CS235 Fall'22 Project Proposal: Group 10 - Customer Acquisition Cost Prediction

ADITYA SRINIVAS KARAMANGALA AMAR, NetID: akara045
ANJANA VENKATESH, NetID: avenk023
KAUSHAL BANDARU, NetID: kband008
MUKESH M KARANTH, NetID: mkara022
VARNITH GONNAGADLA VENKATESHA, NetID: vgonn001

Additional Key Words and Phrases: data mining, regression, supervised learning

ACM Reference Format:
Aditya Srinivas Karamangala Amar, Anjana Venkatesh, Kaushal Bandaru, Mukesh M Karanth, and Varnith Gonnagadla Venkatesha. 2022. CS235 Fall'22 Project Proposal: Group 10 - Customer Acquisition Cost Prediction. 1, 1 (October 2022), 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The cost of acquiring a customer in the current market poses a challenge to many companies worldwide. Companies struggle to predict and allocate an appropriate budget in the right media platform to reach out to their target audience and build their customer base. Unfortunately, we see many organizations shut down for the same shortcoming. Hence, we are planning on analyzing the data from one such company that faced challenges while procuring customers and predicting its potential customers' acquisition costs. The name of the company is Convenient Food Mart (CFM) - a retail food chain in the United States. We plan to achieve this goal with the help of supervised learning by using some regression techniques, along with some data analysis.

1.1 Project Type

This is a software (default) project.

2 PROBLEM DEFINITION

Using Convenient Food Mart's cleaned and labeled data set from Kaggle, we intend to use supervised learning techniques to predict the customer acquisition cost of future customers of Convenient Food Mart.

Authors' addresses: Aditya Srinivas Karamangala AmarNetID: akara045; Anjana VenkateshNetID: avenk023; Kaushal BandaruNetID: kband008; Mukesh M KaranthNetID: mkara022; Varnith Gonnagadla VenkateshaNetID: vgonn001.

3 DATASET DESCRIPTION

We intend to use the labeled data set containing store-specific details, cost of acquisition, and customer data of the Convenient Food Mart retail chain found on Kaggle. It is a cleaned data set that initially had 200,000 records with 75 features but was reduced to around 60,000 records with 40 essential features after pre-processing.

Link for Convenient Food Mart data set - https://www.kaggle.com/datasets/ramjasmaurya/medias-cost-prediction-in-foodmart

4 PROPOSED APPROACH

We will be using the regression suite of supervised learning techniques to predict the acquisition cost of the customer. The following methods will be used :-

- Linear Regression (SGD implementation)
- Perceptron
- Decision Tree
- Bagging ensemble method
- Boosting ensemble method.

5 EVALUATION PLAN

We are using a pre-processed, labeled, and cleaned data set from Kaggle and implementing regression methods. To evaluate our model's performance, we will use the **R squared (R2)** metric. It gives us a standard model to compare with, which no other metrics provide. R2 measures the degree of variability in the dependent variable, so we can use R2 measure to define the performance of our regressor on a scale of 0 to 1. A score closer to 1 is desired. For evaluating the loss function of most regression techniques Mean Squared Error (MSE) metric is commonly used. However, in MSE, the order of loss is more than that of data. Since we cannot directly correlate error with the data set available, we will be using the root of the MSE, which is the **Root Mean Squared Error (RMSE)** metric for evaluation. RMSE will give us the difference between the predicted and actual acquisition costs, showing how the predicted data points are spread out from the mean. We will be using **Mean Absolute Error (MAE)** only if required, where instead of using the square of error as in MSE, we will take the absolute value of error. Like RMSE, the higher the value of MAE, the greater the error, so we will minimize it. However, MAE will only be used if we intend to treat all errors the same, i.e., not penalizing more when there is a more significant prediction error.

With the accurate evaluation measures calculated, we will predict the customer acquisition cost for Convenient Food Mart's future customers. We hope that through this project, we will get detailed insights and draw interesting relations about companies and their customer acquisition costs.

6 PROJECT TEAM & PROJECTED LABOR DIVISION

(1) Aditya Srinivas Karamangala Amar (NetID: akara045): Linear Regression (SGD implementation)
(2) Anjana Venkatesh (NetID: avenk023): Perceptron
(3) Kaushal Bandaru (NetID: kband008): Decision Tree
(4) Mukesh M Karanth (NetID: mkara022): Boosting ensemble method
(5) Varnith Gonnagadla Venkatesha (NetID: vgonn001): Bagging ensemble method