

**IMPLEMENTASI *DATA MINING* PENJUALAN
KOSMETIK DENGAN ALGORITMA APRIORI**

SKRIPSI



**Oleh:
Meta Ulkhairi
150210106**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2020**

**IMPLEMENTASI *DATA MINING* PENJUALAN
KOSMETIK DENGAN ALGORITMA APRIORI**

SKRIPSI

**Untuk memenuhi salah satu syarat
guna memperoleh gelar Sarjana**



**Oleh:
Meta Ulkhairi
150210106**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2020**

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Meta Ulkhairi
Npm : 150210106
Fakultas : Teknik dan Komputer
Program Studi : Teknik Informatika

Menyatakan bahwa Skripsi yang saya buat dengan judul:

IMPLEMENTASI *DATA MINING* PENJUALAN KOSMETIK DENGAN ALGORITMA APRIORI

Adalah hasil karya sendiri dan bukan “duplikasi” dari karya orang lain. Sepengetahuan saya, didalam naskah Skripsi ini tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip didalam naskah ini dan disebutkan didalam sumber kutipan dan daftar pustaka.

Apabila ternyata didalam naskah Skripsi ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia naskah Skripsi ini digugurkan dan Skripsi yang saya peroleh dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dari siapapun

Batam, 25 Juli 2020



Meta Ulkhairi
150210106

IMPLEMENTASI *DATA MINING* PENJUALAN KOSMETIK DENGAN ALGORITMA APRIORI

SKRIPSI

**Untuk memenuhi salah satu syarat
guna memperoleh gelar Sarjana**

**Oleh:
Meta Ulkhairi
150210106**

**Telah disetujui oleh Pembimbing pada tanggal
seperti tertera di bawah ini**

Batam, 24 Juli 2020

**Ellbert Hutabri, S.Kom., M.Kom.
Pembimbing**

ABSTRAK

Seiring dengan perkembangan zaman, data mining saat ini telah diimplementasikan ke berbagai bidang, diantaranya seperti bidang bisnis atau perdagangan, bidang pendidikan dan telekomunikasi. Perkembangan teknologi yang semakin pesat membuat semakin banyaknya pesaing dalam dunia bisnis. Seperti halnya Toko Yati Kosmetik yang menjual berbagai macam *merk* kosmetik, yang masih menggunakan cara manual dalam pendataan barang dan juga tidak mengutamakan *display* (tata letak produk) juga menjadi sebuah kesulitan dalam meningkatkan penjualan produk. Sehingga sulit untuk mengetahui produk (*item*) yang sering terjual dan sulit untuk mengetahui kosmetik yang sudah tidak tersedia. Maka itu diperlukan penerapan algoritma apriori dalam mengetahui pola kombinasi *item sets* yang dapat mempermudah pasar dalam mengolah data penjualan kosmetik. Mengetahui jenis kosmetik yang terjual paling banyak, *stock* kosmetik yang masih tersedia, mengatur *display* produk, serta manfaat data mining dalam sistem penjualan kosmetik bagi penjual dan pembeli. Maka itu diperlukan metode data mining untuk penerapan algoritma apriori dengan bantuan *software tanagra* dalam mengolah data penjualan. Data transaksi penjualan Toko Yati Kosmetik yang berlangsung selama tiga tahun terakhir ini yang akan digunakan untuk menganalisa data pada penelitian ini, yakni pada bulan Januari 2017 sampai Desember 2019. Data yang digunakan sebelumnya adalah data yang sudah melewati tahapan seleksi terlebih dahulu oleh peneliti. Dengan perhitungan algoritma apriori maka diperoleh hasil produk yang paling sering dibeli oleh konsumen yaitu bedak mirabella, bedak inez, bedak maybellin, lipstik pixy, dan lipstik make over dengan nilai *minimum support* 44% dan nilai *minimum confidence* 80% dengan *final rule* yang diperoleh sebanyak 9 *rules* dengan *confidence* 83%, 84%, dan 89%.

Kata Kunci: Data Mining, *Algoritma apriori*, *Tanagra*, penjualan, *display*

ABSTRACT

Along with the times, data mining has been implemented in various fields, including business or trade, education and telecommunications. The rapid development of technology makes more and more competitors in the business world. Like the Yati Cosmetics Shop which sells a variety of cosmetic brands, which still uses manual methods in collecting data and also does not prioritize displays (product layout) also becomes a difficulty in increasing product sales. So it is difficult to find out which products (items) are often sold and it is difficult to find out cosmetics that are no longer available. Therefore, it is necessary to apply a priori algorithm to find out the combination of item sets that can facilitate the market in processing cosmetic sales data. Knowing the types of cosmetics sold the most, cosmetics stocks that are still available, managing product displays, and the benefits of data mining in the cosmetics sales system for sellers and buyers. Then the data mining method is needed for the application of a priori algorithms with the help of tanagra software in processing sales data. Yati Cosmetics Store sales transaction data that lasted for the past three years will be used to analyze the data in this study, namely in January 2017 to December 2019. The data used previously is data that has passed the selection stage in advance by researchers. With the a priori algorithm calculation, the results obtained by the products most frequently purchased by consumers are mirabella powder, inez powder, maybellin powder, pixy lipstick, and make over lipstick with a minimum support value of 44% and a minimum value of 80% confidence with a final rule obtained of 9 rules with 83%, 84% and 89% confidence.

Keywords: *Data Mining, a priori algorithm, Tanagra, sales, display*

KATA PENGANTAR

Dengan mengucapkan puji dan syukur kepada Allah S.W.T yang telah melimpahkan segala rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Teknik Informatika Universitas Putera Batam.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati. Dengan segala keterbatasan, penulis menyadari pula bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Rektor Universitas Putera Batam;
2. Dekan Fakultas Teknik Informatika dan Komputer;
3. Ketua Program Studi Teknik Informatika Universitas Putera Batam Bapak Andi Maslan, S.T., M.SI;
4. Bapak Ellbert Hutabri, S.Kom., M.Kom. selaku pembimbing Skripsi pada Program Studi Teknik Informatika Universitas Putera Batam;
5. Bapak Joni Eka Candra, S.T., M.T. selaku pembimbing akademik selama program studi Teknik Informatika dan seluruh Staff Universitas Putera Batam;
6. Kedua orang tua dan keluarga besar penulis yang telah memberikan motivasi, dukungan, serta doa kepada penulis;
7. Kepada sahabat dan teman-teman seperjuangan Meliani dan Nova Saputra yang telah memberikan motivasi, dukungan, serta masukan kepada penulis dalam menyusun skripsi ini.

Semoga Allah S.W.T membalas kebaikan dan selalu mencurahkan hidayah serta taufik-Nya, Aamiin

Batam, 25 Juli 2020



Meta Ulkhairi

DAFTAR ISI

HALAMAN SAMPUL	
HALAMAN JUDUL	i
SURAT PERNYATAAN ORISINALITAS	ii
HALAMAN PENGESAHAN	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
DAFTAR RUMUS	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang Penelitian	1
1.2 Identifikasi Masalah	4
1.3 Batasan Masalah	4
1.4 Rumusan Masalah	5
1.5 Tujuan Penelitian	5
1.6 Manfaat Penelitian	6
1.6.1 Manfaat Teoritis	6
1.6.2 Manfaat Praktis	6
BAB II TINJAUAN PUSTAKA	8
2.1 <i>Knowledge Discovery in Database (KDD)</i>	8
2.1.1 Proses <i>Knowledge Discovery in Database</i>	9
2.2 Data Mining	11
2.2.1 Tahapan Proses Data Mining	12
2.2.2 Pengelompokan Data Mining	14
2.2.3 Karakteristik Data Mining	15
2.2.4 Langkah-langkah Data Mining	16
2.2.5 Persiapan Data Mining	17
2.2.6 Set Data	18
2.3 Metode Data Mining	19
2.3.1 Association Rule	20
2.3.2 Algoritma Apriori	21
2.4 <i>Software</i> Pendukung	23
2.4.1 <i>Software Tanagra</i>	23
2.4.2 <i>Microsoft Excel</i>	25
2.5 Penelitian Terdahulu	26
2.6 Kerangka Pemikiran	32
BAB III METODE PENELITIAN	33
3.1 Desain Penelitian	33
3.2 Teknik Pengumpulan Data	35
3.2.1 Observasi	35

3.2.2 Wawancara.....	36
3.2.3 Studi Pustaka.....	36
3.3 Operasional Variabel.....	36
3.4 Metode Analisis dan Rancangan Sistem	38
3.4.1 Metode Analisis	38
3.4.2 Rancangan Sistem.....	40
3.5 Lokasi dan Jadwal Penelitian	41
3.5.1 Lokasi Penelitian.....	41
3.5.2 Jadwal Penelitian	41
BAB IV HASIL DAN PEMBAHASAN	43
4.1 Analisa Data	43
4.1.1 Menghitung Pola Iterasi 1 Itemset.....	48
4.1.2 Iterasi Kombinasi 2 <i>Itemset</i>	50
4.1.3 Iterasi Kombinasi 3 itemset	52
4.1.4 Iterasi Kombinasi 4 <i>Itemset</i>	54
4.1.5 Iterasi Kombinasi 5 <i>Itemset</i>	56
4.1.6 Pembentukan Pola Aturan Asosiasi	57
4.2 Hasil Pengujian.....	63
4.2.1 Pembahasan	70
BAB V SIMPULAN DAN SARAN	72
5.1 Simpulan.....	72
5.2 Saran	73
DAFTAR PUSTAKA	74
LAMPIRAN.....	77

DAFTAR GAMBAR

Gambar 2. 1 Tahapan Proses	13
Gambar 2. 2 Tampilan Awal Tanagra	24
Gambar 2. 3 Fitur-fitur Komponen dalam Tanagra.....	24
Gambar 2. 4 Tampilan Halaman Home Excel.....	26
Gambar 2. 5 Kerangka Pemikiran	32
Gambar 3. 1 Desain Penelitian Dikembangkan Oleh Peneliti.....	33
Gambar 3. 2 Metode Rancang Sistem	40
Gambar 4. 1 Format Tabulasi	64
Gambar 4. 2 Halaman Kerja Aplikasi Tanagra	64
Gambar 4. 3 Memasukkan Data ke Aplikasi Tanagra.....	65
Gambar 4. 4 Tampilan Data di Aplikasi Tanagra.....	66
Gambar 4. 5 Proses Memasukkan Atribut Item ke Dalam Tanagra.....	66
Gambar 4. 6 Mengatur Nilai Frequent Itemset.....	67
Gambar 4. 7 Hasil Frequent Itemset.....	68
Gambar 4. 8 Mengatur Nilai Confidence	68
Gambar 4. 9 Hasil Confidence	69
Gambar 4. 10 Final Rules	69

DAFTAR TABEL

Tabel 3. 1 Operasional Variabel.....	38
Tabel 3. 2 Jadwal Penelitian.....	42
Tabel 4. 1 Data Transaksi Penjualan	44
Tabel 4. 2 Tabulasi Penjualan	47
Tabel 4. 3 Pola Iterasi 1 Itemset.....	49
Tabel 4. 4 Iterasi Kombinasi 2 Itemset.....	50
Tabel 4. 5 Iterasi Kombinasi 3 Itemset.....	52
Tabel 4. 6 Iterasi Kombinasi 4 Itemset.....	55
Tabel 4. 7 Iterasi Kombinasi 5 Itemset.....	57
Tabel 4. 8 Pola Aturan Asosiasi	58
Tabel 4. 9 Menghitung Minimum Confidence	61
Tabel 4. 10 Final Rules Asosiasi	63

DAFTAR RUMUS

Rumus 2. 1 Item set Nilai Support.....	21
Rumus 2. 2 Dua Item Set atau Lebih	21
Rumus 2. 3 Nilai Confidence	21

DAFTAR LAMPIRAN

Lampiran 1. Pendukung Penelitian	77
Lampiran 2. Daftar Riwayat Hidup.....	96
Lampiran 3. Surat Keterangan Penelitian	97

BAB I PENDAHULUAN

1.1 Latar Belakang Penelitian

Seiring dengan perkembangan zaman, penerapan data mining telah diimplementasikan keberbagai bidang, seperti bidang dalam bisnis atau perdagangan, hingga bidang pendidikan dan telekomunikasi. Data mining algoritma apriori sangat membantu para pebisnis dalam pengambilan keputusan terhadap persediaan barang seperti penerapannya dalam bidang bisnis.

Perkembangan teknologi yang semakin pesat membuat semakin banyaknya pesaing dalam dunia bisnis. Hal tersebut menuntut segenap pengembang dalam peningkatan penjualan serta pemasaran produk yang dijual. Transaksi penjualan yang dilakukan setiap hari akan terjadi penumpukan data, karena transaksi penjualan yang dilakukan setiap harinya menyebabkan banyaknya transaksi penjualan. Sebuah agen penjualan kosmetik harus bisa mengatur data transaksinya agar bisa menjadi informasi yang berguna. Informasi yang berguna dapat dibuat sebagai kebijakan bisnis data yang dapat digunakan sebagai sebuah strategi bisnis yaitu analisis data penjualan bermanfaat untuk merancang strategi penjualan dan promosi produk.

Toko Yati Kosmetik adalah sebuah toko yang menjual berbagai macam *merk* kosmetik, yang masih menggunakan cara manual dalam pendataan barang. Sehingga sulit untuk melihat *item* yang sering terjual, maupun untuk mengetahui kosmetik yang sudah tidak tersedia. Pemasaran yang dilakukan juga melalui media sosial seperti *Facebook* yang diharapkan dapat memperkenalkan dan meningkatkan

penjualan Yati kosmetik. Sebagai salah satu cara yang dilakukan untuk mempromosikan produk agar dapat memperkenalkan dan meningkatkan penjualannya.

Toko Yati Kosmetik setiap harinya memiliki 30-50 pcs/produk transaksi penjualan. Toko Yati Kosmetik mendapatkan total 200-500 produk penjualan dalam satu bulan. Banyaknya produk yang dijual di Toko Yati Kosmetik, sehingga sulit untuk mengetahui produk (*item*) yang paling sering terjual, dan ini berakibat pada pembelian dan penjualan toko.

Sistem penjualan yang bersifat *in-out* dimana toko hanya memesan setiap produk yang sudah habis atau tidak tersedia. Cara pemesanan tersebut akan mengakibatkan banyak produk lama atau produk yang tidak terjual terabaikan. Selain banyaknya *item* produk yang dijual, cara *display* barang pada Toko Yati Kosmetik juga kurang menarik karena struktur letak produk yang tidak tersusun dengan rapi di atas rak ataupun di dalam rak kaca, dan tidak mengutamakan *display* dalam menata produk juga menjadi sebuah kesulitan dalam meningkatkan penjualan produk.

Banyaknya transaksi penjualan yang terjadi, akan menyebabkan penumpukan data, yang terjadi setiap harinya karena banyaknya transaksi penjualan. Sehingga data ini dapat diolah dengan menggunakan data mining agar dapat mengolah pengetahuan-pengetahuan data dalam skala besar secara efisien dan efektif. Dan menggunakan algoritma apriori sebagai metode yang melihat keterkaitan hubungan antar elemen. Untuk mengetahui hubungan *item set* apa saja yang terjadi pada

proses transaksi dengan menggunakan algoritma apriori merupakan algoritma yang pertama kali digunakan ketika melakukan proses asosiasi.

Menurut Kennedy Tampubolon dalam (Tri Susilo, Sunardi and Waruwu, 2019) algoritma apriori merupakan suatu metode *Association Rules* yang dapat diterapkan bila terdapat beberapa hubungan *item* yang ingin dianalisis. Berdasarkan hal tersebut algoritma apriori dapat menampilkan hubungan antar atribut dan menggali informasi pada data. Oleh karena itu, penelitian ini menerapkan algoritma apriori pada bidang bisnis dan pola pembelian kosmetik.

Menurut Dewantara, H,dkk dalam (Wijayanti, 2017) data mining yang disimpan dalam *database*, *data warehouse*, atau media penyimpanan yang lainnya dapat diproses melalui *interesting knowledge*. Pengolahan dengan menggunakan data mining tersebut akan membantu dalam mengetahui kosmetik apa saja yang dibeli oleh para konsumen dan menganalisis dari kebiasaan membeli konsumen.

Berdasarkan masalah di atas menggerakkan peneliti untuk melakukan penelitian dari persoalan yang terjadi di Toko Yati Kosmetik menggunakan algoritma apriori dalam menyusun perhitungan yang dapat membantu untuk menentukan prosedur penjualan yang nantinya akan membentuk kombinasi *item* yang dapat dikelompokkan berdasarkan nilai parameternya dengan judul **“IMPLEMENTASI DATA MINING PENJUALAN KOSMETIK DENGAN ALGORITMA APRIORI”**

1.2 Identifikasi Masalah

Berlandaskan latar belakang di atas bahwa dapat diidentifikasi sebagian masalah sebagai berikut:

1. Rumitnya mengetahui produk (*item*) yang sering terjual sehingga toko Yati Kosmetik tidak dapat mengetahui kosmetik apa saja yang sudah tidak memiliki persediaan lagi.
2. Kurangnya sistem penjualan dalam memanfaatkan data transaksi pada kosmetik di toko Yati Kosmetik.
3. Kesulitan mengolah informasi data transaksi di toko Yati Kosmetik, membuat kurang efektif pengelolaan data transaksi penjualan kosmetik.
4. Tata letak (*display*) produk yang tidak diutamakan dan tidak tersusun rapi.

1.3 Batasan Masalah

Berdasarkan masalah yang diterangkan dalam penelitian ini begitu luas, agar tidak menyimpang dari maksud utama penulis, maka penulis membatasi masalah-masalah yang akan ditulis, diantaranya yaitu:

1. Penelitian data mining ini memakai data transaksi penjualan kosmetik di toko Yati Kosmetik dalam masa waktu 3 tahun terakhir.
2. Data yang digunakan adalah berdasarkan data penjualan perbulan
3. Aplikasi data mining yang digunakan untuk pengujian data penelitiannya adalah aplikasi tanagra dengan metode algoritma apriori.

1.4 Rumusan Masalah

Berdasarkan gambaran dari latar belakang, maka dapat dirumuskan masalah penelitian yang akan dibahas yaitu:

1. Bagaimana mengaplikasikan algoritma apriori untuk mengetahui pola kombinasi *itemsets* yang dapat memudahkan pasar dalam mengolah data penjualan kosmetik ?
2. Bagaimana cara mengetahui jenis kosmetik yang terjual paling banyak dan *stock* kosmetik yang masih tersedia/ada ?
3. Bagaimana fungsi data mining dalam sistem penjualan kosmetik bagi penjual dan pembeli ?
4. Bagaimana implementasi data mining menggunakan metode algoritma apriori pada produk ?

1.5 Tujuan Penelitian

Berdasarkan gambaran rumusan masalah, maka tujuan dalam melakukan penelitian ini adalah:

1. Untuk mengaplikasikan algoritma apriori dalam mengetahui pola kombinasi *item sets* yang dapat memudahkan pasar dalam mengolah data penjualan kosmetik.
2. Untuk mengetahui jenis kosmetik yang terjual paling banyak dan *stock* kosmetik yang masih tersedia/ada.

3. Untuk mengetahui fungsi data mining dalam sistem penjualan kosmetik bagi penjual dan pembeli.
4. Untuk mengimplementasi data mining menggunakan metode algoritma apriori pada produk.

1.6 Manfaat Penelitian

Manfaat penelitian dapat dibagi menjadi 2, yaitu manfaat teoritis dan manfaat praktis.

1.6.1 Manfaat Teoritis

1. Untuk mengimplementasikan ilmu yang telah dipelajari selama perkuliahan dalam jurusan teknik informatika.
2. Sebagai salah satu acuan untuk mengetahui pemanfaatan data sistem penjualan.
3. Sebagai tambahan acuan bagi pembaca yang akan melakukan penelitian lebih lanjut, khususnya bidang Teknik informatika tentang data mining metode algoritma apriori.

1.6.2 Manfaat Praktis

1. Manfaat sebagai sarana untuk mengelola data sistem penjualan dengan menggunakan aplikasi ini.

2. Memudahkan toko Yati Kosmetik menentukan persediaan barang.
3. Dengan memanfaatkan metode ini, Toko Yati Kosmetik bisa menentukan sistem penjualan dan ketersediaan stok
4. Memanfaatkan metode algoritma apriori dalam mencari tata letak produk

BAB II

TINJAUAN PUSTAKA

2.1 *Knowledge Discovery in Database (KDD)*

Knowledge discovery in database (KDD) adalah istilah data mining untuk menguraikan pengetahuan dalam *database*. Banyak jenis teknik analisa yang dapat digunakan dalam data mining dari definisi data mining yang sangat luas.

Menurut T. L. Seifert, H. Jiawei, dkk dalam (Ependi and Putra, 2019) KDD adalah seperti persimpangan beberapa disiplin ilmu yang mencakup statistik, *database*, AI, visualisasi dan komputer paralel yang mempengaruhi pengetahuan “*interdisciplinary knowledge*”. Sedangkan Y. Ramamohan, dkk, Mewar University and N. Rikhi berpendapat dalam (Ependi and Putra, 2019) KDD juga memiliki istilah seperti proses pemilihan data, pembersihan data, transformasi data, presentasi temuan dan evaluasi.

Knowledge Discovery in Database (KDD) merupakan pengumpulan data, pemakaian data *historis* untuk menemukan pengetahuan, informasi, keteraturan, pola atau hubungan dalam data yang berukuran besar yang merupakan aktivitas dalam KDD (Bulolo, 2020).

Sehingga menurut (Maulana and Fajrin, 2018) mendefinisikan bahwa *knowledge discovery in database (KDD)* untuk mengekstraksi informasi potensial, implisit dan sekelompok data yang tidak dikenal. Dengan menyertakan hasil dari

proses data mining yang merupakan proses pada *knowledge discovery*, lalu informasi tersebut diubah hasilnya secara akurat sehingga mudah dimengerti. KDD adalah proses *non-trivial* untuk mencari dan mengenali pola (*pattern*) dalam data, sehingga bentuk yang diperoleh bersifat baru, bermanfaat dan mudah dipahami.

Berdasarkan pengertian di atas dapat disimpulkan KDD adalah suatu proses untuk mengolah dan mengumpulkan data sehingga menemukan pengetahuan, informasi, keteraturan, pola atau hubungan dalam data yang berukuran besar. Dengan melibatkan proses data mining, lalu mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami.

2.1.1 Proses *Knowledge Discovery in Database*

(Rahmawati and Merlina, 2018), mendefinisikan bahwa istilah data mining dan *knowledge discovery in database* (KDD) untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar, data mining dan KDD sering digunakan secara bergantian. Kedua istilah tersebut memiliki konsep yang berbeda, tetapi saling berkaitan satu sama lainnya. Salah satu tahapan dalam keseluruhan proses KDD adalah data mining. Seperti yang ditunjukkan pada proses berikut:

1. *Data selection*

Sebelum tahap penggalian informasi dalam KDD perlu dilakukan pemilihan data dari sekumpulan data. Yang digunakan untuk proses data

mining adalah data hasil seleksi, disimpan dalam suatu arsip terpisah dari basis data operasional.

2. *Pre-processing/cleaning*

Proses *cleaning* pada data yang merupakan fokus KDD, merupakan awal dari proses sebelum melakukan proses data mining. Membuang penggandaan data, memeriksa data yang *inkonsisten*, dan memperbaiki kesalahan pada data merupakan tahap-tahap proses *cleaning*. Proses “memperkaya” data yang sudah ada dengan data (informasi) lain yang signifikan yang diperlukan untuk KDD, seperti data atau informasi eksternal adalah suatu proses yang dilakukan pada *enrichment*.

3. Integrasi data

Merupakan suatu penggabungan data dari berbagai *database* ke dalam *database* baru.

4. *Transformation*

Transformasi pada data yang telah diseleksi adalah suatu proses *coding*, yang sesuai untuk proses data mining. KDD yang melalui proses *coding* menggambarkan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

5. *Data Mining*

Data mining menggunakan teknik atau metode tertentu dalam menggali pola atau informasi menarik dalam data terpilih. Metode yang terdapat pada data mining sangat beragam. Tujuan dan proses KDD secara keseluruhan akan menentukan metode atau algoritma yang tepat pada proses keseluruhan.

6. *Interpretation/Evaluation*

Menampilkan proses dari suatu pola informasi data mining dalam bentuk yang mudah dimengerti. Tahap ini disebut *interpretation* yang merupakan bagian dari proses KDD, dengan pemeriksaan terhadap pola atau informasi yang ditemukan apakah berbeda dengan fakta atau hipotesis yang ada sebelumnya.

7. Presentasi Pengetahuan (*knowledge*)

Presentasi pengetahuan (*knowledge*), teknik yang diterapkan untuk mendapatkan pengetahuan yang diperoleh pengguna melalui teknik visualisasi dan penyampaian pengetahuan. Dengan merumuskan keputusan atau aksi dari hasil yang didapat merupakan tahap terakhir dari proses data mining.

2.2 Data Mining

Data mining dapat bermanfaat dan dimengerti dalam suatu *database* yang sangat besar, data mining merupakan proses *iterative* dan interaktif untuk menemukan pola atau model baru yang sempurna. (Sikumbang, 2018), mendefinisikan bahwa pengambilan keputusan di waktu yang akan datang dapat dibantu dengan data mining yang berisi pencarian *trend* atau pola yang diinginkan dalam *database* besar, yang mungkin saja menggunakan perangkat pendukung keputusan yang lain dari pola-pola yang dikenali perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan untuk kemudian dapat dipelajari dengan lebih teliti.

Data mining merupakan kegiatan mengekstrak informasi atau pengetahuan (*knowledge*) penting, menggunakan teknik tertentu dari suatu set data berukuran besar. Informasi yang diperoleh dari data mining ini dapat digunakan untuk memperbaiki pengambilan keputusan. Karena proses penemuan informasi atau *knowledge* pada set data seperti orang melakukan kegiatan penambangan oleh karena itu dinamakan data mining atau penambangan data (Santosa and Umam, 2018).

Menurut Zhang, C.dkk dalam (Wijayanti, 2017) ada dua alasan mengapa data mining perlu untuk digunakan:

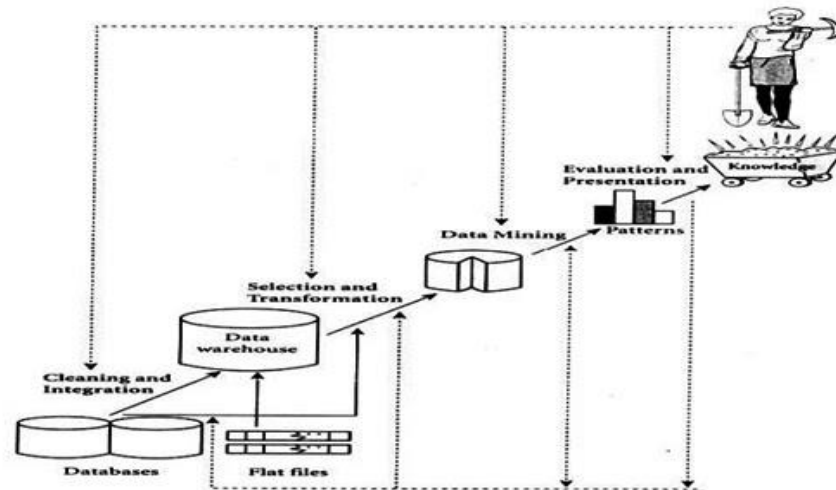
1. Karena fakta pola potensial dalam data sering kali tidak terlihat maka pencarian pola yang terdapat di dalam data dapat mengecewakan bagi para pembuat keputusan yang tidak mempunyai pengalaman.
2. Untuk dianalisis secara manual jumlah data tersebut terlalu besar.

Berdasarkan pengertian di atas dapat disimpulkan data mining adalah suatu pencarian untuk membantu mengambil keputusan di waktu yang akan datang dengan pola yang diinginkan dalam *database* besar. Dengan menggunakan teknik tertentu dari suatu set data berukuran besar data mining dapat mengekstrak informasi atau pengetahuan penting. Untuk memperbaiki pengambilan keputusan dapat diperoleh dari informasi yang dihasilkan data mining tersebut.

2.2.1 Tahapan Proses Data Mining

Menurut Santoso dalam (Amrin, 2018) data mining sering disebut juga *Knowledge Discovery in Database* (KDD), merupakan suatu pengumpulan

pemakaian data *historis* untuk menemukan keteraturan, pola atau hubungan dalam *set* data berukuran besar. Pembuatan aplikasi data mining ditunjukkan pada gambar berikut:



Gambar 2. 1 Tahapan Proses
Sumber: (Amrin, 2018)

Proses tahapan data mining terdiri dari langkah utama yaitu:

1. *Cleaning and Integration*

Pada tahap *cleaning*, yaitu data yang tidak konsisten dan “*noise*” dihilangkan.

Dan pada *integration* merupakan proses penggabungan sumber-sumber data.

2. *Selection and Transformation*

Tahap *selection* adalah proses pengambilan data-data yang relevan untuk dianalisis. Sedangkan konsolidasi data ke bentuk yang sesuai untuk di “*mining*” dan menghasilkan ringkasan atau penggabungan merupakan proses transformasi.

3. Data Mining

Pada tahap ini, proses awal dimana metode pengkajian diterapkan untuk mengekstraksi pola data.

4. *Evaluation and Presentation*

Tahap *evaluation* adalah yang mewakili basis pengetahuan berdasarkan ukuran tertentu dengan mengidentifikasi pola yang menarik. Sedangkan, yang digunakan untuk menampilkan pengetahuan kepada pengguna yaitu pada tahap *presentation*.

2.2.2 Pengelompokan Data Mining

(Bulolo, 2020), menerangkan bahwa data mining terbagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan yaitu:

1. Deskripsi: bertujuan mengidentifikasi/untuk menemukan pola yang sering muncul dan mengubah pola tersebut menjadi aturan yang dapat digunakan untuk mempermudah suatu aktivitas. Seperti pada sebuah Supermarket, pelanggan sering membeli produk A dan produk B secara bersamaan dan berulang, sehingga manajemen supermarket mengubah katalog barang dengan meletakkan produk A dan pada tempat yang berdekatan. Agar pelanggan tidak kesulitan dalam membeli produk tersebut.
2. Klasifikasi: berdasarkan hubungan antara variabel kriteria dengan variabel target yang dikelompokkan. Seperti pengelompokan bencana alam yaitu rusak berat, rusak sedang, atau tidak berdampak.

3. **Prediksi:** prediksi mirip dengan klasifikasi. Memprediksi merupakan salah satu fungsi data mining yang biasa digunakan. Berdasarkan data-data sebelumnya nilai dari hasil prediksi akan digunakan di masa yang akan datang.
4. **Estimasi:** definisi dari estimasi adalah perkiraan/prediksi, sehingga hampir sama dengan klasifikasi, perbedaannya terletak pada bentuk pengelompokkan, dimana estimasi ke arah numerik dan bukan ke arah kategori.
5. **Pengklasteran:** pengklasteran adalah pengelompokkan data yang memiliki persamaan nilai (homogen). Bentuk data yang dapat dikelompokkan dalam pengklasteran yaitu hasil pengamatan, *record* data, atau kelas-kelas dan objek-objek yang memiliki kesamaan.
6. **Asosiasi:** asosiasi merupakan kumpulan, himpunan, persatuan, atau persekutuan. Pencarian *attribute* yang muncul/selalu muncul dalam waktu bersamaan merupakan proses pada data mining, misalnya seperti ketika dibeli produk A maka dibeli produk B, ketika dibeli produk B maka dibeli produk A, ketika dibeli produk A,B, maka dibeli produk C, dan seterusnya.

2.2.3 Karakteristik Data Mining

Menurut C. I. E. J Dahlan Abdullah dalam (Kasus *et al.*, 2018) data mining memiliki karakteristik sebagai berikut:

1. Data mining merupakan pola data tertentu yang tidak diketahui sebelumnya dan penemuan sesuatu yang tersembunyi pada data.
2. Data mining membuat hasil lebih dipercaya karena memakai data yang sangat besar.
3. *Association rule mining* untuk mendefinisikan aturan asosiatif antara suatu kombinasi *item*.
4. *Classification* merupakan model untuk menemukan kegunaan yang menjelaskan konsep atau kelas data.
5. *Decision tree* adalah suatu metode *classification* yang begitu populer karena mudah untuk diinterpretasi. Tiap ujung pohon menyatakan kelas dan setiap percabangan merumuskan kondisi yang terpenuhi.

2.2.4 Langkah-langkah Data Mining

Menurut (Rahmawati and Merlina, 2018) Langkah-langkah data mining terdiri dari beberapa tahap sebagai berikut:

- 1) Tahap pertama: *Precise statement of the problem* (pernyataan tepat terhadap permasalahan) sebelum mengakses perangkat lunak data mining. Pertanyaan yang akan dijawab oleh seorang analis harus memiliki kejelasan tentang pertanyaan tersebut. Harus memiliki formulasi yang tepat untuk problematika yang ada dalam membuat solusinya.

- 2) Tahap dua: *Initial exploration*. Kunci dari tahap ini yaitu diawali dengan mengidentifikasi dan menghilangkan data yang dikodekan salah (*cleaning*), transformasi data, memilih *subset record, data set*, langkah awal seleksi, mendeskripsikan dan memvisualisasikan data.
- 3) Tahap tiga: *Model building and validation*. Tahap ini menentukan yang terbaik bagi performa prediktif dan melibatkan pertimbangan terhadap jenis permodelan.
- 4) Tahap keempat: *Deployment*. Pada tahap ini untuk membuat (*generate*) prediksi yaitu dengan memilih aplikasi yang tepat beserta permodelan.

2.2.5 Persiapan Data Mining

(Sulastri and Gufroni, 2017), mendefinisikan bahwa *Preprocessing Data mining* bisa meningkatkan kualitas data, sehingga data yang diperoleh langsung dari Toko Yati Kosmetik Sagulung, Batam diolah terlebih dahulu melalui tahap-tahap *data cleaning, data integration, data selection, dan data transformation*. Hal tersebut dilakukan agar data yang diolah lebih berkualitas, maksudnya data-data tersebut bersifat objektif, representatif, memiliki *sampling error* yang kecil, terbaharui dan relevan. Persiapan tersebut antara lain:

1. *Data Cleaning*

Data cleaning adalah proses untuk mengatasi nilai yang hilang seperti *noise* dan data yang tidak konsisten.

2. *Data Integration*

Data integration merupakan proses menggabungkan data dari banyak *database*.

3. *Data Selection*

Data selection tetap merepresentasikan data aslinya dengan proses meminimalkan jumlah data yang digunakan untuk proses mining. *Data selection* dapat berupa *sampling*, *denoising*, dan *feature extraction*.

4. *Data Transformation*

Data transformation merupakan proses untuk mengubah bentuk dan format data. Hal tersebut dapat membantu memahami hasil yang didapat dan meringankan pengguna dalam proses mining.

2.2.6 Set Data

Menurut (Nengsih, 2019) Set data (data set) merupakan data yang diolah pada data mining, dataset merupakan kumpulan data, dimana satu dataset merepresentasikan satu tabel pada *database*, atau bisa juga suatu matriks data dimana setiap kolom mewakili variabel tertentu, tiap baris merepresentasikan banyaknya data.

Terdapat 2 jenis dataset yaitu: *private dataset* dan *public dataset*. ***Private dataset***: adalah dataset yang dapat diambil dari organisasi yang kita jadikan obyek penelitian Bank, Rumah Sakit, Industri, Pabrik, Perusahaan Jasa, dan lain-lain. ***Public dataset***: yaitu dataset yang dapat diambil dari repositori publik

(Macmudi, 2018), menerangkan bahwa karakteristik set data seperti atribut, *class*, tipe data dan jumlah instan, merupakan beberapa faktor yang menjadi pertimbangan. Parameter atau faktor yang menyebabkan *class/label/target* terjadi yaitu atribut. *Class* adalah atribut yang dijadikan target, sering juga disebut dengan label. Tipe data untuk variabel pada statistik terbagi menjadi empat: nominal, ordinal, interval, *ratio*, tetapi secara praktis tipe data untuk atribut pada data mining hanya menggunakan dua: Nominal (Diskrit) dan *Numeric* (*Kontinyu* atau Ordinal).

2.3 Metode Data Mining

Pada penelitian ini penulis menggunakan metode asosiasi. (Badrul, Studi and Informasi, 2016), menyatakan bahwa untuk menemukan atribut yang muncul dalam satu waktu disebut asosiasi. Seperti yang akan dibahas pada penelitian ini, merupakan salah satu implementasi dari asosiasi yaitu *market basket analysis* (analisis keranjang belanja).

Menurut (Yanto, R, & Di Kesuma, H., 2017) dalam (Prayitno and Rasim, 2018) metode asosiasi yang memenuhi syarat *minimum support* (nilai penunjang) seperti meng-asosiasikan data dengan menggunakan aturan asosiasi (pengalaman), yaitu kombinasi tiap *item* dalam *database* dan syarat *minimum confidence* (nilai kepastian), yang merupakan kuatnya hubungan antar *item* dalam aturan asosiasi.

2.3.1 Association Rule

(Choiriah *et al.*, 2019), menyatakan bahwa untuk menemukan aturan asosiasi antara suatu kombinasi *item* merupakan teknik dari data mining. Hasil pengolahan data dan perhitungan tertentu yang terdapat suatu ketertarikan (*interesting measure*), adalah untuk menentukan suatu aturan asosiasi. Pada umumnya terdapat dua ukuran ketertarikan dalam aturan asosiasi, yaitu:

1. *Support*, jumlah seluruh transaksi dari probabilitas konsumen membeli beberapa produk secara bersamaan. Dari keseluruhan transaksi yang ada ukuran tersebut menentukan apakah suatu *item/itemset* layak untuk dicari nilai *confidence*-nya. Seberapa besar tingkat dominasi yang menunjukkan bahwa *item X* dan *Y* dibeli bersamaan
2. *Confidence* (tingkat kepercayaan), merupakan produk yang dibeli bersamaan dari beberapa probabilitas kejadian, dimana salah satu produk sudah pasti dibeli (seperti: seberapa sering *item Y* dibeli apabila konsumen membeli *item X*)

Kedua ukuran (*support* dan *confidence*) berfungsi dalam menentukan aturan asosiasi, ditentukan oleh pengguna untuk dibandingkan dengan batasan (*threshold*). Batasan itu seperti batasan minimum dari nilai *support* sebagai *minimum support* dan batasan minimum dari nilai *confidence* sebagai batasan *minimum confidence*.

Menurut L. Daniel. T,dkk dalam (Ependi and Putra, 2019) metode dasar dalam proses analisis asosiasi dapat dilakukan dalam 2 tahap:

1) Analisa pola frekuensi tinggi

Untuk menghitung iterasi 1 *itemset*, mencari nilai syarat minimum dari *item set* nilai *support* dengan rumus:

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

Rumus 2. 1 *Item set* Nilai *Support*

Kemudian untuk menemukan dua *item set* atau lebih:

$$Support(A,B) = P(A \cap B)$$

$$Support(A,B) = \frac{\sum \text{Transaksi yang mengandung A dan B}}{\text{Total transaksi yang telah terjadi}} \times 100\%$$

Rumus 2. 2 Dua *Item Set* atau Lebih

2) Pembentukan aturan asosiatif

Sesudah semua pola dengan frekuensi tinggi ditemukan baru menghitung nilai *confidence*:

$$Confidence(A \rightarrow B) = P(B | A)$$

$$P(B | A) = \frac{\sum \text{Transaksi yang mengandung A dan B}}{\sum \text{Transaksi yang mengandung A}} \times 100\%$$

Rumus 2. 3 Nilai *Confidence*

2.3.2 Algoritma Apriori

(Anjar Wanto, Muhammad Noor Hasan Siregar, 2020), menjelaskan bahwa algoritma apriori merupakan cabang ilmu data mining untuk kategori asosiasi. Algoritma apriori untuk menemukan pola frekuensi tinggi yang cukup terkenal dari salah satu algoritma data mining yang digunakan. *Minimum support* merupakan

nilai *support* di atas ambang batas tertentu yang memiliki pola frekuensi tinggi yaitu pola-pola *item* dalam suatu *database*. Pola frekuensi tinggi ini berfungsi untuk menyusun aturan asosiatif dan juga beberapa teknik data mining lainnya. Sering disebut sebagai *affinity analysis* atau *market basket analysis* yaitu aturan yang menyatakan asosiasi antara beberapa atribut.

Menurut C. N. Parkinson,dkk dalam (Purba and Buulolo, 2020) algoritma apriori berfungsi sebagai sekumpulan data untuk menemukan *frequent itemset* yang dijalankan. Pada iterasi ke-k, akan ditemukan semua *itemset* yang mempunyai k *item*, disebut dengan k-*itemset*. Berikut ini adalah dasar algoritma apriori:

1. Membangun Analisa pola frekuensi tinggi

Tahap ini untuk mencari kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam *database*.

2. Menciptakan aturan assosiatif

Setelah semua pola frekuensi tinggi ditemukan, lalu menghitung *confidence* aturan asosiatif $A \rightarrow B$ untuk dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence*.

3. Analisa penerapan algoritma apriori

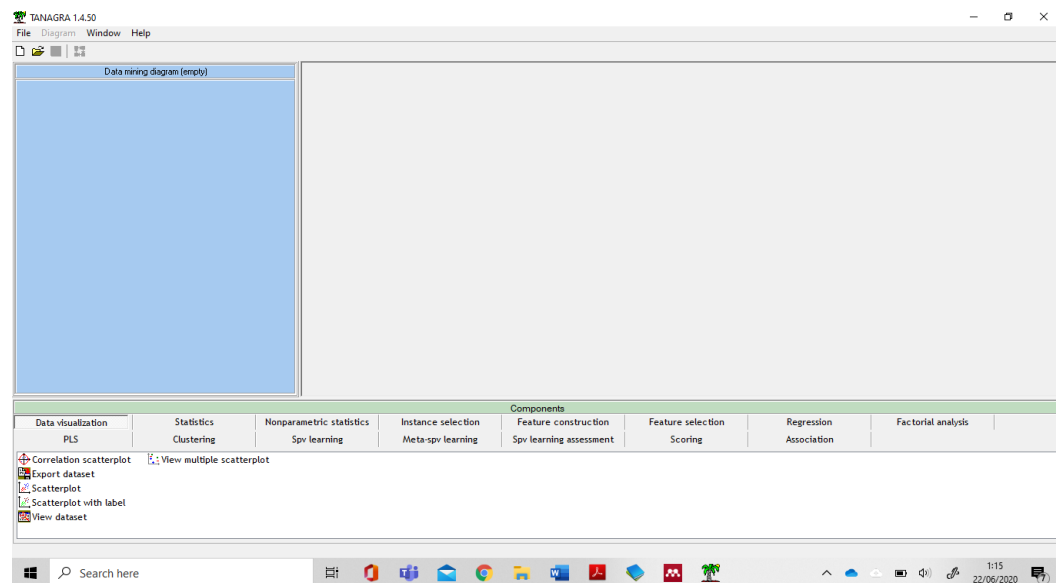
Dalam penentuan pola pembelian serta membantu tata letak barang dalam rak berdasarkan kecenderungan konsumen diperlukan penerapan algoritma apriori untuk membantu dalam membentuk kombinasi barang dengan dilakukan pengujian apakah kombinasi tersebut memenuhi nilai ambang.

2.4 *Software* Pendukung

2.4.1 *Software* Tanagra

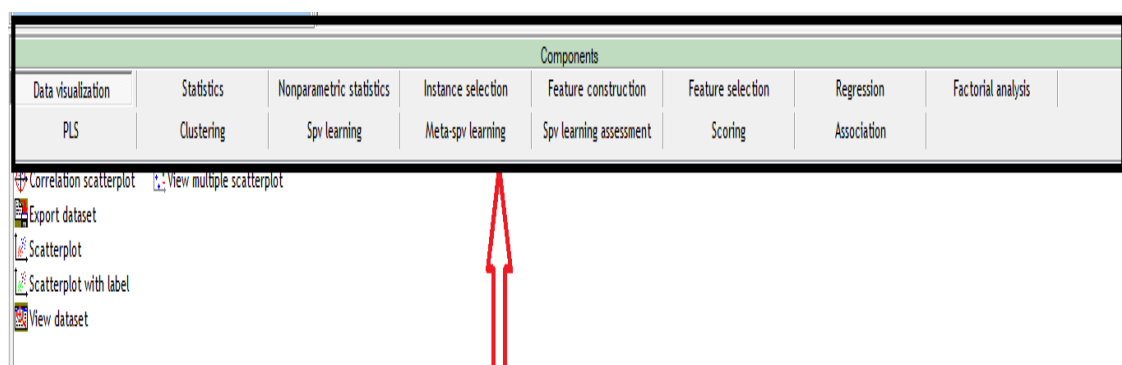
Menurut (Badrul, Studi and Informasi, 2016) tanagra yaitu sebuah *software* untuk tujuan akademik dan penelitian data mining. Beberapa metode data mining yang diusulkan tanagra yaitu analisis eksplorasi data, pembelajaran statistik, pembelajaran mesin dan daerah *database*. Tanagra adalah suatu “proyek *open source*” yang setiap peneliti dapat mengakses kekode sumber, serta memasukkan algoritma sendiri, sesuai dengan lisensi distribusi perangkat lunak dan selama dia setuju. Merupakan suatu Proyek yang bertujuan untuk memberikan kemudahan bagi peneliti dan mahasiswa untuk menggunakan perangkat lunak data mining, memungkinkan untuk menganalisis baik data yang nyata ataupun sintesis sesuai dengan aturan yang disediakan pengembang perangkat lunak dalam domain ini.

Berikut adalah tampilan awal pada aplikasi tanagra:



Gambar 2. 2 Tampilan Awal Tanagra
Sumber: Data Peneliti 2020

Dalam segi fitur tidak hanya terdapat beberapa pembelajaran terkontrol tanagra juga memiliki paradigma lain seperti *clustering*, analisis faktorial, statistik parametrik dan nonparametrik, aturan asosiasi, *feature selection*, dan *construction algorithms*, seperti gambar berikut:

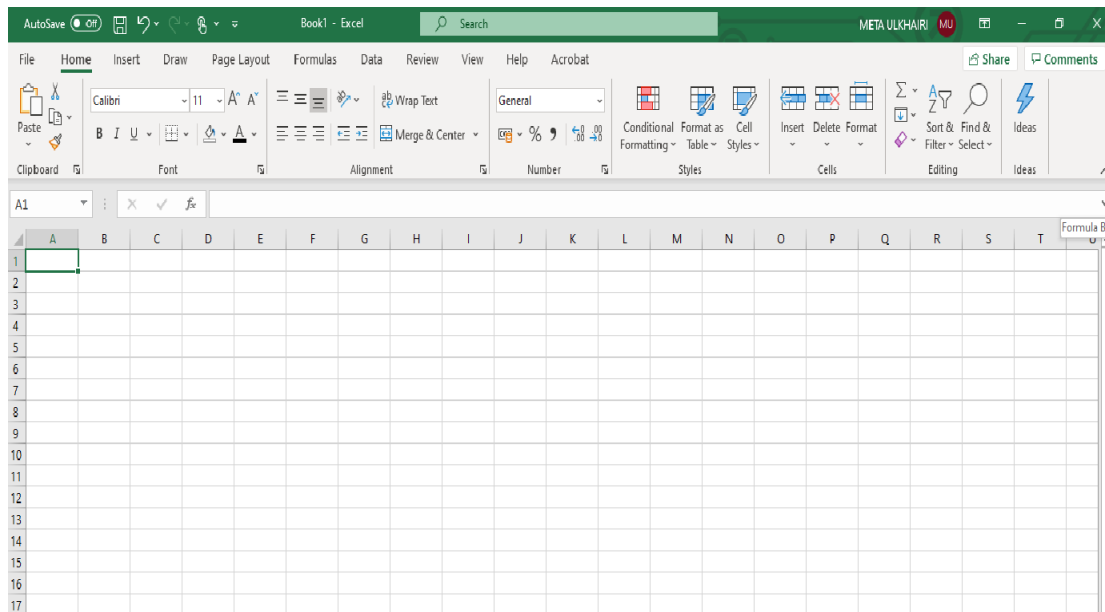


Gambar 2. 3 Fitur-fitur Komponen dalam Tanagra
Sumber: Data peneliti 2020

2.4.2 Microsoft Excel

(Wibowo, 2018), mendefinisikan bahwa sebuah aplikasi lembar kerja *spreadsheet* yang dibuat dan didistribusikan oleh *Microsoft Corporation* yaitu *Microsoft Excel* atau *Microsoft Office Excel* sebagai sistem operasi *Microsoft Windows* dan *Mac OS*. Pada aplikasi ini terdapat pembuatan grafik yang baik dan fitur kalkulasi, dengan menggunakan strategi marketing *Microsoft* yang agresif, menjadikan *Microsoft Excel* sebagai salah satu program komputer yang cukup populer digunakan di dalam komputer mikro hingga saat ini.

Microsoft Excel adalah perangkat lunak untuk mengolah data secara otomatis mulai dari perhitungan dasar untuk berbagai jenis data, melakukan perhitungan keuangan, matematika, statistik, pembuatan grafik hingga manajemen data. Perangkat lunak ini sangat membantu dalam menyelesaikan permasalahan administratif mulai yang paling sederhana sampai yang lebih kompleks.



Gambar 2. 4 Tampilan Halaman *Home Excel*
Sumber: Data Peneliti 2020

2.5 Penelitian Terdahulu

Penelitian terdahulu adalah suatu tumpuan pada penelitian. Berikut merupakan penelitian-penelitian yang dapat digunakan sebagai tumpuan dalam proses penyusunan penelitian ini, yaitu:

1. Penelitian oleh (Indriyani and Irfiani, 2019) Nomor 2 Vol.7 yang berjudul **“CLUSTERING DATA PENJUALAN PADA TOKO PERLENGKAPAN OUTDOOR MENGGUNAKAN METODE *K-MEANS*”**, penelitian ini membahas tentang menjaga kepuasan pelanggan dengan cara tidak membiarkan barang kosong dan menjaga stok persediaan barang. Melalui data laporan penjualan barang dengan menganalisa setiap data barang yang laku dan kurang laku, jika toko tersebut adalah sebuah toko *retail* yang memiliki banyak

data penjualan mulai dari ratusan bahkan ribuan data penjualan setiap bulannya itu akan menjadi sulit atau tidak mudah. Dengan menggunakan algoritma *K-Means Clustering* yang merupakan salah satu teknik dalam data mining. Pengelompokan data penjualan pada Toko Genta Corp memperoleh hasil rekomendasi barang yang laris, kurang laris dan cukup laris melalui penerapan metode *K-Means*. Maka dapat dijadikan rujukan bagi manajemen untuk mengatur stok barang agar toko tidak mengecewakan pelanggan karena barang yang akan dibeli tidak tersedia.

2. Penelitian oleh (Wahyuni, Suherman and Harahap, 2017) Nomor 2 Vol.5 yang berjudul **“IMPLEMENTASI DATA MINING DALAM MEMPREDIKSI STOK BARANG MENGGUNAKAN ALGORITMA APRIORI”**, data mining membantu memprediksi stok barang penjualan sepatu, memprediksi barang-barang atau *brand* apa saja yang terjual ataupun tidak terjual pada suatu toko atau penjualan. Data mining algoritma apriori dapat memperkirakan stok pada penjualan sepatu. Algoritma apriori dapat menentukan dengan mudah barang yang mempunyai hubungan atau kaitan yang lebih dekat, berdasarkan kebutuhan permintaan konsumen apriori memiliki cara menentukan pola persediaan stok sepatu. Metode apriori menentukan persediaan stok sepatu *converse* dengan lebih akurat.
3. Penelitian oleh (Tarigan, 2017) Nomor 2 Vol.12 yang berjudul **“PENERAPAN ALGORITMA APRIORI UNTUK MENENTUKAN BARANG TERLARIS (STUDI KASUS: PT.INDOMARCO PRISMATAMA MEDAN)”**, setiap bulannya perusahaan mengalami

peningkatan pada penjualan barang yang terlaris, dengan bermacam-macam jenis produk yang ditawarkan, dibutuhkan algoritma apriori untuk mengetahui jenis barang yang paling diminati. Dan untuk memperoleh hasil yang lebih baik digunakan *tools tanagra*. Algoritma apriori merupakan jenis aturan asosiasi pada data mining. Seperti analisis pola *frequensi tinggi (frequent pattern mining)* yang merupakan salah satu langkah analisis asosiasi yang menarik perhatian banyak peneliti untuk memperoleh algoritma yang efisien. Seiring dengan berkembangnya usahawan yang membuka bisnis, sehingga untuk menentukan barang yang konsumen butuhkan perusahaan harus mengerti barang apakah yang konsumen minati. Maka itu diperlukan suatu strategi baru yang dapat memberikan suatu informasi yang strategis. Data mining atau sering disebut *Knowledge Discovery in Database* merupakan penguraian penemuan pengetahuan di dalam *database*. *Data Warehous* adalah paradigma baru yang secara spesifik dapat menyediakan informasi yang lebih baik.

4. Penelitian oleh (Purba *et al.*, 2018) Nomor 2 Vol.13 yang berjudul **“IMPLEMENTASI ALGORITMA APRIORI UNTUK Mencari RELASI PADA TRANSAKSI PEMBELIAN ALAT-ALAT KESEHATAN (STUDI KASUS: RS.ESTOMIHI)**, persediaan alat kesehatan dengan penerapan algoritma apriori mempengaruhi kualitas pelayanan rumah sakit, persediaan yang sangat besar akan menyebabkan beberapa resiko, seperti biaya penyimpana yang sangat besar atau pencurian karena itu diperlukan pencatatan seakurat mungkin untuk mengatasi persediaan alat-alat kesehatan. Algoritma apriori adalah suatu kombinasi *item* melalui

mekanisme perhitungan *support* dan *confidence* dari suatu hubungan *item* melalui pengambilan data dengan aturan asosiatif (*association rule*) untuk menentukan hubungan asosiatif, dalam bidang kesehatan serta menganalisa dan penentuan pembelian alat-alat kesehatan algoritma apriori sesuai untuk diterapkan bila terdapat beberapa hubungan *item*. Untuk mengetahui alat kesehatan apa saja yang biasa di beli oleh klinik. Penerapan algoritma apriori dapat membantu dalam membentuk kandidat kombinasi *item*, selanjutnya dilakukan pengujian apakah kombinasi tersebut memenuhi parameter *support* dan *confidence* maka hasil tersebut dapat membantu dalam penentuan pola pembelian alat kesehatan.

5. Penelitian oleh (Prayitno and Rasim, 2018) Nomor 3 Vol.18 yang berjudul **“ANALISIS PENJUALAN PRODUK RETAIL DENGAN METODE DATA MINING ASOSIASI”**, ada beberapa sudut pandang dalam melakukan analisis penjualan, salah satunya yaitu melakukan analisis pada barang yang sering muncul dan dibeli bersamaan oleh konsumen. Hal ini bertujuan supaya membentuk aturan asosiasi atau penyediaan barang memiliki sifat yang terhubung. Setiap barang akan memiliki nilai penunjang dan nilai kepastian. Data mining asosiasi dapat mengatasi masalah ini dengan menerapkan algoritma apriori pada proses pembacaan dan transaksi pada *database*. Untuk membantu mendapatkan hasil akhir yang optimal, digunakan *software Rapid Miner*. Berdasarkan hasil penelitian pada penelitian ini, maka disimpulkan bahwa data tersebut harus bisa membantu dalam proses analisis dan membuat keputusan, bukan hanya disimpan dan dibuatkan laporan secara periodik.

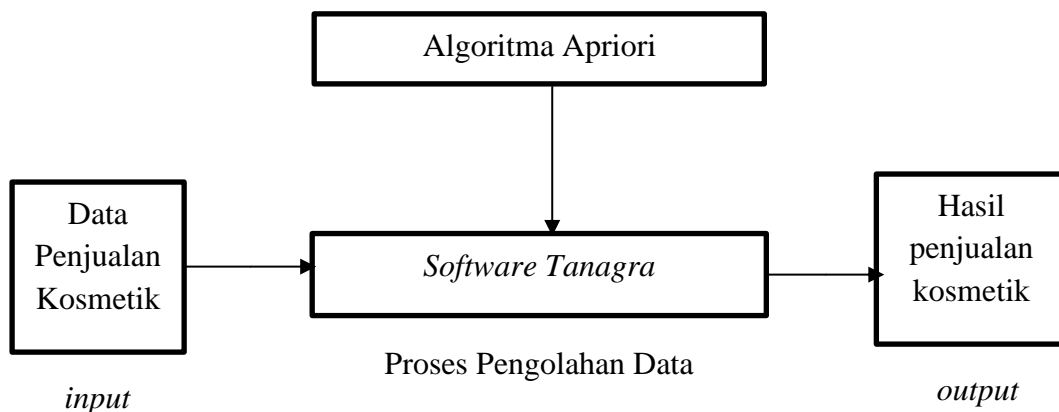
Menentukan keputusan strategi pemasaran dan penjualan bagi perusahaan dengan menggunakan metode data mining asosiasi dapat mengetahui pola pembelian oleh konsumen dan dapat dijadikan *knowledge* untuk perusahaan tersebut.

6. Penelitian oleh (Elisa, 2018) Nomor 2 Vol.2 yang berjudul “**MARKET BASKET ANALYSIS PADA MINI MARKET AYU DENGAN ALGORITMA APRIORI**”, data mining adalah teknik untuk menelusuri informasi baru dari Gudang data, informasi merupakan sesuatu yang sangat berharga atau penting karena dengan menguasai informasi maka dengan mudah untuk meraih sebuah tujuan, hal ini membuat setiap orang berusaha untuk memperoleh informasi, demikian juga pada usaha bisnis seperti minimarket Ayu di Kota Batam. Minimarket ini berlokasi disekitar rumah penduduk, dengan adanya kegiatan penjualan yang tentunya mempengaruhi tingkat penjualan, data transaksi penjualan akan semakin bertambah, menyebabkan penyimpanan data semakin besar. Data transaksi penjualan hanya sebagai arsip dan tidak dimanfaatkan dengan baik. Pada dasarnya kumpulan data memiliki informasi yang sangat bermanfaat. Mencari pola asosiasi sesuai pola belanja yang dilakukan konsumen dengan analisis keranjang pasar melalui algoritma apriori merupakan salah satu metode data mining, sehingga bisa diketahui *item-item* barang apa saja yang dibeli secara bersamaan. Hasil dari penelitian ini menemukan nilai *support* dan *confidence* tertinggi adalah Minyak dan Susu dengan nilai *support* 42.85% dan *confidence* 85.71%.

7. Penelitian oleh (Hezby *et al.*, 2016) Nomor 2 Vol.5 yang berjudul **“ASSOCIATION RULE MINING ON LIBRARY BOOKS LENDING DATA USING COMBINATION OF APRIORI AND JACCARD SIMILARITY”**, UPT Perpustakaan UNS memiliki 37.271 koleksi dan rata-rata 75.316 sirkulasi tahunan buku yang diurus oleh UNSLA (*UNS Library Automation*). Diperlukan analisis untuk menemukan informasi penting yang bisa dimanfaatkan untuk berbagai keperluan. *Association rule mining* merupakan salah satu mekanisme penambangan data untuk menelusuri pola hubungan dalam data keranjang pasar. Algoritma Apriori umumnya digunakan dalam asosiasi *rule mining*. Namun, Apriori memiliki keterbatasan dalam melakukan penambangan aturan asosiasi pada data yang jarang. Algoritma *Jaccard Similarity* digunakan untuk menemukan kesamaan antara dua set. Penerapan *Jaccard* mirip dengan penambangan aturan asosiasi dapat menemukan aturan asosiasi pada data jarang. Penelitian ini dilakukan untuk mengetahui konsistensi aturan asosiasi yang dihasilkan oleh kombinasi apriori dan persamaan *Jaccard* dibandingkan dengan *Apriori reguler* dan Persamaan *Jaccard* pada data peminjaman buku UPT *Library* UNS. Dengan mengamati konsistensi *association rule* yang dihasilkan permasalahan pada penelitian ini dapat terselesaikan. Pada penelitian ini kami menggunakan tiga metode yang berbeda yang telah memberikan hasil analisis konsistensi *association rule*. Semua metode yang telah dilakukan menghasilkan set *association rule* masing-masing. Maka untuk mendapatkan konsistensi kami melakukan perbandingan *association rule* yang dihasilkan pada setiap percobaan.

2.6 Kerangka Pemikiran

Rancang pemikiran yang akan dikembangkan dalam penelitian ini adalah sebagai berikut:



Gambar 2.5 Kerangka Pemikiran

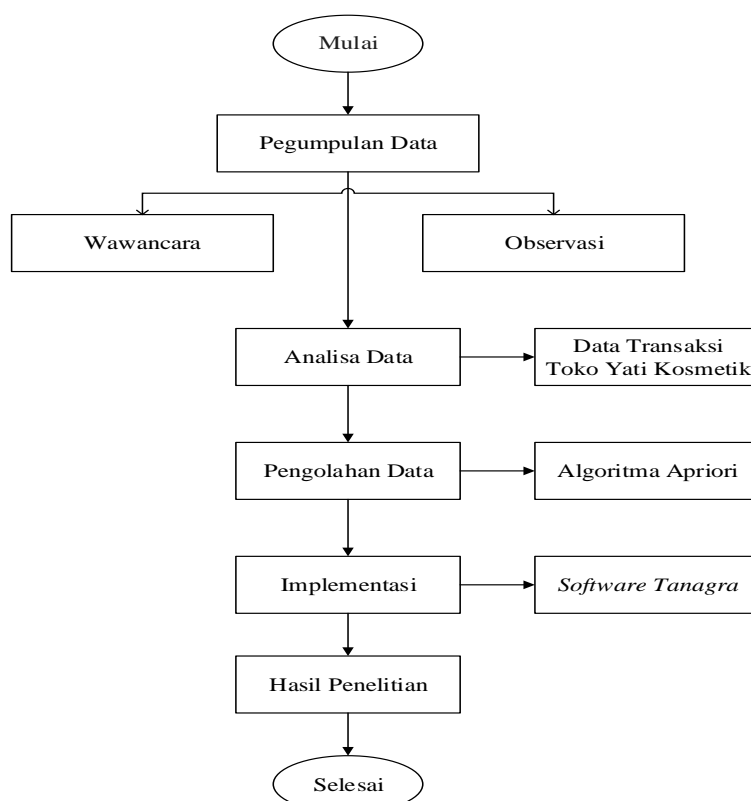
Pada gambar 2.1 bahwa pada proses pertama dari penelitian ini adalah adanya data penjualan kosmetik sebagai proses *input*, dimana data yang diinput berupa produk/item seperti merk dan jenis kosmetik, serta total transaksi yang akan disaring terlebih dahulu. Selanjutnya data ini akan diproses dengan algoritma apriori, setelah data selesai dioalah dan telah diperoleh nilai *minimum support* serta *minimum confidence*-nya. Maka selanjutnya data diuji dengan *software tanagra* untuk memperoleh hasil yang akurat dan menemukan *output* dari penelitian ini yaitu kosmetik yang paling sering dibeli konsumen, sehingga hasil tersebut dapat diterapkan pada toko untuk membantu dalam pemilihan ketersediaan dan *display* produk.

BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Desain penelitian yang dituangkan pada penelitian ini adalah sebagai berikut:



Gambar 3. 1 Desain Penelitian Dikembangkan Oleh Peneliti

Desain penelitian adalah alur penelitian yang akan dilakukan pada penelitian ini. Desain penelitian di atas mendeskripsikan proses dimulai kegiatan hingga selesai. Keterangan dari desain penelitian di atas yaitu:

1. Mulai

Pertama dengan menentukan masalah yang diteliti dan menentukan judul yang diteliti. Pada tahap ini juga terdapat identifikasi masalah yaitu tentang permasalahan penelitian yang akan dilakukan terhadap peneliti. Selanjutnya, yaitu analisa masalah untuk menyusun komponen yang sudah diuraikan dari komponen pembentuknya untuk diteliti dengan mengkaji secara detail pada suatu hal atau benda.

2. Pengumpulam data

Untuk medapatkan informasi yang diperlukan maka dilakukan pengumpulan data dalam upaya mecapai tujuan penelitian. Dengan melakukan pengumpulan data maka peneliti memperoleh data observasi langsung kepada pemilik toko, wawancara secara langsung kepada pihak toko Yati Kosmetik, dan dengan materi-materi studi kepustakaan mengenai data mining dan algoritma apriori yang didapatkan dari referensi jurnal nasional maupun internasional serta dari buku-buku.

3. Pengolahan data

Data yang sudah terkumpul dan di analisa selanjutnya akan diolah dan diproses menggunakan metode data mining yaitu algoritma apriori untuk mengetahui nilai *minimum support* dan nilai *confidience* dari masing-masing *item* produk kosmetik.

4. Implementasi

Implementasi pada penelitian ini adalah penerapan data penjualan menggunakan metode algoritma apriori dengan pengujian menggunakan *software tanagra*.

5. Hasil penelitian

Setelah melalui semua tahap atau proses penelitian maka akan memberikan sebuah hasil. Hasil tersebut kemudian diuji menggunakan perangkat lunak tanagra dengan menggunakan metode apriori, selanjutnya hasil dapat dilihat apakah sesuai dengan hasil pengujian.

6. Selesai

Tahap akhir dari penelitian, pada tahap ini akan di temukan hasil dari penelitian terhadap *item set* yang paling laris dan menentukan *display* produk pada toko.

3.2 Teknik Pengumpulan Data

Teknik pengumpulan data yang dilakukan peneliti yaitu seperti penjelasan berikut:

3.2.1 Observasi

Observasi adalah suatu metode pengumpulan data melalui pengamatan tindakan dalam situasi tertentu lalu mencatat kejadian yang diamati dengan sistematis dan memahami peristiwa yang diamati (Ni'matuzahroh and Prasetyaningrum, 2018). Observasi pada peneltian ini terhadap data pembelian,

persediaan produk dan tata letak adalah dengan langsung datang ke toko Yati Kosmetik.

3.2.2 Wawancara

Wawancara merupakan suatu cara untuk mendapatkan data tentang sesuatu dengan cara bertanya langsung kepada narasumber. Keberhasilan wawancara ditentukan oleh banyak faktor, yaitu: pewawancara, informasi/responden, topik penelitian yang ada dalam daftar pertanyaan dan faktor situasi dimana wawancara sedang berlangsung (*et al.*, 2020). Berdasarkan teori di atas, maka tahap wawancara pada penelitian ini dilakukan secara langsung kepada pemilik toko Yati Kosmetik demi memperoleh data penelitian.

3.2.3 Studi Pustaka

Sebagai penunjang pada penelitian ini, penulis mengumpulkan berbagai referensi dan literatur penelitian berupa buku, jurnal dan artikel yang berasal dari berbagai sumber yang berkaitan dengan objek penelitian.

3.3 Operasional Variabel

Menurut (Nasrudin, 2019), variabel harus didefinisikan secara operasional agar mudah dicari hubungannya antara satu variabel dengan lainnya dan pengukurannya. Tanpa operasional variabel, peneliti akan kesulitan dalam menentukan pengukuran hubungan antar variabel yang masih bersifat konseptual. Dalam penelitian ini operasional variabelnya sebagai berikut:

1. Data transaksi

Data transaksi merupakan total transaksi yang terjadi dalam waktu 36 bulan dari tahun 2017 sampai 2019. Yang dipengaruhi berdasarkan jumlah transaksi setiap produk yang akan menentukan kombinasi-kombinasi *itemset* mana yang paling banyak terjual.

2. Nama produk (*merk* kosmetik)

Nama produk yaitu *merk-merk* kosmetik yang sering dibeli oleh konsumen dan merupakan *item-item* sebagai pemilihan *stock*. *Merk* kosmetik merupakan variabel *input* pada penelitian ini. *Merk-merk* kosmetik apa saja yang paling banyak diminati oleh pembeli dan *merk-merk* yang tersedia di toko Yati Kosmetik.

3. Jenis kosmetik

Jenis kosmetik merupakan jenis-jenis kosmetik yang paling banyak diminati dan dibeli oleh pembeli dan merupakan jenis kosmetik sebagai pemilihan stok. Pada penelitian ini penulis memilih bedak dan lipstik sebagai *item* jenis kosmetiknya.

Tabel 3. 1 Operasional Variabel

No.	Merk Kosmetik	Jenis Kosmetik	Transaksi Penjualan
1	Wardah	Bedak (Bdk)	32
2	Pixy	Lipstik (Lips)	24
3	Make Over	Lipstik (Lips)	29
4	Purbasari	Lipstik (Lips)	24
5	Maybelline	Bedak (Bdk)	36
6	Mirabella	Bedak (Bdk)	19
7	La Tulipe	Lipstik (Lips)	34
8	Ponds	Bedak (Bdk)	17
9	Inez	Bedak (Bdk)	27
10	Viva	Bedak (Bdk)	28

Sumber: Data olahan peneliti 2020

3.4 Metode Analisis dan Rancangan Sistem

3.4.1 Metode Analisis

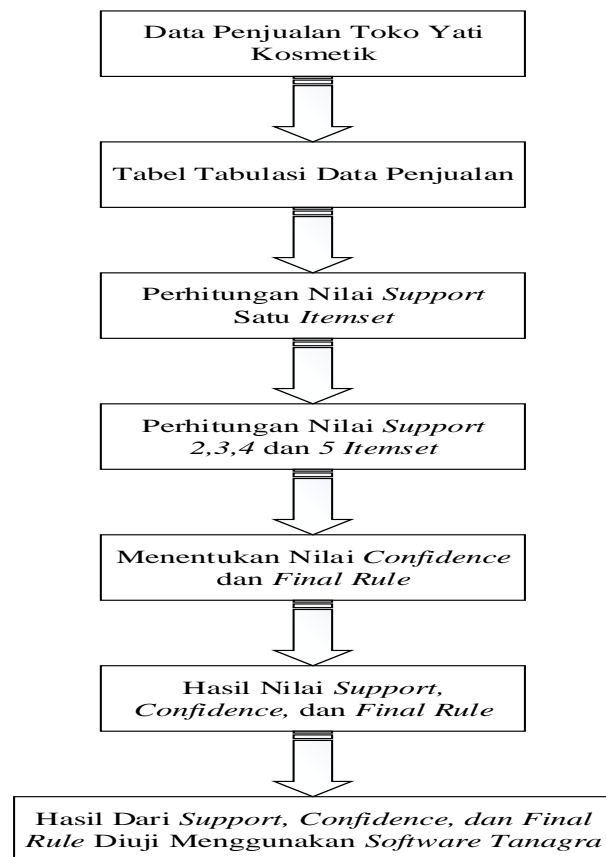
Metode analisis pada penelitian ini adalah metode apriori (asosiasi). *Association* juga disebut sebagai *Market Basket Analysis*, yaitu sebuah *problem* bisnis yang khas dalam menganalisa tabel transaksi penjualan dan mengidentifikasi produk-produk yang sering dibeli bersamaan oleh pelanggan, misalnya apabila orang membeli bedak biasanya juga dia membeli lipstik. Kesamaan yang ada

tersebut dari produk dan kebiasaan apa yang terjadi yaitu untuk kepentingan *cross-selling*.

Menurut (Sikumbang, 2018) analisis asosiasi juga dikenal sebagai salah satu teknik data mining yang menjadi dasar dari berbagai teknik data mining lainnya. Salah satu tahap analisis asosiasi yang menarik untuk menghasilkan algoritma yang efisien adalah analisis pola frekuensi tinggi. Penting atau tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter yaitu *support* dan *confidence*. *Support* (nilai penunjang) adalah persentase kombinasi *item* tersebut dalam *database*, sedangkan *confidence* (nilai kepastian) yaitu kuatnya hubungan antar-*item* dalam aturan asosiasi.

3.4.2 Rancangan Sistem

Tahapan -tahapan perancangan sistem dalam penelitian ini sebagai berikut:



Gambar 3. 2 Metode Rancang Sistem

Dari gambar 3.2 di atas menjelaskan proses awal sampai proses akhir. Proses awal dimulai dari data transaksi penjualan dari Toko Yati Kosmetik yang telah dianalisa terlebih dahulu oleh peneliti, data yang diperoleh yaitu data transaksi penjualan berdasarkan *merk* dan jenis kosmetik di Toko Yati kosmetik dari bulan Januari 2017 sampai bulan Desember 2019. Kemudian data tersebut dituangkan ke

dalam tabel tabulasi penjualan, lalu data-data akan dicari nilai *minimum support* dari 1 *itemset* hingga 5 *itemset*, setelah nilai *minimum support* diperoleh maka tahap selanjutnya yaitu menentukan nilai *confidence* hingga memperoleh nilai *final rule* yang dilakukan secara manual menggunakan metode algoritma apriori. Selanjutnya, tahap akhir yaitu pengujian menggunakan *software* Tanagra guna memperoleh hasil yang sama antara manual dengan *software* Tanagra.

3.5 Lokasi dan Jadwal Penelitian

Lokasi dan jadwal penelitian yang dilakukan adalah sebagai berikut:

3.5.1 Lokasi Penelitian

Penelitian ini dilakukan di Toko Yati Kosmetik yang berlokasi di Pasar Perumnas Sagulung Batu Aji, Batam.

3.5.2 Jadwal Penelitian

Penelitian ini dilakukan selama 6 bulan (selama satu semester) sesuai jadwal dari Universitas Putera Batam dalam kalender akademik. Penelitian ini dilakukan pada bulan Maret 2020 hingga bulan Agustus 2020. Jadwal yang dilakukan peneliti dapat dilihat pada tabel berikut:

Tabel 3. 2 Jadwal Penelitian

Kegiatan	Waktu Kegiatan																							
	Maret 2020				April 2020				Mei 2020				Juni 2020				Juli 2020				Agustus 2020			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Pengajuan Judul	■	■	■																					
Penyusunan BAB I				■	■	■	■																	
Penyusunan BAB II								■	■	■	■													
Penyusunan BAB III												■	■	■	■									
Penyusunan BAB IV																■	■	■	■	■				
Penyusunan BAB V																					■	■		
Pengumpulan Skripsi																							■	■

Sumber: Data olahan peneliti 2020