Pilot Study Proposal On Customer Purchase Prediction

1. Introduction

Customer behaviour prediction and identifying their buying patterns play a vital role in understanding the potential customers for the Supermarkets and Ecommerce. It helps the decision makers and managers to understand the customer interests and their buying psychology. It increases customer satisfaction and sales, by facilitating an increase of customer experience through personalization, recommendations and special offers. By utilizing the past purchase history and additional customer data, data analytics and data mining solutions can be used to target the potential buyers. This data and information can be fed to the machine learning algorithm to carry out predictions ranging from customer classification, purchase prediction and recommender systems.

2. Research Problem

Categorizing whether a customer visiting the supermarket will end in a purchase or not, is a real-event use case in the context of machine learning. In case the customer makes a purchase, classifying whether the customer is prone to buy a few expensive products and predicting the amount of money they usually spends in a month is a challenging problem where we have to leverage the power of both classification and regression models to arrive at a solution.This categorization is followed by the display of gift cards to non-purchasing customers, to convince them of a purchase and to attract the purchasing customers to spend more on expensive items.
This results in the following research questions:
i) How can a customer visiting the supermarket be categorized as a buying or no buying customer?
ii) Will the potential buyers buy a few expensive products or not?

3. Research Objectives

The main aims of this project are as follows:

i) to investigate the feasibility of using machine learning to predict whether a new customer belongs to the first class or not.

ii) to identify the appropriate data requirements needed for modelling the algorithm..

iii) to conduct an analysis on the provided datasets to select the appropriate features that have a relation with categorizing a potential buyer..

iv) to experiment with different classification models and compare their performance of the solution before deploying it.

4. Feasibility Study

A feasibility study, as the name suggests, is designed to reveal whether a project plan is feasible or not. It is an assessment of the practicality of a proposed project plan.
This study analyzes machine learning models to predict a pur-chase, which is a relevant use case as applied by a famous supermarket chain "Tosco & Spency". I would like to check whether a project plan is feasible or not based upon these requirements:

i) Training Data:

We can assume from the problem statement that our modelling task is supervised and formulating a supervised prediction algorithm requires you to have access to training data. For creating a prediction algorithm to categorize whether a new customer belongs to the first class or not, we need access to historical data of past customers. Since, we are provided with the historical data of past customers along with additional customer details, we can say that the proposed solution is feasible.

ii) Predictive Features :

Having training data with no quality information about the subject matter is like a waste. Most machine learning projects fail due to lack of quality information and features in the data. If the data does not have the useful features that drive the learning algorithm to predict the target, then we have to redefine our data requirements and data collection strategies. A good dataset is a dataset with quality information about the predictive features. Predictive features are those features which have a correlation with our target outcome. Since, we are provided with the historical sales data of past customers who visited the Supermarket often along with information about several features of the customers like. age, average income and nationality, we are sure that we could make an accurate prediction algorithm with these features as customers having high income are more likely to buy expensive products.

iii) Learning Algorithm:

Categorizing whether a customer visiting the supermarket will buy an expensive product or not is a classification task. There are various classification based learning algorithms we can choose to model this task. Algorithms like Support Vector Machines, Decision Trees, Logistic Regression and Random Forest are used very often for doing the classifying task. Random Forest Algorithms have proven to give better accuracy on many classification problems and datasets.

iv) Model Evaluation And Production:

Evaluating a prediction model and getting the best model requires you to experiment with many different algorithms and compare the results. We can use different evaluation techniques like

cross validation and make use of different evaluation metrics to select the best performing model. Evaluation metrics like ‘accuracy’, ‘recall’ and ‘precision’ can be used to evaluate the models and the model with best ‘accuracy’,’recall’ and ‘precision’ score is finally deployed for production.

5. Conclusion/Recommendation

After analyzing all the possible feasibility criterias of the proposed plan, it looks promising that the proposed solution is feasible and I would strongly suggest continuing working in the development phase of the project.