

PROGRAM BANTU PEMILIHAN PAKAIAN DAN BAHAN BATIK BAGI KONSUMEN DENGAN PENDEKATAN DECISION TREE Studi Kasus : Toko InBATIK

Abet Alpha P
Yetli Oslan

Abstrak

Ketepatan pemilihan rekomendasi produk bagi konsumen merupakan hal yang cukup penting untuk diperhatikan dalam sebuah proses penjualan batik karena seringkali menjadi hal yang cukup sulit ditangani oleh pihak pemilik usaha dan karyawan. Hal ini terjadi karena proses pemilihan produk yang terbatas pada pengetahuan pemilik usaha dan karyawan serta banyak faktor yang perlu dipertimbangkan dalam meninjau kriteria produk yang tepat bagi konsumen. Pada akhirnya kegiatan ini membutuhkan waktu yang cukup lama dan rangkaian proses yang sulit, dan seringkali menghasilkan keputusan yang kurang sesuai dengan kebutuhan konsumen.

Rekomendasi produk melalui sistem diwujudkan dengan cara memberikan beberapa macam contoh rekomendasi berdasarkan inputan kriteria produk pilihan konsumen. Disamping itu, secara efisien sistem juga dapat menentukan urutan kriteria berdasarkan perhitungan Algoritma C4.5. Algoritma C4.5 mengklasifikasikan data dan membentuk pola dari data transaksi penjualan serta setup data barang baru. Aturan keputusan pengurutan kriteria menurut perhitungan decision tree, melalui proses learning dengan melakukan perhitungan nilai information gain tertinggi dari setiap atribut.

Penelitian ini akan menghasilkan sebuah sistem yang dapat memberikan rekomendasi produk sesuai dengan kebutuhan konsumen. Dengan adanya sistem tersebut, proses pemilihan produk tidak berbelit-belit, efisiensi waktu dan hasil rekomendasi akan sesuai dengan harapan konsumen.

Kata Kunci : Rekomendasi Pakaian dan Bahan Batik, Decision Tree, Algoritma C45

1. Pendahuluan

InBatik merupakan salah satu toko batik di Yogyakarta yang menjual berbagai macam batik. Kegiatan untuk menentukan batik yang tepat bagi konsumen menjadi proses yang cukup sulit. Hal ini dikarenakan proses pemilihan batik terbatas pada tingkat pengetahuan pemilik usaha atau karyawan serta banyak faktor yang perlu dipertimbangkan dalam meninjau kriteria produk yang tepat bagi konsumen. Selain itu, penilaian dan evaluasi pada setiap produk perlu dilakukan agar memperoleh produk yang tepat bagi konsumen. Pemilik usaha atau karyawan perlu melihat terlebih dahulu dari beberapa aspek, seperti motif, harga, warna, dan model yang dibutuhkan konsumen, dan lain-lain. Kegiatan ini membutuhkan waktu yang cukup lama dan rangkaian proses yang sulit, dan seringkali menghasilkan keputusan yang kurang sesuai dengan kebutuhan konsumen.

Peneliti menyarankan untuk mengembangkan suatu sistem yang dapat membantu Toko Inbatik untuk memberikan rekomendasi pemilihan batik yang sesuai dengan kriteria yang dibutuhkan konsumen. Proses pemilihan kriteria batik dilakukan dengan mengimplementasikan perhitungan algoritma C4.5 untuk mengurutkan kriteria pertanyaan yang diajukan kepada pelanggan sehingga pengguna tidak perlu menanyakan semua kriteria kepada pelanggan. Dengan adanya sistem tersebut, diharapkan proses pembuatan keputusan untuk pemilihan batik dapat berjalan dengan efektif, efisien, dan menghasilkan rekomendasi pilihan batik yang sesuai dengan kebutuhan konsumen.

2. Tinjauan Pustaka

2.1. Data Mining

Data mining adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting. Secara teknis, *data mining* dapat disebut sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan *field* dari sebuah relasional *database* yang besar (Mabrur, Ginanjar dan Riani, 2012).

2.2. Decision Tree

(Andriani, 2012) Pada *decision tree* terdapat 3 jenis node, yaitu:

- Root Node*, merupakan *node* paling atas, pada *node* ini tidak ada input dan bisa tidak mempunyai output atau mempunyai output lebih dari satu.
- Internal Node*, merupakan *node* percabangan, pada *node* ini hanya terdapat satu input dan mempunyai output minimal dua.

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

2.3. Algoritma C4.5

Algoritma C4.5 dan pohon keputusan merupakan dua model yang tidak dapat terpisahkan, karena untuk membangun sebuah pohon keputusan dibutuhkan *algoritma C4.5*.

Menurut (Nasari, 2014) Secara umum alur proses algoritma C4.5 untuk membangun pohon keputusan dalam *data mining* adalah:

- Pilih atribut sebagai simpul akar.
- Buat cabang untuk tiap-tiap nilai.
- Bagi kasus dalam cabang.
- Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Pemilihan atribut sebagai simpul, baik akar (*root*) atau simpul internal didasarkan pada nilai *Gain* tertinggi dari atribut-atribut yang ada.

Untuk Menghitung nilai *entropy* digunakan rumus:

$$Entropy(S) = \sum_{i=1}^n -p_i \log_2 p_i$$

Keterangan:

S : himpunan kasus.

n : jumlah partisi S

Pi : proporsi Si terhadap S

Kemudian hitung nilai *informationgain* menggunakan rumus:

$$Gain(S, A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * entropy(S_i)$$

Keterangan:

S : himpunan kasus

A : fitur

n : jumlah partisi atribut A

|S1| : proporsi Si terhadap S

|S2| : jumlah kasus dalam S

Pada penelitian ini *Algoritma C4.5* digunakan untuk perhitungan nilai *information gain* pada setiap parameter penentu keputusan pemilihan batik. Untuk melakukan perhitungan nilai *entropy* dan *information gain* tertinggi dari setiap parameter, sistem akan menghitung berapa banyak jumlah “Ya” dan “Tidak” dari setiap parameter. Berikut ini merupakan contoh kasus penjualan batik untuk memudahkan penjelasan mengenai *algoritma C4.5*.

Tabel 1.
Status Penjualan Batik

No	Model	Ukuran	Warna	Motif	Rentang Harga	Terjual
1	Formal	XXL	Hijau	Kupu	>50000-100000	Ya
2	Formal	L	Hijau	Bunga	>100000-150000	Ya
3	Formal	XXL	Putih	Bunga	>100000-150000	Ya
4	Casual	S	Putih	Bunga	>100000-150000	Ya
5	Casual	XXL	Ungu	Kupu	>50000-100000	Ya
6	Casual	S	Hijau	Kupu	>100000-150000	Ya
7	Casual	L	Hijau	Bunga	>50000-100000	Ya
8	Casual	S	Ungu	Bunga	>50000-100000	Ya
9	Casual	XXL	Putih	Bunga	>50000-100000	Ya
10	Casual	S	Putih	Kupu	>50000-100000	Ya
11	Formal	L	Ungu	Bunga	>50000-100000	Tidak
12	Formal	L	Ungu	Kupu	>100000-150000	Tidak
13	Formal	XXL	Ungu	Bunga	>100000-150000	Tidak
14	Formal	XXL	Putih	Kupu	>100000-150000	Tidak

Langkah-langkah dalam pembentukan pohon keputusan dengan menggunakan *algoritma C4.5* untuk menyelesaikan permasalahan tabel 1:

- a. Menjadikan *information gain* tertinggi dari hasil perhitungan nilai-nilai atribut yang dipilih oleh pengguna sebagai akar.

Menghitung seluruh jumlah kasus, jumlah kasus “Ya” dan “Tidak”, dan *entropy* dari seluruh kasus yang dibagi berdasarkan atribut **MODEL**, **UKURAN**, **WARNA**, **MOTIF**, dan **RENTANGHARGA**. Setelah itu lakukan perhitungan *information gain* dari setiap atribut. Hasil perhitungan dapat dilihat pada tabel 2.

Tabel 2.
Perhitungan Node 1

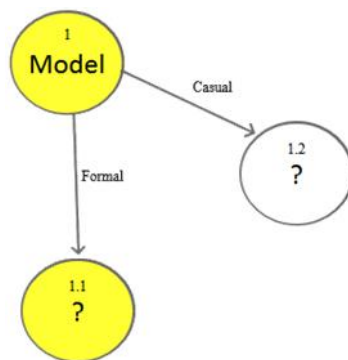
Node	Model	Ukuran	Jumlah Kasus	Tidak	Ya	Entropy	Gain
1	Total		14	4	10	0,863120569	
	Model						
		Formal	7	4	3	0,985228136	
		Casual	7	0	7	0	
	Ukuran						0,183850925
		S	4	0	4	0	
		L	4	2	2	1	
		XXL	6	2	4	0,918295834	
	Warna						0,258521037
		Hijau	4	0	4	0	
		Putih	5	1	4	0,721928095	

		Ungu	5	3	2	0,970950594	
	Motif						0,005977711
		Bunga	8	2	6	0,811278124	
		Kupu	6	2	4	0,918295834	
	Rentang Harga						0,074670111
		>100000-150000	7	3	4	0,985228136	
		>50000-100000	7	1	6	0,591672779	

Dari hasil perhitungan pada tabel 1 dapat diketahui atribut dengan *information gain* tertinggi yang akan menjadi node akar adalah **MODEL** dengan nilai *information gain* 0,370506501:

- b. Membuat cabang dari masing-masing nilai.
MODEL memiliki 2 nilai atribut, yaitu **FORMAL** dan **CASUAL**. Dengan demikian **FORMAL** dan **CASUAL** merupakan cabang dari nilai atribut **MODEL**.
- c. Membagi kasus dalam bentuk percabangan.
Tiap nilai atribut akan menghasilkan perhitungan *entropy*, *information gain* dan bentuk pohon keputusan yang berbeda. Dalam contoh kasus ini pengguna memilih nilai atribut **MODEL = FORMAL** sebagai *node* akar. Bentuk pohon keputusan sementara dapat dilihat pada gambar 1.
- d. Menyaring lebih lanjut pada tabel 1 untuk mempermudah analisis. Diambil data dengan **Model = Formal**, sehingga jadilah tabel 3 seperti berikut ini.

Pada penelitian ini *Algoritma C4.5* digunakan untuk perhitungan nilai *information gain* pada setiap parameter penentu keputusan pemilihan batik. Untuk melakukan perhitungan nilai *entropy* dan *information gain* tertinggi dari setiap parameter, sistem akan menghitung berapa banyak jumlah “Ya” dan “Tidak” dari setiap parameter. Berikut ini merupakan contoh kasus penjualan batik untuk memudahkan penjelasan mengenai *algoritma C4.5*.



Gambar 1. Pohon Keputusan Hasil Perhitungan Node 1

Tabel 3.
Status Penjualan Batik Model = Formal

No	Model	Ukuran	Warna	Motif	Rentang Harga	Terjual
1	Formal	L	Ungu	Bunga	>50000-100000	Tidak
2	Formal	L	Hijau	Bunga	>100000-150000	Ya
3	Formal	XXL	Putih	Bunga	>100000-150000	Ya
5	Formal	L	Ungu	Kupu	>100000-150000	Tidak
7	Formal	XXL	Ungu	Bunga	>100000-150000	Tidak
11	Formal	XXL	Putih	Kupu	>100000-150000	Tidak

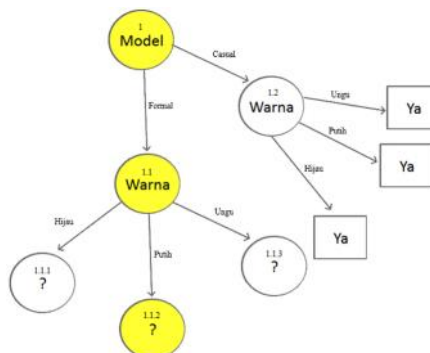
14	Formal	XXL	Hijau	Kupu	>50000-100000	Ya
----	--------	-----	-------	------	---------------	----

- e. Melakukan proses yang sama untuk masing-masing cabang.
Menghitung seluruh jumlah kasus, jumlah kasus “Ya” dan “Tidak”, dan *entropy* dari seluruh kasus yang dibagi berdasarkan atribut **UKURAN**, **WARNA**, **MOTIF**, dan **RENTANGHARGA** yang dapat menjadi *node* akar dari atribut **FORMAL**. Setelah itu lakukan perhitungan *information gain* dari setiap atribut. Hasil perhitungan dapat dilihat pada tabel 4.

Tabel 4.
Perhitungan Node 1.1.

Node			Jumlah Kasus	Tidak	Ya	Entropy	Gain
1.1	Model-Formal		7	4	3	0,985228136	
	Ukuran						0,020244207
		S	0	0	0	0	
		L	3	2	1	0,918295834	
		XXL	4	2	2	1	
	Warna						
		Hijau	2	0	2	0	
		Putih	2	1	1	1	
		Ungu	3	3	0	0	
	Motif						0,020244207
		Bunga	4	2	2	1	
		Kupu	3	2	1	0,918295834	
	Rentang Harga						0,005977711
		>100000-150000	5	3	2	0,970950594	
		>50000-100000	2	1	1	1	

Dari tabel 4 dapat diketahui bahwa atribut dengan nilai gain tertinggi adalah **WARNA**, yaitu dengan nilai gain sebesar 0,69951385. Dengan demikian warna dapat menjadi nilai cabang dari atribut **FORMAL**. **WARNA** memiliki 3 nilai atribut yaitu **HIJAU**, **PUTIH**, dan **UNGU**. Dalam contoh kasus ini pengguna memilih nilai atribut **WARNA = PUTIH** sebagai *node* akar untuk perhitungan *information gain* tertinggi selanjutnya. Dari hasil tersebut dapat digambarkan dalam pohon keputusan sementara seperti gambar 2.



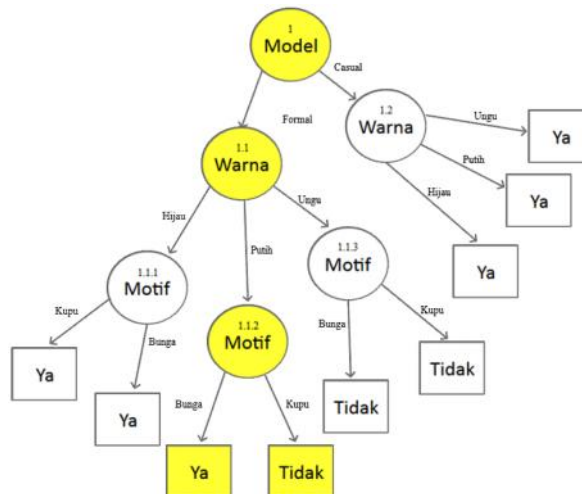
Gambar 2. Pohon Keputusan Hasil Perhitungan Node 1.1

- f. Menghitung seluruh jumlah kasus, jumlah kasus “Ya” dan “Tidak”, dan entropy dari seluruh kasus yang dibagi berdasarkan atribut **UKURAN**, **MOTIF**, dan **RENTANGHARGA** yang dapat menjadi *node* akar dari atribut **PUTIH**. Setelah itu lakukan perhitungan *information gain* dari setiap atribut. Hasil perhitungan dapat dilihat pada tabel 5.

Tabel 5.
Perhitungan Node 1.1.2

Node		Jumlah Kasus	Tidak	Ya	Entropy	Gain
1.1.2	Model Formal dan Warna Putih	7	4	3	0,985228136	
	Ukuran					0
	S	0	0	0	0	
	L	0	0	0	0	
	XXL	2	1	1	1	
	Motif					
	Bunga	1	0	1	0	
	Kupu	1	1	0	0	
	Rentang Harga					0
	>100000-150000	2	1	1	1	
	>50000-100000	0	0	0	0	

Dari tabel 5 dapat diketahui bahwa atribut dengan nilai *information gain* tertinggi adalah **MOTIF**, yaitu dengan nilai *information gain* sebesar 1. Dengan demikian **MOTIF** dapat menjadi nilai cabang dari nilai atribut **PUTIH**. **MOTIF** memiliki 2 nilai atribut yaitu **BUNGA** dan **KUPU**. Nilai atribut **BUNGA** sudah diklasifikasikan menjadi 1, yaitu keputusan “Ya” dan nilai atribut **KUPU** sudah diklasifikasikan menjadi 1 kelas, yaitu keputusan “TIDAK”. Sehingga tidak perlu dilakukan perhitungan lanjut. Dari hasil tersebut dapat digambarkan dalam pohon keputusan seperti gambar 3.



Gambar 3. Pohon Keputusan Hasil Perhitungan Node 1.1.2

Dari gambar 3 dapat dilihat bahwa semua kasus sudah masuk dalam kelas. Jadi Pohon keputusan pada gambar 3 merupakan pohon keputusan terakhir yang terbentuk dan tidak perlu dilakukan perhitungan lebih lanjut pada nilai atribut **RENTANGHARGA**. Jika semua kasus sudah membentuk kelas tetap saat dilakukan perhitungan *entropy* dan *information gain*, maka nilai yang dihasilkan adalah nilai 0, yang artinya pohon keputusan tidak akan terbentuk lagi.

2.4. Batik

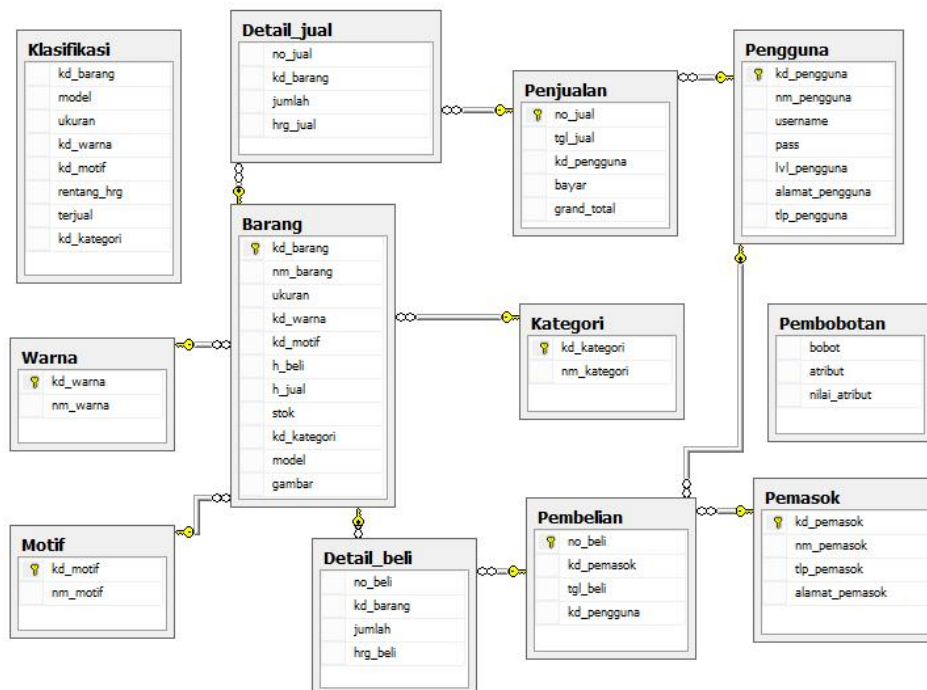
Batik adalah proses penulisan gambar atau ragam hias pada media apapun dengan menggunakan lilin (*wax/malam*) sebagai alat pewarna. Definisi batik ini telah disepakati pada konvensi Batik Internasional di Yogyakarta pada tahun 1997.

Batik merupakan warisan budaya nusantara (Indonesia) yang memiliki nilai dan perpaduan seni yang tinggi, sarat dengan makna filosofis dan simbol penuh makna yang memperhatikan cara berfikir pembuatnya (Muftukhah, 2011).

3. Perancangan Sistem

3.1. Kebutuhan Alat

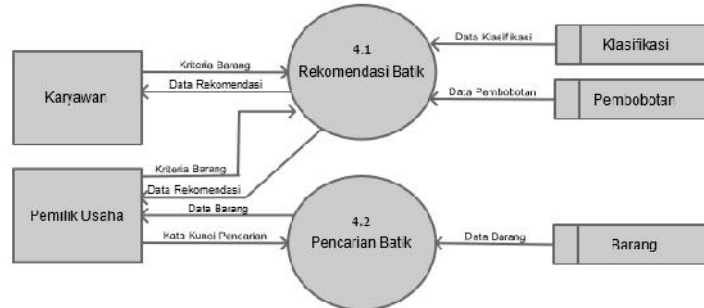
Alat yang digunakan dalam penelitian ini adalah perangkat lunak untuk membangun sistem. Perangkat lunak yang digunakan antara lain *Visual Studio 2010* dan *Microsoft SQL 2008*. *Visual Studio 2010* sebagai *SoftwareDevelopmentTools* yang digunakan untuk membangun sistem informasi. Bahasa pemrograman yang digunakan adalah *VB.NET*. *Microsoft SQL 2008* digunakan untuk membangun database.



Gambar 4. Rancangan Basis Data InBatik

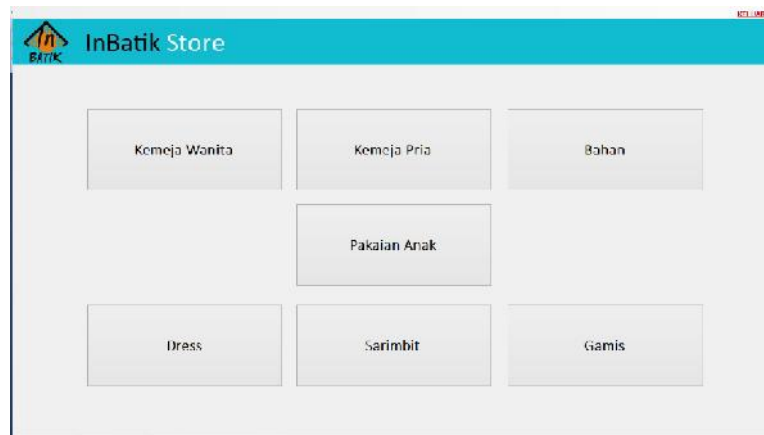
3.2. Data Flow Diagram

Tabel yang terlibat dalam pembuatan aplikasi ini adalah : Klasifikasi, Pembobotan, dan Barang.



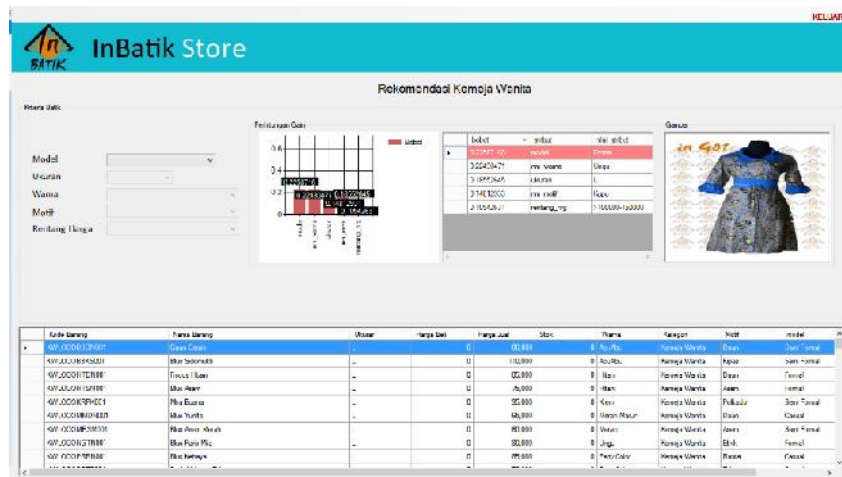
Gambar 5. Rancangan Data Flow Diagram

4. Hasil Dan Pembahasan



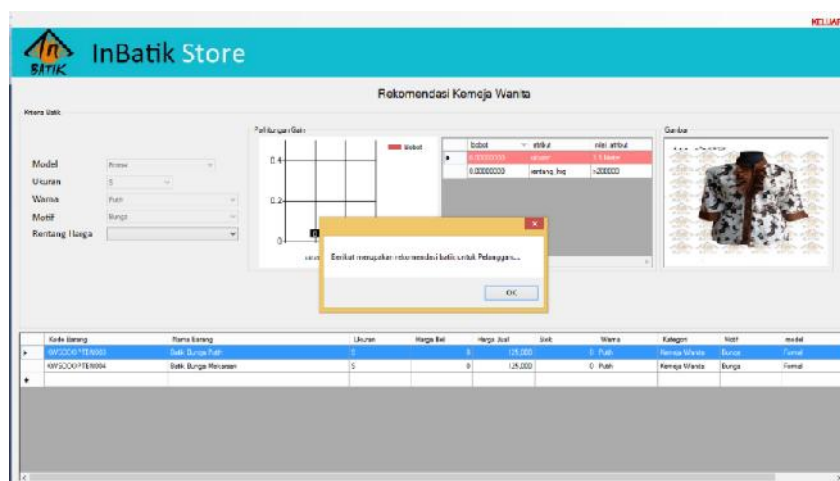
Gambar 6. Halaman Menu Rekomendasi Batik

Proses penyaringan data berdasarkan kategori batik dilakukan sebelum pengguna diarahkan ke halaman rekomendasi batik. Penyaringan data berdasarkan kategori batik dilakukan dengan menggunakan fungsi *filter* pada modul *modul utama* dengan tipe data string dan di atur nilainya pada halaman menu rekomendasi batik. Semua nilai variabel global di atur pada kode program sesuai dengan nama tombol yang mengikuti nama kategori batik yang tersedia.



Gambar 7. Halaman Rekomendasi Batik

Pada Gambar 7 berlangsung proses untuk menghitung *information gain* dari yang tertinggi hingga yang terendah. Agar data yang digunakan *valid*, perhitungan *information gain* berdasarkan transaksi penjualan yang pernah terjadi di Toko InBatik Yogyakarta. Pengkodean dibagi menjadi 2 bagian. Bagian pertama merupakan fungsi untuk menghitung *entropy* tiap *field*. Pada fungsi ini sistem akan melakukan perhitungan jumlah kasus yang memiliki keputusan “Ya” dan “Tidak” pada tiap *field* yang ada, jika salah satu dari *field* memiliki keputusan “Ya” dan “Tidak” bernilai 0 maka nilai *entropy* akan bernilai 0. Fungsi hitung *entropy* memasukkan nilai ke dalam variabel *totale*. Pada bagian kedua sistem melakukan perhitungan *information gain* untuk membentuk pohon keputusan. Sistem akan melakukan perhitungan *information gain* tertinggi untuk penentu keputusan. Dalam perhitungan *information gain*, sistem menghitung satu persatu dari setiap *field* yang ada. Setelah menghitung *information gain* sistem akan memasukkan nilai *information gain* ke dalam variabel **totalGain** dan disimpan ke dalam basis data.



Gambar 8. Rekomendasi Batik

Sistem melakukan perhitungan *information gain* tertinggi untuk mendapatkan urutan pertanyaan kriteria yang dipilih untuk menentukan rekomendasi batik. Pembeli menentukan pilihan kriteria seperti apa yang diinginkan, sehingga dapat diperoleh rekomendasi batik sesuai dengan keinginan pembeli.

Sistem akan melakukan penyaringan data barang berdasarkan inputan dari pengguna. Pada gambar 7 pengguna melakukan input kriteria berdasarkan Model = "Formal", Ukuran = "S", Warna = "Putih" dan Motif = "Bunga" maka pada *datagridview* hanya tampil 2 macam pilihan barang, sehingga mempermudah pembeli untuk menentukan barang yang akan dibeli. Sistem melakukan *multiple filter* untuk mempersempit ruang pada proses pemilihan barang. Ketika rekomendasi barang yang tampil ≤ 3 maka akan ditampilkan pesan barang yang tampil merupakan rekomendasi bagi pembeli. Ketika data barang yang ditampilkan berjumlah lebih dari 3 maka sistem akan melakukan perhitungan gain untuk membuat pohon keputusan selanjutnya. Perhitungan gain dilakukan sampai maksimal rekomendasi barang berjumlah 3.

5. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan maka dapat disimpulkan hal-hal sebagai berikut:

- Algoritma C4.5 dapat digunakan untuk mengurutkan kriteria pertanyaan yang akan diajukan kepada pelanggan.
- Sistem dapat menangani perhitungan *information gain* untuk pembobotan penentu pemilihan kriteria batik.

Adapun saran untuk mengembangkan sistem serupa yaitu:

- Sebaiknya sistem dapat melakukan penambahan kriteria baru untuk penentuan rekomendasi batik, sehingga kriteria yang digunakan tidak bersifat statis, yang hanya dapat menghitung model, ukuran, warna, motif, dan rentang_harga.
- Sebaiknya sistem dapat melakukan perhitungan ulang pada pemilihan kriteria sebelumnya yang telah dipilih oleh pengguna, sehingga ketika pengguna melakukan kesalahan pada saat pemilihan kriteria, pengguna tidak perlu melakukan load ulang pada sistem.
- Sebaiknya perlu dilakukan penelitian dengan metode yang lain sehingga dapat dilakukan perbandingan dan dapat dihasilkan metode yang paling sesuai untuk menangani kasus pemberian rekomendasi batik ini.
- Sebaiknya perlu dipertimbangkan untuk membangun sistem berbasis *webapplication* yang terintegrasi dengan website promosi toko agar pembeli dapat mendapatkan rekomendasi batik dan melakukan transaksi pembelian secara online.

Daftar Pustaka

- Andriani, Anik (2012). "Penerapan Algoritma C4.5 pada Program Klasifikasi Mahasiswa Dropout". Prosiding Seminar Nasional Matematika
- Mabrur, Angga Ginanjar dan Riani Lubis (2012). "Penerapan Data Mining untuk Memprediksi Kriteria Nasabah Kredit". Jurnal Komputer Informatika. Edisi 1, Vol. 1.
- Nasari, Fina (2014). "Analisa Faktor Penyebab Tingginya Listrik Rumah Tangga Menggunakan Algoritma C4.5". UPI YPTK. Padang