A MACHINE LEARNING BASED APPROACH TO PREDICT TRAVEL EXPERIENCE BASED ON TOURISTS' RATING REVIEWS

*Emam Hossain and Sowmitra Das
Department of Computer Science and Engineering,
Port City International University, Chattogram.

ABSTRACT

The objective of this work is to evaluate and analyze the performance of Decision Tree, Support Vector Machine, Neural Network, Random Forest and Naïve Bayes algorithms for predicting google user review rating on travel experience. This research can help the tourists to determine which places to travel and which places not to. The experiment was conducted based on the rating the users provided on 24 key attributes of tourist spots. The performance of the five algorithms was calculated using confusion matrix, area under curve (AUC) and kappa matrix. The results indicate that among the five algorithms, Support Vector Machine provides better results than other algorithms in terms of all three matrices.

Keywords: travel experience, travel recommendation, tourists' ratings, travel assistant, machine learning.

INTRODUCTION

Nowadays, many people like to explore the world for various reasons like, to spend their holidays, to hang out with friends, to escape from the work pressure, or just to enjoy the beauty of the nature. But sometimes they have to face unpleasant situation in unknown places about which they don't have any prior idea. Many such unwanted situations can be reduced if the traveler has some idea about the place he wants to go. Almost every traveler gives a google review rating wherever they go. Those review are quite helpful for other travelers for various understanding and planning processes. Although, user rating is available to everyone but prediction using classification algorithm for overall tour experience using user review is hard to find and also, people cannot decide whether to visit a tourist spot based on simple review of a fixed place. In this research, a machine learning based tour place decider is proposed using Google reviews of tourist places across the Europe. Five supervised machine learning algorithms were used for classification, namely Random Forest, Support Vector Machine, Neural Network, Decision Tree and Naïve Bayes and then compare performance of these algorithms. The main aim of this research is to build an intelligent system to decide whether the tourists' experience of a particular place was good or bad based on their google reviews. To achieve this aim, secondary objectives are:

- Collecting and preprocessing the Google review dataset of tourist places across the Europe
- Building a model to determine tourists' experience based on their rating given in each of 24 attributes
- Comparing the performance Support Vector Machine, Decision Tree, Random Forest, Naïve Bayes and Neural Network in terms of Accuracy, Area Under Curve and Kappa matrices.

Pang *et al.* proposed a method for classifying documents by overall sentiment, but not by topic (Pang, Lee, and Vaithyanathan, 2002). They used standard movie review as training data in Maximum Entropy Classification, Support Vector Machine and Naïve Bayes. Their results showed that machine learning techniques predict better than human-produced baselines.

Corresponding Author: Emam Hossain; *ehfahad01@gmail.com*