ISM 6136 Project: In-Vehicle Recommendations

Group Team Members: Jennifer Baker, Yanjing Ren, Safrin Patil, Devan Kreitzer, Patrick Maggio, Sravani Yadavalli, Rachana Nagaraj Banakar

URL: <https://archive.ics.uci.edu/dataset/603/in+vehicle+coupon+recommendation>

The food services industry is worth hundreds of billions of dollars. Large brands compete for market share, along with smaller shops that may offer higher quality products or personalized services. Customers are quick to change their preferences based on brand recognition, cost, and quality of product. Because of this, customer acquisition and retention are critical to the growth of a food services company. Coupons are one of the methods used to entice new customers or encourage existing customers to become frequent users. However, excess coupons can lead to a negative brand image, or annoy the customer to the point where they would accept any product if it were not from the offending company. Our model aims to provide the benefits of increasing the customer base and sales without the negative effects of spamming coupons. This will be done by predicting likely users based on several individual characteristics. To that end, we planned to build a model optimized for a custom F1 score with Beta = 2 (further referenced as F2 score) that could optimize recall (failing to give a coupon to someone who would use it) while still considering the value of precision (giving a coupon to someone who does not want it).

The advent of apps for each restaurant, rewards, and customer specific data collections means that location services and existing customer data could be leveraged with our model to deliver push notifications at appropriate times to increase sales. The size of the USA coffee industry is about $340 Billion and the fast casual market, valued at $190 Billion is predicted to grow at a rate of 12.4% ([source](https://www.fastcasual.com/news/fast-casual-industry-growing-strong/#:~:text=The%20fast%20casual%20restaurant%20market,in%20South%20America%20and%20MEA.)). These figures highlight the potential for a targeted system to boost sales when customers are most receptive. Within these respective markets, targeted coupons will be effective in converting customers from one brand to another. For example, a customer on the way to Dunkin for coffee due to the low cost may change their mind and go to Starbucks when a 20% off coupon is provided.

This 2017 survey on Amazon Mechanical Turk (a crowd-sourcing platform for surveying and other temporary tasks) was based on descriptions of different driving scenarios including: various destinations, passengers, times, weather. There are 25 attributes describing the customer and situation. Participants were asked whether they would redeem a coupon in those conditions. The coupon could be for assorted establishments: coffee shops, cheap restaurants, expensive restaurants, takeout, or a bar. All the features of the dataset are categorical in nature with even numerical data grouped into buckets.

This data was referenced in the paper by Wang et al, 'A Bayesian framework for learning rule sets for interpretable classification.' The Journal of Machine Learning Research 18, no. 1 (2017): 2357-2393. They were utilizing the data to apply rule-set models to provide predictions with the descriptions of a class – the reasoning – of the model. The authors are attempting to understand and develop context-aware recommender systems. This is interesting because we, as a team and class, utilize “black box” methods rather than trying to understand ruleset or Bayesian framework models. We do not necessarily need the descriptive analytics that the paper’s authors are researching to provide a business value.

Our team is proposing to develop a model for this dataset because personalized recommendations based on profiles to provide encouragement or engagement with local products and services is a market that we feel is strong enough to warrant predictive analysis on different metrics for coupon redemption. If we know that someone will use a coupon, that means that we know that person will be a customer, which leads to more profit, even with the coupon’s value. The coupon is a form of advertisement for the goods and services available based on the driver’s profile, and a way to interact with potential customers that may not think to otherwise purchase or experience our goods and services. Customers are bombarded with ads everywhere. However, customers favor coupons because of the adventurous feeling they get by trying new experiences without paying the full cost as well as when they get to save on items that they already want or need.

The dataset we have selected has provided a foundation for numerous scholarly inquiries, one of which is detailed in the paper titled “Comparative Study of J48 Decision Tree Classification Algorithm, Random Tree, and Random Forest on In-Vehicle Coupon Recommendation Data” by Hermawan et al. Leveraging the same data set, the researchers embarked on an exploration of three decision tree classification algorithms – employing 10-fold cross validation for their approach. They found that the Random Tree algorithm, despite being the most time-efficient at 0.14 seconds, lagged in accuracy at 67.38%. Meanwhile, the J48 algorithm improved accuracy to 72.79% but required double the processing time. The Random Forest algorithm emerged as the most accurate at 77.09%, albeit at a considerably higher computational cost of 3.1 seconds. Building upon this knowledge, we are keen to explore how the inclusion of other sophisticated models such as Logistic Regression, K-Nearest Neighbors (KN), Support Vector Classifier (SVC), Neural Networks, and XGBoost may further enhance the predictive ability of coupon utilization and specifically, we intend to utilize an F2-Score.

The reasoning behind using the F2-score to evaluate our models is tied to customers’ favor. F2 score optimizes recall which reduces failing to give a coupon to someone who would use it while still considering the value of precision which lowers giving a coupon to someone who does not want it. Bulk-sending of coupons may be wide across the industry, but we want to focus on customers that will use them and prevent customers from becoming annoyed with bulk notifications of coupons and then muting the notifications. Currently, companies will send coupons in bulk to all users of the app. This is effective in helping customers that would already use the coupon, but it will cause customers to turn notifications off if they are annoyed at the intrusion, potentially preventing them from ever going to the restaurant. By tuning models to the F2-Score, we take a different approach than others who have researched this dataset.

Furthermore, we can create an in-house app or sell the model to a company since this model will provide substantial value to those that use it. Our model will allow us to give coupons to demographics that we expect to use the coupon, giving a positive impression while avoiding brand harm by pestering those that would not be interested. Our model will leverage existing data collection by the restaurant industry and deploy it in a way that provides a competitive advantage through intelligent advertising.