

BAB 2

LANDASAN TEORI

2.1 *Data Mining*

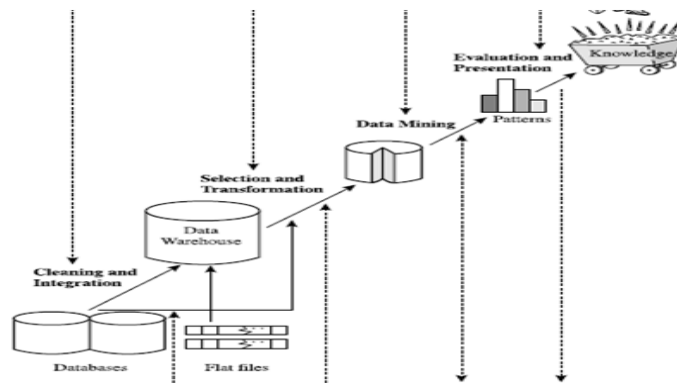
Data Mining sebuah proses pengolahan data menggunakan teknik statistik, matematika, kecerdasan buatan, *meachine learning*, yang bertujuan untuk ekstraksi dan identifikasi sebuah informasi dari kumpulan data (Arifin, 2015). Definisi lain dari *Data Mining* adalah sebuah proses pencarian informasi dan penganalisisan pola menggunakan sebuah metode ataupun teknik tertentu (Gunadi dan Sensuse, 2016). *Data Mining* adalah penggabungan teknik dari pembelajaran mesin, pengenalan pola, database, statistik, dan visualisasi yang termasuk dalam bidang keilmuan komputer untuk pengambilan informasi database yang besar, Larose (2005) dalam (Ardani dan Fitriana, 2016). *Data Mining* adalah proses pencarian pola dan informasi menarik yang terdapat dalam data, dilakukan dengan *data mining* yang memiliki teknik serta metode yang bervariasi. Metode algoritma dapat dipilih secara tepat tergantung tujuan dan proses dalam KDD secara keseluruhan.

Berdasarkan definisi-definisi dari *Data Mining* Tampubolon, Saragih, Reza, Epicentrum, dan Asosiasi (2013), hal hal terpenting yang terkait dengan *data mining* adalah,

1. Pemrosesan otomatis terhadap data yang telah tersedia.
2. Data yang dapat digunakan dalam pemrosesan adalah data dalam jumlah besar.
3. Menemukan hubungan dan pola yang diharapkan mampu memberikan indikasi yang bermanfaat merupakan tujuan utama *Data Mining*.

2.1.1 *Proses Data Mining*

Proses *data mining* dapat dilihat pada Gambar 2.1.



Gambar 2.1. Tahapan proses KDD

Tahapan-tahapan proses KKD (Anggraeni, Saputra, dan Noranita, 2013).

1. Pembersih data (*data cleaning*)
Menghilangkan *noise* serta data yang tidak konsisten atau tidak relevan merupakan kegiatan dalam pembersihan data. Data dari database yang memiliki atribut dan item yang tidak lengkap dan data yang hilang, juga termasuk data yang tidak valid atau data yang hanya sekedar salah pengetikan. Data yang tidak relevan itu juga disarankan untuk dihilangkan. Hasil data dan aktualifitas dari teknik data *mining* sangat bergantung pada proses pembersihan data, karena bila adanya kesalahan data yang ditangani akan berkurang jumlah dan kompleksitasnya.
2. Integrasi data (*data integration*) Kumpulan data yang digabungkan menjadi sebuah database baru. Pengerjaan integrasi data harus dilakukan secara cermat dan teliti untuk menghindari kesalahan yang dapat mengganggu hasil pengambilan keputusan.
3. Seleksi data (*data selection*) Pemilihan data yang sesuai dalam proses analisis yang digunakan dalam database, proses ini dilakukan dikarenakan beberapa data dalam database seringkali tidak terpakai.
4. Transformasi data (*data transformation*)
Pengubahan data kedalam bentuk format yang sesuai untuk pemrosesan. Sebagai contoh salah satu metode standar seperti analisis asosiasi dan *clustering* hanya bisa menerima *input* data *kategorikal*. Karenanya data berupa angka *numerik* yang berlanjut perlu dibagi-bagi menjadi sebuah interval. Proses ini sering disebut transformasi data.
5. Proses *mining*
Penerapan metode yang bertujuan untuk menemukan pengetahuan berharga dan tersembunyi, yang merupakan proses utama dalam penelitian ini.
6. Evaluasi pola (*pattern evaluation*)
Proses pengidentifikasian yang nantinya hasil dari tahapan ini dapat berupa pola-pola yang unik dalam bentuk analisis *knowledge based* maupun model prediksi dan evaluasi untuk menilai benarkah *hipotesa* yang ada memang tercapai atau sebaliknya.
7. Presentasi pengetahuan (*knowledge presentation*)
Metode yang dilakukan untuk memperoleh pengetahuan dalam penyajiannya secara visualisasi. Tahap terakhir adalah proses menerjemahkan keputusan maupun aksi dari hasil analisis yang didapat. Setelahnya diharapkan penyajian kedalam bahasa pengetahuan dapat dipahami masyarakat umum sebagai salah satu tahapan yang diperlukan. Dalam penyajian ini, penyajian

ini juga diharapkan dapat membantu menyajikan hasil data *mining*.

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas atau pekerjaan yang dapat dilakukan Larose dalam (Gunadi dan Sensuse, 2016).

1. Deskripsi
Deskripsi dari pola dan tren yang sering terjadi pendeskripsian dalam data mining harus dilakukan secara transparan, dan nantinya hasil yang didapat diharapkan dapat menjelaskan pola secara deskriptif.
2. Estimasi
Dalam estimasi, record dapat membantu pembangunan model yang nantinya dapat menghasilkan nilai berdasarkan variabel target untuk penentuan nilai prediksi. Dan untuk peninjauan selanjutnya estimasi nilai berdasarkan variabel target dibentuk berdasarkan nilai dari variabel prediksi.
3. Prediksi
Sama layaknya klasifikasi dan estimasi, terkecuali dengan prediksi, nilai dari hasil prediksi akan digunakan dimasa mendatang.
4. Klasifikasi
Dalam klasifikasi, terdapat hal yang harus dicapai untuk variabel kategori. Sebagai contoh penggolongan pendapatan dapat dipisahkan dalam tiga kategori, contohnya; kategori makanan, kategori minuman, dan kategori snack.
5. Pengklasteran (*Clusterring*)
Clustering merupakan teknik penambangan data yang berupaya mengidentifikasi data yang memiliki kemiripan. Proses ini membantu untuk memahami perbedaan dan persamaan di antara data.
6. Asosiasi Teknik penambangan data ini membantu untuk menemukan hubungan antara dua item atau lebih. Ini menemukan pola tersembunyi di set data.

2.2 Association Rule Mining

Analisis asosiasi atau *association rule mining* adalah salah satu teknik dalam data mining yang menentukan aturan asosiatif antara suatu kombinasi item dengan item lainnya. Seperti salah satu aturan asosiatif dari analisis pembelian di suatu pasar swalayan dapat diketahui berapa besar kemungkinan seorang pelanggan membeli roti bersama dengan susu. Dengan pengetahuan yang dihasilkan, pemilik pasar swalayan dapat mengatur penempatan barang secara bersama. *Association Rule Mining* dikenal juga sebagai *market basket analysis*, Erwin (2009) dalam (Fatihatul dkk., 2011). *Association rule* juga merupakan proses data *mining* untuk menentukan semua aturan asosiatif yang memenuhi syarat minimum untuk support (*min-sup*) dan confidence (*minconf*) pada sebuah database. Kedua syarat tersebut akan di-

gunakan untuk *interesting association rules* dengan dibandingkan dengan batasan yang telah ditentukan, yaitu *minsup* dan *minconf*. Berdasarkan definisi tersebut didapatkan pencarian pola kaidah asosiasi menggunakan dua buah pengukuran nilai yaitu dukungan (*support*) dan kepercayaan (*confidence*) yang memiliki nilai antara 0%-100%. Berikut penjelasan mengenai *support* dan *confidence*.

Dalam menentukan suatu *association rule*, terdapat ukuran yang menyatakan bahwa suatu informasi atau *knowledge* dianggap menarik (*interestingness measure*). Ukuran ini didapatkan dari hasil pengolahan data dengan perhitungan tertentu. Untuk mengukur *interestingness measure*, dapat digunakan variable berikut ini (Sumangkut, Lumenta, dan Tulenan, 2016).

2.2.1 Analisa Pola Frekuensi Tinggi Support

Pada tahapan ini dilakukan pencarian kombinasi item yang memenuhi syarat minimum dari nilai support didalam basis data. Nilai *support* pada sebuah item diperoleh dengan menggunakan rumus sebagai berikut (Pane, 2013).

Nilai *Support* dari *Item* diperoleh dengan menggunakan rumus Persamaan 2.1.

$$Support(A) = \frac{JumlahTransaksiMengandungA}{TotalTransaksi} \quad (2.1)$$

Sedangkan pada nilai *Support* dari 2 *item* diperoleh dengan menggunakan rumus Persamaan 2.2.

$$Support(A \cap B) = \frac{\sum JumlahTransaksiMengandungA}{\sum TotalTransaksi} \quad (2.2)$$

2.2.2 Pembentukan Aturan Asosiasi Confidence

Pencarian aturan asosiasi dilakukan dengan memenuhi syarat minimum untuk *confidence* yang menghitung *confidence* aturan asosiatif $A \rightarrow B$. Nilai *confidence* dari aturan $A \rightarrow B$ diperoleh dengan rumus sebagai berikut (Pane, 2013).

Nilai pada *Confidence* diperoleh dengan menggunakan rumus Persamaan 2.3.

$$Confidence = \frac{\sum JumlahTransaksiMengandungAdanB}{\sum JumlahTransaksiMengandungA} \quad (2.3)$$

2.3 Algoritma Frequent Pattern Growth

Frequent Pattern Growth algoritma yang paling sering dipakai dalam menentukan hubungan asosisasi pada data sejumlah kumpulan dengan melihat pola

yang sering muncul secara bersamaan (Sijabat, 2015). Algoritma *Frequent Pattern Growth* FP-Growth merupakan sebuah algoritma pengembangan untuk menutupi kekurangan dari algoritma apriori. Dalam proses pengerjaannya algoritma Apriori memakan waktu yang sangat besar saat proses pemindaian database. Tentunya ini sangat berpengaruh pada efisiensi waktu dalam proses data mining. Algoritma FP-Growth diusulkan untuk mengatasi kelemahan pada algoritma apriori, karena disebut tidak perlu untuk menghasilkan *generate candidate item sets* dan dianggap lebih efisien. FP-Growth dengan membaca kumpulan data satu transaksi dalam waktu dan pemetaan setiap transaksi ke dalam bagian FP-Tree yang akan dibangun (Wang, Lee, dan Pang, 2010).

Karakteristik algoritma FP-Growth adalah struktur data yang digunakan *tree* yang disebut dengan FP-Tree. Dengan menggunakan FP-Tree, algoritma *Frequent Pattern Growth* FP-Growth dapat langsung mengekstrak *frequent itemset* dari FP-Tree. FP-tree adalah struktur penyimpanan data yang dikompresi ke dalam visualisasi penggambaran. FP-tree dibuat dengan proses memetakan satu persatu data transaksi ke menjadi setiap lintasan dalam FP-tree. Pada setiap transaksi yang telah dipetakan, terdapat kemungkinan transaksi yang memiliki *item* yang sama, lintasannya memungkinkan untuk saling tertimpa. Semakin banyak data transaksi yang memiliki *item* yang sama, maka proses kompresi dengan struktur data FP-tree akan semakin efektif. Penggalian *itemset* yang *frequent* dengan menggunakan algoritma FP-Growth akan dilakukan dengan cara membangkitkan struktur data *tree* FP-tree.

Dalam FP-Tree terdapat berbagai simpul yang harus memiliki tiga *field* (Ghozali, Sugiharto, dkk., 2017).

1. *Item-name*: sebuah simbol yang akan dipresentasikan harus bisa menginformasikan item dengan jelas
2. *Count*: Simpul yang akan dilewati harus dipresentasikan dengan jumlah transaksi
3. *Node-Link*: Simpul-simpul yang terhubung dengan item-name yang serupa dengan penghubung yang ada, atau jumlah nilai yang dihitung *null* jika kosong.

2.3.1 Tahapan Algoritma *Frequent Pattern Growth*

Algoritma FP-Growth dibagi menjadi tiga langkah utama (Ikhwan dan Nofriansyah, 2015).

1. Tahap Pembangkitan (*Conditional pattern base*.) Kumpulan *subdatabase* yang terdiri dari *prefix path* (lintasan *prefix*) dan *suffix pattern* (pola akhiran), dan didapatkan melalui FP-Tree yang sebelumnya telah dibangun.

2. Tahap pembangkitan (*Conditional FP-Tree*) Tahapan ini *support count* dari item pada *conditional pattern base* akan dijumlahkan, dan setiap *item* yang memiliki jumlah *support count* lebih besar dan sama dengan *minimum support count* kemudian akan dibangkitkan dengan *conditional FP-tree*.
3. Tahap pencarian (*Frequent itemset*) Jika (*Conditional FP-tree*) merupakan lintasan tunggal (*single path*), akan didapatkan *frequent itemset* dengan dilakukan kombinasi item untuk setiap *conditional FP-tree*. apabila tidak merupakan lintasan tunggal, langkah selanjutnya akan dilakukan pembangkitan *FP-Growth* secara *rekursif*.

2.4 Pengertian WEKA 3.8.3

Weka merupakan *tools machine learning* yang praktis. “*Waikato Environment for Knowledge Analysis*” atau yang dikenal dengan WEKA dibuat di *Universitas Waikato, New Zealand*, yang dikhususkan untuk menunjang bidang penelitian, pendidikan dan berbagai pengaplikasiannya dalam Data mining. *Software* dibangun menggunakan *class java* dengan metode berorientasi objek dan dapat dijalankan hampir pada semua *platform*, dalam penggunaannya Weka sangat mudah diterapkan pada beberapa tingkatan yang berbeda. Weka menyediakan implementasi algoritma pembelajaran *state of the art* yang dapat diterapkan pada *dataset* dari *command line*.

Dalam WEKA terdapat *tools* yang berguna untuk *preprocessing data, klasifikasi, regresi, clustering, aturan asosiasi, dan visualisasi*. Penggunaannya dapat dilakukan *preprocess* pada data, memasukkan dalam sebuah skema pembelajaran, dan menganalisis *classifier* yang dihasilkan oleh performanya, semua dikerjakan tanpa menulis kode program. Salah satu contoh penggunaan WEKA dengan menerapkan sebuah metode pembelajaran kedalam dataset serta menganalisis hasilnya untuk memperoleh informasi tentang data, atau menerapkan beberapa metode dan membandingkan performanya untuk dipilih.

Definisi lain dari Weka adalah alat yang dapat digunakan untuk *preprocessing dataset* dan membuat *user* dapat terfokus pada algoritma yang akan digunakan tanpa terlalu memperhatikan detail seperti pembacaan data dari file, implementasi algoritma *filtering*, dan penyediaan kode untuk evaluasi hasil (Purnamasari, Henharta, Sasmita, Ihsani, dan Wicaksana, 2013).

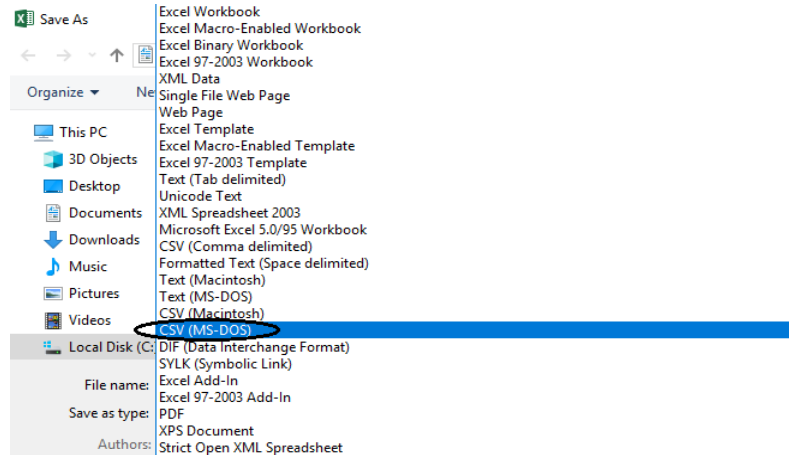
2.5 Penerapan *Frequent Pattern Growth* Dengan Weka

Penerapan *FP-Growth* pada weka sebagai berikut:

1. Langkah pertama jika *variabel* data yang akan diproses adalah *numeric* diganti dalam *variabel* nominal. Hal ini dikarenakan Weka hanya dapat mem-

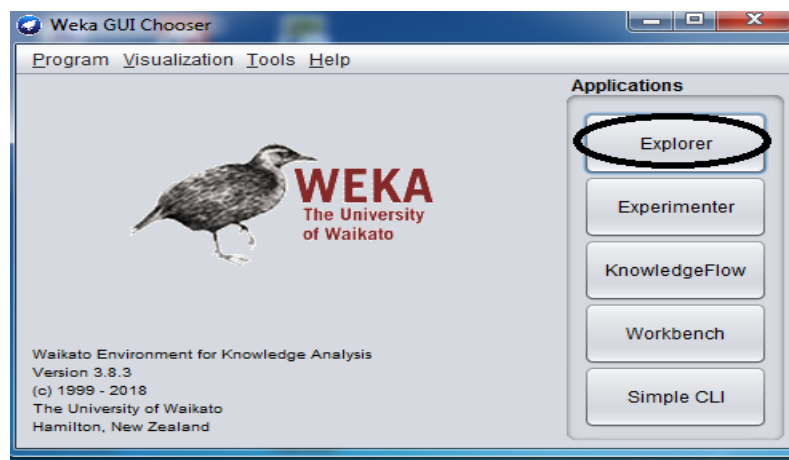
baca tipe *variabel* nominal.

- Langkah selanjutnya mengubah format *file* kedalam *.csv yang terlihat seperti pada Gambar 2.2.



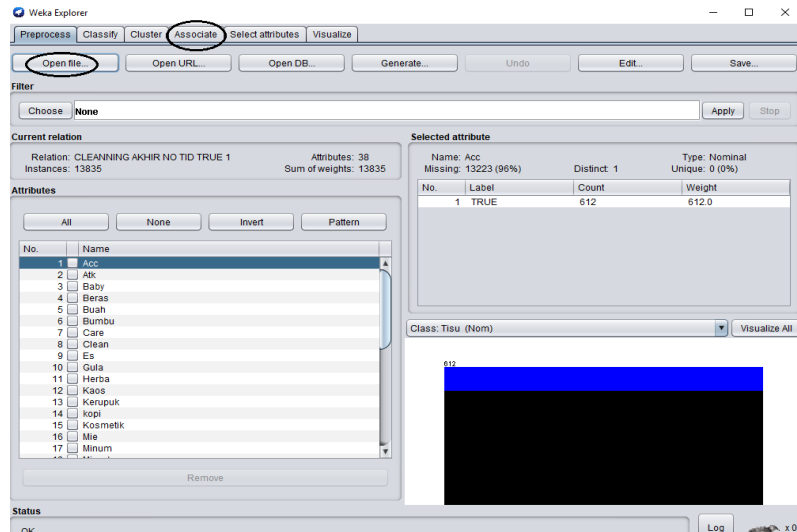
Gambar 2.2. Ubah format file

- Langkah selanjutnya *Open software* pada *tools* Weka dan pilih menu *explorer* seperti yang terlihat pada Gambar 2.3.



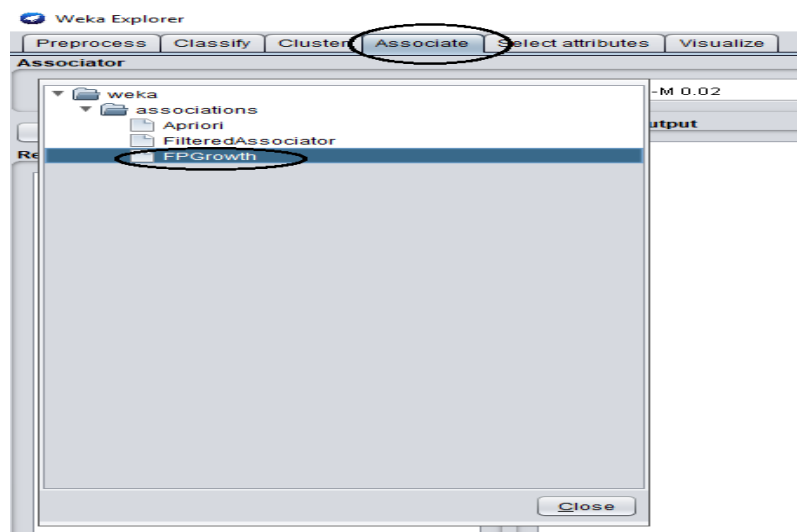
Gambar 2.3. Halaman utama plikasi Weka

- Open file* dan pilih menu untuk Metode *association* dalam aplikasi weka seperti yang terlihat pada Gambar 2.4.



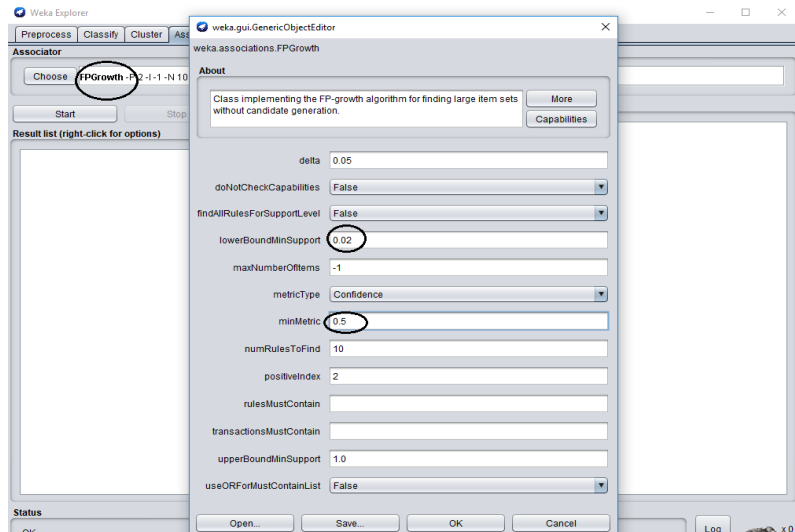
Gambar 2.4. Preproses data di Weka

5. Klik menu *choose* agar dapat memilih algoritma yang akan digunakan, seperti yang terlihat pada Gambar 2.5.



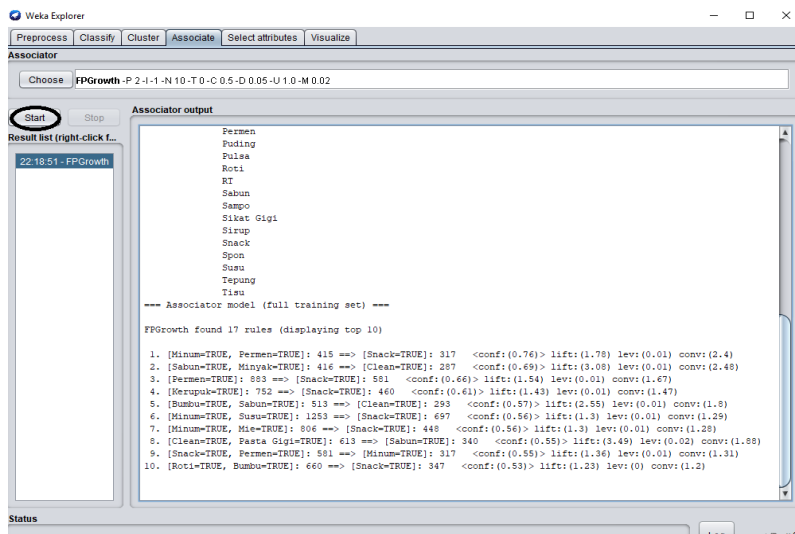
Gambar 2.5. Pemilihan Algoritma

6. Saat akan Melakukan *start* sebelumnya arahkan *kursor* pada kolom dibagian menu *choose* kemudian klik, lalu atur nilai *support* dan *confidence* seperti yang terlihat pada Gambar 2.6.



Gambar 2.6. Menentukan *Support* dan *Confidence*

7. Tahapan terakhir klik perintah *start* dan akan muncul hasil seperti yang terlihat pada Gambar 2.7.



Gambar 2.7. Hasil Perhitungan

2.6 212Mart

212Mart adalah *brand minimarket* Koperasi Syariah 212. Modal kepemilikan berjamaah yang diusung oleh 212Mart, dikelola secara profesional dan terpusat untuk menjaga daya saingnya baik dari sisi jaringan distribusi, produk, harga maupun promo. 212Mart mengusung Semangat 212, gerakan kebangkitan umat yang dipicu oleh perjuangan umat pada Aksi Bela Islam (ABI) ketiga yang fenomenal pada 02 Desember 2016. Sedangkan koperasi syariah 212 sendiri adalah perjuangan umat pada aksi 212 tersebut, perjuangan dibidang ekonomi. 212Mart

cabang Pekanbaru adalah toko yang didirikan oleh komunitas perempuan, dengan jumlah *owner* atau pemegang saham 183, koperasi ini bekerja sama dengan Pt. Intan Payung Berkah. Kisah dibalik berdirinya 212Mart Pekanbaru dilandaskan pada Iman kepada Allah SWT dan bertujuan membangkitkan ekonomi umat, yang berbasis Syariah. Peresmian 212Mart bertepatan pada reuni 212 di Monas, Jakarta, yaitu pada tanggal 02 Desember 2017.

2.7 Tata Letak

Tata ruang merupakan sebuah konsep tatanan secara desain fisik, menjadi bagian bagian terminal kerja dengan macam dukungan peralatan dan perlengkapan yang menunjuk kepada suatu prose berjalannya bisnis atau produksi, Sumayang (2003) dalam (Prakoso, 2015).

Perencanaan Tata letak memiliki peran penting dalam menentukan sebuah bisnis atau aktivitas dalam jangka panjang. Tata letak menjadi penentu strategis dalam tata letak untuk menentukan daya saing sebuah perusahaan dalam segi kapasitas, proses fleksibilatis, biaya, kualitas dalam lingkungan kerja, pelayanan pada pelanggan, serta citra perusahaan. Tata letak yang efektif diharapkan mampu membantu perusahaan dalam pencapaian strategi untuk menunjang diferensiasi, biaya rendah ataupun proses cepat suatu perusahaan. Tujuan utama dalam menentukan strategi tata letak adalah diharapkan dapat menciptakan tata letak yang ekonomis sehingga dapat memenuhi kebutuhan persaingan antar perusahaan, Heizer & Reinder (2009) dalam (Tarigan, 2014).

Perencanaan dalam menentukan tata letak merupakan langkah langkah dalam desain suatu sistem produksi seperti barang maupun jasa. Desain tata letak harus dipertimbangkan sehingga dapat mencapai hal – hal berikut, Heizer dan Reinder (2009:532) dalam (Tarigan, 2014).

1. Bagaimana memperoleh penggunaan yang tinggi pada masing-masing ruangan merupakan perumpaan layout yang baik.
2. Diharapkan dapat membantu memperbaiki arus informasi, bahan baku, dan orang.
3. Diharapkan dapat memperbaiki moral pekerja dan menciptakan kondisi kerja yang lebih aman dan nyaman.
4. Dapat memperbaiki interaksi pelanggan dan klien, dan memahami keinginan pelanggan.
5. Fleksibilitas, tergantung dalam kebutuhan perusahaan.

2.8 Penelitian Terdahulu

Analisis keranjang pasar menggunakan apriori dan FP- *Growth* untuk pola pengeluaran analisis konsumen di Berkah Mart Dipecanbaru Riau, Menunjukkan bahwa penerapan algoritma FP-*Growth* dan Apriori untuk analisis pola belanja konsumen di Berkah Mart dapat meningkatkan pendapatan keseluruhan atau keuntungan, namun disarankan untuk menggunakan algoritma FP-*Growth* (Ilham dkk., 2018).

Selanjutnya penelitian dilakukan Maulana dan Fajrin (2018) tentang Hasil penjualan dari *spare part* motor *sport* yang paling banyak terjual di TBDamar CV.TJAHAJA BARU Padang bisa diketahui dengan menggunakan algoritma FP-*Growth*. *Spare part* yang memenuhi *minimum support* dan *minimum confidence* serta yang banyak terjual adalah *screw valve adjusting*, *oil seal*, *battery assy*, *axle*, *gasket cylinder*, dan *cable clutch*.

Selanjutnya penelitian yang dilakukan Gunadi dan Sensus (2016) tentang penerapan metode *data mining market basket analysis* pada data penjualan produk menggunakan algoritma Apriori dan *frequent pattern growth* FP-*Growth* Studi Kasus: Percetakan PT.Gramedia. Data yang dipilih dari 1 Januari 2005 hingga 1 Agustus 2010. Hasil analisa yang didapat menunjukkan perbandingan pada dua metode asosiasi tersebut, oleh algoritma Apriori memiliki tingkat kekuatan lebih tinggi dibandingkan dengan yang dihasilkan oleh algoritma FP-*Growth*.

Selanjutnya penelitian yang dilakukan oleh Ghozali dkk. (2017) analisa pola belanja menggunakan algoritma FP-*Growth*, *Self Organizing Map* (SOM) dan *K Medoids*, hasil analisis eksperimen ditetapkan nilai *minimum support* adalah 10%-100% dan nilai *minimum confidence* 10%-100%, yang menjelaskan data transaksi bahwa produk A dan B adalah produk yang dibeli secara bersamaan.

Selanjutnya penelitian yang dilakukan Triyanto (2014) tentang penentuan rekomendasi promosi produk menggunakan *Association Rule Mining*, hasilnya dapat disimpulkan bahwa penerapan *Association Rule Mining* menggunakan algoritma FP-*Growth* sangat membantu dalam penentuan rekomendasi produk secara tepat. penelitian tersebut menghasilkan 3 rule dengan menggunakan *minimum support* 40% dan *minimum confidence* 80%.

Selanjutnya penelitian yang dilakukan Larasati dkk. (2015) tentang analisis dan implementasi algoritma FP-*Growth* dari hasil analisis dapat disimpulkan semakin besar *minimum support* 0,002 dan *minimum confidence* 0,5 menghasilkan hasil rule yang hampir sama dengan beberapa bulan yang berbeda.