

IMPLEMENTASI DATA MINING UNTUK MENGLASIFIKASIKAN PRODUK PADA SEBUAH SUPERMARKET MENGGUNAKAN ALGORITMA ID3 PADA ORANGE

Nengah Widya Utami, I Wayan Budi Suryawan

Sistem Informasi Akuntansi, Teknik Informatika

STMIK Primakara, Jalan Tukad Badung NO 135, Denpasar Selatan

KATA PENGANTAR

Dengan berkembangnya zaman di era digital ini menyebabkan ikutnya meningkatnya kebutuhan sehari-hari setiap individu. Dengan demikian setiap toko maupun supermarket pun berlomba-lomba menyediakan berbagai jenis kebutuhan sehari-hari, mulai dari produk yang berupa makanan dan minuman, barang elektronik, produk kesehatan, dsb.

Namun untuk menjual suatu produk, perlunya melakukan sebuah analisis untuk mengetahui jenis produk yang dibutuhkan dan digemari oleh pasar atau pelanggan yang nanti akan menjadi target dalam memasarkan produk yang akan dijual.

Dengan mengetahui produk yang dibutuhkan oleh pelanggan, maka sebuah toko ataupun supermarket dapat menentukan sebuah produk yang nantinya akan dijual sehingga dapat menarik pelanggan untuk berbelanja pada toko atau pun supermarket mereka.

Berdasarkan permasalahan diatas, pada artikel ini akan melakukan analisis pada penjualan sebuah supermarket untuk mengetahui jenis produk yang sering oleh pelanggan dengan menggunakan algoritma ID3.

TINJAUAN TEORITIS

Produk Supermarket

Produk yang terjual pada supermarket di setiap tempat menyediakan berbagai jenis barang yang tersedia. Namun terkadang pada beberapa supermarket yang memiliki lebih dari satu cabang terkadang memiliki kemungkinan pada setiap cabangnya tidak menyediakan beberapa barang yang sama, sehingga terkadang pelanggan yang berbelanja pada cabang supermarket yang ada di daerahnya memiliki kemungkinan barang yang ingin dibeli

tidak tersedia. Dengan demikian perlunya setiap supermarket untuk mendistribusikan produk yang dijual dengan merata di setiap cabangnya. Untuk mengetahui jenis produk yang dibutuhkan oleh para pelanggan, sebuah supermarket perlunya melakukan analisis pada jenis produk yang digemari oleh pelanggan untuk dibeli pada sebuah supermarket. Analisis yang dilakukan dapat berupa, dengan mendata jenis produk yang sering dibeli oleh pelanggan, produk yang memiliki pembelian yang lebih tinggi daripada produk sejenis lainnya, dsb. Dengan mengetahui setiap dari kebutuhan pelanggan tersebut dapat digunakan sebagai acuan untuk sebuah supermarket dalam memasok jenis produk dengan tepat sehingga dapat menghindarkan kehabisan stok dari produk yang dijual.

ALGORITMA ID3

Algoritma ID3 (*Iterative Dichotomiser 3*) merupakan sebuah algoritma dasar yang digunakan untuk membangun sebuah pohon keputusan (*decision tree*). Proses pencarian yang terjadi pada algoritma ini dilakukan secara menyeluruh (*greedy*) pada setiap kemungkinan pada sebuah pohon keputusan. Pohon keputusan (*decision tree*) digunakan untuk mengklasifikasi sampel data yang nantinya dapat digunakan sebagai pedoman pada masa mendatang. Pohon keputusan (*decision tree*) merupakan sebuah diagram alir yang berbentuk pohon, yang dimana memiliki sebuah *internal node* mengimplementasikan setiap atribut yang akan diuji, pada setiap cabangnya akan mempresentasikan hasil dari atribut yang diuji dan *leaf node* akan mempresentasikan setiap kelas yang ada. Pada algoritma ID3 menghasilkan sebuah

pohon keputusan secara *top-down* (dari atas ke bawah) dengan melakukan prediksi pada atribut yang akan digunakan sebagai *root node* (node akar) sebagai pedoman awal. Berikut merupakan ringkasan singkat dari algoritma ID3:

- a. Pemilihan atribut dengan menggunakan *Information Gain*.
- b. Atribut dipilih berdasarkan pada nilai terbesar pada *Information Gain*.
- c. Buat simpul yang berisi atribut tersebut.
- d. Penghitungan dengan menggunakan *Information Gain* dilakukan terus hingga data telah termasuk dalam kelas yang sama. Atribut yang telah dipilih tidak diikuti lagi dalam perhitungan nilai *Information Gain*.

Pemilihan atribut pada algoritma ID3 menggunakan *Information Gain*. *Information Gain* digunakan untuk mengukur dari keefektifitasan suatu atribut dalam mengklasifikasi data. Dengan memilih data dengan nilai tertinggi, yang bertujuan untuk mendefinisikan *gain*. Kemudian melakukan pengukuran jumlah dari informasi yang ada dengan menggunakan *entropy*.

Untuk dapat menghitung sebuah *Information Gain* menggunakan rumus sebagai berikut:

$$\text{Gain (A)} = \text{entropy(S)} - \sum_{i=1}^k \frac{|S_i|}{|S|} \times \text{entropy}(S_i)$$

Berikut merupakan keterangan pada rumus: A = Atribut pada data

S = Jumlah seluruh sampel data

i = Menyatakan nilai yang mungkin untuk atribut

|S_i| = Jumlah sampel data pada nilai i

METODE

Langkah-langkah yang digunakan pada penelitian ini adalah sebagai berikut.

1. Persiapan

Persiapan yang dilakukan yaitu dengan menentukan topik yang akan dibahas yaitu adalah penjualan pada sebuah supermarket, menyiapkan dataset yang dibutuhkan dalam penelitian dan menginstal Orange *data mining*

2. Pengumpulan Data

Data yang menggunakan *dataset* yang didapat dari situs Kaggle yang dimiliki oleh user Aung Pyae. Data yang digunakan pada dataset tersebut berasal dari data dari beberapa supermarket yang terdapat pada kota yang berasal dari negara Myanmar.

3. Implementasi Algoritma ID3

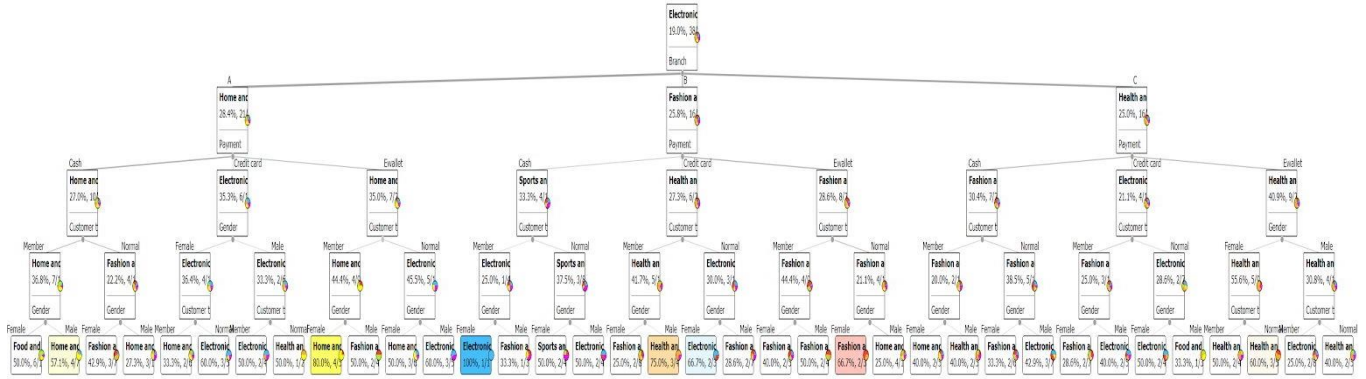
Implementasi dengan algoritma ID3 menggunakan pada aplikasi Orange data mining. Dimulai dengan melakukan import file dari dataset yang digunakan dengan menggunakan file bertipe csv. Kemudian melakukan penggantian *role* pada atribut *Produk Line* yang merupakan jenis produk yang dijual, dari *feature* menjadi target sehingga data dapat disalurkan pada widget *decision tree* dan kemudian dapat dilakukan testing data.

4. Testing

Testing yang dilakukan adalah dengan menerapkan data pada widget data sampler dan kemudian mengambil data pada *dataset* sebanyak 20% proporsi data. Setelah menentukan jumlah data yang dibutuhkan untuk testing, lalu data tersebut akan masuk pada *widget tree* dan *tree viewer* untuk diproses untuk menjadi sebuah *decision tree* dan selanjutnya data akan dievaluasi berdasarkan data dari *widget tree* dan *data sampler*.

TEMUAN DAN PEMBAHASAN

Pada penelitian ini didasarkan pada data yang terdapat pada sebuah supermarket dengan berfokus pada atribut data yaitu jenis produk yang sering dibeli oleh pelanggan dengan menggunakan algoritma ID3. Berdasarkan dari data yang didapatkan lalu akan diproses untuk mendapatkan hasil sebuah hasil berupa *decision tree* sebagai berikut:



Gambar 1. Decision Tree Penjualan pada Sebuah Supermarket Berdasarkan decision

tree diatas diketahui

bahwa jenis produk yang memiliki nilai Information Gain adalah kelas Electronic accessories sehingga menjadi root node. kemudian dilanjutkan melakukan pencarian pada leaf node yang dicari berdasarkan semua data yang dari atribut Product line sehingga

dapat ditemukan banyak kemungkinan dari atribut lainnya berdasarkan cara pembayaran sampai dengan jenis kelamin dari pelanggan. Kemudian berdasarkan hasil dari decision tree diatas sehingga didapat hasil nilai prediksi berdasarkan 20% dari total data 1000 yang menjadi 200 data acak sehingga didapat hasil sebagai berikut:

Model	AUC	CA	F1	Precision	Recall
Tree	0.526	0.210	0.196	0.193	0.210

Gambar 2. Hasil Skor Testing Data

Pada data diatas, terdapat beberapa data diantaranya yaitu AUC(Area under ROC) yang merupakan nilai area dibawah kurva operasi penerima, CA(Classification Accuracy) adalah proporsi yang diklasifikasi secara benar, F1 adalah rata-rata dari Precision yang merupakan proporsi positif benar di antara contoh yang diklasifikasikan sebagai positif dan Recall yang merupakan proporsi positif benar di antara semua contoh positif dalam data. Dengan data yang didapat tersebut, didapatkan juga data berupa Confusion Matrix sehingga didapat data sebagai berikut:

	Electronic accessories	Fashion accessories	Food and beverages	Health and beauty	Home and lifestyle	Sports and travel	Σ
Electronic accessories	8	13	3	1	10	3	38
Fashion accessories	12	11	4	4	5	2	38
Food and beverages	3	10	6	4	1	1	25
Health and beauty	6	10	7	5	5	2	35
Home and lifestyle	7	4	5	3	12	3	34
Sports and travel	9	8	3	3	7	0	30
Σ	45	56	28	20	40	11	200

Gambar 3. Confusion Matrix

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dapat ditarik kesimpulan bahwa dalam klasifikasi data pada dataset supermarket dengan menggunakan algoritma ID3 mendapat beberapa kemungkinan berdasarkan atribut Product line, didapatkan berbagai jenis pembayaran secara Cash, Credit Card, dan E-wallet, dari setiap jenis pembeli yaitu pelanggan Member dan Normal. Kemudian dilanjutkan dari kemungkinan berdasarkan dari Gender pembeli yaitu Male dan Female. Dari pembentukan decision tree, ditemukan bahwa nilai information gain dengan nilai terbesar yaitu berada pada kelas "Electronic accessories" dimana didapatkan beberapa kemungkinan, pada setiap 3 cabang supermarket ditemukan bahwa setiap pelanggan memiliki beberapa kriteria pembayaran yang dilakukan mulai dari menggunakan cash, credit card dan e-wallet. Kemudian pelanggan juga dibagi menjadi dua tipe yaitu normal dan member dan memiliki dua kemungkinan jenis kelamin yaitu male dan female.

REFERENCES

- [1] Khaerul Ma'mur, "Analisis Penerapan Algoritma ID3 dalam Mendiagnosis Kesuburan Pria" *Jurnal Informatika Universitas Pamulang*, vol. 4, no. 2, pp:35-42. 2019.
- [2] Broto Poernomo T.P. dan Rina Dewi Indah Sari, "Penerapan Data Mining Untuk Prakiraan Cuaca Di Kota Malang Menggunakan Algoritma Iterative Dichotomiser 3 (ID3)" *JOUTICLA*, vol. 3, no.2, pp: 101-108. 2017.
- [3] Yusuf Elmande dan Prabowo Pudjo Widodo, "Pemilihan Criteria Splitting Dalam Algoritma Iterative Dichotomiser 3 (ID3) Untuk Penentuan Kualitas Beras : Studi Kasus Pada Perum Bulog Divre Lampung" *Jurnal TELEMATIKA MKOM* vol.4 no.1, pp:73-82. 2012.
- [4] Avia Enggar Tyasti, Dwi Ispriyanti dan Abdul Hoyyi, "Algoritma Iterative Dichotomiser 3 (ID3) Untuk Mengidentifikasi Data Rekam Medis (Studi Kasus Penyakit Diabetes Mellitus Di Balai Kesehatan Kementerian Perindustrian, Jakarta)" *Jurnal Gaussian*, vol. 4, no. 2, pp:237-246. 2015.
- [5] M. Safii, "Implementasi Data Mining Dengan Metode Pohon Keputusan Algoritma ID3 Untuk Menentukan Status Mahasiswa" *Jurnal Mantik Penusa*, vol. 2, no. 1, pp:82-86. 2018.