BAB II

TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penulis mengamati penelitian-penelitian yang cukup relevan terhadap topik penelitian ini.

Veronica (2015), melakukan penelitian dengan menggunakan 3 algoritma yang berbeda yaitu *Naive Bayes*, algoritma ID3 dan Regresi Linear. Dalam penelitian tersebut penulis menggunakan studi kasus dari Sekolah Menengah Atas Negeri 6 Surakarta untuk memprediksi siswa yang akan menerima berdasarkan prestasinya. Penulis menggunakan sampel sekitar 305 siswa bahwa algoritma ID3 lebih baik daripada algoritma yang lain. Namun dalam penelitian ini penulis hanya membandingkan *accuracy*, *precision*, dan *recall*. Tetapi hasil yang berupa angka bisa dibilang nilai yang mutlak.

Folorunsho (2012), melakukan penelitian dengan menggunakan *Artificial Neural Network*(ANN) dan *Decision Tree Algorithm*. Data yang digunakannya adalah data meteorologi dari tahun 2000 hingga 2009 di kota Ibadan, Nigeria. Kesimpulan yang didapat oleh penulis adalah paramater tersebut mempengaruhi cuaca yang diamati selama periode tersebut. ANN dapat mendeteksi antara hubungan variabel *input* dengan *output* berdasarkan pola yang melekat pada data tanpa perlu menggunakan software pembantu. ANN dapat mendeteksi parameter cuaca dan menggunakannya sebagai prediksi cuaca di masa depan. Parameter yang dapat diprediksi yaitu kecepatan angin, evaporasi, radiasi, temperatur, dan curah di pada bulan disetiap tahunnya.

Jo Ting (2006), melakukan penelitian dengan menggunakan algoritma Apriori dan Time Series Mining. Data yang digunakan adalah data biasa yang hanya berupa huruf seperti A, B dan C. Penulis mencoba mengevaluasi kinerja

dari algoritma-algoritma tersebut dan keefektifannya, penulis hanya berkonsentrasi pada harga dan data stok yang berpengaruh sehingga dapat dipertimbangkan untuk proses konversi numerik yang nantinya akan menghasilkan aturan asosiatif.

Tabel 2. 1. Perbandingan Penelitian

No	Item Pembanding	Veronica	Folorunsho	Jo Ting	Erick
1	Algoritma	Naive Bayes, ID3, Regresi Linear	Apriori, FP-Growth	Apriori, Time Series	Apriori, FP-Growth
2	Pokok Penelitian	Performa Algoritma	Performa Algoritma	Performa Algoritma	Tren Kenaikan
3	Klasifikasi pada	Prediksi Prestasi Mahasiswa	Prediksi Cuaca	Data Kecil	Harga Barang Ban dan Oli

2.2. Landasan Teori

2.2.1. Definisi Data Mining

Data mining adalah sebuah ilmu untuk mengumpulkan, membersihkan, memproses, menganalisis, dan mendapatkan ilmu yang berguna (Charu C. Aggarwal,2015). Dalam data mining terdapat beberapa metode yang bisa digunakan.Metode tersebut digunakan untuk menemukan pengetahuan yang nantinya digunakan.Metode yang ada seperti *classification, clustering, regression, dependency modeling, deviation change detection,* dan *summarization*.

2.2.2. Peramalan

Peramalan yang biasa dikenal dalam data mining

dengan*forecasting*atau *prediction*merupakan suatu hal yang dapat memperkirakan yang akan terjadi di masa depan. Namun pada dasarnya makna kedua kata tersebut tidak sama, istilah *forecasting* merupakan sebuah kegiatan yang meramal dengan metode kuantitatif sedangkan istilah *prediction* merupakan kegiatan yang meramal dengan metode kualitatif.

2.2.3. Data, Informasi dan Pengetahuan

Menurut Ralston dan Reilly (Chamidi, 2004: 314), data dapat didefinisikan sebagai fakta atau apa yang dikatakan sebagai sebuah hasil dari sesuatu observasi terhadap fenomena yang terjadi. Hasil observasi terhadap fenomena yang terjadi maksudnya adalah data dapat berupa gambar maupun tulisan.Dapat dikatakan juga bahwa data masih berupa bentuk mentah karena tidak mengandung informasi apapun.

Sedangkan informasi menurut George H. Bodnar (2000:1), informasi adalah data yang diproses dan diolah sehingga dapat dijadikan sebuah dasar untuk mengambil keputusan yang tepat.Dan juga menurut Gordon B. Davis (1991:28), informasi adalah data yang diolah menjadi bentuk yang berarti bagi penerimanya dan bisa bermanfaat untuk pengambilan keputusan saat ini atau mendatang.Jadi dapat disimpulkan bahwa informasi mengandung sebuah pengetahuan yang didapatkan dari pengalaman atau pembelajaran tentang suatu peristiwa tertentu.Sepertinya contohnya adalah berita, biasanya berita menyampaikan suatu peristiwa yang terjadi dan hal yang disampaikan tersbut adalah informasi.

Sumber dari informasi tersebut ialah data. Data tersebut diolah dengan suatu metode atau algoritma sehingga dapat menghasilkan sebuah informasi yang kemudian sang penerima informasi mendapatkan informasi tersebut sehingga dapat melalukan suatu tindakan atau sebuah keputusan yang menjadikan sejumlah data kembali.

Untuk pengetahuan adalah hal yang dimiliki manusia untuk memahami sesuatu berdasarkan informasi yang telah diterimanya. Setiap

manusia memiliki pengetahuan yang berbeda-beda berdasarkan informasi yang sama. Sehingga informatsi dapat menjadi sarana untuk meningkatkan kegiatan dibidang ilmu pengetahuan. Menurut Nitecki (Pendit, 1992:81) hubungan informasi dengan pengetahuan lebih ditekankan kepada sebuah proses yang bersambungan. Informasi tidak dianggap berhubungan dengan pengetahuan dikarenakan informasi adalah bagian dari sebuah hubungan yang disadari oleh manusia.Sehingga konsep ini merujuk ke suatu hubungan yang berlanjut antara informasi yang baru diporoleh dengan pengetahuan yang masih statis ketika informasi belum diterima.

2.2.4. Metode Algoritma Apriori

Algoritma Apriori adalah sebuah algoritma yang cukup terkenal untuk menemukan sebuah pola yang memiliki frekuensi tinggi algoritma ini pertama kali diperkenalkan oleh Agrawal dan Shrikant pada tahun 1994. Algoritma Apriori juga banyak diimplementasikan untuk data mining. Rumus yang digunakan untuk mencari *support* adalah:

$$Support\ (A) = \frac{Jumlah\ Transaksi\ mengandung\ A}{Transaksi\ Total}$$

Sedangkan untuk rumus matematisnya memiliki beberapa langkah yang pertama adalah menentukan *minimum support,* pada iterasi pertama hitung item dari support atau transaksi yang memuat seluruh item dengan memindai basis data untuk 1-*itemset,* setelah 1-*itemset* didapatkan maka dilihat apakah kondisi dari 1-*itemset* diatas *minimum support* atau tidak, apabila telah memenuhi minimum support, maka 1-*itemset* tersebut menjadi frekuensi tertinggi.

Kemudian pada iterasi kedua untuk mendapatkan 2-*itemset*, kembali dilakukan kombinasi dari k-*itemset* sebelumnya, kemudian dipindai lagi dari basis data untuk dihitung item-item yang memenuhi *support*.*Itemset* yang memenuhi *minimum support* dipilih sebagai pola

frekuensi tertinggi. Setelah itu menetapkan nilai k-itemset dari *support* yang telah memenuhi minimum support dari k-itemset. Proses iterasi berlanjut hinga tidak ada lagi k-itemset yang tidak memenuhi *minimum support*.

Setelah semua pola frekuensi tertinggi telah ditemukan, maka langkah selanjutnya mencari aturan asosiatif yang memenuhi syarat minimal untuk *confidence* dengan menghitung *confidence* yang memiliki aturan A->B yang diperoleh dari rumus:

$$Confidence = P(B|A) = \frac{Jumlah\ Transaksi\ mengandung\ A\ dan\ B}{Jumlah\ Transaksi\ mengandung\ A}$$

Seperti contoh misalnya ketika sebuah toko memiliki data transaksi sebagai berikut.

Tabel 2.2 Contoh Data Uji

ID	***Items***
1	A,B,C
2	B,C,D,E
3	D,E,F
4	B,C
5	B,C,D,E,G

Lalu tentukan *minimum support* yaitu yang bernilai 2.Maka untuk iterasi 1 hitung dan pindai *database* agar mendapatkan sebuah pola dari *support*.

Tabel 2.3 Contoh 1-itemset

Itemset	***Count***	***Support %***
A	1	20
B	4	80

C	4	80
D	3	60
E	3	60
F	1	20
G	1	20

Setelah itu hapus data yang tidak memenuhi *minimum support* menjadi seperti berikut.

Tabel 2.4 Contoh Pola Frekuensi 1-*itemset*

Itemset	***Count***	***Support %***
B	4	80
C	4	80
D	3	60
E	3	60

Selanjutnya kita memulai iterasi kedua, iterasi sebelumnya pola frekuensi dari support telah didapat dari 1-*itemset*.Maka untuk mencari k-itemset menggunakan kombinasi seperti berikut.

Tabel 2.5 Contoh Kombinasi 2-itemset

Itemset
B,C
B,D
B,E
C,D
C,E
D,E

Barulah melakukan iterasi kedua yang dinamakan C2 yang dihitung

masing-masing frekuensi item dan didapatkan hasil sebagai berikut

Tabel 2.6 Contoh *2-itemset*

Itemset	***Count***	***Support %***
B,C	4	80
B,D	2	40
B,E	2	40
C,D	2	40
C,E	2	40
D,E	3	60

Seletah itu memangkas k-itemset dengan menghitung *support* dari *itemset*, apabila salah satu *itemset* tidak muncul {D,F} maka dihapus agar dapat menghemat memori dan lakukan iterasi ketiga sebagai berikut.

Tabel 2.7 Contoh Kombinasi 3-*itemset*

Itemset
B,C,D
B,C,E
C,D,E

Kandidat dari 3-itemset yang telah memenuhi minimum support maka akan menjadi panutan k-itemset selanjutnya.

Tabel 2.8 Contoh 3-*itemset*

Itemset	***Count***	***Support %***
B,C,D	2	40
B,C,E	2	40
C,D,E	2	40

Setelah itu lakukan iterasi ke 4 karena jumlah support yang sama maka hasilnya sebagai berikut.

Tabel 2.9 Contoh 4-*itemset*

Itemset	***Count***	***Support %***
B,C,D,E	2	40

Karena tidak ada kombinasi lagi yang dapat dibentuk untuk k-itemset selanjutnya, maka proses dihentikan dan telah ditemukan pola frekuensi tertinggi yaitu B,C,D,E.

Selanjutnya membentuk sebuah aturan asosiasi atau *association rules*dengan menghitung *confidence* yang mempunyai aturan A->B. pembentukan aturan asosisatif sebagai berikut.

Tabel 2.10 Aturan Asosiatif

Aturan Asosiatif	***Support (AUB)***	***Support (A)***	***Confidence***
{B,C,D} -> {E}	40%	60%	66%
{E,C,D} -> {B}	40%	80%	50%
{B,E,D} -> {C}	40%	80	50%
{E,C,B} -> {D}	40%	60%	66%
{E,C} -> {D,B}	40%	40%	100%

Pembentukan aturan asosiatif sangatlah penting agar mendapat nilai confidence.Penggunaan algoritma apriori sebenernya cukup memboroskan memori dikarenakan banyaknya iterasi sehingga menghabiskan banyak waktu juga untuk memindai basis data.

2.2.5. Metode Algoritma Frequent Pattern Growth (FP-Growth)

Algoritma Frequent Pattern Growth (FP-Growth) merupakan

pengembangan dari algoritma Apriori, sehingga kekurangan-kekurangan dari algoritma apriori sebelumnyatelah diperbaiki oleh algoritma FP-Growth.FP-Growth adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul dalam sebuah kumpulan data. Pada Algoritma Apriori diperlukan cara*generate candidate* untuk mendapatkan sebuah itemset. Hal tersebut menyebabkan algoritma FP-Growth lebih cepat prosesnya daripada Algoritma Apriori.

Karakteristik algoritma FP-Growth adalah struktur data yang digunakan adalah pohon atau *tree* yang disebut dengan *FP-Tree.*Dengan menggunakan FP-Tree, algoritma FP-Growth dapat langsung mengekstrak frekuensi itemset dari FP-Tree. Penggalian itemset dilakukan menggunakan FP-Growth dengan cara membangkitkan struktur data *tree*.

Metode FP-Growth terbagi menjadi 3 tahap (Han et al. 2006). Ketiga tahapan tersebut adalah :

a) Tahap pembangkitan *conditional pattern base*

   Tahap ini merupakan *subdatabase* yang isinya adalah *prefix path*(lintasan himpunan yang berurutan) dan suffix pattern (pola akhiran). Pembangkitan ini dapat dilakukan melalui FP-Tree yang sudah dibangun sebelumnya.

b) Tahap pembangkitan *conditional* FP-tree

   Tahap ini merupakan *support count*pada setiap item *conditional pattern base* yang akan dijumlahkan kemudian item yang mempunyai *support count* yang lebih besar sama dengan *minimum support* akan dibangkitkan melalui *conditional* FP-Tree.

c) Tahap pencarian *frequent itemset*

   Tahap ini jika conditional FP-Tree adalah *single path*, maka akan didapatkan *frequent itemset* dengan melakukan kombinasi item pada setiap *conditional* FP-Tree. Dan apabila bukan merupakan single path maka, akan dilakukan pembangkitan

FP-Growth secara rekursif.

2.2.6. Lift Ratio

Biasa digunakan untuk melihat kekuatan aturan asosiatif. Cara kerjanya adalah dengan cara *confidence* dibagi dengan *expected confidence*. Untuk rumus *expected confidence* sebagai berikut.

$$Expected\ Confidence = \frac{transaksi\ mengandung\ B}{jumlah\ transaksi}$$

Maka *lift ratio* dihitung dengan membandingkan *confidence* pada salah satu aturan asosiatif dibagi dengan *expected confidence*. Rumus lift ratio sebagai berikut.

$$Lift\ Ratio = \frac{Confidence}{Expected\ Confidence}$$

Apabila nilai *Lift Ratio* lebih dari angka 1 maka aturan asosiasi tersebut memiliki manfaat dan kekuatannya sangat besar (Santosa 2007).