



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB II

LANDASAN TEORI

2.1 Association Rules

Association Rule Mining merupakan bagian dari *Frequent Pattern Mining*. *Frequent Pattern Mining* merupakan salah satu *task data mining* yang sangat penting. Task ini mencari hubungan/relasi, asosiasi, dan korelasi dalam data. Pengetahuan yang dihasilkan juga sangat berguna untuk klasifikasi, *clustering*, dan *task data mining* yang lain (Prasetyo, 2006). Selain *Association Rule Mining*, masih ada *Sequential Pattern*, dan *Structured Pattern* yang termasuk dalam *Frequent Pattern Mining*. *Association Rule Mining* dapat juga disebut *Frequent Itemset Mining* karena pola yang dihasilkan adalah pola item yang sering muncul bersamaan dalam sebuah *database*.

2.2 Attendance

Attendance adalah tingkat kehadiran suatu karyawan dalam sebuah perusahaan (Steers & Rhodes, 1978).

2.3 Data

Data adalah fakta berupa angka, karakter, simbol, gambar, tanda-tanda, isyarat, tulisan, suara, bunyi yang merepresentasikan keadaan sebenarnya yang selanjutnya digunakan sebagai masukan suatu Sistem Informasi (Ghofar, 2010).

2.3.1 Nominal

Data *nominal* adalah ukuran yang paling sederhana, dimana angka yang diberikan kepada objek mempunyai arti sebagai label saja, dan tidak menunjukkan tingkatan apapun (Nazir, 1985). Ciri-ciri data nominal adalah hanya memiliki atribut, atau nama, atau diskrit. Data nominal merupakan data diskrit dan tidak memiliki urutan.

2.3.2 Binomial

Data *Binomial* adalah data probabilitas diskret jumlah keberhasilan dalam n percobaan ya/tidak (berhasil/gagal) yang saling bebas, dimana setiap hasil percobaan memiliki probabilitas p . Eksperimen berhasil/gagal juga disebut percobaan bernoulli. Data binomial merupakan dasar dari uji binomial dalam uji signifikansi statistik (Faizal, 2016).

2.3.3 Polinomial

Dalam matematika, *polinomial* atau suku banyak (juga ditulis sukubanyak) adalah pernyataan matematika yang melibatkan jumlahan perkalian pangkat dalam satu atau lebih variabel dengan koefisien (Polinomial, 2017).

2.4 Microsoft Office Excel

Microsoft Office Excel adalah sebuah program aplikasi lembar kerja *spreadsheet* yang dibuat dan didistribusikan oleh Microsoft Corporation (Microsoft Excel, 2017).

2.5 Fingerprint Machine

Fingerprint Machine adalah sebuah perangkat elektronik yang digunakan untuk menangkap gambar digital dari pola sidik jari. Gambar tersebut disebut pemindaian hidup. Pemindaian hidup adalah pemrosesan digital untuk membuat sebuah template biometrik yang disimpan dan digunakan untuk pencocokan. Ini merupakan ikhtisar dari beberapa sidik jari yang lebih umum digunakan sensor teknologi (Pemindai Sidik jari, 2017).

2.6 Market Basket Analysis

Market Basket Analysis adalah teknik matematis yang biasa digunakan oleh marketing profesional untuk menyatakan kesamaan antara produk individu atau kelompok produk (Wen-xiu, 2010).

2.6.1 Support

Support adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu *item/itemset* dari keseluruhan transaksi. Ukuran ini menentukan apakah suatu *item/itemset* layak untuk dicari *confidencenya* (Wen-xiu, 2010).

2.6.2 Confidence

Confidence adalah suatu ukuran yang menunjukkan hubungan antara 2 item secara *conditional*(Wen-xiu, 2010).

2.6.3 Lift

Lift Ratio adalah parameter penting selain *support* dan *confidence* dalam *association rule*. *Lift Ratio* mengukur seberapa penting rule yang telah terbentuk berdasarkan nilai *support* dan *confidence*. *Lift Ratio* merupakan nilai yang menunjukkan kevalidan proses transaksi(Wen-xiu, 2010).

2.7 Metode Knowledge Discovery Data Mining

Knowledge Discovery Data Mining adalah proses yang dibantu oleh komputer untuk menggali dan menganalisis sejumlah besar himpunan data dan mengekstrak informasi dan pengetahuan yang berguna(Fayyad, 2001). *Data mining tools* memperkirakan perilaku dan tren masa depan, memungkinkan bisnis untuk membuat keputusan yang proaktif dan berdasarkan pengetahuan. *Data mining tools* mampu menjawab permasalahan bisnis yang secara tradisional terlalu lama untuk diselesaikan. *Data mining tools* menjelajah *database* untuk mencari pola tersembunyi, menemukan informasi yang prediktif yang mungkin dilewatkan para pakar karena berada di luar ekspektasi mereka.

2.7.1 Data Collection

Data Collection adalah proses pengumpulan data untuk nantinya akan diolah(Fayyad, 2001).

2.7.2 Data Selection & Cleansing

Data Selection & Cleansing adalah proses pemilahan data berdasarkan kebutuhan analisis atau klasifikasi dan setelah itu dibersihkan dengan cara eliminasi *noise* dan data yang inkonsisten(Fayyad, 2001).

2.7.3 Data Transformation

Data Transformation adalah penggabungan data ke dalam bentuk yang sesuai untuk penggalian lewat operasi *summary* atau *aggregation*(Kotu & Despande, 2015).

2.7.4 Data Mining/Pattern Discovery

Data Mining adalah proses esensial untuk mengekstrak pola dari data dengan metode cerdas yang setelahnya diidentifikasi menjadi pola yang menarik dan merepresentasikan pengetahuan berdasarkan *interestingness measures*(Kotu & Despande, 2015).

2.7.5 Evaluation

Evaluation adalah tahapan akhir dimana hasil dari *data mining* dinilai kembali, apakah sesuai dengan yang diharapkan dan dapat memproyeksikan keadaan lapangan(Kotu & Deshpande, 2015).

2.8 Microsoft SQL Server

Microsoft SQL Server adalah sebuah sistem manajemen basis data relasional (RDBMS) produk Microsoft. Bahasa *query* utamanya adalah Transact-SQL yang merupakan implementasi dari SQL standar ANSI/ISO yang digunakan oleh Microsoft dan Sybase(Mistry, 2008).

2.8.1 Database

Database adalah kumpulan data yang disimpan secara sistematis di dalam komputer yang dapat diolah atau dimanipulasi menggunakan perangkat lunak (program aplikasi) untuk menghasilkan informasi(Mistry, 2008). Pendefinisian Database meliputi spesifikasi berupa tipe data, struktur data dan juga batasan-batasan pada data yang akan disimpan. *Database* merupakan aspek yang sangat penting dalam sistem informasi karena berfungsi sebagai gudang penyimpanan data yang akan diolah lebih lanjut.

2.8.2 Datatable

Datatable adalah tabel berisi data yang terdapat pada *database*(Mistry, 2008).

2.8.3 Row

Row adalah baris data yang terdapat pada *datatable*(Mistry, 2008).

2.8.4 Column

Column adalah baris kolom yang terdapat pada *datatable*(Mistry, 2008).

2.8.5 Record

Record adalah suatu data individu yang terdiri dari 1 *row* dan beberapa *columns*(Mistry, 2008).

2.9 Rapidminer

Rapid Miner merupakan *software* untuk pengolahan *data mining*. RapidMiner Sebuah lingkungan untuk *machine learning*, *data mining*, *text mining* dan *predictive analytics*(Kotu & Deshpande, 2015). *Machine learning* Algoritma di mana perilaku komputer ber-evolusi berdasarkan data empiris, seperti sensor atau *database*. *Data mining process* mengekstrak pola-pola dari data set yang besar dengan mengombinasikan metoda statistika, kecerdasan buatan dan *database*.

2.9.1 Import to Rapidminer

Proses *import* data dari excel atau SQL kedalam rapidminer(Kotu & Deshpande, 2015).

2.9.2 Select Attributes

Proses pemilihan *attribute* pada *record* data untuk diolah pada rapidminer(Kotu & Deshpande, 2015).

2.9.3 Conversion Nominal to Binomial

Proses konversi data berjenis *nominal* menjadi *binomial*(Kotu & Deshpande, 2015).

2.9.4 Conversion Nominal to Polinomial

Proses konversi data berjenis *nominal* menjadi *polynomial*(Kotu & Deshpande, 2015).

2.9.5 Example Set

Tampilan contoh sebagian hasil pengolahan rapidminer(Kotu & Deshpande, 2015).

2.9.6 Frequent Pattern Growth

Frequent Pattern Growth (FP- Growth) adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data(Kotu & Deshpande, 2015).

2.10 SQL Server Management Studio

SQL Server Management Studio adalah *tools* SQL dengan fungsi untuk mengembangkan, menyebarkan, dan atasi masalah *database* SQL Server, serta perangkat tambahan untuk fungsi sebelumnya(Mistry, 2008).

2.10.1 Query

Query adalah kemampuan untuk menampilkan data dari *database* untuk diolah lebih lanjut yang biasanya diambil dari tabel tabel dalam *database*. Pengertian *query* yang lain adalah pertanyaan (*question*) atau permintaan (*order*) informasi tertentu dari sebuah *database* yang tertulis dalam format tertentu. *Query* dapat didefinisikan sebagai perintah-perintah untuk mengakses data pada *database* atau basis data. Sehingga secara garis besar, Pengertian *query* adalah bahasa yang digunakan untuk memanipulasi, mengubah, menambahkan, mengatur sesuatu atau data dalam *database*(Mistry, 2008).

UMMN