

BAB II LANDASAN TEORI

2.1. Tinjauan Pustaka

Beberapa penelitian terdahulu yang dijadikan acuan dalam penelitian ini dapat dilihat pada

Tabel 2.1:

Tabel 2. 1 Tinjauan Pustaka

No	Detail	Keterangan
1	Judul	Analisa Transaksi Penjualan Menggunakan Metode Algoritma Apriori (Studi Kasus Toko Busana Sakato, Gajah Mada Plaza B-19 Malang).
	Penulis (Tahun)	Tida and Zubair (2022)
	Metode	Algoritma Apriori
	Permasalahan	Untuk mengetahui jenis pakaian apa saja yang sering dibeli oleh konsumen toko masih menggunakan cara manual
	Hasil	Hasil penelitian ini pihak toko dapat menggunakannya sebagai acuan untuk pola penempatan barang. Analisis ini dilakukan menggunakan Microsoft Excel dan aplikasi tambahan Rapidminer versi 9.9.
2	Judul	Analisis Keranjang Belanja Pada Transaksi Penjualan Menggunakan Algoritma Apriori.
	Penulis (Tahun)	Sari and Hayuningtyas (2021)
	Met	Tabel 2.1 Tinjauan Pustaka
	Permasalahan	Kegiatan masih dilakukan secara manual
	Hasil	Hasil analisis yaitu dapat mengetahui perilaku pembeli pada keranjang belanja untuk mengetahui menu makanan pada restoran yang diminati. Analisis keranjang belanja ini dilakukan dengan mengimplementasikan algoritma apriori untuk mengetahui aturan asosiasi pada setiap itemset, pada

No	Detail	Keterangan
		penelitian ini dilakukan dengan melakukan aturan kombinasi 3 itemset.
3	Judul	Implementasi Data Mining Untuk Analisis Data Penjualan Dengan Menggunakan Algoritma Apriori (Studi Kasus Dapoerin's)
	Penulis (Tahun)	Prasetyo, Musyaffa and Sastra (2020)
	Metode	Algoritma Apriori
	Permasalahan	Penjualan produk kudapan pada DAPOERIN'S belum memanfaatkan algoritma data mining yang dapat membantu menganalisis data transaksi untuk mengoptimalkan penjualan dan juga agar dapat mengurangi banyak sisa kudapan yang tidak terbeli oleh pembeli.
	Hasil	Hasil analisis data transaksi menggunakan data mining dengan metode algoritma apriori, menggunakan algoritma tersebut dapat diketahui produk kudapan yang paling banyak terjual, sehingga dapat menyiapkan persediaan bahan baku yang diprioritaskan dan menyusun strategi memasarkan produk dengan jenis kudapan yang lain dengan mengkaji keunggulan produk kudapan satu dengan dan lainnya yang sering banyak terjual.
4	Judul	Analisa Pola Frekuensi Keranjang Belanja dengan Algoritma Apriori (Studi kasus: Minimarket Adi).
	Penulis (Tahun)	Wijaya (2017)
	Metode	Algoritma Apriori
	Permasalahan	Kegiatan persediaan barang sesuai minat pelanggan masih dilakukan secara manual
	Hasil	Hasil analisis pola data dengan menerapkan algoritma apriori yang ditunjukkan untuk frekuensi keranjang belanja dapat membantu dalam ketersediaan barang

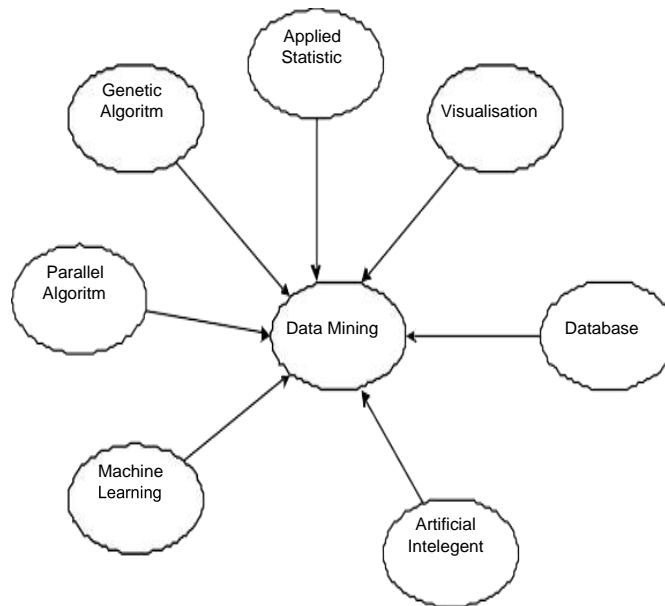
No	Detail	Keterangan
5	Judul	Implementasi Algoritma Apriori Untuk Analisis Penjualan Konter Berbasis Web.
	Penulis (Tahun)	Gumilang (2020)
	Metode	Algoritma Apriori
	Permasalahan	Toko belum melakukan penerapan analisis penjualan
	Hasil	Hasil analisa dan pengujian telah dilakukan ujicoba sistem sebanyak tiga kali menggunakan data transaksi Konter G Cell selama periode bulan Desember 2019 sampai November 2020 (1 Tahun) dengan merubah parameter minimum support dan minimum confidence maka didapat hasil kombinasi menu item yang dapat di buat untuk proses pengembangan promosi.

2.2. Data Mining

Menurut (Saputra, Wasiyanti and Nugraha, 2020) Data mining adalah proses pengumpulan dan pengolahan data yang bertujuan untuk mengekstrak informasi penting pada data. Proses pengumpulan dan ekstraksi informasi tersebut dapat dilakukan menggunakan perangkat lunak dengan bantuan perhitungan statistika, matematika, ataupun teknologi *Artificial Intelligence* (AI). Data mining sering disebut juga *Knowledge Discovery in Database* (KDD). Tujuan utama dari proses KDD adalah memprediksikan nilai-nilai yang berguna dari variabel-variabel yang ada atau menemukan pola-pola dari sebuah gugusan data yang dapat diinterpretasikan oleh manusia. Sesuai dengan tujuan tersebut, maka proses dalam mengenali informasi baru dan penemuan pola tersebut perlu diaplikasikan dengan *data*

mining. Sehingga sebenarnya *data mining* merupakan suatu bagian yang tidak dapat dilepaskan dari proses KDD.

Perlu diketahui bahwa *data mining* merupakan salah satu bidang yang cukup banyak didukung oleh cabang ilmu lain di dalam teknologi informasi yaitu statistik, teknologi basis data, *machine learning*, sistem pakar, algoritma paralel, algoritma genetika, pengenalan pola, visualisasi data, dan lain-lain.



Gambar 2. 1 *Data mining* merupakan bidang multidisipliner

Ada beberapa faktor yang menjadi alasan utama mengapa menggunakan *data mining*:

1. Banyaknya data yang terkumpul sehingga memerlukan waktu yang sangat lama dan tenaga ahli yang cukup banyak untuk menganalisisnya.
2. Komputer menjadi salah satu pilihan utama karena kemampuannya dalam kecepatan, ketepatan, tidak pernah lelah dan mudah dioperasikan.
3. Tekanan dari kompetisi bisnis yang terus menguat sehingga menjadikan informasi menjadi sangat penting dan harus segera dimiliki.

4. Mampu menemukan suatu pola yang tidak terpikirkan sama sekali.

Menurut (Widiartha, 2019) *data mining* merupakan salah satu aktifitas dibidang perangkat lunak yang dapat memberikan ROI (*Return of Investment*) yang tinggi.

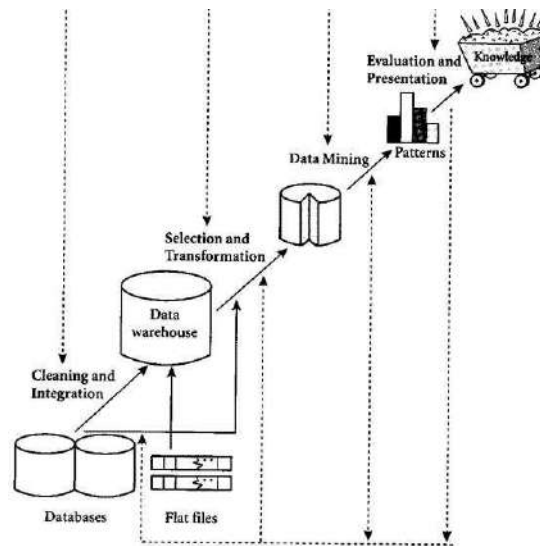
Hal yang perlu diperhatikan adalah bahwa *data mining* berbeda dengan *query tools*. *Query* dan *data mining* merupakan dua hal yang saling melengkapi. Keberadaan *data mining* bukan untuk menggantikan *query* tetapi menambahkan beberapa tambahan yang berarti. Jika menggunakan *query* sederhana maka informasi yang dapat diakses sekitar 80% dari data yang ada dalam basis data sedangkan 20% lagi akan menjadi informasi tersembunyi yang memerlukan teknik-teknik khusus dalam mengaksesnya.

2.2.1. Tahap-Tahap Data Mining

Karena *data mining* adalah suatu rangkaian proses maka dibagi menjadi beberapa tahap antara lain :

- a. Pembersihan data: untuk membuang data yang tidak konsisten dan noise.
- b. Integrasi data: untuk menggabungkan data dari beberapa sumber.
- c. Transformasi data : untuk mengubah data menjadi bentuk yang sesuai untuk *data mining*.
- d. Aplikasi teknik *data mining*.
- e. Evaluasi pola yang ditemukan : untuk menemukan informasi yang menarik ataupun bernilai.
- f. Presentasi pengetahuan dengan teknik visualisasi.

Tahap-tahap diatas dapat digambarkan sebagai berikut :



Gambar 2. 2 Tahap-tahap dalam data mining

2.2.2. Teknik Metode Data Mining

Secara umum, terdapat beberapa metode yang digunakan untuk melakukan data mining.

Berikut ini adalah metode data mining:

1. *Association*

Teknik yang pertama adalah association. Association adalah metode berbasis aturan yang digunakan untuk menemukan asosiasi dan hubungan variabel dalam satu set data. Biasanya analisis ini terdiri dari pernyataan “if atau then” sederhana. Association banyak digunakan dalam mengidentifikasi korelasi produk dalam keranjang belanja untuk memahami kebiasaan konsumsi pelanggan. Sehingga, perusahaan dapat mengembangkan strategi penjualan dan membuat sistem rekomendasi yang lebih baik.

2. *Classification*

Selanjutnya classification, ia adalah metode yang paling umum digunakan dalam data mining. *Classification* adalah tindakan untuk memprediksi kelas suatu objek.

3. *Regression*

Regression adalah teknik yang menjelaskan variabel dependen melalui proses analisis variabel independen. Sebagai contoh, prediksi penjualan suatu produk berdasarkan korelasi antara harga produk dengan tingkat pendapatan rata-rata pelanggan.

4. *Clustering*

Terakhir, metode clustering. Clustering digunakan dalam membagi kumpulan data menjadi beberapa kelompok berdasarkan kemiripan atribut yang dimiliki. Contoh kasusnya adalah Customer Segmentation. Ia membagi pelanggan ke dalam beberapa grup berdasarkan tingkat kemiripannya.

2.3. *Market basket Analysis*

Market basket analysis adalah analisis yang sering dipakai untuk menganalisa isi keranjang belanja konsumen dalam suatu pasar swalayan. Contoh penerapan dari aturan asosiatif adalah analisa pembelian produk pada sebuah toko alat tulis, pada analisa itu misalkan dapat diketahui berapa besar kemungkinan seorang pelanggan membeli pensil bersamaan dengan membeli penghapus. Penerapan aturan asosiasi dalam kasus tersebut dapat membantu pemilik toko untuk mengatur penempatan barang, mengatur persediaan atau membuat promosi pemasaran dengan menerapkan diskon untuk kombinasi barang tertentu (Choiriah *et al.*, 2019)

Fungsi *association rules* seringkali disebut dengan analisis keranjang belanja (Market Basket Analysis) yang digunakan untuk menemukan relasi atau korelasi di antara himpunan item-item. Analisis keranjang belanja adalah analisis dari kebiasaan membeli customer

dengan mencari asosiasi dan korelasi antara item-item berbeda yang diletakkan customer dalam keranjang belanjanya.

Data keranjang belanja dalam bentuknya yang paling mentah adalah data transaksi pembelian oleh pelanggan, yang mengindikasikan hanya barang-barang yang dibeli bersamaan. Data ini menantang karena beberapa hal :

1. Jumlah datanya besar (kegiatan transaksi penjualan dengan data yang banyak)
2. Pembagiannya luas (setiap keranjang belanja hanya mengandung sebagian kecil dari benda-beda yang dijual)
3. Heterogenitas (orang-orang dengan selera yang berbeda cenderung sekelompok barang yang spesifik)

2.3.1. Tujuan Market Basket Analisis

Tujuan dari Market Basket Analisis adalah untuk menganalisa atas perilaku konsumen secara spesifik dari suatu golongan atau suatu kelompok tertentu. Selain itu tujuan dari market basket adalah untuk mengetahui produk-produk mana yang mungkin akan dibeli secara bersamaan. Analisis data transaksi dapat menghasilkan pola pembelian produk yang sering terjadi. (Jumisah *et al.*, 2017)

2.4. Association rules

Association rule adalah aturan asosiasi mengungkap item yang sering dikaitkan bersama". Pada awalnya algoritma aturan asosiasi itu dikembangkan dalam konteks pasar analisis keranjang untuk mempelajari perilaku pembelian pelanggan yang dapat digunakan untuk pemasaran (Bukovský, 2017).

Association rule berguna untuk menemukan hubungan penting antar item dalam setiap transaksi, hubungan tersebut dapat menandakan kuat tidaknya suatu aturan dalam asosiasi,

Tujuan association rule adalah untuk menemukan keteraturan dalam data. Association rule dapat digunakan untuk mengidentifikasi item-item produk yang mungkin dibeli secara bersamaan dengan produk lain, atau dilihat secara bersamaan saat mencari informasi mengenai produk tertentu. Dalam pencarian *association rule*, diperlukan suatu variabel ukuran kepercayaan (*interestingness measure*) yang dapat ditentukan oleh user, untuk mengatur batasan sejauh mana dan sebanyak apa hasil output yang diinginkan oleh user.

Association rule adalah salah satu teknik *data mining* yang sudah mulai dipopulerkan oleh Rakesh Agrawal dan Ramakrishnan Srikant untuk analisis keranjang pasar pada tahun 1993 (Muchlis *et al.*, 2019) Aturan asosiasi sering dipakai dalam penggalian data transaksi. Proses pencarian aturan asosiasi bertujuan untuk menemukan pola yang sering muncul, asosiasi antara suatu kombinasi *item* yaitu dengan membuat korelasi antara *item* data yang dikelompokkan ke dalam transaksi kemudian mengambil kesimpulan berdasarkan hubungan yang terbentuk dari beberapa *item* data. Hal-hal tersebut dapat merepresentasikan informasi penting yang ingin diketahui pada data yang ada.

Proses pencarian *frequent itemset* merupakan pra-syarat dan membutuhkan waktu sangat banyak, sehingga banyak algoritma dikembangkan untuk lebih mengefisienkan proses ini. Ada beberapa faktor yang mempengaruhi efisiensi pencarian *frequent itemset* diantaranya adalah bila basis data besar (jumlah transaksi yang banyak), *item* yang sangat banyak dan nilai *support* yang rendah.

Penting tidaknya suatu aturan asosiasi dapat diketahui dengan dua parameter, *support* yaitu prosentase kombinasi *item* dalam database dan *confidence* yaitu kuatnya hubungan antar *item* dalam aturan asosiasi (Saputra, Wasiyanti and Nugraha, 2020) Pencarian *association rule* bertujuan untuk menemukan semua aturan yang terdapat pada database dengan *minsup* (*minimum support*) dan *minconf* (*minimum confidence*) yang melebihi batas tertentu.

2.4.1. Support

Support dari suatu *association rule* adalah presentasi kombinasi item tersebut dalam database, dimana jika mempunyai item A dan item B maka support adalah proporsi dari transaksi dalam database yang mengandung A dan B. Support (dukungan) merupakan suatu ukuran yang menunjukkan seberapa besar dominasi suatu item atau itemset dari keseluruhan transaksi. (Kusumo, Bijaksana, & Darmantoro, 2016)

2.4.2. Confidence

Confidence dari *association rule* adalah ukuran ketepatan suatu rule, yaitu presentasi transaksi dalam *database* yang mengandung A dan mengandung B. Dengan adanya *confidence* kita dapat mengukur kuatnya hubungan antar item dalam *association rule*. (Kusumo, Bijaksana, & Darmantoro, 2016)

$$S = \frac{\sum (Ta+Tc)}{\sum (T)} \dots\dots\dots (2.1)$$

Keterangan :

S = *Support*

$\sum (Ta+Tc)$ = Jumlah transaksi yang mengandung *antecedent* dan *consequent*.

$\sum (T)$ = Jumlah Transaksi

$$C = \frac{\sum (Ta+Tc)}{\sum (Ta)} \dots\dots\dots (2.2)$$

Keterangan :

C = *Confidence*

$\sum (Ta+Tc)$ = jumlah transaksi yang mengandung *antecedent* dan *consequent*.

$\sum (T)$ = jumlah transaksi yang mengandung *antecedent*

2.4.3. Analisa Pola Frekuensi Tinggi

Pada tahap ini, melakukan pencarian kombinasi item yang memenuhi syarat *minimum* dari nilai *support* dalam *database*. Nilai *support* item diperoleh dengan Persamaan

1.

$$Support (A) = \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}} \dots\dots\dots(2.3)$$

Sedangkan Nilai *Support 2 Item* diperoleh dari persamaan 2.

$$Support (A,B) = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\dots\dots\dots}(2.4)$$

Total Transaksi

2.4.4. Pembentukan Aturan Asosiatif

Pada tahap ini, setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiatif yang memenuhi syarat *minimum* untuk nilai *confidence* dengan menghitung nilai *confidence* dalam aturan asosiatif. Nilai *confidence* dari aturan diperoleh dari Persamaan 3.

$$\text{Confidence} = P(A|B) = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Total Transaksi Mengandung A}} \dots\dots\dots(2.5)$$

2.5. Algoritma Apriori

Algoritma apriori digunakan untuk mencari *frequent itemset* yang memenuhi *minsup* kemudian mendapatkan *rule* yang memenuhi *minconf* dari *frequent itemset* tadi. Algoritma ini mengontrol berkembangnya kandidat *itemset* dari hasil *frequent itemset* dengan *support-based pruning* untuk menghilangkan *itemset* yang tidak menarik dengan menetapkan *minsup*. Prinsip dari apriori ini adalah bila *itemset* digolongkan sebagai *frequent itemset*, yang memiliki *support* lebih dari yang ditetapkan sebelumnya, maka semua subsetnya juga termasuk golongan *frequent itemset* dan sebaliknya. Algoritma Apriori adalah algoritma analisis keranjang pasar yang digunakan untuk menghasilkan aturan asosiasi, dengan pola if-then. Pada algoritma Apriori menentukan kandidat yang mungkin muncul dengan cara memperhatikan minimum support dan minimum *confidence* (Jumisah *et al.*, 2017) .

Dalam metode ini terdapat salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien yaitu analisis pola frekuensi tinggi (*frequent pattern mining*). Suatu asosiasi dikatakan penting dengan memperhatikan dua tolak ukur yaitu: *support* dan *confidence*. *Support* yaitu persentase kombinasi item tersebut dalam

database (nilai penunjang). *Confidence* yaitu kuatnya hubungan antar item dalam aturan asosiasi (nilai kepastian). Proses awal dalam algoritma Apriori adalah analisis pola frekuensi tinggi merupakan cara mencari kombinasi item yang memenuhi syarat minimum dari nilai *support* dalam basis data. *Frequent itemset* menunjukkan *itemset* yang memiliki frekuensi kemunculan lebih dari nilai minimum yang ditentukan. Tahap berikutnya adalah pembentukan aturan asosiasi, yaitu setelah semua pola frekuensi tinggi ditemukan, kemudian dicari aturan asosiasi yang memenuhi syarat *minimum confidence* dengan menghitung *confidence* $A \rightarrow B$. Untuk menentukan aturan asosiasi yang akan dipilih maka harus diurutkan berdasarkan *support x confidence*. Aturannya diambil sebanyak n aturan yang memiliki hasil terbesar. Semakin kecil minimum *support* dan *confidence* yang ditentukan, semakin banyak pula *rules* yang dapat dihasilkan oleh aplikasi, dengan konsekuensi waktu proses akan lebih lama dibandingkan minimum *support* yang lebih besar.

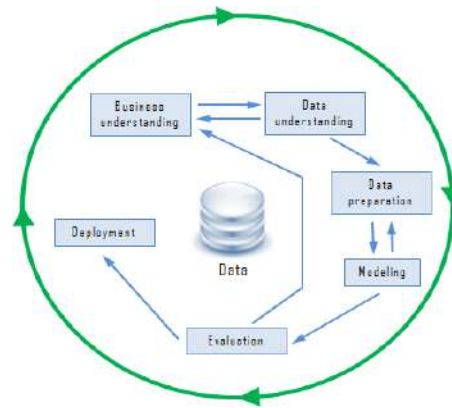
Untuk membentuk kandidat itemset ada dua proses utama yang dilakukan algoritma apriori (Han and Kamber, 2016):

1. *Join Step* (Penggabungan) Pada proses ini setiap item dikombinasikan dengan item lainnya sampai tidak terbentuk kombinasi lagi.
2. *Prune Step* (Pemangkasan) Pada proses ini, hasil dari item yang dikombinasikan tadi kemudian dipangkas dengan menggunakan minimum support yang telah ditentukan oleh user.

2.6. CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM) merupakan suatu standar yang telah dikembangkan pada tahun 1996 yang ditujukan untuk melakukan proses analisis dari

suatu industri sebagai strategi pemecahan masalah dari bisnis atau unit penelitian (Chapman, 2015). Untuk data yang dapat diproses dengan *CRISP-DM* ini, tidak ada ketentuan atau karakteristik tertentu, karena data tersebut akan diproses kembali pada fase-fase di dalamnya. Terdapat enam fase dalam *CRISP-DM* ini yakni dijelaskan pada gambar berikut :



Gambar 2. 3 Fase *CRISP-DM*
Sumber : (Chapman, 2015)

a. Fase Pemahaman Bisnis (*Business Understanding Phase*)

- 1) Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan.
- 2) Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan data mining.
- 3) Menyiapkan strategi awal untuk mencapai tujuan.

b. Fase Pemahaman Data (*Data Understanding Phase*)

- 1) Mengumpulkan data, jika data berasal dari lebih dari satu *database* maka dilakukan proses integrasi data atau *Data Integration*.
- 2) Mengembangkan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal.
- 3) Mengevaluasi kualitas data, memeriksa data dan membersihkan data yang tidak *valid* atau proses *Data Cleaning*.
- 4) Jika diinginkan, pilih sebagian kecil grup data yang mungkin mengandung pola dari

permasalahan.

c. Fase Pengolahan Data (*Data Preparation Phase*)

- 1) Siapkan data awal, kumpulan data yang akan digunakan untuk keseluruhan fase berikutnya atau proses *Data Selection*.
- 2) Pilih kasus dan variabel yang akan dianalisis, sesuai dengan analisis yang akan dilakukan.
- 3) Lakukan perubahan pada variabel jika diperlukan.
- 4) Siapkan data awal sehingga siap untuk perangkat permodelan atau *Data Transformation*

d. Fase Permodelan (*Modelling Phase*)

- 1) Pilih dan aplikasikan teknik permodelan yang sesuai.
- 2) Kalibrasi aturan model untuk mengoptimalkan hasil.
- 3) Dapat menggunakan beberapa teknik yang sama untuk permasalahan yang sama
- 4) Dapat kembali ke fase pengolahan data jika diperlukan untuk menjadikan data ke dalam bentuk kebutuhan tertentu

e. Fase Evaluasi (*Evaluation Phase*)

- 1) Mengevaluasi satu atau lebih model yang digunakan dalam fase permodelan atau proses *Evaluation Pattern*.
- 2) Menetapkan apakah model tadi sudah sesuai dengan tujuan pada fase awal.
- 3) Menentukan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik.
- 4) Mengambil keputusan berkaitan dengan penggunaan hasil dari data mining.

f. Fase Penyebaran (*Deployment Phase*)

Menggunakan model yang dihasilkan dan di persentasikan atau proses *knowledge persentation*.

2.7. RapidMiner Studio

RapidMiner merupakan perangkat lunak yang bersifat terbuka (*open source*). *RapidMiner* adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi. Berbagai teknik deskriptif dan prediksi digunakan *RapidMiner* untuk memberikan kepada pengguna sehingga dapat membuat keputusan yang paling baik. Terdapat kurang lebih 500 operator *data mining* yang dimiliki *RapidMiner* termasuk operator untuk *input*, *output*, *datapreprocessing* dan *visualisasi*. *RapidMiner* merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin *data mining* yang dapat diintegrasikan pada produknya sendiri (Afifi, Nastiti and Aini, 2020). *RapidMiner* ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi *RapidMiner* memiliki beberapa sifat sebagai berikut:

1. Ditulis dengan bahasa pemrograman Java sehingga dapat dijalankan di berbagai sistem operasi.
2. Konsep multi-layer untuk menjamin tampilan data yang efisien dan menjamin penanganan data.
3. Memiliki GUI, command line mode, dan Java API yang dapat dipanggil dari program lain.

Beberapa Fitur dari *RapidMiner*, antara lain:

1. Banyaknya algoritma *data mining*, seperti *decision tree* dan *self-organization map*.
2. Bentuk grafis yang canggih, seperti tumpang tindih diagram histogram, *tree chart* dan 3D *Scatter plots*.
3. Banyaknya variasi *plugin*, seperti *text plugin* untuk melakukan analisis teks.
4. Menyediakan prosedur *data mining* dan *machine learning* termasuk: ETL (*extraction, transformation, loading*), *data preprocessing*, *visualisasi*, *modelling* dan evaluasi.

5. Proses *data mining* tersusun atas operator-operator yang *nestable*, dideskripsikan dengan XML, dan dibuat dengan GUI
6. Mengintegrasikan proyek data mining Weka dan statistika R.