Udacity Machine Learning Engineer Nanodegree

Capstone Proposal

Mahananda Bose

Table of Contents

Domain Background:

Starbucks needs no introduction. It is one of the biggest companies in the world. A world renowned coffee-shop chain which has stores worldwide. Everyday, those stores deal with thousands of customers. It offers a free app for online order, sending special offers for better experience of the customers. This project is to have a better understanding of user demands and improve business.

Problem Statement:

Our goal here is to create a model which can identify which offer is suitable for which customers. Based on various attributes of the customers, identifying the special offers which will mostly attract them.

Dataset and Inputs:

Starbucks gives offers to its app users. The whole data is divided into three json files

portfolio.json - It contain data regarding all the offers rolled out by Starbucks
profile.json - It contains data about all the users or customers who use Starbucks app.
transcript.json - It contains data about all transactions, offer received, offer viewed, offer completed.

- ## Portfolio.json:

id - Offer Id
offer_type - Type of Offer
difficulty - minimum amount to spent to avail the offer
reward - reward given to complete the offer
duration - time limit of the offer, in days
channels - media of offer

- ## Profile.json:

age - Age of the customer
became_member_on - date when the customer created account in the app
gender - Gender of the customer

id - customer id
income - customer's income

- **Transcript.json:**

event - Transaction type
person - Id of the customer
time - time in hours in start od=f test
value - offerid or amount depend on type of transaction

Solution Statement:

I want to create a model which will identify which offer suits best to which customer. For this, age and gender of the customer, Response to a particular offer, most responded offers will be considered.

Benchmark Model:

A simple classifier model is ideal for a benchmark model. I am considering here KNeighboursClassifier as a benchmark model as it is very simple and at the same time accurate for classification purposes.

Evaluation Metrics:

In this project, F1 score will be the evaluation metrics on basis of which model's performance will be evaluated. F1 score is considered as it incorporates both precision and recall value. Higher F1 is considered as better performance.

Project Design:

Following steps are followed to design this project:

1. Establishing a Jupyter Notebook for the project.
2. Evaluating and understanding all the data available.

3. Cleaning all the data as required by the model.
4. Merging all data into a single dataset.
5. Analyzing the new merged dataset.
6. Cleaning and Processing the merged data for training the models.
7. Splitting the data into Train and Test dataset.
8. Training the models with Train dataset.
9. Testing the models prediction with Test dataset.
10. Evaluating F1 score of all the models to understand their performance.