

KAJIAN TEORI

2.1 Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) untuk menguraikan pengetahuan dalam *database* yang juga di sebut dengan istilah data mining. Banyak jenis teknik analisa yang dapat diterapkan dalam data mining menurut definisi data mining yang sangat luas.

Menurut T. L. Seifert, H. Jiawei,dkk dalam (Ependi and Putra 2019) KDD adalah seperti "*interdisciplinary knowledge*" yaitu persimpangan dari berbagai bidang ilmu yang mencakup statistik, *database*, AI, visualisasi dan komputer parallel yang mempengaruhi pengetahuan. Sedangkan Y. Ramamohan,dkk, Mewar University and N. Rikhi berpendapat dalam (Ependi and Putra 2019) KDD juga memiliki pengertian sebagai suatu proses temuan dan evaluasi, pemilihan data, penyaringan data, transformasi data, dan presentasi.

2.2 Data Mining

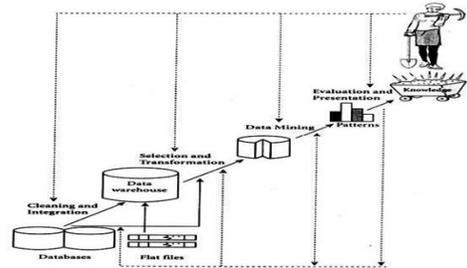
Data mining adalah suatu teknik tertentu dari data yang berukuran besar yang dilakukan untuk mengekstrak informasi atau pengetahuan (*knowledge*) penting. Sehingga menghasilkan informasi atau *knowledge* yang dapat digunakan untuk memperbaiki pengambilan keputusan. Proses penemuan informasi atau *knowledge* pada set data seperti orang melakukan kegiatan penambangan oleh karena itu disebut dengan penambangan data (Santosa and Umam 2018).

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

1. Fakta pola potensial dalam data sering kali tidak terlihat sehingga mencari pola yang terdapat di dalam data akan mengecewakan para pembuat keputusan yang tidak mempunyai pengalaman.
2. Jumlah data yang akan dianalisis secara manual terlalu besar.

2.3 Tahapan Proses Data Mining

Menurut Santoso dalam (Amrin 2018) data mining biasa disebut sebagai *Knowledge Discovery in Database*, merupakan pola atau hubungan set data berukuran besar, yang meliputi pengumpulan pemakaian historis. Pembuatan aplikasi data mining ditunjukkan pada gambar berikut:



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

Proses tahapan data mining terdiri dari langkah utama yaitu:

1. *Cleaning and Integration*
Pada tahap *cleaning*, yaitu melakukan penghilangan "noise" dan data yang tidak sesuai. Dan pada penggabungan sumber-sumber data dilakukan pada proses *integration*.
2. *Selection and Transformation*
Tahap *selection* adalah proses pengambilan data-data yang relevan untuk dianalisis. Sedangkan *transformation* adalah teknik transformasi dan penggabungan data kebentuk "mining" dan menghasilkan ringkasan atau penyatuan.
3. *Data Mining*
Pada tahap ini, di mulai dengan metode pengkajian diterapkan untuk mengekstraksi pola data.
4. *Evaluation and Presentation*
Tahap *evaluation* merupakan proses mengidentifikasi basis pengetahuan berdasarkan ukuran tertentu yang mewakili pola yang menarik. Selanjutnya, pada tahap *presentation* adalah proses untuk menampilkan pengetahuan kepada pengguna.

2.4 Langkah-langkah Data Mining

(Sulastri and Gufroni 2017), mendefinisikan bahwa kualitas data

dapat ditambah melalui *preprocessing data mining*, yang didapatkan langsung dari Toko Yati Kosmetik Sagulung, Batam dengan diolah menggunakan tahap-tahap *data cleaning*, *data integration*, *data selection*, dan *data transformation*. Data-data tersebut bersifat objektif, representatif, memiliki *sampling error* yang kecil, karena hal tersebut dilakukan agar data yang diolah lebih berkualitas.

1. *Data Cleaning*

Data cleaning merupakan proses untuk mengatasi data yang tidak konsisten, nilai yang hilang, dan *noise*.

2. *Data Integration*

Merupakan penggabungan data dari banyak *database*.

3. *Data Selection*

Data selection merupakan proses untuk tetap mempresentasikan data aslinya dengan meminimalkan jumlah data yang digunakan. *Data selection* terdiri dari *sampling*, *denoising*, dan *feature extraction*.

4. *Data Transformation*

Data transformation membantu meringankan pengguna saat proses mining dan memahami hasil yang diperoleh. *Data transformation* digunakan dalam perubahan bentuk dan format data.

2.5 Metode *Data Mining*

Pada penelitian ini penulis menggunakan metode asosiasi. (Badrul, Studi, and Informasi 2016), menyatakan bahwa atribut yang ditemukan atau muncul dalam satu waktu merupakan cara dari asosiasi data mining. *Market basket analysis* (analisis keranjang belanja) adalah salah satu implementasi dari asosiasi, seperti yang akan dibahas pada penelitian ini.

1. *Association Rule*

(Choiriah et al. 2019), menyatakan bahwa aturan asosiasi merupakan cara untuk mendapatkan aturan asosiasi antara suatu kombinasi (gabungan) *item*. Untuk hasil olahan data dan perhitungan tertentu terdapat suatu ukuran ketertarikan (*interestingness measure*), dalam

menentukan suatu aturan asosiasi, yaitu:

- a) *Support*, merupakan probabilitas total seluruh transaksi pelanggan saat membeli produk secara bersamaan. Layak atau tidaknya suatu *item/itemset* untuk dicari nilai *confidence*-nya (misalkan dengan total transaksi yang ada) berdasarkan ukuran dari probabilitas tersebut. Dapat dilihat dari seberapa besar *item X* dan *Y* memiliki tingkat dominasi dibeli bersamaan.
- b) *Confidence* (tingkat kepercayaan), adalah dimana suatu produk sudah pasti dibeli atau kemungkinan peluang kejadian beberapa produk yang dibeli bersamaan. (seperti: seberapa sering *item Y* dibeli apabila konsumen membeli *item X*)

Ukuran *support* dan *confidence* berfungsi untuk menentukan pola asosiasi, sebagai batasan (*threshold*) atau pembanding dengan yang telah ditetapkan oleh pengguna. Batasan tersebut terdiri dari batasan minimum dari nilai *support* (*minimum support*) dan batasan minimum dari nilai *confidence* (*minimum confidence*).

2.6 Algoritma Apriori

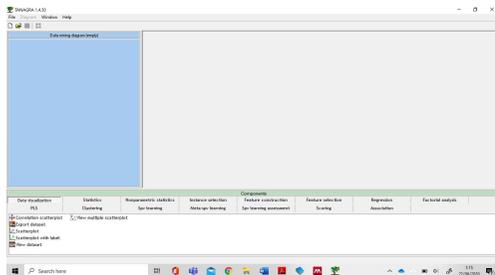
Menurut C. N. Parkinson, dkk dalam (Purba and Buulolo 2020) algoritma apriori berfungsi untuk menemukan *frequent itemset* pada sekelompok data. Dengan iterasi ke-*k*, semua *itemset* akan ditemukan yang mempunyai *k item*, disebut dengan *k-itemset*. Berikut ini adalah dasar algoritma apriori:

1. Membangun Analisa pola frekuensi tinggi
2. Pembentukan aturan asosiatif
3. Analisa penerapan algoritma apriori

2.7 *Software Tanagra*

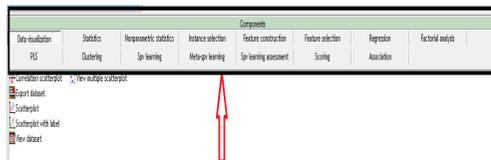
Menurut (Badrul et al. 2016) tanagra yaitu sebuah *software* untuk tujuan akademik dan penelitian data mining. Beberapa metode data mining yang diusulkan tanagra yaitu analisis eksplorasi data, pembelajaran statistik, pembelajaran mesin dan daerah *database*.

Berikut adalah tampilan awal pada aplikasi tanagra:



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

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



Gambar 3. fitur-fitur Komponen dalam Tanagra (Sumber: Data Peneliti 2020)

2.8 Penelitian Terdahulu

Penelitian terdahulu merupakan suatu tumpuan pada penelitian. Berikut merupakan penelitian-penelitian yang 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

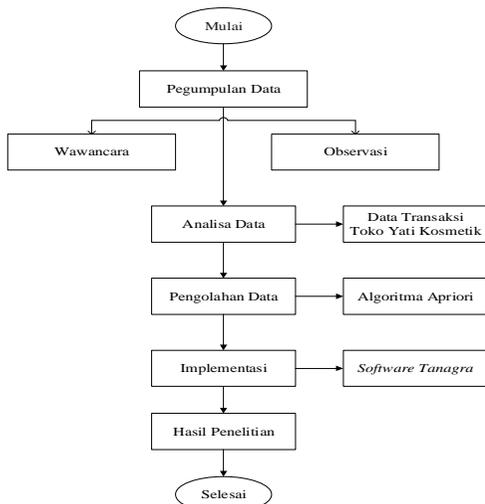
mulai dari ratusan bahkan ribuan data penjualan setiap bulannya itu akan menjadi sulit atau tidak mudah. Dengan menggunakan algoritma *K-Means Clustering* yang merupakan salah satu teknik dalam data mining. Pengelompokan data penjualan pada Toko Genta Corp memperoleh hasil rekomendasi barang yang laris, kurang laris dan cukup laris melalui penerapan metode *K-Means*.

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.

METODE PENELITIAN

3.1 Desain Penelitian

Desain penelitian yang dituangkan dalam penelitian ini adalah sebagai berikut:



Gambar 4. Desain penelitian
(Sumber: Data Peneliti 2020)

Desain penelitian merupakan tahap atau kegiatan dalam penelitian yang dilakukan dalam penelitian ini. Desain penelitian di atas merupakan gambaran proses dimulai kegiatan hingga selesai. Penjelasan dari desain penelitian di atas sebagai berikut:

1. Mulai

Tahap awal dalam menetapkan judul dan masalah yang diteliti. Pada tahap ini juga terdapat identifikasi masalah yaitu tentang permasalahan dalam penelitian yang dilakukan oleh peneliti. Selanjutnya, yaitu analisa masalah untuk menyusun komponen yang sudah diuraikan dari komponen pembentuknya untuk diteliti dengan mengamati secara detail pada suatu hal atau benda.

2. Pengumpulan data

Dalam upaya mencapai tujuan penelitian dilakukan pengumpulan data agar memperoleh informasi yang dibutuhkan. Data yang diterapkan seperti data wawancara secara langsung kepada pemilik toko Yati Kosmetik, pengamatan langsung kepada pihak toko, dan dengan materi-materi studi kepustakaan yang diperoleh 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 untuk mengetahui nilai *minimum*-nya yaitu *minimum support* dan *minimum confidence* dari masing-masing *item* produk kosmetik menggunakan metode data mining yaitu algoritma apriori.

4. Implementasi

Pengujian melalui *software tanagra* pada data penjualan dengan implementasi data ke dalam metode algoritma apriori.

5. Hasil penelitian

Setelah melalui semua tahap atau proses penelitian maka akan memberikan sebuah hasil. Untuk melihat apakah hasil yang diteliti sesuai dengan pengujian, hasil tersebut akan diuji dengan *software tanagra* metode apriori.

6. Selesai

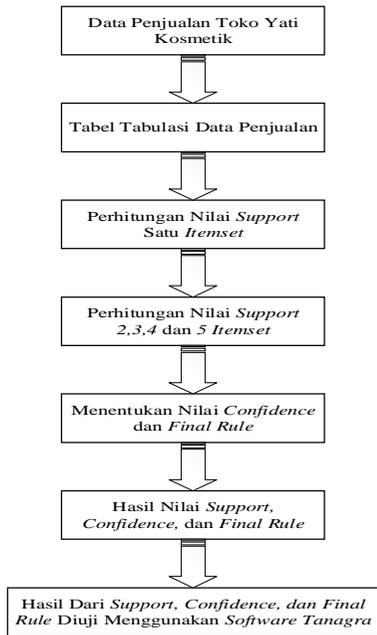
Pada tahap ini akan di temukan hasil dari penelitian terhadap *item set* yang paling laris dan menentukan *display* produk pada toko.

3.2 Metode Analisis

Metode analisis adalah metode apriori (asosiasi). *Market Basket Analysis* atau yang juga disebut *association*, menganalisa tabel transaksi dan mengidentifikasi produk-produk yang sering dibeli bersamaan oleh pelanggan termasuk dalam sebuah *problem* bisnis, misalnya apabila orang membeli bedak biasanya juga dia membeli lipstik.

3.3 Rancangan Sistem

Tahapan-tahapan perancangan sistem dalam penelitian ini sebagai berikut:



Gambar 5. Metode Rancangan Sistem
(Sumber: Data Peneliti 2020)

HASIL DAN PEMBAHASAN

4.1 Analisis Data

Data transaksi penjualan kosmetik 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, agar dapat mempermudah menemukan hasil analisa dari data transaksi penjualan kosmetik pada toko Yati Kosmetik maka dapat dilihat pada tabel berikut:

Tabel 1. Pola Iterasi 1 Itemset

No.	Nama Produk Kosmetik	Jumlah	Support (%)
1	Bdk Wardah	32	88%
2	Lips Pixy	24	66%
3	Lips MakeOver	29	80%
4	Lips Purbasari	24	66%
5	Bdk Maybellin	36	100%
6	Bdk Mirabella	19	52%
7	Lips La Tulipe	34	94%
8	Bdk Ponds	17	47%
9	Bdk Inez	27	75%
10	Bdk Viva	28	77%

(Sumber: Data Peneliti 2020)

Tahap pertama dalam proses pencarian nilai *minimum support* ialah mencari nilai iterasi dari 1 *itemset* berdasarkan tabel dari item-item produk kosmetik pada tabel 4.1 transaksi penjualan (data merk dan jenis kosmetik keluar) dan pada tabel 4.2 tabulasi penjualan dengan nilai *minimum support* yang telah ditentukan yaitu 0,44 atau 44% terlihat seperti yang terlihat pada tabel berikut ini :Perhitungan nilai *minimum support* berdasarkan rumus 2.1 dengan iterasi 1 *itemset* sebagai berikut:

1. $Support (Bdk Wardah) = \frac{32}{36} \times 100\% = 88\%$
2. $Support (Lips Pixy) = \frac{24}{36} \times 100\% = 66\%$

3. $Support (Lips MakeOver) = \frac{29}{36} \times 100\% = 80\%$
4. $Support (Lips Purbasari) = \frac{24}{36} \times 100\% = 66\%$
5. $Support (Bdk Meybellin) = \frac{36}{36} \times 100\% = 100\%$

Pembentukan Pola Aturan Asosiasi Setelah hasil ditemukan berdasarkan kombinasi *itemset* yang telah dihitung dengan nilai *minimum support* 44%, maka selanjutnya akan menghitung nilai *minimum confidence* untuk memperoleh *confidence rules* dari hasil perhitungan kombinasi 2 *itemset*, 3 *itemset*, 4 *itemset* dan 5 *itemset* dapat dilihat pada tabel berikut:

Tabel 2. Pola Aturan Assosiasi

No.	Nama Produk Kosmetik	Jumlah	Support (%)
1	Bdk wardah – Lips pixy	21	58%
2	Bdk wardah – Lips makeover	25	69%
3	Bdk wardah – Lips purbasari	20	55%
4	Bdk wardah – Bdk meybellin	32	88%
5	Bdk wardah – Bdk mirabella	17	47%
6	Bdk wardah – Lips latulip	30	83%
7	Bdk wardah – Bdk Inez	24	66%
8	Bdk wardah – Bdk viva	24	66%
9	Lips pixy – Lips makeover	20	55%
10	Lips pixy – Bdk meybellin	24	66%
11	Lips pixy – Lips latulip	22	61%
12	Lips pixy – Bdk inez	20	55%
13	Lips pixy – Bdk viva	17	47%
14	Lips makeover – Lips purbasari	19	52%
15	Lips makeover – Bdk meybellin	29	80%

(Sumber: Data Peneliti 2020)

Berdasarkan perhitungan nilai *minimum confidence* diatas dengan nilai *confidence* yang telah ditentukan yaitu 0,80 atau 80% terdapat 9 *rules*, dapat dilihat pada tabel berikut:

Tabel 3. Final Rules Assosiasi

No	Nama Produk Kosmetik	Jumlah	Support (%)	Confidence (%)
1	Bdk mirabella – bdk inez	16/19	44%	84%
2	Bdk meybelli, bdk mirabella – bdk inez	16/19	44%	84%
3	Bdk mirabella – bdk meybellin, bdk inez	16/19	44%	84%
4	Lips pixy – bdk inez	20/24	55%	83%
5	Lips pixy – bdk mirabella, bdk inez	20/24	55%	83%
6	Bdk meybellin, lips pixy – bdk inez	20/24	55%	83%
7	Bdk mirabella – bdk meybellin, lips makeover	17/19	47%	89%
8	Bdk mirabella – lips makeover	17/19	47%	89%
9	Bdk meybellin, bdk mirabella – lips makeover	17/19	47%	89%

(Sumber: Data Peneliti 2020)

4.2 Hasil Pengujian

Memanfaat metode algoritma apriori dengan menentukan nilai *minimum support* dan *minimum confidence* berdasarkan data-data yang telah diolah menggunakan teori data mining maka menghasilkan sebanyak 5 iterasi kombinasi nama produk kosmetik serta menghasilkan *final rules* sebanyak 9 *rules* dari Toko Yati Kosmetik. Selanjutnya data-data tersebut dibuktikan kedalam

aplikasi TANAGRA dengan langkah-langkah sebagai berikut:

1. Membuat format tabulasi data produk kosmetik kedalam *Ms.Excel*



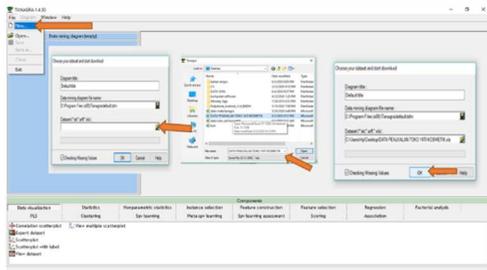
Gambar 6. Format Tabulasi

2. Halaman Kerja Aplikasi TANAGRA



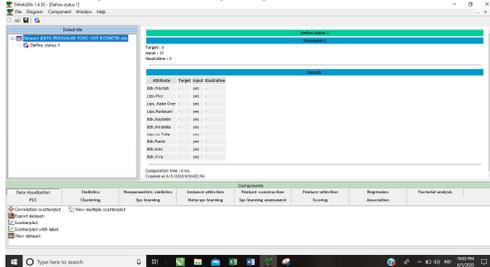
Gambar 7. Halaman Kerja Aplikasi TANAGRA

3. Memasukan Data Ke Aplikasi TANAGRA



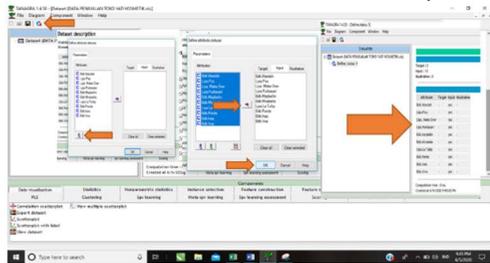
Gambar 8. Memasukkan Data ke Aplikasi TANAGRA

4. Tampilan data penjualan Toko



Gambar 9. Tampilan Data Di Aplikasi TANAGRA

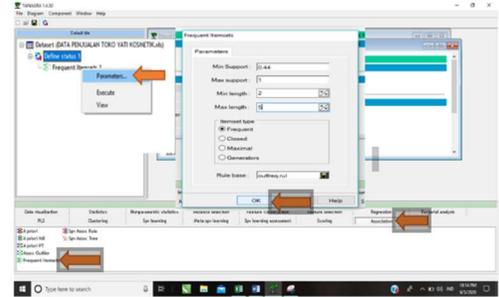
5. Memindahkan atribut ke kotak input



Gambar 11. Proses Memasukkan Atribut Item ke Dalam TANAGRA

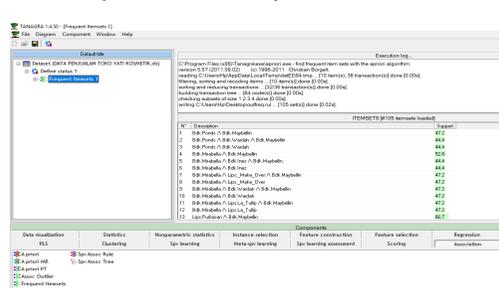
Meta Ulkhaire

6. Mengatur frequent itemset.



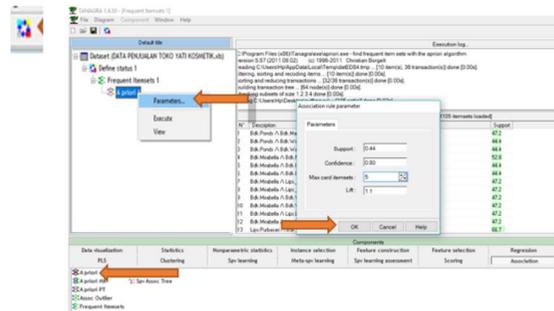
Gambar 12. Mengatur Nilai Frequent Itemset

7. Tampilan inilah frequent itemset



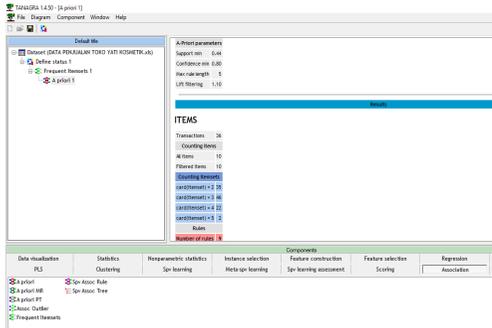
Gambar 13. Hasil Frequent Itemset

8. Menentukan nilai confidence.



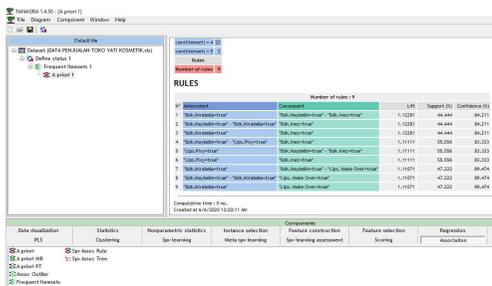
Gambar 14. Mengatur Nilai Confidence

9. Tampilan nilai confidence dengan jumlah atribut sebanyak 10 items dengan 36 transaksi.



Gambar 15. Hasil Confidence

10. Hasil final rules terdapat 9 rules dengan 2 itemset terdapat 35 rules, 3 itemset dengan 46 rules, 4 itemset dengan 22 rules dan 5 itemset dengan 2 rules



Gambar 16. Final Rules

4.3 Pembahasan

Setelah hasil diuji dengan aplikasi TANAGRA maka diperoleh hasil yang memenuhi minimum support 44% dan minimum confidence 80% sebanyak 9 rule, yaitu:

1. Jika membeli bedak mirabella maka membeli bedak inez
2. Jika bedak meybelling dan bedak mirabella maka membeli bedak inez
3. Jika membeli bedak mirabella maka membeli bedak meybelling dan bedak inez
4. Jika membeli lipstik pixy maka membeli bedak inez
5. Jika membeli lipstik pixy maka membeli bedak mirabella dan bedak inez
6. Jika membeli bedak meybelling dan lipstik pixy maka membeli bedak inez
7. Jika membeli bedak mirabella maka membeli bedak meybelling dan lipstik makeover

Meta Ulkhaire

8. Jika membeli bedak mirabella maka membeli lipstik makeover
9. Jika membeli bedak meybelling dan bedak mirabella maka membeli lipstik makeover.

Dengan hasil yang sudah didapat seperti *item* produk kosmetik yang paling sering dibeli konsumen, maka owner Toko Yati Kosmetik dapat mengatur display produk kosmetik berdasarkan hasil yang diperoleh dengan metode *algoritma apriori* dan pengujian dengan aplikasi TANAGRA yaitu bedak mirabella, bedak inez, bedak meybelling, lipstik pixy dan lipstik makeover secara berdekatan guna memudahkan konsumen saat membeli serta dapat mengatur strategi dalam menentukan ketersediaan stok kosmetik guna menjaga kestabilan penjualan kosmetik yang paling dibutuhkan konsumen serta dapat meningkatkan penjualan produk kosmetik yang sangat sedikit diminati berdasarkan itemset yang diperoleh agar penjualan produk kosmetik yang masih ada stok bisa terjual.

SIMPULAN

Kesimpulan yang dapat penulis jabarkan berdasarkan hasil dari penelitian ini, yaitu sebagai berikut:

1. Pengimplementasian algoritma apriori pada Toko Yati Kosmetik, dengan mengolah hasil transaksi seperti nama produk (*merk* kosmetik), jenis kosmetik, dan data transaksi. Dari kosmetik yang paling banyak terjual maka dapat ditentukan penyusunan letak dari kosmetik tersebut, serta akan memudahkan penjual dalam pemilihan *stock* produk dan meringankan pembeli dalam membeli.
2. Produk yang memenuhi minimum support 44% dan minimum confidence 80% dengan menggunakan metode algoritma apriori sebagai metode untuk mengetahui produk yang paling sering dibeli oleh konsumen toko Yati Kosmetik, dengan hasil barang yang paling sering dibeli konsumen yaitu bedak mirabella, bedak inez, bedak meybelling, lipstik pixy, dan lipstik makeover.

3. Hasil final *rule* yang diperoleh dari nilai *minimum support* dan *minimum confidence* yang telah ditentukan adalah 9 *rules* dengan *confidence* 83%, 84%, dan 89%.

DAFTAR PUSTAKA

Amrin, Amrin. 2018. "Aplikasi Diagnosa Penyakit Tuberculosis Menggunakan Algoritma Naive Bayes." *Jurikom* 5(5):498–502.

Badrul, Mohammad, Program Studi, and Sistem Informasi. 2016. "ALGORITMA ASOSIASI DENGAN ALGORITMA APRIORI UNTUK ANALISA DATA PENJUALAN | Jurnal Pilar Nusa Mandiri." (2):121–29.

Choiriah, Wirdah, Sistem Informasi, Sistem Informasi, and Universitas Lancang Kuning. 2019. "PENERAPAN DATA MINING MENGGUNAKAN METODE ASSOCIATION PENDAHULUAN Perkembangan Teknologi Komputer Mengakibatkan Hampir Seluruh Aktivitas Kehidupan Manusia Menggunakan Bantuan Komputer , Hal Ini Berdampak Pada Peningkatan Data Komputer Secara Dalam Operasi." V(2):193–98.

Ependi, Usman, and Ade Putra. 2019. "Solusi Prediksi Persediaan Barang Dengan Menggunakan Algoritma Apriori (Studi Kasus: Regional Part Depo Auto 2000 Palembang)." *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)* 5(2):139.

Purba, Citra Verawati, and Efori Buulolo. 2020. "Implementasi Algoritma Apriori Untuk Menentukan Pola Data Penyakit Pada Anak Usia Dini (Studi Kasus : RS . Estomihi)." 7(2):308–13.

Sulastri, Heni, and Acep Irham Gufroni. 2017. "Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia." *Jurnal Nasional Teknologi Dan Sistem Informasi* 3(2):299–305.

Wahyuni, Sri, Suherman, and lumalo portibi Harahap. 2017. "Implementasi Data Mining Dalam

Memprediksi Stok Barang Menggunakan Algoritma Apriori." *Prosiding SINTAK 2017* 2(2):31–39.

Wijayanti, Aris Wijaya. 2017. "Analisis Hasil Implementasi Data Mining Menggunakan Algoritma Apriori Pada Apotek." *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)* 3(1):60.

	<p>Penulis pertama, Meta Ulkhairi,, merupakan mahasiswa Prodi Teknik Informatika Universitas Putera Batam.</p>
	<p>Penulis kedua, Ellbert Hutabri, S.Kom., M.Kom, merupakan Dosen Prodi Teknik Informatika Universitas Putera Batam.</p>