



CONSUMER PURCHASE PREDICTION

Using Random Forest Regression

¹Preethi B, ²Anu C.S, ³Dr. Nirmala C.R

^{1,2}Assistant Professor, CS&E Dept. B.I.E.T College, Davangere

³Head of Dept., CS&E Dept. B.I.E.T College, Davangere

ABSTRACT—Nowadays, customers have become more interested in the quality of service (QoS) that organizations can provide them. Services provided by different vendors are not highly distinguished which increases competition between organizations to maintain and increase their QoS. Customer Relationship Management systems are used to enable organizations to acquire new customers, establish a continuous relationship with them and increase customer retention for more profitability. CRM systems use machine-learning models to analyze customers' personal and behavioral data to give organization a competitive advantage by increasing customer retention rate. Those models can predict customers who are expected to churn and reasons of churn. Predictions are used to design targeted marketing plans and service offers. This paper tries to compare and analyze the performance of different machine-learning techniques that are used for churn prediction problem. Facilitated by statistical and machine learning models the study seeks to predict the purchase decisions based on adaptive or dynamic pricing of a product. Different data sources which capture visit attributes, visitor attributes, purchase history, web data, and context understanding, lays a strong foundation to this framework.

Keywords— Recommendation, Machine Learning, Prediction system

1. INTRODUCTION

Purchase decision process describes the sequence of actions performed by a customer when deciding to purchase a particular product or service. It can also be described as a process of problem solving, where a consumer satisfies

his needs after thoughtful consideration. The outcome of a purchase decision process is a decision whether a customer will buy a given product or service or not. There are several factors affecting buying behaviour, such as cultural, social and personal decision elements. Cultural factors include cultural context and belonging to a certain social class or subculture. Social factors are defined with position and role of the individual, his family and reference groups, which have a direct or indirect impact on buying behavior. Personal factors are determined with individuals lifestyle, occupation, property status, personality and self-esteem.

2. LITERATURE SURVEY

Research [1] indicates that the cost of retaining a customer is less than attracting new ones. This is due to marketing costs required to appeal to new customers. For this reason, together with the increase of competition it has become pivotal that the current customers base is retained. Normally, customers churn gradually and not abruptly. This means that by analyzing customers historic buying patterns one can adopt a proactive approach in predicting churn. Since all transactions are inserted through POS and stored in databases, understanding customers' needs and patterns is possible as data is accessible.

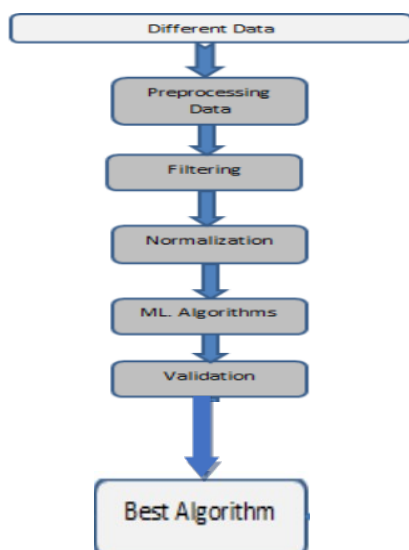
According to [2], executives are dedicating marketing budgets to focus on customer retention campaigns. Various models designed to predict churn focus on statistical and renowned machine learning algorithms including Random Forest and Logistic Regression. This paper focuses on two aspects when predicting churn within the grocery retail industry. The first is based on the features which will be passed on to the model. Instead of

using customers buying trends to cluster the individuals, these values will be created as features and are passed to the model. Therefore, for each customer various features are created to allow the model to learn and identify patterns per individual. For this reason, two datasets are created to test and evaluate how data should be represented to predict churn. The second aspect is the implementation of the algorithms. The novelty of this study is the use of deep learning to predict churn within the grocery industry. To our knowledge, this is the first study which implements deep learning within this industry. The strength of using deep learning is that it can reveal hidden patterns within the available dataset. An important aspect within the business is to have a good understanding of customers' needs, whereby holistic views of their patterns may be analyzed. When customers are satisfied with the service or products, customer loyalty increases [3].

3. EXPERIMENTAL DATASET & METHODOLOGY

In the present study first data used is Black Friday dataset with 11 columns. All information procured from machine learning storehouse of Github Website.

The methodology for the classification of these datasets is displayed in Figure 1. The analysis has been performed in jupyter notebook platform running on (Intel i5 processor) with 3 GB RAM installed. The dataset is taken as input for feature extractor and classification algorithm. The dataset is passed through a



sequence of pre-processing blocks.

Figure 1: Methodology diagram

Theoretical analysis:

While selecting the algorithm that gives an accurate prediction we gone through lot of algorithms which gives the results abruptly accurate and from them we selected only one algorithm for the prediction problem that is Random Forest Regression, it assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature that is how the prediction work great with the Random Forest Regression Algorithm.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FT+FN}$$

Dataset description:

In this paper, the dataset we used is derived from

<https://github.com/shwetachandel/Black-Friday-Dataset> It contains more than 550070 users with 12 attributes. After that, the missing values are filled in by means of mean interpolation, and the duplicate or meaningless attributes are deleted, finally we have retained to 11 attributes. Those attributes were shown below in the screenshot of the data set we used.

Preprocessing:

After the gathering of data next phase is to perform the preprocessing on the collected data. It is the technique that changes the raw data to the understandable format.

Filtering:

After preprocessing of data next phase is to filter the data.

Normalization:

It is a stage in which all the values are changed to values

reason for standardization is to attract alternate scale

Machine learning algorithms:

After normalization use different algorithms used such as Random Forest, Simple linear regression, Multi linear regression and Decision tree

4. RESULT AND DISCUSSION:

In this paper, the Random Forest Regression algorithm is used to predict its performance, and compared with another three machine learning methods namely the decision tree, simple linear regression, multi linear regression. The obtained results are displayed in Table below. The results show that, the performance of decision tree and Random forest regression have comparable performance than that of linear regression and multi linear regression, but random forest

regression still performs the best, with an accuracy of 70%, higher than the decision tree with an accuracy of 43%.

5.CONCLUSION:

The study of consumer behaviour basically is to mould consumer behaviour and decisions by marketing and to avoid failure of their product, promote new products and for sales promotion. The science at times is misused and to protect consumers there are a number of enactments both in India and other countries. Consumer behavior analysis has emerged as an important tool to understand customers. By looking into consumer psychology and the forces behind customer buying behavior, companies can craft new products, marketing campaigns and increase profitability.

REFERENCES

- [1] C. Jie, Y. Xiaobing, and Z. Zhifei, "Integrating OWA and data mining for analyzing customers churn in ecommerce," The Editorial Office of JSSC and Springer- Verlag Berlin Heidelberg, vol. 28, pp. 381-391 2015.

- [2] The science behind customer churn. [Online]. Available: <http://finance.inbusinesslife.info/thesciencebehind-customer-churn/>

- [3] S. Neslin, S. Gupta, W. Kamakura, L. Junxiang, and C. H. Mason, "Defection detection: Measuring and understanding the predictive accuracy of customer churn models," Journal of Marketing Research, vol. 43, pp. 204211, 2006.