Assignment: Design and Application of a Machine Learning System for a Practical Problem

REPORT

Reg# 2101142

Word count: 738

Table of Contents

ABSTRACT:

The purpose of this project is to use the machine learning to predict whether a new hotel opened in each location will be profitable or not. For this we collected relevant data and apply different classification and regression algorithm to build a model and then we compared these models to determine which one is best and use this model to predict about the new hotel. From which it is predicted that new hotel opened will be profitable.

INTRODUCTION:

Humans have become increasingly reliant on data and information in society over the last two decades, and as a result, technologies for data storage, analysis, and processing on a large scale have emerged. Data Mining and Machine Learning have used them to not only gain information and find new things, but also to uncover hidden patterns and concepts that have led to the prediction of future occurrences. Machine learning models using classification algorithms are widely utilized in a variety of domains, including data analytics, image classification, text classification, computer vision, exploratory analysis, and gaming AI, to name a few. In this case, the model must be extremely accurate, versatile, and efficient to successfully complete the work at hand.

The key step in hotel opening is to predict the future demand of the hotel in that region. We have been investigated the possible features that could be helpful in this type of analysis to collect data. We have to analyze the algorithm that should be used like Random forest, k-means, KNN, linear regression, decision tree, and SVMs to evaluate the performance of system.

DATA AND METHOD:

These are the important steps that should be followed in this project.

Data collection:

With the given information we proceed to collect relevant data. For all the hotels given in the given Geographical region we collect these features.

- The number of customers in the given region
- The frequency of visit to hotel
- The average length of stay
- The average budget of the customer
- The types of rooms and services they are providing.
- The market position of hotel with respect to competing hotels
- The quality of food

For every hotel we will find the values of these features for each hotel and class of this data is yes or no value to predict whether hotel is successful or not.

Data preprocessing:

For the preprocessing of data, we will do EDA on data, we observe dataset then find any missing values, then Categorize values, Find the shape of dataset by plotting to Identify relationships in dataset.

Train-test split:

We will split the data in training data and test data to prepare model by dividing the (70/75/80) % of whole data in training data and (20/25/30) % in test data.

Algorithm implementation:

By analyzing the data, it is predicted that it is classification task so we will implement some of the classification algorithms on it. These algorithms include:

- SVM
- Logistic regression
- Random forest
- Decision tree
- K-Nearest Neighbours
- LightGBM Classifier

Comparison of results:

The results from each of the method is compared by comparing the accuracy, precision, Recall, F1 score and area under the curve to predict the best model for this task.

The best model is then used to predict whether a new hotel opened in this location will be profitable or not by providing all the features to the model.

CONCLUSION:

The objective of this project is to prepare a machine learning model to predict whether the new hotel opened will be profitable. For this purpose, we collected data and do data preprocessing and then apply classification algorithms such as Logistic regression, Random Forest, Decision tree, K-Nearest Neighbors, LightGBM Classifier and the highest accuracy is of random forest which is above 88% and then we find it to be profitable.

REFERENCES:

[1] Yang, Yang. (2015). Hotel location evaluation: A combination of machine learning tools and web GIS. International Journal of Hospitality Management. 47. 10.1016/j.ijhm.2015.02.008.

[2] M. Zhang, Hotel Occupancy Prediction Based on EEMD-ARIMA, Shaanxi Normal University, Xi'an, China, 2016.

[3] J. Yu, Model Established and Analysis of Passenger Flow Forecast in Hotel Management Decision Support System, Dalian University of Technology, Dalian, China, 2008.