Final Project Proposal - Coupon Recommendation (In vehicle)

Aarushi Verma

27 October, 2021

Overview

This dataset is used in the research paper - "A bayesian framework for learning rule sets for interpretable classification". The paper proposes a machine learning algorithm comprising of small short rules to build classifiers for a in-vehicle coupon recommendations. This analysis aims to explore the various factors which may impact the likelihood of whether the driver will or will not accept the coupon.

Research Questions

The questions of interest in this project are: * What factors influence the driver to accept a coupon? * How does the acceptance of the coupon change based on demographic factors of the driver? * Does the type of coupon affect the likelihood of the driver accepting the coupon? * What factor, if controlled for can provide the highest likelihood of a driver accepting a coupon?

Data

The data is being sourced from the UCI Machine Learning Repository. The data was collected via a survey on Amazon Mechanical Turk. The survey described different driving scenarios including the destination, current time, weather, passenger, etc., and then asked the person whether they will accept the coupon if they are the driver. This data was used in the paper 'A Bayesian framework for learning rule sets for interpretable classification'. The dataset contains observations for 12,684 surveyed respondents. Some relevant variables in the dataset are:

- Coupon: Restaurant(<$20), Coffee House, Carry out & Take away, Bar, Restaurant($20-$50)
- Destination: No Urgent Place, Home, Work
- Weather: If the weather is either Sunny, Rainy or Snowy
- Time: 2PM, 10AM, 6PM, 7AM, 10PM
- Coupon: Restaurant(<$20), Coffee House, Carry out & Take away, Bar, Restaurant($20-$50)
- Gender: Female, Male
- Age: 21, 46, 26, 31, 41, 50plus, 36, below21
- MaritalStatus: Unmarried partner, Single, Married partner, Divorced, Widowed
- Has_Children:1, 0
- Education: Some college - no degree, Bachelors degree, Associates degree, High School Graduate, Graduate degree (Masters or Doctorate), Some High School

The complete data dictionary is included in the Github Repo

Project Plan

The response vairable in this problem looks like a binary classification problem. We are looking to classify whther a driver will or will not accept the coupon. It might be a good decision to use logisitic regression to analyze which factors affect the odds of coupon acceptance. EDA will be a major chunk of the project, since there are multiple variables to explore in this dataset. The plan to complete this project is as follows - * EDA - 7 days * Model Building and Validation - 7 days * Validation and Final Resuls - 2 days * Report writing and presentation - 3 days