

PREDIKSI PRODUK E-AUCTION YANG AKAN MENGUNTUNGKAN MELALUI KLASIFIKASI MENGGUNAKAN DECISION TREE (STUDI KASUS : DATA MINING CUP 2006) PREDICTION OF PROFITABLE E-AUCTION'S PRODUCT BY CLASSIFICATION USING DECISION TREE (CASE STUDY : DATA MINING CUP 2006)

Muhammad Doris^{1, -2}

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Semakin tinggi teknologi, maka manusia semakin dipermudah. Hal ini terjadi pada internet, apalagi dunia bisnis. Misalnya pelelangan, biasanya pelelangan barang dilakukan dalam suatu ruangan, dengan internet, pelelangan sudah bisa dilakukan di dunia maya yang tanpa ruang. Di pelelangan, pasti pemilik barang lelang ingin mendapatkan harga yang menguntungkan. Tapi, pada kenyataannya kadang menguntungkan, kadang tidak. Jadi alangkah bagusnya jika si pemilik barang bisa memprediksi terlebih dahulu barangnya yang mana saja yang akan menguntungkan. Cara menyelesaikan permasalahan diatas adalah dengan mempelajari faktor-faktor yang mempengaruhi terlelangnya suatu produk, baik dengan harga yang maksimal ataupun minimal. Faktornya bisa dari sebentar atau lamanya dilelangkan, rendah atau tingginya harga awal, kesudahan pelelangan pada akhir minggu, kategori produk, jenis pelelangan, dan lain-lain. Untuk itu diperlukan penggalian informasi pada data pelelangan yang lalu. Data mining adalah suatu ilmu yang berupaya menemukan pola, aturan, dan informasi berharga dan belum diketahui sebelumnya dari sekumpulan besar data. Dataset e-auction yang ada sudah memiliki kelas maka sebaiknya menggunakan teknik klasifikasi. Klasifikasi berusaha membangun model yang merumuskan kelas dari rekor-rekor data pada training set dan test set serta data baru yang belum diketahui kelasnya. Pada klasifikasi terdapat banyak metode, namun dalam tugas akhir ini hanya akan menggunakan metode decision tree umumnya, algoritma C4.5 khususnya.

Kata Kunci : data mining, klasifikasi, decision tree, algoritma C4.5, training set, test set.

Abstract

Higher technology, people is getting easier. It happens to internet, moreover business world. For example auction, usually it conducts in one room., by internet, it can be done in illusion world without room. In auction, the product's owners hope to get profitable price. But in fact, sometime got satisfiable price, sometime didnot. So, its good if product's owners know which product that will be gainful, before launch the products on auction. The solution of this problem is by learn the factors that influence sales whether with maximum price or minimum price. The factor can be from duration days of auction, low or high start price, ending of auction on weekend, category of product, kind of auction, etc. To solve this problem need mining information from auction data that have been passed. Data mining is one of scince that try to discover pattern, rule, and valuable information that didnot know yet before from a hugh data. Because of e-auction dataset that we have use class, so its better to use classification technique. Classification try to build model on training set which formulate classes from data records from training set and test set and also new eauction that donot have a class yet. In classification there are so many method, but in this final project only use decision tree method commonly, C4.5 algorithm especially.

Keywords : data mining, classification, decision tree, C4.5 algorithm, training set, test set.

BAB I

PENDAHULUAN

I.1 Latar Belakang

E-auction merupakan suatu wadah penjualan barang melalui internet dengan sistem lelang. Pelelangan merupakan penentuan harga suatu barang yang diserahkan pada penawar tertinggi. Biasanya pelelangan dihadiri oleh beberapa orang dalam suatu ruangan dan masing-masing orang yang berminat pada suatu barang mengajukan harga sampai akhirnya tidak ada lagi orang yang mengajukan harga, maka barang jatuh pada penawar tertinggi. Namun dengan adanya *e-auction*, akan memberi peluang kepada para pembeli dari seluruh penjuru dunia untuk ikut menawar barang dengan harga yang sesuai dengan kemampuan masing-masing.

Pada kenyataannya ada barang yang terjual dengan harga yang memuaskan, tapi ada juga yang terjual dengan harga yang murah, sehingga mengakibatkan kerugian besar pada penjual barang. Jadi sebaiknya sebelum meluncurkan barang untuk dilelang, kita harus bisa memprediksi apakah barang tersebut akan mendapatkan harga yang lebih tinggi dari rata-rata hasil penjualan atau tidak. Kita harus bisa memanfaatkan tumpukan data barang *e-auction* yang telah terjual, untuk bisa menemukan suatu pola dan aturan kriteria barang yang kira-kira bisa menghasilkan harga yang tinggi, sehingga layak diluncurkan. Hal ini untuk menghindari kerugian pihak penjual barang. Jadi bisa disimpulkan ada dua kelas yaitu harga optimal dan harga non-optimal. Sehingga proses penggalan informasi yang terbaik untuk data ini adalah dengan metode klasifikasi.

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Ada banyak metode pengklasifikasian yaitu berupa aturan “jika-maka”, berupa *decision tree*, *memory based reasoning*, formula matematis, *naïve bayes* and *bayesian belief networks*, *neural network* atau *support vector machines*. Ilustrasi dari cara pengklasifikasian pada *data mining* adalah terdapat *training set* yang terdiri atas beberapa atribut dan *class*, *training set* tersebut kemudian

diinduksikan dengan algoritma yang ada sehingga terbentuk suatu model sesuai data yang ada, dari model tersebut diaplikasikan ke dalam *test set*.

Penyusun memilih *decision tree* sebagai metode klasifikasi karena mudah untuk diinterpretasi oleh manusia. *Decision tree* adalah salah satu metode klasifikasi yang paling populer. Algoritma *decision tree* yang paling terkenal adalah C4.5. Algoritma C4.5 ini terlebih dahulu membangun *decision tree* dari *training set*, lalu melakukan *prunning* untuk mengurangi kompleksitas dari *tree*, dan mengubah *tree* yang dihasilkan dalam beberapa *rule*. Jumlah *rule* sama dengan jumlah *path* yang mungkin dapat dibangun dari *root* sampai *leaf node*.

I.2 Perumusan Masalah

Rumusan masalah dari penelitian ini adalah:

- Masalah utama/prioritas adalah bagaimana memprediksi data barang pada *test set e-auction* yang akan menguntungkan jika diluncurkan dalam lelang.
- Bagaimana membentuk pola pengklasifikasian dalam bentuk *decision tree*.

I.3 Tujuan

Berdasarkan pada masalah yang telah diidentifikasi di atas, maka tujuan Tugas Akhir ini adalah:

- Memprediksi produk *e-auction* yang akan menguntungkan dengan memperhatikan tingkat keakurasian dan skornya.

I.4 Batasan Masalah

Dalam pembuatan tugas akhir ini, untuk mengatasi permasalahan yang ada maka penyusun membatasi permasalahan sebagai berikut :

1. *Dataset (training set dan test set)* yang akan digunakan sebagai kajian hanyalah data *e-auction* yang diperlombakan pada *Data mining Cup 2006*.
2. Memprioritaskan kualitas prediksi (akurasi) dan skor.

I.5 Metodologi Penelitian

Pendekatan sistematis/metodologi yang akan digunakan dalam merealisasikan tujuan dan pemecahan masalah di atas adalah dengan menggunakan langkah-langkah sebagai berikut:

1. Studi Literatur

Mempelajari literatur – literatur tentang *e-auction*, atributnya, tipe atributnya, jenis recordnya, konsep, dan teknik *data mining* khususnya klasifikasi *decision tree* (algoritma C4.5).

2. Pengumpulan data

Mengumpulkan data yang diperlukan, berupa training set dan test set serta melakukan preproses agar bisa digunakan oleh aplikasi.

3. Perancangan dan Implementasi

Implementasi membangun suatu perangkat lunak untuk melakukan klasifikasi objek-objek ke dalam kelas-kelas tertentu dengan menggunakan algoritma C4.5. Perangkat lunak dibangun menggunakan PHP 5.0.3 sebagai bahasa pemrograman dan MySQL sebagai *database*.

4. Pengujian

Melakukan pengujian perangkat lunak dengan memasukkan data produk *e-auction* yang akan dievaluasi serta mencatat hasil keluaran program.

5. Analisis

Membuat analisis dari hasil implementasi yang telah dibuat dengan memperhatikan keakurasian dan skornya.

Telkom
University

I.6 Sistematika Penulisan

Tugas Akhir ini akan disusun berdasarkan sistematika penulisan sebagai berikut :

BAB I PENDAHULUAN

Menguraikan mengenai latar belakang dari pembuatan perangkat lunak, rumusan masalah, batasan dari masalah yang timbul, tujuan, dan sistematika penulisan yang digunakan.

BAB II LANDASAN TEORI

Menguraikan berbagai teori yang mendukung dan mendasari penulisan tugas akhir ini seperti *data mining*, klasifikasi, *decision tree*, dan algoritma C4.5.

BAB III ANALISIS DAN PERANCANGAN SISTEM

Menguraikan analisis spesifikasi kebutuhan perangkat lunak, perancangan struktur data dan proses dari perangkat lunak yang dibuat.

BAB IV IMPLEMENTASI DAN EVALUASI

Menguraikan implementasi terhadap algoritma C4.5 pada kasus e-auction menggunakan bahasa pemrograman PHP dengan editor Dreamweaver. Selain itu, terdapat evaluasi terhadap hasil prediksi yang didapat dan tingkat keakurasiannya.

BAB V KESIMPULAN DAN SARAN

Berisi tentang kesimpulan dari sistem yang dibuat serta saran untuk pengembangan sistem yang lebih baik.

Telkom
University

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari hasil evaluasi prediksi produk lelang dapat ditarik kesimpulan sebagai berikut :

1. Perangkat lunak ini berhasil memprediksi produk lelang yang akan menguntungkan dengan keakurasian dan skor yang cukup tinggi dan dilengkapi dengan level prediksi sebagai nilai tambah.
2. Keakurasian yang dicapai oleh perangkat lunak ini untuk skor tertinggi (4808) adalah 53.46 %. Skor tersebut jika diurutkan dalam daftar pemenang, berada di peringkat ke-19.
3. Level prediksi akan membantu menguatkan keyakinan pengguna bahwa produknya akan benar-benar menguntungkan jika dilelang.
4. Algoritma C4.5 mampu membentuk decision tree yang kemudian diubah ke dalam bentuk rule.
5. Pembentukan rule yang memiliki akurasi yang tinggi dipengaruhi oleh pengolahan data training dan data testing serta jumlah masing-masing data.
6. Bisa saja terjadi adanya data produk lelang yang tidak bisa diprediksi karena tidak ada satupun rule yang sesuai dengan data tersebut.
7. Algoritma C4.5 kurang baik untuk data yang sangat besar, yaitu dari segi kecepatan.

5.2 Saran

Perangkat lunak ini masih memiliki kekurangan sehingga disarankan sebagai berikut:

1. Untuk proses prediksi yang lebih cepat maka lebih baik digunakan memori yang lebih besar dan processor yang lebih cepat.
2. Perangkat lunak ini bisa diubah ke Bahasa Inggris kemudian diupload ke internet, sehingga siapa saja yang akan melelang produknya (hanya untuk jenis Audio & Hi-Fi:MP3-Player) bisa memprediksi dulu apakah akan laku dengan harga yang memuaskan atau tidak.

DAFTAR PUSTAKA

- [1] **Berry, M.J., Linoff, G.,** *Data Mining Techniques For marketing, Sales, and Customer Support*, Wiley, 1997.
- [2] **Han, Jiawei dan Kamber, Micheline.,** *Data Mining: Concepts and Techniques*. San Francisco, CA : Morgan Kaufmann, 2001.
- [3] **Pramudiono, Iko.,** *Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data*. IlmuKomputer.com, 2003
- [4] **Quinlan, John Ross.,** *C4.5: Programs for Machine Learning*. San Mateo, CA : Morgan Kaufmann, 1993.
- [5] **Quinlan, John Ross.,** Simplifying decision trees. *International Journal of Man-Machine Studies*, 27:221-234, 1987.
- [6] **Roiger , Richard J dan Michael W Geatz.,** *Data mining A Tutorial-Based Primer*. Addison-Wesley Publishing, 2003.
- [7] **T. Imielinski dan A. Virmani.,** *M-sql: Query language for database mining*. Technical report, Rutgers University, January 1996.
- [8] **Tan, P., Steinbach, M., Kumar, V.,** *Introduction to Data Mining*. Addison-Wesley, 2004.
- [9] www.data-mining-cup.com
- [10] www.kdnuggets.com