

Bab II

Tinjauan Pustaka

2.1 Pakaian

Berdasarkan KBBI pakaian berasal dari kata “pakai” yang di imbah akhiran “an” adalah benda yang dipakai (baju dan celana). Selain itu juga berfungsi untuk melindungi tubuh manusia dari objek asing, dan dibutuhkan untuk menutupi tubuh dan terutama aurat. Seiring berjalanya waktu, fungsi pakaian menjadi lebih banyak, tidak hanya untuk melindungi tubuh manusia saja. Pakaian dapat digunakan sebagai penambah kepercayaan diri bagi pengguna, pakaian dapat menambah kecantikan/ketampanan, modis, dan menunjukkan derajat sosial dari pemakai. Terdapat berbagai jenis pakaian yang termasuk baju pria (Kho dan Loembie, 2013) adalah sebagai berikut:

1. *Dress Shirt*

Dress Shirt merupakan pakaian yang umumnya digunakan di kantor dan kegiatan yang formal. Pakaian ini ada juga varian berlengan pendek untuk lebih *casual*.

2. *Camp Shirt*

Camp Shirt merupakan pakaian yang umumnya digunakan untuk berkemah. Pakaian ini bersifat tidak terlalu formal dan *casual*.

3. *Guayabera Shirt*

Guayabera Shirt merupakan varian dari pakaian kemah. Pakaian ini dipopulerkan oleh Charlie Sheen. Pakaian ini cocok digunakan untuk acara formal.

4. *T-Shirt*

T-Shirt merupakan pakaian yang wajib dimiliki pria. Pakaian ini bersifat fleksibel untuk atasan. Pakaian ini cocok dipadukan dengan kemeja sebagai dalaman.

5. *Long T-Shirt*

Long T-Shirt merupakan varian dengan pada bagian tangan lebih panjang. Pakaian ini cocok digunakan dengan kemeja sebagai dalaman untuk menciptakan penampilan yang bersifat rapi.

6. *T-Shirt (Ringer)*

T-Shirt (Ringer) merupakan varian *T-Shirt* dengan hem pada lengan dan kerah berbeda dari kaosnya. Pakaian ini cocok digunakan untuk acara *casual* karena pakaian ini berkesan santai dan cocok untuk digunakan kegiatan olahraga.

7. *Raglan*

Raglan merupakan pakaian didesain bagi orang dengan bahu yang tidak terlalu lebar. Pakaian ini umumnya cocok digunakan untuk kegiatan yang bersifat *casual*.

8. *Polo Shirt*

Polo Shirt atau merupakan pakaian yang disebut sebagai *for that smart and casual look*. Hal ini disebabkan pakaian ini terkesan *smart* dan *fun*.

9. *Henley Shirt*

Henley Shirt merupakan *Polo Shirt* tanpa kerah yang memberi kesan lebih santai. Pakaian ini cocok digunakan untuk acara non-formal.

10. *Rugby Shirt*

Rugby Shirt identik dengan pemain *rugby* dan dikenal oleh logo klub dan pola bergaris-garis. Pakaian ini cocok digunakan untuk kegiatan olahraga.

11. *Sweet Shirt*

Sweet Shirt cocok digunakan untuk yang ingin tampil *sporty*. Pakaian ini cocok digunakan saat berada diluar seperti kemah, mendaki, berlari, dll.

12. *Sleeveless Shirt*

Pada umumnya *Sleeveless Shirt* berbahan denim, namun terdapat juga yang berbahan kemeja. Pakaian ini cocok digunakan saat berada diluar untuk menghindari cuaca panas.

2.2 Perilaku Konsumen

Perilaku konsumen merupakan pembelajaran perilaku individu, kelompok dan organisasi ketika memilih, membeli, menggunakan dan menempatkan barang, jasa, ide atau pengalaman untuk memuaskan keinginan dan kebutuhan konsumen (Kotler dan Keller, 2008). Perilaku konsumen sangat tidak mudah untuk dipahami, hal ini disebabkan karena dalam membeli suatu produk setiap konsumen memiliki pendapatan, umur, pendidikan, dan selera yang berbeda-beda. Dalam perilaku konsumen terdapat empat faktor yang memengaruhi adalah faktor budaya, faktor sosial, faktor pribadi, dan faktor psikologi.

2.3 Data Mining

Data Mining merupakan istilah yang digunakan untuk menguraikan penemuan pengetahuan dalam basis data. Tujuan dari *Data Mining* adalah untuk mendapatkan hubungan atau pola yang berkemungkinan memberikan indikasi yang bermanfaat. *Data Mining* merupakan teknik yang menggunakan statistika, matematika, dan kecerdasan buatan untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dari penelitian yang berbasis data besar (Kusrini dan Taufiq, 2009). *Data Mining* berisikan pencarian pola yang diinginkan dalam database besar untuk membantu mengambil keputusan dimasa yang datang. *Data Mining* memiliki kegunaan dan tugas untuk mengspesifikasikan pola yang harus ditemukan dalam proses *Data Mining*. Secara tugas, *Data Mining* dibagi menjadi dua kategori yaitu prediktif dan deskriptif. Tugas prediktif adalah memprediksi nilai dari atribut tertentu berdasarkan nilai dari atribut lain. Atribut yang telah diprediksi disebut sebagai *variable* tidak bebas atau target. Sedangkan tugas deskriptif adalah menurunkan pola-pola yang meringkat hubungan yang pokok dalam tugas. Pola-pola dalam tugas deskriptif adalah korelasi, tren, kluster, dan anomali.

Menurut Larose (2005) *Data Mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilaksanakan yaitu:

1. Deskripsi

Deskripsi atau *Description* dilakukan oleh peneliti dan analisis secara sederhana untuk mencari cara menggambarkan pola dan kecenderungan dalam data. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola dan kecenderungan.

2. Estimasi

Estimasi atau *Estimation* mirip dengan klarifikasi, yang membedakan adalah target variabel estimasi lebih ke arah numerik daripada kategori. Model dibuat dengan menggunakan rekaman lengkap mengenai nilai dari *variable target* sebagai nilai prediksi. Langkah selanjutnya estimasi nilai dari *variable target* dibuat berdasarkan nilai variabel prediksi. Metode yang terdapat dalam estimasi seperti *Linear Regression*, *Neural Network*, *Deep Learning*, *Support Vector Machine*, dan *Generalized Linear Model*.

3. Prediksi

Prediksi atau *Forecasting* mirip dengan klasifikasi dan estimasi, yang membedakan adalah prediksi nilai dari hasil akan ada dimasa mendatang. Metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat digunakan untuk prediksi, disaat yang tepat. Metode yang terdapat dalam prediksi seperti *Linear Regression*, *Neural Network*, *Deep Learning*, *Support Vector Machine*, dan *Generalized Linear Model*.

4. Klasifikasi

Klasifikasi atau *Classification* dalam *Data Mining* merupakan proses menemukan definisi kesamaan karakteristik dalam suatu kelas. Klasifikasi merupakan salah satu metode yang paling sering digunakan. Klasifikasi dilakukan dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Ketika menggunakan klasifikasi, untuk mendapatkan model perlu dilakukan analisis terhadap *training set*. Sedangkan *test set* digunakan untuk mengetahui tingkat akurasi dari model yang dihasilkan. Metode yang terdapat dalam klasifikasi seperti *Decision Tree (ID3, C4.5)*, *Naive Bayes*, *Linear Discriminant Analysis*, dan *Logistic Regression*.

5. Pengklusteran

Pengklusteran atau *Clustering* merupakan pengelompokan *record*, pengamatan dan membentuk kelas objek yang terlihat mirip. Clustering berbeda dari klasifikasi, dimana tidak adanya *variable target* dalam *clustering*. Metode yang dapat digunakan dalam pengklusteran seperti *K-Means*, dan *K-Medoids*.

6. Asosiasi

Asosiasi atau *association* bertugas untuk menemukan atribut yang muncul dalam satu waktu. Implementasi dari asosiasi adalah *market basket analysis* atau analisis keranjang belanja. Metode yang dapat digunakan dalam asosiasi seperti Algoritma apriori, *Chi Square*, dan *FP-Growth*.

2.4 Klasifikasi *Data Mining*

Klasifikasi *Data Mining* merupakan metode yang digunakan untuk menemukan model yang digambarkan dengan perbedaan kelas data atau konsep yang berfungsi untuk memprediksi kelas dari objek yang sudah diketahui (Annur, 2018). Proses penemuan model merupakan proses membedakan dan menggambarkan kelas data atau konsep yang bertujuan memprediksi kelas dari objek yang label kelasnya tidak diketahui. Dalam klasifikasi terjadi dua proses yaitu yang pertama adalah *process learning* dimana algoritma klasifikasi diciptakan untuk menganalisa *data training* untuk kemudian dipresentasikan ke dalam bentuk *rule* klasifikasi, proses kedua adalah klasifikasi dimana *data testing* dipakai untuk memperkirakan akurasi dari klasifikasi (Gorunescu, 2010). Menurut Gorunescu (2010) terdapat komponen yang mendasari proses klasifikasi yaitu:

1. *Class*

Class merupakan variabel dependen yang memiliki ketergantungan berupa data kategorikal yang mempresentasikan suatu label yang melekat pada objek.

2. *Predictor*

Predictor merupakan variabel dependen yang memiliki ketergantungan yang direpresentasikan oleh atribut tertentu.

3. *Training dataset*

Training data set merupakan sekumpulan data yang mengandung nilai dari dua komponen sebelumnya yang digunakan untuk memilih kelas yang tepat berdasarkan *predictor*.

4. *Testing dataset*

Testing dataset merupakan sekumpulan data baru yang akan diproses dengan menggunakan model yang ditentukan, dan hasil tersebut akan dievaluasi keakuratannya.

2.5 *Decision Tree*

Decision Tree atau pohon keputusan merupakan salah satu dari algoritma klasifikasi. Metode *decision tree* mengubah fakta yang sangat besar menjadi *decision tree* yang merepresentasikan aturan. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi di antara sejumlah calon *variable input* dan *variable target* (Gorunescu, 2010). *Decision tree* merupakan *flowchart* seperti struktur pohon, dimana setiap internal *node* menunjukkan sebuah *test* pada sebuah atribut, tiap cabang menunjukkan hasil dari *test*, dan *leaf node* menunjukkan *class distribution*. Pada *decision tree* terdapat tiga jenis *node* yaitu:

1. *Rood Node*

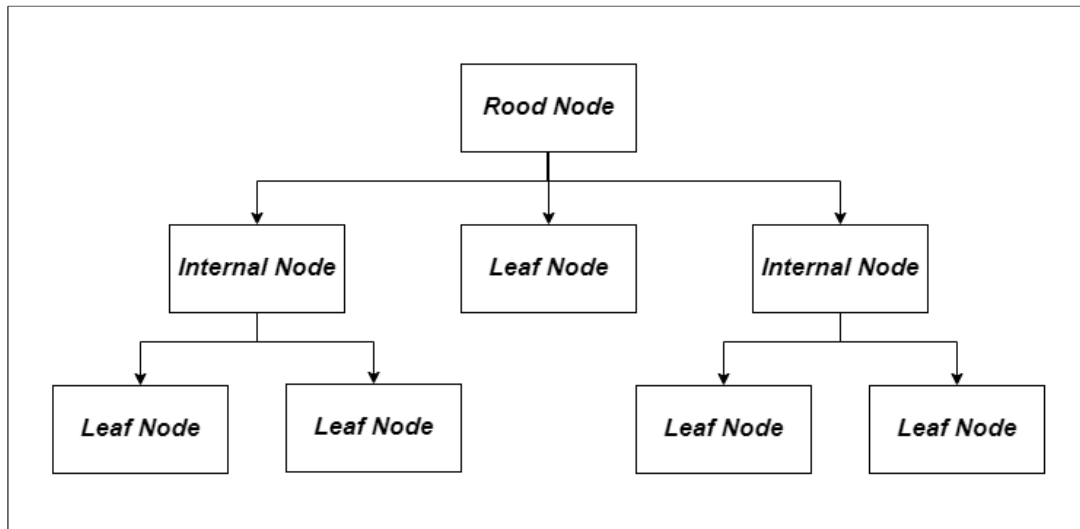
Rood Node merupakan *node* paling atas, dan pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.

2. *Internal Node*

Internal Node merupakan *node* percabangan, dan pada *node* ini hanya satu *input* dan mempunyai *output* minimal dua.

3. *Leaf Node*

Leaf Node merupakan *node* akhir, dan pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.



Gambar 2.1 *Decision Tree*

2.6 *Association Rule*

Association Rule merupakan salah satu metode *Data Mining* yang bertujuan untuk mencari pola yang sering muncul diantara semua transaksi, dimana setiap transaksi terdiri dari beberapa *item*, sehingga metode ini mendukung sistem rekomendasi melalui penemuan pola antar *item*. Metodologi dasar dari *Association Rule* dibagi menjadi dua tahap (Kusrini dan Taufiq, 2009) adalah sebagai berikut:

1. Analisis Pola Frekuensi Tinggi

Pada tahap ini dilakukan pencarian kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam *database*. Nilai *support* atau penunjang merupakan persentase *item* atau kombinasi *item* yang ada pada *database*. Nilai *support* atau penunjang sebuah *item* diperoleh dengan menggunakan rumus:

Untuk *itemset* yang hanya memiliki 1 jenis *item* diperoleh dengan menggunakan rumus:

$$Support(A) = \frac{\sum \text{Jumlah Transaksi Mengandung A}}{\sum \text{Total Transaksi}} \quad (2.1)$$

Sedangkan nilai dari support 2 *item* diperoleh dengan menggunakan rumus:

$$Support(A, B) = P(A \cap B)$$

$$Support(AB) = \frac{\sum \text{Jumlah Transaksi Mengandung A dan B}}{\sum \text{Transaksi}} \quad (2.2)$$

2. Pembentukan Aturan Asosiatif

Langkah selanjutnya setelah semua pola frekuensi tinggi ditemukan adalah mencari aturan asosiatif yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiatif A U B. Nilai *confidence* jika A maka B dapat diperoleh dengan menggunakan rumus:

$$Confidence P(B|A) = \frac{\sum \text{Jumlah Transaksi Mengandung A dan B}}{\sum \text{Transaksi Mengandung A}} \quad (2.3)$$

2.7 Algoritma Apriori

Pada tahun 1994 Algoritma Apriori pertama kali diajukan oleh Agrawal dan Srikan. Penemuan tersebut dilakukan untuk menemukan *frequent Itemsets* pada aturan asosiasi Boolean. Algoritma Apriori adalah untuk mencari *frequent itemset* dari basis data transaksi, lalu untuk menghilangkan *itemset* dengan frekuensi yang rendah berdasarkan level *minimum support* yang sudah ditentukan sebelumnya, dan selanjutnya membangun aturan asosiasi dari *itemset* yang memenuhi nilai *minimum confidence* dalam basis data (Agrawal dan Srikan, 1994). Algoritma Apriori termasuk jenis aturan asosiasi dari *Data Mining*. Aturan yang menyatakan asosiasi antar atribut disebut sebagai *affinity analysis* atau *market basket analysis*.

2.8 Populasi dan Sampel

Populasi didefinisikan sebagai wilayah yang memiliki karakteristik atas objek maupun subjek yang disamaratakan oleh peneliti (Sugiyono, 2015). Populasi tidak hanya manusia saja melainkan termasuk benda mati dan benda hidup lainnya. Populasi terdiri dari gabungan elemen peristiwa pada hal atau barang yang memiliki karakteristik sesuai dengan perhatian peneliti. *Sample* merupakan pecahan dari jumlah serta karakteristik yang dimiliki populasi (Sugiyono, 2015). Sampel merupakan bagian dari populasi yang karakteristiknya ingin diselidiki serta menjadi perwakilan dari keseluruhan populasi.

Dalam mengumpulkan sampel dibagi menjadi 2 yaitu terdapat dua macam sampel yang dapat dilakukan yaitu probabilitas sampel dan non-probabilitas sampel. Probabilitas sampel adalah pengambilan sampel dimana semua populasi mendapat peluang sama. Non-probabilitas adalah pengambilan sampel dimana tidak semua memiliki peluang untuk dipilih menjadi sampel. Teknik yang digunakan pada penelitian ini adalah teknik kuota sampling. Kuota sampling menentukan sampel dari populasi yang memiliki ciri-ciri tertentu sampai jumlah yang dibutuhkan (Sugiyono, 2015). Dengan menggunakan rumus Slovin (Sevilla, 1993):

$$n = \frac{N}{1 + Ne^2} \quad (2.4)$$

2.9 Penelitian Terdahulu

Terdapat beberapa orang yang telah melakukan penelitian terlebih dahulu mengenai analisis perilaku pembelian oleh konsumen dengan menggunakan *Data Mining*.

Penelitian pertama dilakukan oleh Dewi, Anif, dan Fenty (2016) dengan judul “Penerapan Metode Asosiasi Menggunakan Algoritma Apriori pada Aplikasi Analisa Pola Belanja Konsumen”. Pada penelitian ini dilakukan di Toko Buku. Analisis dilakukan pada informasi penjualan yang dikumpulkan dan disimpan pada *data base*. Dilakukan *association rule* untuk mengetahui kebiasaan buku-buku yang sering dibeli secara bersamaan. Proses pencarian asosiasi dilakukan menggunakan bantuan algoritma apriori untuk menghasilkan pola kombinasi *item* dan *rules* sebagai ilmu pengetahuan informasi penting dari data transaksi penjualan. Hasil penelitian ini berupa aplikasi untuk menganalisa pola belanja yang dapat dijadikan sebagai rekomendasi dalam menentukan strategi penjualan oleh pihak Toko Buku.

Penelitian kedua dilakukan oleh Alfansa dan Algifanri (2018) dengan judul “Algoritma Klasifikasi *Data Mining* untuk Memprediksi Siswa dalam Memperoleh Bantuan Dana Pendidikan”. Tujuan penelitian tersebut untuk menyusun Bantuan Dana Pendidikan bagi siswa yang kurang mampu secara ekonomi. Penelitian ini dilakukan dengan menggunakan algoritma klasifikasi *Data Mining* yaitu Algoritma C4.5 untuk

memprediksi siswa dalam memperoleh bantuan dana pendidikan. Dari hasil pengujian diperoleh nilai *accuracy* algoritma C4.5 sebesar 98.8%, nilai presisi sebesar 98.02%, dan nilai sensitivitas sebesar 99%. Dengan demikian algoritma C4.5 merupakan algoritma dan teknik terbaik untuk memprediksi siswa dalam memperoleh bantuan dana pendidikan.

Dari dua contoh terdahulu pada paragraf sebelumnya dengan penelitian ini memiliki persamaan dari tujuan penelitian yaitu menganalisis perilaku konsumen dengan menggunakan metode *Data Mining*. Selain persamaan terdapat juga perbedaan dari dua penelitian terdahulu dengan penelitian ini yaitu perbedaan tempat penelitian, waktu penelitian, serta objek.