ASSIGNMENT
PILOT – STUDY PROPOSAL

Joel John

University of Essex

Aim:

To investigate the feasibility of using machine learning (ML) techniques to predict the chance of the business is profitable. If that is successful, it will further extend the project to predict the expected achievable annual profit. The prediction and feasibility study will be done based on the data provided by the client. The accuracy of the used method will be determined based upon the provided data for the finalization of the algorithm.

Data Collection:

For the successful implementation of ML algorithms, the undeniable factors that directly depend on data to be predicted are the accuracy and feature the client will select for the data to be collected upon. So, for just reference, the following are the suggested informative features that could be added to the data collection process:

1. Location Features
    - Distance from the nearest tourist spots
    - Number of hotels nearest to the location providing similar amenities
    - Distance from the nearest competitor
    - Distance from the nearest town center
    - Number of local Festivities that attract customers in a year
2. Monetary Features
    - Average expenditure per customer
    - charges applied by the competitor
    - Average employee Salary in the region
    - The annual profit of competitor
3. Accessibility
    - Number of shopping facilities available
    - Methods of travel facilities available
4. Number of stores and shopping complexes and other attractions
5. The maximum number of customers the establishment can handle at a time.

These are a few features that the client can consider while collecting the data.

Proposal:

The business's target is to maximize profitability, whether it be a hotel, textile, or fast-food chain. So, by using Machine Learnings techniques, we will be investigating the feasibility of opening a branch in a new location and its success.

To achieve this project's aim, we will be using Supervised ML techniques. To predict whether the establishment will be profitable, we will use different classifiers. This method is selected because the result may depend on a lot of factors and the expected to predict only one class, that is, whether the establishment will be profitable or not. We will be investigating four different well-known classification methods in practice, they are:

1. Random Forest Classifier

2. Decision Tree Classifier

3. K-Nearest Neighbor

4. Support Vector Machines

Depending on the accuracy calculated by training and predicting the model on previously collected data, we will select one of these trained models to predict future cases. This model will be the suggested model by us for future references.

Method of Evaluation:

The system's performance will be evaluated based upon the accuracy of the classifier to determine which one we will be using. For this purpose, the data collected after the successful investigation will be split into two portions. The first set will be used to train each model. All will be trained on the same data set, and the second set will be used to determine the accuracy. For this, we will be using tools that are available in python.

This method can help determine how the trained model may act with new set inputs outside the trained model.

Finalization:

The model will be demonstrated to the client to get an acceptable accuracy. If it meets the standards that the client is looking for, this will be used to predict other scenarios he presents at the time (New locations and values).