

Customer Stratification using Data Mining

Tanvi Vinayaraj¹, Darshan Patel², Suraj Kakad³

^{1,2,3}Student, Department of Computer Engineering, Rajiv Gandhi Institute of Technology, Mumbai, India

Abstract: Classifying customers in an enterprise makes it easier to make appropriate decisions. Outlier data includes important customer information. In order to study customer classification problems based on customer assets, a customer classification model based on outlier customer data and their asset was constructed successfully. The model is based on variables in three dimensions including time period, transaction amount and age of the customer. Data mining algorithm like Naive Bayes is used for classifying customers into two categories which are exclusive and economic. The companies and their details that will be recommended to the customers would also be based into two categories. These two categories will be determined by the customers ratings given for that company or services. This categorization is done with the help of collaborative filtering. Based on the ratings companies are recommended to the customers.

Keywords: Customer classification, Data Mining, Naive Bayes, Collaborative filtering, Assets.

1. Introduction

Economy forces the enterprises to adapt their behaviour towards the different circumstances of the market. Filtering data from mass data and applying data mining algorithms to it helps us develop a model. This model makes data analysis and further prediction easier. Data mining algorithms help improve customer relationship management which is quite essential.

Online marketing helps the client trade on the world-wide platform. Identifying and owning excellent customers, developing and maintaining customers in a targeted manner, which avoids the waste of resources and higher costs caused by the decentralization of energy. We intend to build a customer classification model based on customer's assets. Three dimensions of indicators including transaction amount, time period and customer age are used to classify customers into two categories i.e. economic and exclusive customers. The main idea behind this is to make sure that the right services or recommendations are provided to the customer based on their investment amount.

After the customers are categorized, the products and the related services are also classified based on the user ratings which is done with the help of Collaborative filtering. The customer then gets a list of recommended schemes or options based on their product selected. This reduces a lot of valuable time and efforts of going to a counselor. As far as customer assets are concerned, it aims to help enterprises understand customers more deeply, grasp customer needs, and intend to influence customer behavior by managing customers, so as to

promote customers to make greater value contribution to enterprises.

2. Literature review

Classification is mainly used in analysis and statistics. It is mainly identifying to which set of categories a new observation belongs and this is done on the basis of a training data set. Our classification model analysis is done on a big data set which includes data related to our customers. Any enrolment or registration of a new customer who opts to our services and benefits is a new observation to our model.

The major factor or asset of our classification model is Customer relationship management (CRM). It is basically an approach to manage, deal and interact with our current and potential customers. In our model data analysis is done based on customer history, their interactions and transactions to improve business relationships with them. In the paper of prediction analysis based on Data Mining titled as "Customer Classification of Discrete Data Concerning Customer Assets Based on Data Mining", the authors Zuo Lie and Guo Junfeng describes data analysis as, prediction and analysis is utilized to achieve high success rate to increase rate of success of a particular Business. According to them useful information under big data can help many companies and enterprises to classify clients and customers accurately. They suggested a model in which is based on variables in four dimensions which includes transaction frequency, types of products or trading of services, transaction amount and customer's age. By using classification after clustering which divides twenty-five types of customer data into four categories.

Depending on these categories of customers, any organization or service related company can provide services to customers or recommend the related services to them. This would also help companies or organizations to create a good relationship with customers and also customers will also be able to know the company. Also, one of the management strategies of the authors was that the purpose of customer relationship management was to achieve the balance between maximizing both customer value and enterprise revenue. Because different customers will be having different relationship values and the company should always be focusing on the most valuable customer or the client. The similar approach has been followed by us which focuses on the problem of how to filter data from mass data and how to use data mining algorithms to complete the process of purification from customer data and find

important customers.

3. Methodology

With Data Mining we can analyze a customer’s area of interest in a particular domain. However, data from test trials usually include training data that require a quite different approach to analysis which is by using various algorithms mainly “Naïve Bayes Classifier Algorithm”. Using this algorithm our model follows a set of rules which divides the customers into two categories. These two categories are “Economic Customers” and “Exclusive Customers”. This categorization is done based on some factors which are taken as input from Customer. During the registration process the customers, personal data and field of interest is collected. This field may include Stocks, Mutual Funds, Shares, Policies etc. From the personal data collected, there are three main dimensions created. These dimensions are Customer Age, Customer Amount and Time period. Based on these three dimensions our customers are classified and then after our model follows a set of rules which groups customers into two categories, “Economic” and “Exclusive”.

After performing Classification using Naïve Bayes Classifier, products and services are recommended to our customers. These recommendations are done by applying Collaborative filtering technique on the products or the services. In Collaborative Filtering technique items or products are filtered based on the reviews and ratings provided by similar customers given to the same type of products in the past. So products or the services are divided into two categories, the r1 category and the 2 category. The r1 category includes all the products and services with ratings between 1 to 3 and r2 category includes all the products and services having ratings 4 and 5. The r1 category of services will be recommended to our Economic customers. The r2 category of products will be

recommended to our Exclusive customers. The end result will be companies related to Stock market, Policies, Fixed Deposits, Mutual Funds etc. which will be recommended to our customers according to their category. According to the three dimensions which are transaction amount, customer age and the time period for which the amount will be invested by the customers, the customers can be divided into 2 categories. Table 1 shows below the details of all the customers which will be then classified. The age for classification is less than 40, amount less than 4,00,000 and time period is less than 18 months.

After classification of the customers by Naive Bayes and by our classifier the customers are categorized which is shown in the table below. The model that our classifier uses consists of various rules through which the customers are classified and the clustered into two different groups. It is also possible that the customer may change the amount of value which will be invested by him or her, or the time period for which the amount will be invested depending on the changes done by the customer, he or she may move into different categorized groups and may get different recommendations the best suitable for them according to their needs. Our rules are specially designed for this recommendation model so that organizations can make greater value contributions for customers. The rules which our classified model applies after the classification of customers mainly depends on the three dimensions.

The rules which our classified model applies after the classification of customers mainly depends on the three dimensions.

This is our classification and recommendation model which has been developed using Python. As far as customer assets are concerned, it aims to help enterprises understand customers more deeply, grasp customer needs, and intend to influence customer behaviour by managing customers, so as to promote customers to make greater value contributions to enterprises as

Table 1
Customer details

Customer	Age	Amount	Time Period (Months)
Customer 1	25	4,00,000	11
Customer 2	40	2,25,000	48
Customer 3	36	3,00,000	22
Customer 4	21	1,00,000	15
Customer 5	58	6,80,000	12
Customer 6	62	52,00,000	10
Customer 7	33	4,00,000	8
Customer 8	28	2,50,000	6

Table 2
Customer categorized

Customer	Age = Low	Amount = Less	Time Period = Less	Category
Customer 1	✓	×	✓	Economic
Customer 2	×	✓	×	Exclusive
Customer 3	✓	✓	×	Exclusive
Customer 4	✓	✓	✓	Economic
Customer 5	✓	×	×	Exclusive
Customer 6	×	×	✓	Economic
Customer 7	×	×	×	Exclusive
Customer 8	✓	✓	✓	Economic

well as for themselves.

4. Conclusion and future scope

The proposed system was developed taking in mind the benefits of the customers. In this work we presented a framework where a customer can get the best recommendations possible as per their needs. This system saves a lot of the client's time and resources by processing everything quickly. We also aim to develop and maintain customers in a targeted manner, which not only avoids the waste of resources, but also reduces the huge risk brought by blind marketing of enterprises. As the market changes the customers will be updated of their profits and losses and as when needed they can change their investments. The main aim here is to provide best strategies to clients and also always persevere in providing them with increasingly perfect services. In the result we can clearly observe that the system not only adapts to customers' needs but also user friendly systems which keep personal data secure in organization.

Acknowledgment

We wish to express our sincere gratitude to Dr. Udhav. V.

Bhosle, Principal and Dr. Satish. Y. Ket, H.O.D of Computer Department of Rajiv Gandhi Institute of Technology for providing us an opportunity to do our project work on "Customer Stratification using Data Mining". This project bears an imprint of many peoples. We sincerely thank our project guide Prof. Dilip Dalgade for his guidance and encouragement in carrying out this project work. Finally, we would like to thank our colleagues and friends who helped us in completing the project work successfully.

References

- [1] Zuo Lei, and Guo Junfeng, "Customer Classification of Discrete Data Concerning Customer Assets Based on Data Mining."
- [2] Zheng Ruiying. "Customers' Classification based on Attributes Reduction of Rough Set."
- [3] Kaveri Roy, Aditi Choudhary and J. Jayapradha, "Product Recommendations using Data Mining and Machine Learning Algorithms."
- [4] P. S. Badase; G. P. Deshbhratar, "Classification and analysis of clustering algorithms for large datasets."
- [5] Deuk Hee Park, Hyea Kyeong Kim, "A Review and Classification of Recommender Systems Research."