

ABSTRACT

The focus of the study is to classify the characteristics of category's buying behavior using CHAID algorithm and implement the result using location-based service. To do so, several consecutive phases were conducted; they are pre-processing, labeling, classification, Analysis, and implementation.

The dataset originally contained record more than 24,000 customers and more than 32,000 transactions around 1st August, 2015 – 30th October 2015 in Grand Indonesia, Jakarta. In preparing the data, the model used 6 features such as age group, gender, credit card group, amount of transaction group, buying frequency in the previous 2-3 months and buying behavior in the last 1 month. Then, CHAID algorithm was applied to the datasets. As the result whitelists are made and ready to be implemented through the LBA core.

Experiment shows that the accuracy of the classification depends on many factors such as number of features, total dataset, transactional log period, and varieties of transaction. This conclusion can be seen though validation process where groceries category has the biggest accuracy (94.64%), followed by Food and Beverage category (75.759%) and fashion (73.454%). The model created in this research worked in both telecommunication and bank system.

Keywords: data mining, classification, CHAID Decision Tree Algorithm, Location-Based Service, Transaction Historical Dataset.

INTISARI

Fokus pada penelitian ini adalah mengklasifikasikan karakteristik pembelian kategori menggunakan algoritma CHAID dan menerapkan hasilnya menggunakan promosi berbasis lokasi (LBA). Untuk melakukannya, dilakukan beberapa fase yang berurutan, yaitu pengolahan awal, pelabelan, klasifikasi, analisis, dan implementasi.

Kumpulan data berisi data lebih dari 24.000 pelanggan dan lebih dari 32.000 transaksi sekitar 1 Agustus 2015 - 30 Oktober 2015 di Grand Indonesia, Jakarta. Dalam menyiapkan data, model menggunakan 6 fitur seperti kelompok umur, jenis kelamin, kelompok kartu kredit, jumlah kelompok transaksi, membeli frekuensi dalam 2-3 bulan sebelumnya dan perilaku pembelian dalam 1 bulan lalu. Kemudian, algoritma CHAID diaplikasikan pada dataset. Sebagai hasil, whitelist akan terbentuk dan siap diimplementasikan melalui LBA core.

Percobaan menunjukkan bahwa ketepatan klasifikasi tergantung pada banyak faktor seperti jumlah fitur, jumlah dataset, periode transaksi, dan varietas transaksi. Kesimpulan ini dapat dilihat meskipun proses validasi pada kategori bahan makanan memiliki akurasi terbesar (94,64%), diikuti oleh Makanan dan Minuman (75,759%) dan Fashion (73,454%). Model yang dibuat dalam penelitian ini bekerja dalam dua intitusi yaitu, telekomunikasi dan sistem perbankan.

Kata kunci: *data mining*, Klasifikasi, CHAID Decision Tree Algorithm, Lokasi Berbasis Service, Kumpulan Riwayat Transaksi.