

Pilot Study Proposal

Word count: 657 words

Topic:

Pilot study proposal to check the feasibility of Machine Learning algorithms applied for predicting whether a particular new hotel will incur profit or loss. This proposal will help to get a clear idea if use of algorithms will aid in prediction regarding the new hotel and if yes, which algorithms or models should be used to get optimum results along with the justifications.

Significance:

The owner wants to decide if the new hotel under the chain that he/she wants to start will be a good or bad decision. The manager can provide the historical data of rewarding and unsuccessful hotels that have opened under the chain's brand along with additional information about the locations and neighbouring status of the hotels. This will give assistance to determine the deciding factors for the hotel's profit or loss. The proposal will analyse the best possibility to build a Machine Learning (ML) model to solve the problem and it can also be decided if the use of ML is fruitful or not.

Predictive task to be performed:

The predictive task that I would like to recommend for this problem is **Classification**. The manager has provided with the previous data which will help to train the classification model. This data will help the model to perform the learning process and then for new data it will predict if it belongs to **category of "profit" or "loss"**. As the predicting categories are discrete, we do not use regression as it is used for continuous outcomes. Rule mining is not used as, at the end, it helps to find the most important relationships from the data base however, it won't be rewarding in predicting task in this case.

Examples of important features to be provided:

Some of the basic vital features that I would expect to be provided are as follows:

- 1) **Distance**: The distance of hotels from the location of the new hotel to be set up.
- 2) **Popularity**: List of all popular and unpopular hotels from the chain
- 3) **Decision Factor**: The main reason why a particular hotel is successful or unsuccessful. For e.g., "less crowd" or "customer service" can be deciding factors for unsuccessful hotels.
- 4) **Age group**: The general age group range that visits the hotels from the chain.
- 5) **Location**: The list of crucial spots in proximity of hotel. Spots may include tourist spots, localities, schools, etc.

6] **Marketing strategy**: The methods that are implemented through which crowd can be attracted. For e.g. Use of pamphlets, discounts and offers, etc.

7] **Years**: The number of years for which the hotel branch is opened.

8] **Closed hotels**: The count of hotels that were closed and its reason. for e.g. Understaff, no locality in vicinity, etc

9] **Profits or loss**: The amount of profits the hotel incurs in a particular span of time.(weekly, monthly, per Anum)

10] **Transportation**: Frequency and availability of transportation services at the location of hotel.

Learning procedure:

The predictive task of profit or loss has **distance or location as deciding factor** rather than monetary profit or loss. Thus, the learning procedure I would like to suggest is **k Nearest Neighbour (kNN) algorithm**. Another reason for choosing kNN algorithm is that we have **sufficient** socio- economic and historical **knowledge** about the hotel chain. This algorithm does not form any assumptions about the available or underlying data. The main task to be done carefully in this algorithm is proper selection of value of k in order to get accurate results.

Performance Evaluation:

The primary evaluation factor will be **accuracy and confusion matrix** obtained while training data. The data would be split in 80:20 or 70:30 ratio where the large split will be used for training and the small split for testing the performance of the trained model before handling new data. The small data split acts as validation data set. Then, I will perform **feature engineering** and **altering the parameters** of the model to increase the accuracy.

References:

- 1) <https://www.javatpoint.com/regression-vs-classification-in-machine-learning>
- 2) <https://blog.bismart.com/en/classification-vs.-clustering-a-practical-explanation>
- 3) <https://www.ibm.com/docs/en/ias?topic=knn-usage>
- 4) https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_knn_algorithm_finding_nearest_neighbors.htm
- 5) <https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn>
- 6) <https://www.gcu.ac.uk/library/pilot/researchskills/writingaresearchproposal/writingtheresearchproposal/>