dengan *scan* seluruh *database* sebanyak *k-itemset* terpanjang.

#### **C. Tetapkan pola frekuensi tinggi**

Pola frekuensi tinggi yang memuat *k* item atau *k-itemset* ditetapkan dari kandidat *k-itemset* yang *Support*-nya lebih besar dari *minimum Support*. Kemudian dihitung *Confidence* masing-masing kombinasi item. Iterasi berhenti ketika semua item telah dihitung sampai tidak ada kombinasi item lagi.

Gambar 2.3 merupakan *flowchart* analisis *data mining* dengan algoritma apriori. Penjelasan *flowchart* tahapan dari analisis pola pembelian konsumen menggunakan algoritma Apriori dijelaskan sebagai berikut:

1. *Input Itemset, Input Transaksi, Input Fk dan Input Maks\_k\_itemset.*

Pada tahapan ini untuk memasukan data yang dibutuhkan. *Itemset* merupakan himpunan *item* atau produk yang berada dalam himpunan *item* yang dijual atau produk yang dijual. *Fk* adalah nilai ambang batas dalam proses atau nilai minimum transaksi. Sedangkan *maks\_k\_itemset* adalah maksimal barang yang dikombinasikan dalam analisis. Dalam penelitian ini penulis menggunakan  $maks\_k\_itemset = 2$ .

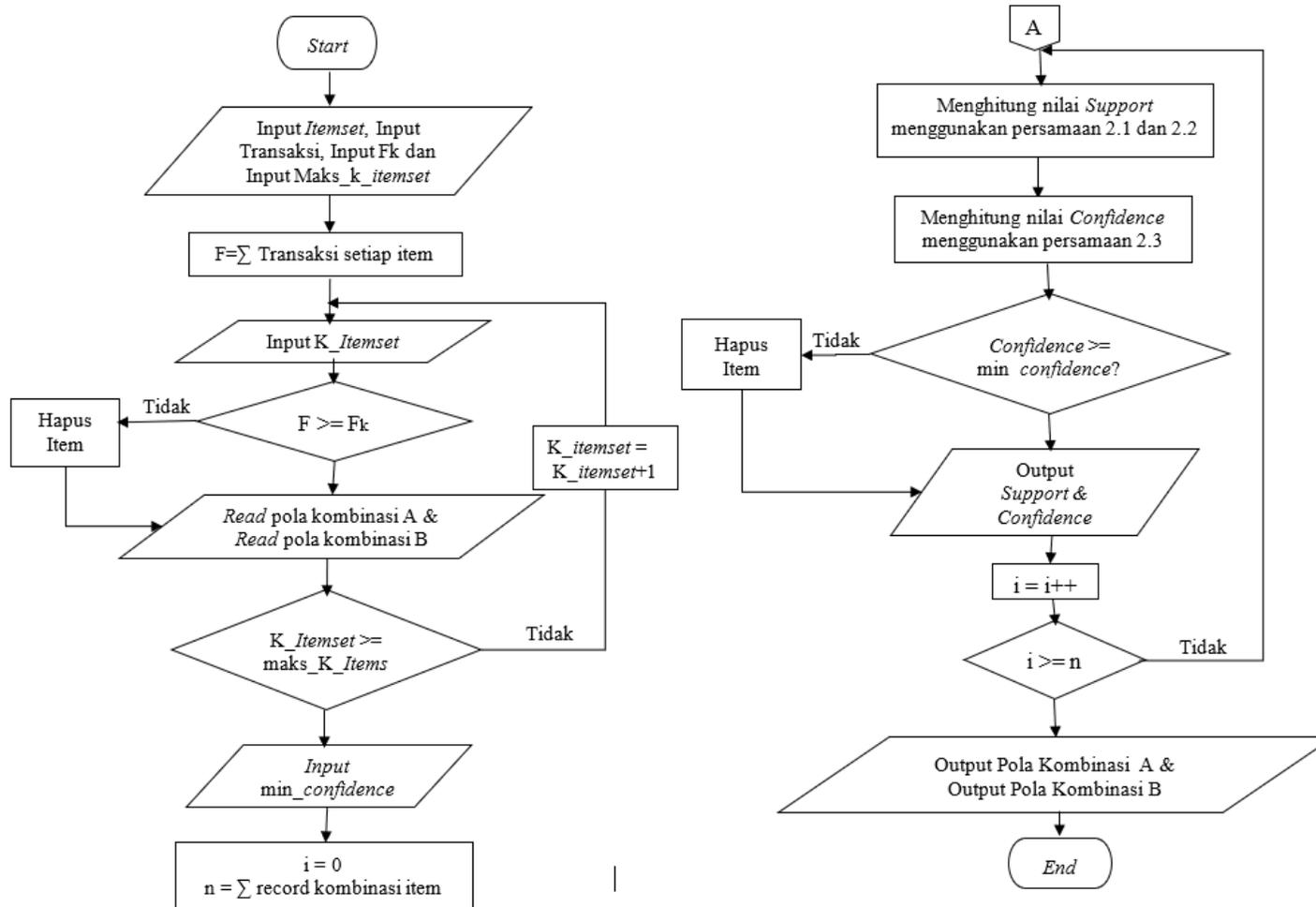
2.  $F = \sum$  Transaksi setiap item

Dari setiap item atau produk, dihitung kemunculan jumlah transaksinya dari seluruh transaksi yang dilakukan.

3. *Input K\_Itemset*

*K\_itemset* yang pertama adalah dengan *1\_itemset* di mana setiap *1\_itemset* memiliki jumlah transaksi atau frekuensi.

4.  $F \geq \text{Nilai Fk}$   
 Apabila frekuensi dari item di bawah ambang batas, maka item dihapus.
5. *Read* pola kombinasi A & *Read* pola kombinasi B  
 Menghasilkan kombinasi item di atas ambang batas frekuensi.
6.  $K\_Itemset \geq \text{maks\_K\_Itemset?}$   
 Apabila  $k\_itemset$  belum maksimal maka  $k\_itemset$  ditambah 1 sampai  $k\_itemset$  sudah pada  $\text{maks\_k\_itemset}$ .
7. *Input min\_confidence*  
 Memasukan minimum *confidence* untuk menentukan nilai ambang batas kuatnya hubungan antar item.
8.  $i = 0 \ \& \ n = \sum \text{record}$  kombinasi item  
 Proses untuk memeriksa seluruh *record* dalam *database* diperiksa.
9. Nilai *Support* didapat dari jumlah transaksi dari kombinasi item dibagi dengan seluruh total transaksi.
10. Nilai *Confidence* didapat dari jumlah transaksi kombinasi item dibagi dengan jumlah transaksi kondisi item.
11.  $\text{Confidence} \geq \text{minimum\_confidence?}$   
 Apabila *confidence* di bawah ambang batas maka kombinasi item dihapus dan akan menghasilkan nilai *Support* dan *Confidence*.
12. Bila seluruh kombinasi item sudah menghasilkan nilai *Support* dan *Confidence* maka akan menghasilkan pola kombinasi item.



Gambar 2.3 Flowchart analisis data mining dengan algoritma Apriori