

Customer Classification Based on The Historical Purchase Data

Problem Statement

Classification of the customer can be helpful in targeting specific customer for the specific product. This can remove those irrelevant ads that do not suit for a customer or for which particular customer will not be interested at all. Also, irrelevant ads floated to the customer might lead to losing a customer as a customer might irritate of this. Identification of most valuable person will lead to business luxury. Customer classification is useful for insurance companies for selling insurance. In financial marketing, it is useful for classifying customer based on the type of service they used. In banking, it can be helpful in identifying customers and deciding whether credit card should be issued to a specific customer or whether the loan should be given or to detect a suspected customer. Customer classification can also be used for dividing network bandwidth based on customer performance. Such classification will aid in identifying most profitable customers.

Background

Customers are categorized into different classes based on their purchase history. In food market customer can be classified based on the type of food they eat: Vegetarian, eggetarian or Non-Vegetarian, type of food they buy: Uses Fresh vegetable or Frozen vegetable, based on eating style: Chinese, south Indian, north Indian. In cloth market, they can be classified based on fabric they used, type of dress they use like: Indian dress, western dress, Indo-westerns. In general, they can be classified based on time like: regular, sessional, occasional, monthly, yearly or seldom buyer, based on willingness to buy: tightwads, spendthrift or average spenders. Customers can be classified based on the risk or product groups. Dividing them based on risk will help in developing a customer-oriented business approach.

The RFM classification of customers is based on recency, frequency, and money. Recency means assigning different scores to the customer based on their last purchase period. Assign more score if the purchase is more recent. Frequency refers to how often customer came and made a purchase. And money refers to the price of an item a customer buys.

Many visitors visit websites or come at the shop but only some of them made a purchase. Identifying these parameters will help in deciding the class to which customer belongs. Identifying the class of customer will provide the detail of future demand. It will also help in attracting and recommending customer offers that will influence that class customer to buy. Past studies show that getting a dissatisfied customer back incurs most cost than getting a new customer and getting a new customer is costlier than keeping an existing customer.

In past smartphone users has been classified based on the application usage sequences. Some has segmented the customer based on customer lifetime value. Some has used classifier model using SPSS tool for customer basket data. Some has used Fragmentation-Coagulation Process that splits during fragmentation and merges the customer groups during coagulation step depending on their behavior.

The data available for classification is the personal detail of customer, their transaction history, the frequency of their visit, their click history, their session detail. The commonly used approaches are HMM (Hidden Markov Model), Decision Tree, EM, SVM based customer classification.

Methodology

Identifying right customer and providing right service at right time and treating different types of customers differently is the key to success in business. So, a classifier model will be used to classify customers into different categories based on their purchase history. Then the model can be used to predict category of future coming customer. Here, we will use Deep learning techniques with Keras for classification of the customer.

Step 1 Collecting Customer Data (Purchase History):

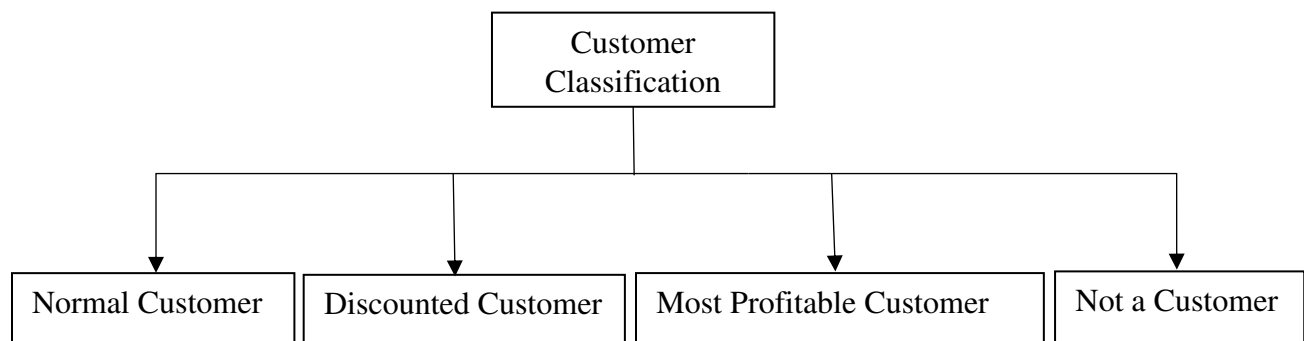
This step involves the collection of purchase data from shopping vendors.

Step 2 Analyzing Data and Training Model:

In this step, the collected data will be analyzed and features will be extracted for training the model. The classifier will be trained for four categories using supervised neural network (RNN) based method.

Step 3 Testing and Classifying Data (Customer):

In this step, the new tuple (customer) will be classified in one of four categories and will be tested against the ground truth.



Evaluation Measures

The performance of a classifier can be evaluated by calculating the accuracy, precision, recall, specificity.

Software and Hardware Requirements

Software

Anaconda python

Deep learning python libraries

Hardware

The model will be trained using NVIDIA GPUs

Dataset

- <https://archive.ics.uci.edu/ml/datasets/wholesale+customers>
- <https://www.kaggle.com/datafiniti/womens-shoes-prices>
- <https://www.kaggle.com/kyanyoga/sample-sales-data>
- <https://archive.ics.uci.edu/ml/datasets/Online+Retail>

Reference

- [1] Ling Luo, Bin Li et. al., Tracking the Evolution of Customer Purchase Behavior Segmentation via a Fragmentation-Coagulation Process, Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)
- [2] XIANG-BIN YAN, YI-JUN LI, Customer Segmentation based on Neural Network with Clustering Technique, 5th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, Madrid, Spain, February 15-17, 2006 (pp265-268)