Project Proposal

CSC 219
Machine Learning
Prof. Dr. Meiliu Lu

PROBLEM STATEMENT:

Prediction and Comparative Study of the Annual Restaurant Revenue based on the objective measurements.

Background: While opening a new restaurant, there is a huge amount of investment involved. Crucial decisions like where to open the restaurant, what kind of restaurant should we open, etc. need well-founded answers. These answers play a vital role in determining the revenue of the restaurant. In order to get statistically proven results, we can use Data-mining techniques by which the revenue can be predicted scientifically!

Main Objective: Our main objective is to predict 100,000 restaurants revenue from the big data collected from kaggle. Our ultimate goal is to predict and compare precisely the revenue of the restaurant using the different machine learning prediction models..

DATASET EXPLORATION:

There are 2 sets of data: Training Dataset and Testing Dataset.
The dataset is in the excel .csv format
Training Dataset Size: 137 restaurants with given Revenues
Test Dataset size: 100,000 restaurant.
Each Restaurant has 43 attribute values to it.

METHODOLOGY:

Key Steps:

1. Feature Engineering
    a. Data Cleaning:
        i. Removing problematic objects/attributes to improve accuracy.
        ii. Removing Outliers and variables with low correlation.
    b. Transformation of attributes' values
        i. Normal Distribution of Data

2. Modelling and Prediction:
    a. Use of R as programming language for development.
    b. Random Forest, a R package, is used to build model after feature engineering.
    c. Support Vector Machine
    d. K- Nearest Neighbour
    e. 10-fold Cross Validation for evaluation

References:

1. Kaggle.com : https://www.kaggle.com/c/restaurant-revenue-prediction