# BAB II

## LANDASAN TEORI

### Tinjauan Pustaka

Dari berbagai aspek dan sudut pandang yang berbeda-beda mencakup sesuai dengan hal-hal yang berkaitan dengan penelitian sebagai berikut:

#### Algoritma Apriori

*Apriori is a seminal algorithm proposed by R. Agrawal and R. Srikant in 1994 for mining frequent itemsets for Boolean association rules. The name of the algorithmis based on the fact that the algorithm uses prior knowledge of frequent itemset properties, as we shall see later. Apriori employs an iterative approach known as a level-wisesearch, where k-itemsets are used to explore.* [ Han dkk, 2012: 248]

Berdasar pengertian diatas disimpulkan Apriori adalah algoritma yang diusulkan oleh R. Agrawal dan R. Srikant pada tahun 1994 untuk menambang *item itemset* yang sering muncul untuk aturan asosiasi Boolean. Nama algoritma didasarkan pada kenyataan bahwa algoritma menggunakan pengetahuan sebelumnya tentang item yang sering muncul. Apriori menggunakan pendekatan berulang di mana k-itemset digunakan untuk menjelajahi .

“Algoritma apriori termasuk jenis aturan asosiasi pada *data mining*. Analisa asosiasi adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item.” [Kusrini dan Luthfi, 2009: 149]. Metodologi dasar analisis asosiasi terbagi menjadi dua tahap, yaitu:

1. Analisis pola frekuensi tinggi

Tahap ini mencari kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam database. Nilai *support* sebuah item diperoleh dengan rumus berikut:

Sumber: Kusrini (2009: 150)

Gambar II.1

Rumus Nilai *Support*

Sementara itu, nilai *support* dari 2 *item* diperoleh dari rumus 2 berikut.

Sumber: Kusrini (2009: 150)

Gambar II.2

Rumus Nilai *Support* 2 *item*

Berikut contoh *database* dari transaksi belanja pasar swalayan

Tabel II.1

Transaksi

| **Transaksi** | **Item yang dibeli** |
| --- | --- |
| 1 | Susu,Teh,Gula |
| 2 | Teh,Gula,Roti |
| 3 | Teh,Gula |
| 4 | Susu, Roti |
| 5 | Susu,Gula,Roti |
| 6 | Teh,Gula |
| 7 | Gula,Kopi,Susu |
| 8 | Gula,Kopi,Susu |
| 9 | Susu,Kopi,Roti |
| 10 | Gula,Teh,Kopi |

Sumber : Kusrini (2009: 151)

Tabel II.2

Representasi Data Transaksi dalam *Database* Transaksional

|  |  |
| --- | --- |
| **Transaksi** | **Item yang dibeli** |
| 1 | Susu |
| 1 | Teh |
| 1 | Gula |
| 2 | Teh |
| 2 | Gula |
| 2 | Roti |
| 3 | Teh |
| 3 | Gula |
| 4 | Susu |
| 4 | Roti |
| 5 | Susu |
| 5 | Gula |
| 5 | Roti |
| 6 | Teh |
| 6 | Gula |
| 7 | Gula |
| 7 | Kopi |
| 7 | Susu |
| 8 | Gula |
| 8 | Kopi |
| 8 | Susu |
| 9 | Susu |
| 9 | Kopi |
| 9 | Roti |
| 10 | Gula |
| 10 | Teh |
| 10 | kopi |

Sumber: Kusrini (2009: 151- 152)

Tabel II.3

Format Tabular Data Transaksi

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Transaksi** | **Teh** | **Gula** | **Kopi** | **Susu** | **Roti** |
| 1 | 1 | 1 | 0 | 1 | 0 |
| 2 | 1 | 1 | 0 | 0 | 1 |
| 3 | 1 | 1 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 1 | 1 |
| 5 | 0 | 1 | 0 | 1 | 1 |
| 6 | 1 | 1 | 0 | 0 | 0 |
| 7 | 0 | 1 | 1 | 1 | 0 |
| 8 | 0 | 1 | 1 | 1 | 0 |
| 9 | 0 | 0 | 1 | 1 | 1 |
| 10 | 1 | 1 | 1 | 0 | 0 |

Sumber: Kusrini (2009: 152)

Misalkan D adalah himpunan transaksi yang direpresentasikan dalam tabel III.1, di mana setiap transaksi T dalam D mempresentasikan himpunan *item* yang berada dalam I. I adalah himpunan *item* yang dijual {Teh, Gula, Kopi, Susu,Roti}. Misalkan kita memiliki himpunan item A (misal Susu dan Gula) dan himpunan *item* lain B (misal Kopi). Kemudian aturan asosiasi akan berbentuk:

*Jika A, maka B (A → B)*

Tidak berarti

*Jika B, maka A.*

Definisi ini tidak berlaku untuk aturan *trivial* seperti:

*Jika beans dan squash, maka beans*

Seorang analis mungkin hanya akan mengambil aturan yang memiliki *support* dan atau *confindence* yang tinggi. Aturan yang kuat adalah aturan-aturan yang melebihi kriteria *support* dan/atau *confidence minimum.* Misalnya seorang analis menginginkan aturan yang memiliki *support* lebih dari 20% dan *confidence* lebih dari 35%.

Sebuah *itemsets* adalah himpunan *item-item* yang ada dalam I, dan k-*itemset* adalah itemset yang berisi k *item*. Misalnya {Teh, Gula} adalah sebuah 2-*itemsets* dan {Teh, Gula, Roti} merupakan 3-*itemsets.*

*Frequent Itemset* menunjukan *itemset* yang memiliki frekuensi kemunculan lebih dari nilai minimum yang telah ditentukan (Ø). Misalkan Ø = 2, maka semua *itemset* yang memiliki frekuensi kemunculannya lebuh dari atau sama dengan 2 kali disebut *frequent.* Himpunan dari *frequent k-itemset* dilambangkan dengan Fk.

Berikut ini menunjukan calon *2-itemset* dari data transaksi pada tabel III.1

Tabel II.4

Calon 2-*itemsets*

|  |  |
| --- | --- |
| Kombinasi | Jumlah |
| Teh, Gula | 5 |
| Teh, Kopi | 1 |
| Teh, Susu | 1 |
| Teh, Roti | 1 |
| Gula, Kopi | 3 |
| Gula, Susu | 4 |
| Gula, Roti | 2 |
| Kopi, Susu | 3 |
| Kopi, Roti | 1 |
| Susu, Roti | 3 |

Sumber: Kusrini (2009: 154)

Dari data tersebut di atas jika ditetapkan nilai Ø = 3, maka:

F2 = {{Teh, Gula}, {Gula, Kopi}, {Gula, Susu}, {Gula, Roti},

{Kopi, Susu}, {Susu, Roti}}

Tabel II.5

Calon 3-*itemsets*

|  |  |
| --- | --- |
| Kombinasi | Jumlah |
| Teh, Gula, Kopi | 1 |
| Teh, Gula, Susu | 1 |
| Gula, Susu, Kopi | 2 |
| Gula, Susu, Roti | 0 |
| Gula, Kopi, Roti | 0 |
| Kopi, Susu, Roti | 1 |

Sumber: Kusrini (2009: 154)

Kombinasi dari *itemset* adalah F2 dapat kita gabungkan menjadi calon 3-*itemsets.* Itemset-itemset dari F2 yang dapat digabungkan adalah *itemset-itemset* yang memiliki kesamaan dalam k-1 *item* pertama. Calon 3-itemset yang dapat dibentuk dari F2 tampak pada tabel II.5

Dengan demikian F3= {{Gula, Susu, Kopi}}, karena hanya kombinasi ini adalah yang memiliki frekuensi >= Ø

1. Pembentukan Aturan Asosiasi

Pembentukan Aturan Asosiasi adalah mencari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiatif A→ B. Nilai *confidence* dari aturan A→ B diperoleh dari rumus berikut. [Kusrini, 2009: 155]

Sumber: Kusrini (2009: 155)

Gambar II.3

Rumus Nilai *Confidence*

Dari F3 yang telah ditemukan, dapat dilihat besarnya nilai *support* dan *confidence* dari calon aturan asosiasi seperti tampak tabel pada tabel II.6

Tabel II.6

Calon aturan asosiasi dari F3

|  |  |  |
| --- | --- | --- |
| **Aturan** | ***Confidence*** | |
| Jika membeli gula dan susu, maka akan membeli kopi | 2/4 | 50% |
| Jika membeli gula dan kopi, maka akan membeli susu | 2/3 | 67% |
| Jika membeli kopi dan susu, maka akan membeli gula | 2/3 | 67% |

Sumber: Kusrini (2009: 155)

Misalkan ditetapkan nilai *confidence minimal* adalah 60%, maka aturan yang bisa

terbentuk adalah aturan dengan dua *antecendent* berikut.

1. Jika membeli gula dan kopi, maka akan membeli susu.
2. Jika membeli kopi dan susu, maka akan membeli gula.

Calon aturan asosiasi F2 bisa dilihat pada tabel II.7

Tabel II.7

Calon aturan asosiasi dari F2

| Aturan | *Confidence* | |
| --- | --- | --- |
| Jika membeli teh, maka akan membeli gula. | 5/5 | 100% |
| Jika membeli gula, maka akan membeli teh. | 5/8 | 62.5% |
| Jika membeli gula, maka akan membeli kopi | 3/8 | 37.5% |
| Jika membeli kopi, maka akan membeli gula | 3/4 | 75% |
| Jika membeli gula, makan akan membeli susu | 4/8 | 50% |
| Jika membeli susu, maka akan membeli gula | 4/6 | 67% |
| Jika membeli gula, maka akan membeli roti | 2/8 | 25% |
| Jika membeli roti, makan akan membeli gula | 2/4 | 50% |
| Jika membeli kopi, maka akan membeli susu | 3/4 | 75% |
| Jika membeli susu, makan membeli kopi | 3/6 | 50% |
| Jika membeli susu, maka akan membeli roti | 3/6 | 50% |
| Jika membeli roti, makan akan membeli susu | 3/4 | 75% |

Sumber: Kusrini (2009: 155- 156)

Aturan asoiasi final terurut berdasarkan *support x confident* terbesar dapat dilihat pada tabel II.8

Tabel II.8

Aturan asosiasi final

| **ATURAN** | ***SUPPORT*** | ***CONFIDENCE*** | ***SUPPORT X CONFIDENCE*** |
| --- | --- | --- | --- |
| Jika membeli teh, maka akan membeli gula | 50% | 100% | 50.0% |
| Jika membeli gula, maka akan membeli teh | 50% | 62.50% | 31.3% |
| Jika membeli susu, maka akan membeli gula | 40% | 67% | 26.8% |
| Jika membeli kopi, maka akan membeli gula | 30% | 75% | 22.5% |
| Jika membeli kopi, maka akan membeli susu. | 30% | 75% | 22.5% |
| Jika membeli roti, maka akan membeli susu | 30% | 75% | 22.5% |
| Jika membeli gula dan kopi, maka akan membeli susu | 20% | 67% | 13.4% |
| Jika membeli kopi dan susu, maka akan membeli gula | 20% | 67% | 13.4% |

Sumber: Kusrini (2009: 156)

#### Pengertian Analisa Aturan Asosiasi

Analisa asosiasi atau *association rule mining* adalah teknik *data mining* untuk menemukan aturan asosiatif antara suatu kombinasi item. Analisa asosiasi dikenal juga sebagai salah satu teknik *data mining* yang menjadi dasar dari berbagai teknik data mining lainnya. Secara khusus, salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien adalah pola frekuensi tinggi ( *frequent pattern mining*). [Nofriansyah, 2014: 78]

“Secara umum association rules memiliki dua bagian penting yaitu antecedent (sebab) yang biasa diistilahkan dengan IF dan cosequent (akibat) yang biasa diistilahkan dengan THEN. “ [Adinugroho dan Sari, 2018: 129]

#### Pengertian *Support*

“Suatu ukuran yang menunjukan seberapa besar tingkat dominasi suatu item/ itemset dari keseluruhan transaksi. Ukuran ini menentukan apakah suatu item/ itemset layak untuk dicari confidence tersebut. “[Nofriansyah, 2014: 77]

#### Pengertian *Confidence*

“Suatu ukuran yang menunjukan hubungan antar 2 item secara conditional (misal, seberapa sering item B dibeli jika orang membeli item A).” [Nofriansyah, 2014: 77]

#### Pengertian Lift

*“Lift* adalah ukuran yang menunjukkan adanya tingkat kekuatan *rules* kejadian acak dari *antecedent* dan *consequent.”* [Tampubolon, 2013: 105]

*“The improvement brought by a rule, by comparison with a random response, is called the lift (or simply the ‘improvement’), and is as follows:*

Sumber: Tuffery (2011: 288)

Gambar II.4

Rumus Nilai *Lift*

*When the ‘result’ is independent of the ‘condition’, the lift is clearly equal to 1. If the lift is less than 1, the rule does not help.”* [Tuffery, 2011: 288]

Berdasar pengertian diatas disimpulkan jika nilai *lift* 1 maka aturan tersebut dapat membantu, jika kurang dari 1 aturan tidak dapat membantu.

#### Menentukan Nilai *Support* dan *Confidence*

*“Typically, association rules are considered interesting if they satisfy both a minimum support threshold and a minimum confidence threshold. These thresholds can be a set by users or domain experts. Additional analysis can be performed to discover interesting statistical correlations between associated items.”* [ Han dkk, 2012: 245]

Berdasar pengertian diatas disimpulkan dalam menentukan nilai *support* dan *confidence,* bisa ditentukan oleh pengguna atau seorang ahli.

#### Pengertian Penjualan

“Penjualan adalah suatu transaksi yang bertujuan untuk mendapatkan suatu keuntungan, dan mendapatkan suatu jantung dari suatu perusahaan .” [ Himayati, 2008: 123]

#### Pengertian Tanagra

*Tanagra is a free data mining software for academic and research purposes. It proposes several data mining methods from exploratory data analysis, statistical learning, machine learning and databases area. Tanagra is an "open source project" as every researcher can access to the source code, and add his own algorithms, as far as he agrees and conforms to the software distribution license.The main purpose of Tanagra project is to give researchers and students an easy-to-use data mining software, conforming to the present norms of the software development in this domain (especially in the design of its GUI and the way to use it), and allowing to analyse either real or synthetic data*. [Tanagra: 2008].

Berdasarkan pengertian diatas dapat disimpulkan Tanagra adalah perangkat lunak data mining gratis untuk tujuan akademik dan penelitian yang mengusulkan beberapa metode penambangan data dari analisis data eksplorasi, pembelajaran statistik, pembelajaran mesin dan area database. Tujuan utama proyek Tanagra adalah memberi para peneliti dan mahasiswa perangkat lunak penambangan data yang mudah digunakan, sesuai dengan norma-norma sekarang dari pengembangan perangkat lunak dan memungkinkan untuk menganalisis data nyata atau sintetis.

#### Pengertian *Data Mining*

“*Data mining is the process of discovering useful patterns and trends in large*

*data sets. “*[Larose, 2015: 4]

Berdasarkan pengertian diatas dapat disimpulkan Data Mining adalah proses menemukan pola dan tren yang bermanfaat dalam data yang luas.

“*Data Mining* adalah serangkaian proses ekstraksi atau penggalian data dan informasi yang besar yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari *database* yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting.” [Indrajani, 2011: 289]

“Istilah *data mining* dan *knowledge discovery in databases* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain.” [Kusrini dan Luthfi, 2009: 7]

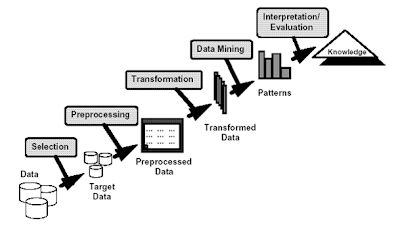
Berdasarkan pengertian-pengertian di atas dapat disimpulkan bahwa Data Mining adalah proses menggali dan mengekstrak informasi yang selama ini tidak diketahui atau tersembunyi dari suatu basis data untuk menemukan pengetahuan, yang penting untuk kepentingan perusahaan atau organisasi tersebut.

* + 1. **Proses Tahapan *Data Mining***

“*Data Mining* merupakan salah satu dari rangkaian *Knowledge Discovery in Database (KDD)*. KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, intepretasi dan visualisasi dari pola-pola sejumlah data. Serangkaian proses tersebut memiliki tahapan sebagai berikut:” [Retno, 2017: 2]

1. Pembersihan data (untuk membuang data yang tidak konsisten dan *noise*)
2. Integrasi data (penggabungan data dari beberapa sumber )
3. Transformasi Data (data diubah menjadi bentuk yang sesuai untuk di-*mining*)
4. Aplikasi teknik *Data Mining*, proses ekstrasi pola daridata yang ada.
5. Evaluasi pola yang ditemukan (proses intepretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan)
6. Presentasi pengetahuan (dengan teknik visualisasi)

Langkah terkahir KDD adalah mempresentasikan pengetahuan dalam bentuk yang mudah dipahami pengguna.



Sumber: Retno (2017: 33)

Gambar II.5

Tahapan *Knowledge Discovery in Database* (KDD)

* + 1. **Pengelompokan *Data Mining***

*Data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu: [Kusrini dan Luthfi, 2009:10-12]

1. **Deskripsi**

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari data untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menentukan keterangan atau fakta bahwa siapa yang tidak cukup professional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelesan untuk suatu pola atau kecenderungan. [Kusrini dan Luthfi, 2009:10-12]

1. **Estimasi**

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik dari pada kearah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel predikasi. Sebagai contoh akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya. [Kusrini dan Luthfi, 2009:10-12]

1. **Prediksi**

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam predikasi nilai dari hasil akan ada dimasa mendatang. Contoh prediksi bisnis dan penelitian adalah:

a. Prediksi harga beras dalam tiga bulan yang akan datang.

b. Prediksi persentasi kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikkan.

Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi. [Kusrini dan Luthfi, 2009:10-12]

1. **Klasifikasi**

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh penggolongan pendapatan dapat dipisahkan dalam tiga kategori , yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Contoh lain klasifikasi dalam bisnis danpenelitian adalah:

a. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau tidak.

b. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.

c. Mendiagnosis penyakit seorang pasien untuk mendapatkan termasuk kategori penyakit apa. [Kusrini dan Luthfi, 2009:10-12]

1. **Pengklusteran**

Pengkluteran merupakan pengelompokan *record,* pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogeny), yang mana kemiripan dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal. Contoh pengklusteran dalam bisnis dan penelitian adalah:

a. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari satu suatu produk bagi perusahaan yang tidak memiliki dana pemesaran yang besar.

b. Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku financial dalam baik dan mencurigakan.

c. Melakukan pengklusteran terhadap ekspresi dari gen, untuk mendapatkan kemiripan perilaku dari gen dalam jumlah besar. [Kusrini dan Luthfi, 2009:10-12]

1. **Asosiasi**

Tugas asosiasi dalam data mining adalah menemukan attribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja. Contoh asosiasi dalam bisnis dan penelitian adalah:

a. Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respon positif terhadap penawaran upgrade layanan yang diberikan.

b. Menentukan barang dalam supermarket yang dibeli secara bersamaan dan yang tidak pernah dibeli secara bersamaan. [Kusrini dan Luthfi, 2009:10-12]

### Penelitian Terkait

Beberapa penelitian yang terkait dalam penulisan skripsi ini adalah sebagai berikut:

1. Kecelakaan merupakan suatu kejadian yang tidak terencana begitupun pada sebuah proyek konstruksi dimana kecelakaan sering terjadi hal ini disebabkan oleh berbagai faktor. Kita lihat pada Industri jasa konstruksi yang merupakan salah satu sektor industri yang memiliki risiko kecelakaan kerja yang cukup tinggi. Banyaknya kecelakaan kerja yang terjadi tidak terlepas dari faktor Human Error, tentunya berdampak pada kinerja dan pekerjaan yang dilaksanakan, Metode yang digunakan dalam analisis ini adalah Algoritma C4.5 yang merupakan salah satu algoritma modern untuk melakukan Data Mining, Algoritma C4.5 disebut juga dengan pohon keputusan (decision tree) yaitu merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon, dan pada setiap node merepresentasikan atribut,cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas, Konsep dari pohon keputusan ini adalah dengan mengumpulkan data selanjutnya dibuatkan decision tree yang kemudian akan dihasilkan rule-rule solusi permasalahan. dari hasil penelitian faktor-faktor yang menjadi penyebab terjadinya kecelakaan kerja kontrusksi yang sering terjadi adalah Lingkungan Tempat Kerja, Rambu-Rambu Keselamatan dan Pekerja dan Cara kerja. [Elisa, 2017:36].
2. Lemahnya pengawasan dalam proses pemberian kredit kepada karyawan PT. X Group menyebabkan tingginya kredit macet. Dalam menyalurkan kreditnya, PT. X Group haruslah pintar dalam menilai para nasabah dimasa yang akan datang apakah akan menguntungkan atau tidak. Faktor ini sangatlah penting bagi pihak perusahaan karena hal ini akan menunjukkan bahwa layak atau tidaknya suatu usaha atau individu yang akan diberikan pinjaman atau kredit, pada penelitian ini digunakan teknik data mining klasifikasi dengan metode C4.5 untuk mengetahui apakah nasabah tergolong nasabah lancar ataupun tidak. Dengan penelitian ini diharapkan dapat membantu pihak perusahaan dalam membaca pola pembayaran dari nasabahnya sehingga dapat menentukan apakah nasabah tersebut layak mendapatkan kredit atau tidak dan menghasilkan rule dari pohon keputusan yang diterapkan pada implementasi sistem klasifikasi data nasabah kredit di PT. X Group. [Pratama dkk, 2018: 121].
3. Kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan pinjam meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam melunasi hutangnya setelah jangka waktu tertentu dengan pemberian bunga, pada koperasi permasalahan kredit merupakan permasalahan manajemen, dimana jika banyak nasabah yang menunggak dalam pembayaran maka akan mengganggu system keuangan yang ada, untuk itulah penelitian ini menerapkan proses analisa kredit nasabah terlebih dahulu sebelum diambil sebuah keputusan pemberian kredit, analisa keputusan memberikan kredit menggunakan algoritma klasifikasi C4.5 dan Naïve Bayes dimana kedua algoritma tersebut dilakukan penilaian, mana algoritma yang paling akurat dalam menganalisa kemampuan nasabah dalam membayar kredit, analisa berdasarkan data history. Hasil yang didapatkan dari perbandingan kedua algoritma tersebut, bahwa tingkat akurasi yang lebih baik adalah menganalisa menggunaka algoritma klasifikasi C4.5 yaitu 88.90 % sedangkan untuk tingkat akurasi menggunakan algortima klasifikasi Naïve Bayes yaitu 80.00%.[Marsipah, 2016: 187]
4. Data rekam medis adalah catatan khusus pasien yang sedang, seringkali data rekam medis hanya menjadi data yang menumpuk dan tidak dilakukan penelusuran untuk menghasilkan pengetahuan yang berguna bagi rumah sakit. Penelitian ini bertujuan mengolah tumpukan data rekam medis untuk mengklasifikasikan jenis penyakit yang terjadi pada rumah sakit berdasarkan pada kode penyakit internasional (International Classification of Disease) ICD-10. Metode yang digunakan dalam penelitian ini adalah metode algoritma C4.5 dengan menggunakan atribut kode penyakit internasional sebagai atribut label tujuan sebanyak 21 kelompok penyakit internasional, yaitu: A00-B99 sampai dengan Z00-Z99. Penilitian ini menghasilkan pohon keputusan dari 21 berupa value atribut tujuan kode penyakit, algoritma C4.5 dapat merepresentasikan sebanyak 14 value atribut tujuan kode penyakit dan persentasi data yang terbaca lebih dari 66%. Kesimpulan dari penelitian ini yaitu algoritma C4.5 membantu mengklasifikasikan kode penyakit internasional berdasarkan ICD-10 dan membentuk pohon keputusan yang dapat memberikan informasi penyakit apa saja yang sering terjadi pada rumah sakit [Fiandra Dkk, 2018:82]

Berdasarkan penelitian-penelitian diatas dapat disimpulkan bahwa data transaksi perusahaan maupun organisasi masih belum dimanfaatkan secara maksimal sehingga menghasilkan tumpukan data yang memakan memori penyimpanan. Jika dianalisa dengan baik kumpulan data tersebut dapat menghasilkan informasi yang bermanfaat. Salah satunya analisa menggunakan algoritma c.45 yang dapat menemukan informasi bermanfaat.