Nate Bitting
CIS435 - Individual Assignment 1
April 20, 2014

Executive Summary

The Wisconsin Dells city conducted a survey in order to learn the preferences of visitors based on all of the activities available including water parks, shopping, fine dining, bungee jumping, among many others. The objective of gathering this data is to develop a targeted marketing strategy in order to increase business for each of the business owners in the Wisconsin Dells area. The campaign will focus targeted online advertising on the Wisconsin Dells website. To track the return on investment of this campaign, site visitors will be given unique coupons they can leverage at different businesses they visit while vacationing in Wisconsin Dells.

Key Components of the Marketing Campaign

In order to better understand the preferences of the visitors, we leverage the demographic information from the survey to create decision tree models for each of the activities to determine which visitors would be interested in each activity. Once created, the decision tree models would be implemented on the Wisconsin Dells website to provide targeted advertisements of activities the site visitors might consider while visiting the area. The key advertising components of the marketing strategy include the following:

1. Online advertisements and coupons on the Wisconsin Dells website leveraging cookie data for people just browsing the site
2. After customers book a vacation on the Wisconsin Dells website, we then leverage the data from their booking to create unique coupons they can use while visiting

By creating these unique coupons for site visitors browsing the website as well as the coupons for the visitors who actually book a vacation, we can have a means of calculating the return on investment (ROI) of implementing this marketing strategy over time. In order to improve the

decision tree models in the future, we can then send follow up emails with the survey after the people have visited Wisconsin Dells and update the models to further improve accuracy.

Methodology

After conducting some initial exploratory data analysis (EDA) on the survey data, it was obvious the higher the number of nights the respondents stayed, the more activities in which they participated. Another visual representation which provided interesting insight was calculating the sum and mean for each of the activities to determine which activities were the most popular. As it turns out, shopping was the highest activity for individuals from specific regions. While this information was interesting, creating decision tree models that could be implemented on the Wisconsin Dells website seemed like the best opportunity to implement a machine learning algorithm for targeted marketing.

The first step in creating the decision tree classification models was to first determine what set of input variables (demographic data) would be used predict the target variable (each activity). The input variables included all of the variables except for the ID variable and the target variable was each of the activities (binary) provided in the survey data. We then ran each of the decision models through a 10 fold cross-validation to determine the accuracy of each decision tree.

Attachments

- Average cross-validation results for each decision tree model (Avg CV Results.png)
- Graphical representation of 33 decision tree classification models (~/trees/*.dot)