**PENERAPAN DATA MINING UNTUK PREDIKSI PENJUALAN WALLPAPER MENGGUNAKAN ALGORITMA C4.5**



**Oleh :**

**Nira Agustina**

**09031181520034**

**SISTEM INFORMASI**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS SRIWIJAYA**

**2018**

Knowladge Discovery In Database (KDD) adalah proses menemukan pengetahuan yang berguna dari kumpulan data. Teknik penambangan data yang banyak digunakan ini adalah proses yang meliputi persiapan dan pemilihan data, pembersihan data, menggabungkan pengetahuan sebelumnya pada set data dan menafsirkan solusi akurat dari hasil yang diamati. Area aplikasi KDD utama termasuk pemasaran, deteksi penipuan, telekomunikasi dan manufaktur.

**Tahap-tahap KDD :**

a. Data selection adalah Pemilihan data baru sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining disimpan dalam suatu berkas, terpisah dari basis data operasional.

b. Preprocessing/Cleaning adalah Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak.

c. Transformasi yaitu Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan cari dalam basis data.

d. Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.

e. Evaluasi adalah Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.

**PENERAPAN DATA MINING UNTUK PREDIKSI PENJUALAN WALLPAPER MENGGUNAKAN ALGORITMA C4.5**

 Saat ini, untuk memperindah dinding tidak selalu menggunakan cat dinding, namun ada berbagai alternatif telah tersedia. Dalam suatu rumah, dinding merupakan bagian yang paling dominan yang selalu mengelilingi kita. Sehingga, bagian ini pula yang paling sering kita lihat saat berada di rumah. Dinding dalam bangunan rumah tidak hanya dipandang sebagaipemisah antar ruang, tetapi juga menjadi media bagi penghuni untuk mengekspresikan diri. Salah satu alternatif, untuk penutup dinding yang digemari saat ini adalah wallpaper atau kertas dinding.

Dalam kehidupan sehari-hari, dengan sadar maupun tidak sadar manusia saat ini sudah terpengaruh oleh berbagai produk dan desain yang ada disekitarnya, misalnya saja desain wallpaper yang dapat membuat seseorang merasa betah didalam suatu ruangan. Banyak faktor yang dapat menjadi pengaruh jika berbicara mengenai pemilihan sebuah desain wallpaper untuk dinding rumah, konsumen dihadapkan pada berbagai macam pilihan merk dengan variasi warna dan desain yang berbeda-beda. Selain itu untuk menghitung biaya yang dibutuhkan, diperlukan perhitungan yang cukup rumit. Hal tersebut mengakibatkan konsumen seringkali mengalami kesulitan dalam memilih wallpaper. Perkembangan teknologi komputer yang selalu berkembang dalam usaha memenuhi kebutuhan manusia di bidang komunikasi, informasi dan pengolahan data. Sebagai contoh penjualan wallpaper yang berkembang dan tumbuh sangat cepat. Dengan semakin meningkatnya konsumen yang membutuhkan wallpaper, maka semakin banyak pula permasalahan yang timbul di dalam memberikan pelayanannya kepada pelanggan. Sistem persediaan wallpaper harus disiapkan dengan maksimal agar penjual dapat memenuhi keinginan konsumen. Untuk mempermudah penjual memilih wallpaper mana yang banyak diminati konsumen agar disediakan stok untuk wallpaper tersebut, maka perlu diprediksi untuk penjualan wallpaper terbanyak dengan metode klasifikasi menggunakan metode data mining algoritma c4.5.

Knowledge Discovery in Databases (KDD) adalah penerapan metode saintifik pada data mining. Dalam konteks ini data mining merupakan satu langkah dari proses KDD. Beberapa teknik dan sifat data mining adalah sebagai berikut:

 a. Classification [Predictive]

 b. Clustering [Descriptive]

 c. AssociationRule Discovery [Descriptive]

 d. Regression [Predictive]

e. DeviationDetection [Predictive]

Cara algoritma C4.5 untuk membangun pohon keputusan yaitu:

 a. Pilih atribut yang akan digunakan sebagai akar

b. Buatlah sebuah cabang untuk setiap nilai

c. Bagilah kasus dalam sebuah cabang

d. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama. Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 .

1. Menyiapkan data training. Data ini diambil dari data yang sudah pernah ada sebelumnya dan sudah dikelompokan kedalam kelas tertentu.

2. Setalah itu tentukan akar dari pohon. Pilih akar dari atribut, cara adalah dengan menghitung nilai gain dari semua atribut, yang menjadi akar pertama adalah nilai gain yang paling. Sebelum menentukan nilai gain, terlebih dahulu hitung nilai entropy.

Untuk menentukan nilai entropy gunakan rumus :



3. Setelah itu tentukan nilai gain menggunakan rumus:



4. Setelah itu ulangilah langkah ke-2 sampai semua record terpartisi secara sempurna.

5. Proses partisi pohon keputusan akan berhenti saat :

 a. Semua record dalam simpul N mendapat kelas yang sama.

 b. Tidak ada atribut di dalam record yang dipartisi lagi.

c. Tidak ada record di dalam cabang yang kosong.

Pohon (tree) adalah sebuah struktur data yang terdiri dari simpul (node) dan rusuk (edge). Simpul pada sebuah pohon dibedakan menjadi tiga, yaitu simpul akar (root node), simpul percabangan/ internal (branch/ internal node) dan simpul daun (leaf node), [6]. Pohon keputusan merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas berhingga, dimana simpul internal maupun simpul akar ditandai dengan nama atribut, rusuk-rusuknya diberi label nilai atribut yang mungkin dan simpul daun ditandai dengan kelas-kelas yang berbeda.

Tabel .1 data walpaper



Pada tabel di atas terdapat 30 penilaian dari 6 penilaian yang menentukan minat konsumen terhadap wallpaper.

Tabel.2 Nilai Gain dan Nilai Entropy



Tabel di atas dapat diketahui bahwa Gain tertinggi adalah jumlah motif yaitu sebesar 0.419973. Dengan demikian, jumlah motif dapat menjadi node akar. Ada dua nilai atribut dari jumlah motif, yaitu banyak dan sedikit, sehingga perlu dilakukan perhitungan lagi.



Gambar 1 Decision Tree

Pada gambar di atas dapat diambil keputusan bahwa jika jumlah motif banyak maka minat konsumen banyak. Jika jumlah motif sedikit maka minat konsumen sedikit.

**Kesimpulan :**

Berdasarkan hasil penelitian yang dilakukan, maka penulis dapat menarik kesimpulan bahwa pembelian wallpaper dengan menggunakan metode Data Mining khususnya Algoritma C4.5 akan bermanfaat sekali dalam proses pengambilan keputusan dalam pembelian wallpaper.

1. Yang menjadi faktor tertinggi yang mempengaruhi penjualan adalah faktor jumlah motif wallpaper.

2. Faktor Harga, Ukuran, Kualitas Bahan, dan Warna tidak mempengaruhi pembelian karena wallpaper dengan harga mahal, ukuran yang kecil, kualitas bahan yang baik, dan warna yang sedikit ternyata masih diminati oleh pelanggan.