

Project Synopsis

Customer Classification Prediction Model For Online Store

Prepared by

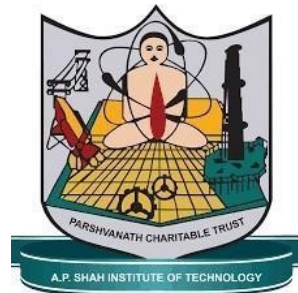
Arhanti Gawde

Arpita Hirlekar

Snehal Surve

Under the guidance of

Prof. Pravin Adivarekar



Bachelor of Engineering (BE)

Department of Computer Engineering

A. P. Shah Institute of Technology, Thane

University of Mumbai

2019-2020

Customer Classification Prediction Model For Online Store

Abstract

This paper deals with classification of the customers as a part of the marketing strategy in order to expand the business profit by understanding the relationship between the customers and the products that they purchase. Classification of the customer proves to be helpful since it targets the specific customer for the specific product. This will help the company heighten the level of information to the appropriate customers. On the basis of what customers are buying, adding to cart or viewing, we will cluster the customers depending upon the similar parameters. Also due to this the business managers and analysts will be further able to provide recommendations to the customers belonging to a particular cluster only. While launching new products, services or releasing new versions of existing products / services the business can reach potential customers only. However, a real time big data will be used to perform the analysis on. This model will serve the purpose of identifying the customers who will most likely respond to the recommendations by the company based on their past purchasing history. This system would be useful for the companies in putting a marketing tactic for promotion of their new products. In this manner the business will stay focused and targeted.

1. Introduction

Identifying the potential customers of certain products and promoting the products accordingly is an important concept of almost all businesses nowadays. There are two methods for advertisements and promotions, namely mass marketing and direct marketing. Mass marketing is a procedure wherein the promotion of the products is carried out by targeting the general public through media such as newspapers, radios and televisions. This results in high wastage and low response rate from the customers who will actually buy the product. Hence this is not an effective method to be applied in today's competitive market. Therefore rather than going forward with such an unreliable method, the marketers are now shifting their focus from traditional Mass Marketing to Direct Marketing. The tool used for this project is the Elasticsearch Logstash Kibana (ELK).

Elasticsearch:

Raw data flows into Elasticsearch from a variety of sources, including logs, system metrics, and web applications. Data ingestion is the process by which this raw data is parsed, normalized, and enriched before it is indexed in Elasticsearch. Once indexed in Elasticsearch, users can run complex queries against their data and use aggregations to retrieve complex summaries of their data.

Logstash:

Logstash is a free and open server-side data processing pipeline that ingests data from a multitude of sources, transforms it, and then sends it to your favorite "stash." Logstash

dynamically ingests, transforms, and ships your data regardless of format or complexity. Logstash, one of the core products of the Elastic Stack, is used to aggregate and process data and send it to Elasticsearch.

Kibana:

Kibana is a free and open user interface that lets you visualize your Elasticsearch data and navigate the Elastic Stack. Kibana is an open source visualization tool which provides a beautiful web interface to visualize the Elasticsearch data. Kibana allows us to create real-time dashboards in browser based interfaces. Kibana has different visualization effects like bar charts, graphs, pie charts, maps, tables etc. It allows to save, edit, delete and share the dashboards.

2. Literature Review

- A. Title: A Comparison of Different Classification Techniques for Bank Direct Marketing

Author: K Wisaeng

About The Paper: In this paper, the comparison of different classification techniques in open source data mining software which consists of a decision tree methods and machine learning for a set of bank direct marketing dataset are presented. All decision tree methods tested are J48-graft and LAD tree while machine learning tested are radial basis function network and support vector machine. The experiment results show are about classification sensitivity, specificity, accuracy, mean absolute error and root mean squared error.

- B. Title: Bank Direct Marketing Based on Neural Network Author: Hancy.A.Elsalamony, Alaa.M.Elsayad

About The Paper: This paper introduces applications of recent and important models of data mining; Multilayer perceptron neural network (MLPNN) and Ross Quinlan new decision tree model (C5.0). The objective is to examine the performance of MLPNN and C5.0 models on a real-world data of bank deposit subscription. The experimental results demonstrate, with higher accuracies, the success of these models in predicting the best campaign contact with the clients for subscribing deposit. The performances are measured by three statistical measures; classification accuracy, sensitivity, and specificity.

- C. Title: Data Mining Framework for Direct Marketing Author: Lilian Sing'oei, Jiayang Wang

About The Paper: The objective of this paper is to provide a comprehensive framework to guide research efforts focusing on direct marketing strategy and aid practitioners in their quest to achieve direct marketing success using data mining methods like the Decision Tree Algorithm. The framework builds on the literature from direct marketing concepts and data mining methods that provides a systematic approach to users who have little knowledge in data mining in order to carry out effective marketing campaigns. A case study on bank marketing campaigns was used for evaluating the feasibility of the framework.

- D. Title: A Data Mining Based Response Model for Targeting Selection in Directing Marketing

Author: Eniafe Festus Ayetrian, Adesesan Barnabas Adeyemo

About The Paper: In this paper, using historical purchase data, a predictive response model with data mining techniques was developed to predict the

probability that a customer in Ebedi Microfinance bank will respond to a promotion or an offer. To achieve this purpose, a predictive response model using customers' historical purchase data was built with data mining techniques. The data were stored in a data warehouse to serve as management decision support system. The response model was built from customers' historic purchases and demographic dataset. Bayesian algorithm precisely Naive Bayes algorithm was employed in constructing the classifier system. Both filter and wrapper feature selection techniques were employed in determining inputs to the model.

3. Problem Statement

To generate a model that will accept a big data and classify customers on some constraints and make clusters of customers who portray some or the other common characteristics. Clusters will be having different customers on the basis of their past purchases, their time of purchase, the brand they purchase, etc. Also this model will provide a good and detailed report about the performed analysis. This process of classification is an integral part of the direct marketing scheme which is trending globally nowadays.

4. Objectives

The objective of this project is to build a model that will accept a big data and classify customers based on the basis of their purchase history, purchase time, brand, etc. and give the business a good and detailed report of the analysis. After customers are classified, the company or the business will not have to face the excess of expenditure in providing irrelevant recommendations for new products that will be launched on the website. The company will be benefited because it can send recommendation only to customers who are likely to revert to the recommended product.

5. Scope

Direct marketing deals with an important issue of identifying customers who are more likely to respond to the new offers. Data mining is an extensively used method to carry out the process of target selection i.e. identifying the potential customers. This project develops predictive analysis using an analysis software known as ELK (Elasticsearch, Logstash and Kibana) in order to predict the bifurcate the customers according to their shopping time, history, choice, etc. The steps such as importing the dataset into the database, analysis of it, choosing parameters and accordingly creating certain visualizations and ultimately creating a detailed report of the analysis will be carried out using the software. The objective of this project is to provide a comprehensive framework for the efforts taking place in direct marketing strategy. The results obtained will help the store/organization/company to plan effective marketing of their products and services by obtaining a guiding report on the status of their customers which will go a long way in assisting management in saving significant amount of money that could have been spent on wasteful promotional campaigns.

6. Benefits for environment

Website creation, being a part of this project, is an important area which will help to lessen the environmental issues. Opting for a darker background for the website can surely be valued. Darker backgrounds are known to save as much as 4% of the total electrical energy as compared to lighter or white background. It also supports visual

hierarchy thus reducing eye strain if browsed for hours. It offers depth in the reflection of the presented content. Using cool dark backgrounds for a website instantly transforms it into a night-mode one. Nowadays most of the devices feature an option which is called night-mode and turns the lighter colours on the device to black. Because of this particular technique the device produces less light due to which energy is saved which also prevents the end-user's eyes from getting sore quickly. Since this project is typically based on software and analysis, the probability of producing e-waste or the hardware waste has been brought down.

7. Benefits for society (If applicable)

NA

8. Applications

The classification of customers will help to convey specific, helpful information about clothes to customers by giving them recommendations so that they get to know about new and trending products and our sale increases by attracting them towards such products. An online presence helps in terms of building and fostering a rapport among your customers. Providing special offers to particular group of customers on the website lets them know we appreciate them for visiting and shopping from our website. Customers will get recommendations, different recommendations to different people, which will give customers products they might like resulting in increasing sales and thereby giving profit. This technique of direct marketing will prove to be helpful for not only online stores but also the ground shops. It can be applied to a variety of stores, industries, banks, etc.

9. Technology stack

The software used for this project is the ELK. ELK is an acronym for Elasticsearch, Logstash and Kibana. These are the three popular open source softwares which help you load your dataset into it, store it, analyze it, perform queries on it, visualize it and ultimately also provide you with an interface which will lend a helping hand in creating the reports. The Elasticsearch basically acts like a database which stores the data grabbed from the user, the data could be of any format. The Logstash is like a pipeline which helps transfer the user data into Elasticsearch. It actually ships the data from the source mentioned by the user to the database which the Elasticsearch portrays. The Kibana is nothing but the major part of the software since it is the visualization interface which helps analyze data and creates reports out of it.

References

- T. K. Das, "A Customer Classification Prediction Model Based on Machine Learning Techniques"
- K. Wisaeng, "A Comparison of Different Classification Techniques for Bank Direct Marketing"
- Hany. A. Elsalamony, Alaa. M. Elsayad, "Bank Direct Marketing Based on Neural Network"
- Lilian Sing'oei1, Jiayang Wang, "Data Mining Framework for Direct Marketing: A Case Study of Bank Marketing"

- Eniafe Festus Ayetiran, "A Data Mining-Based Response Model for Target Selection in Direct Marketing"
- Karim Masud, Rashedur M. Rahman, "Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing"
- Paulo Cortez, "Using Data Mining For Bank Direct Marketing: An Application Of The Crisp-DM Methodology"
- Customer Classification Based on The Historical Purchase Data - Leading India Paper

Arhanti Gawde

Arpita Hirlekar

Snehal Surve