In-Vehicle Coupon Recommendation System

Su23 – Data Science Programming – Final Project

Vishal Anand Gupta, Trevor Allison, Jyotis Joy,
Laurenz Pehl, Pritesh Singh

Introduction:

Coupon marketing is one of the oldest and most effective business marketing methods, in which special offers on products and services are sent to prospective customers. The success of any coupon marketing campaign depends not only on targeting the right set of customers but also targeting them at the right time and in the right situation when they are most likely to be convinced to make a purchase.

Description of Project Goals:

The goal of the project is to propose a framework that would act as a coupon recommender system for in-vehicle users. The framework is created using machine learning models that try to understand optimal scenarios that influence whether a person will utilize a coupon that has been presented to them. In essence, this discovery has to do with optimal marketing conditions that in turn drive customer traffic and revenue. The dataset we have chosen for this purpose was collected via a survey on Amazon Mechanical Turk. The survey describes different driving scenarios including the destination, current time, weather, passenger, etc., and then asks the person whether he/she will accept the coupon.

The importance of this framework lies within the practical applications of the insights that can be gleaned from it. For example, if Starbucks were to send out a coupon on their app, should they do it on a sunny or rainy day, at 10am or 6pm? These micro factors often coalesce into determining whether someone makes a purchase or not. Any food outlet/chain or marketing agency could utilize our findings to optimize their advertising approach which would in turn lead to increased customer engagement, boost in sales and revenue, reduction in costs and competitive advantage.

Exploratory Analysis:

The dataset we had chosen had 12,684 instances and 26 variables to begin with. The variables are summarized below under 6 broad categories:

- **Target variable** – whether the coupon is accepted or not

- **Demographics** – Age, Gender, Marital status, Education, Occupation, Income etc.

- **Propensity to eat out** – How often does the respondent goes to a bar / coffee house / restaurant / gets carry away food

- **Day Attributes** – Weather, Temperature & Time during journey

- **Journey Attributes** – Destination & Passengers in the car

- **Coupon Attributes** – Coupon type, expiration timeline, time to reach outlet and direction of the outlet

Note that we took out 1 variable – "car" which had a lot of missing data.

Moreover, we removed instances that were missing data in one or more variables. Eventually, we were left with 12,079 instances and 25 variables.

We then moved to viewing some basic statistics and easily recognizable patterns to gain a better understanding of our dataset. We plotted the raw probability of accepting the coupon by each of the variables assuming independence among them.

1. **Demographics**: Younger people, unmarried folks, and high school or college graduates have higher likelihood of accepting the coupon. *(Refer to Fig 1.1)*

2. **Day Attributes**: Afternoons, sunny days and moderate temperatures around 80 degrees work best for acceptance of coupon. *(Refer to Fig 1.2)*

3. **Journey Attributes**: Higher likelihood of acceptance when friends are co-passengers, and they don't have to reach anywhere urgent. *(Refer to Fig 1.3)*

4. **Propensity To Eat Out**: Higher likelihood of acceptance when the person generally eats out more often. *(Refer to Fig 1.4)*

5. **Coupon Attributes**: Carry out and budget restaurant coupons have higher chance of acceptance. *(Refer to Fig 1.5)*

Solution and Insights:

We tried different classification models and compared the accuracy among them and vs **baseline accuracy (56%)** to come up with the best predictions.

Train: Test Split

- In case of all models the data was split in the ratio of 70:30, i.e. 70% to train the model and 30% to test it

Comparison of Models across all parameters

Model Type	Parameters	Precision	Recall	F1 Score	Accuracy
Naïve Bayes Classifier	N/A	68%	72%	70%	65%
Decision Tree	max depth = 10	71%	76%	74%	69%
Bagging	N/A	77%	74%	75%	73%
Random Forest	# estimators = 27	75%	81%	78%	74%
Gradient Boosting	# estimators = 500 max depth = 6	78%	80%	79%	76%

(Refer to Fig 3-6 to see how the parameters for each of the model were selected)

(Refer to Fig 7 for the precision vs recall curve for the ensemble methods)

Based on the table above, we can see that **Gradient Boosting model** gives the best predictions.

(Refer to Fig 8-11 to see feature importance across models)

Recommendations

Based on the feature importance list, we derived the following recommendations for food outlets and marketing agencies:

- **Coupon Type** - The market agency should focus on coupons pertaining to take out food or restaurants where per person cost is < $20 as they are likely to yield better results

- **Low Yield** - Coupons pertaining to coffee house or a restaurant where per person cost is $20-$50 are likely to yield low results

- **Journey Attributes** - Notifications should be sent when people are travelling with friends and are not going anywhere urgent

- **Distance from outlet** - Notifications should be when the vehicle is within a radius of 15 minutes from the outlet

- **Day Attributes** - Weather and temperature should be moderate, and time should preferably be in the afternoon

- **Save Efforts** - Efforts should be minimized on people who never or rarely go to a coffee house or bar or get carry away food

Limitations

There are also some limitations of these findings.

- **Data Collection** - It is difficult to collect data in such detail like how frequently does the person goes to restaurants or gets take away food

- **Precision Targeting** - High precision is required with respect to sending the notification as there is a golden window of time during which the campaign is effective

- **Limitations of survey data** - The data is based on a survey. Responses cannot be taken at face value. While it does identify the scenarios where the campaign may be most effective, the success rate within those scenarios may not be 100%

Appendix:

Figure 1.1: EDA on Demographics variables

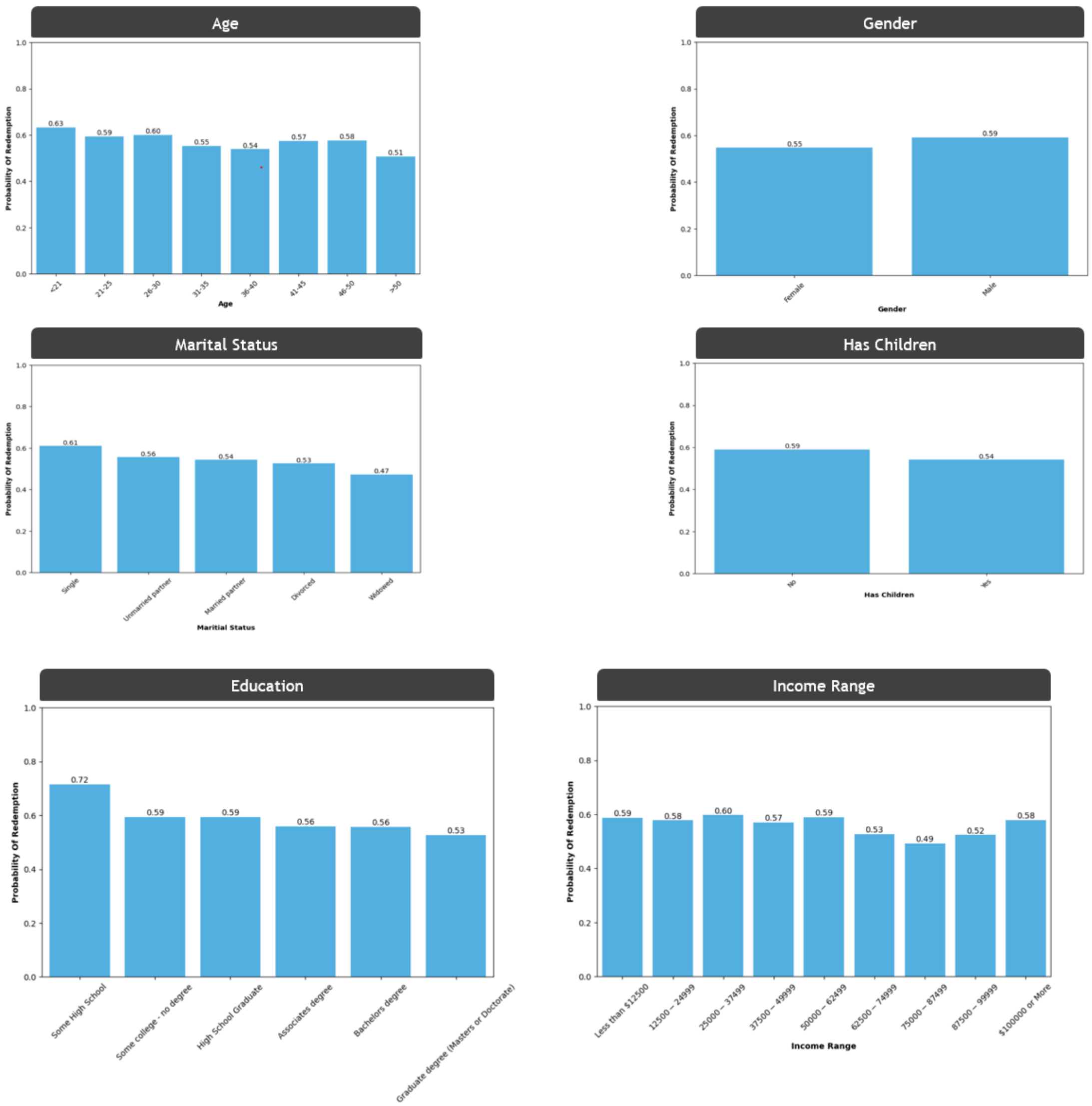

Figure 1.2: EDA on Day Attributes

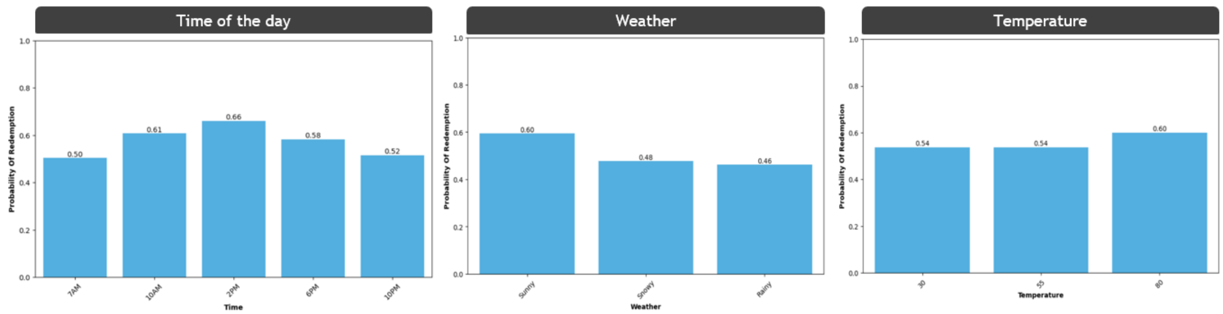

Figure 1.3: EDA on Journey Attributes

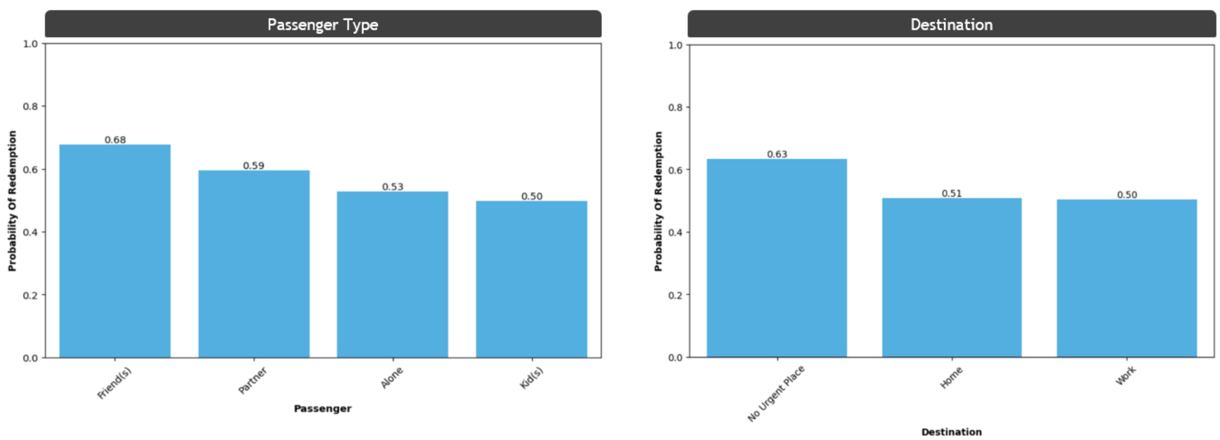

Figure 1.4: EDA on Propensity to eat out

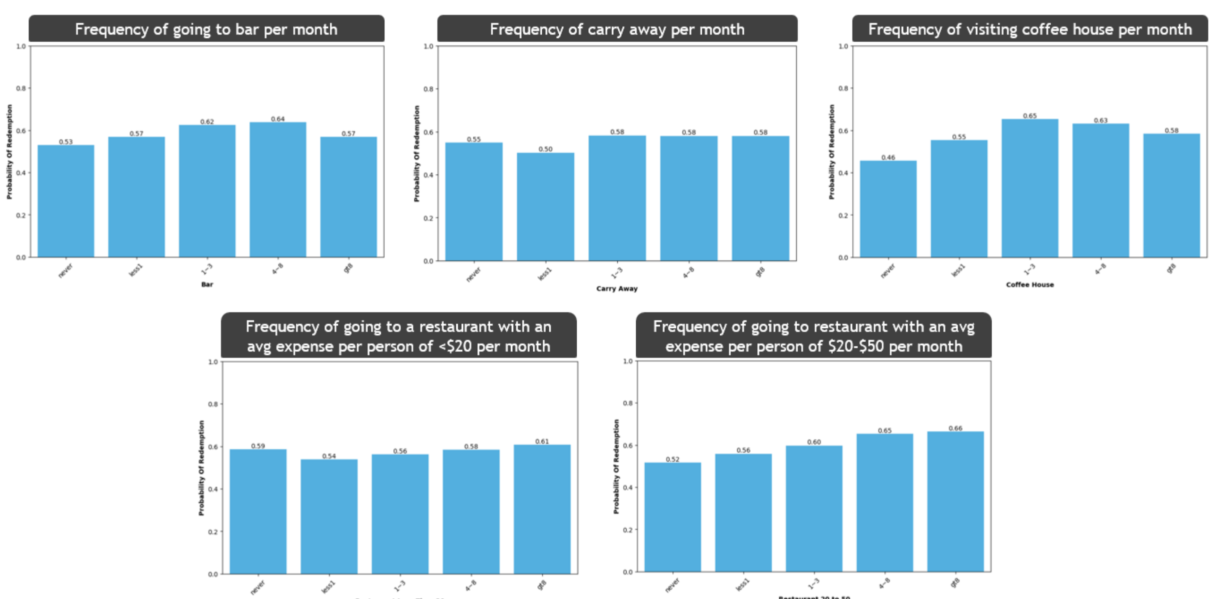

Figure 1.5: EDA on Coupon Attributes

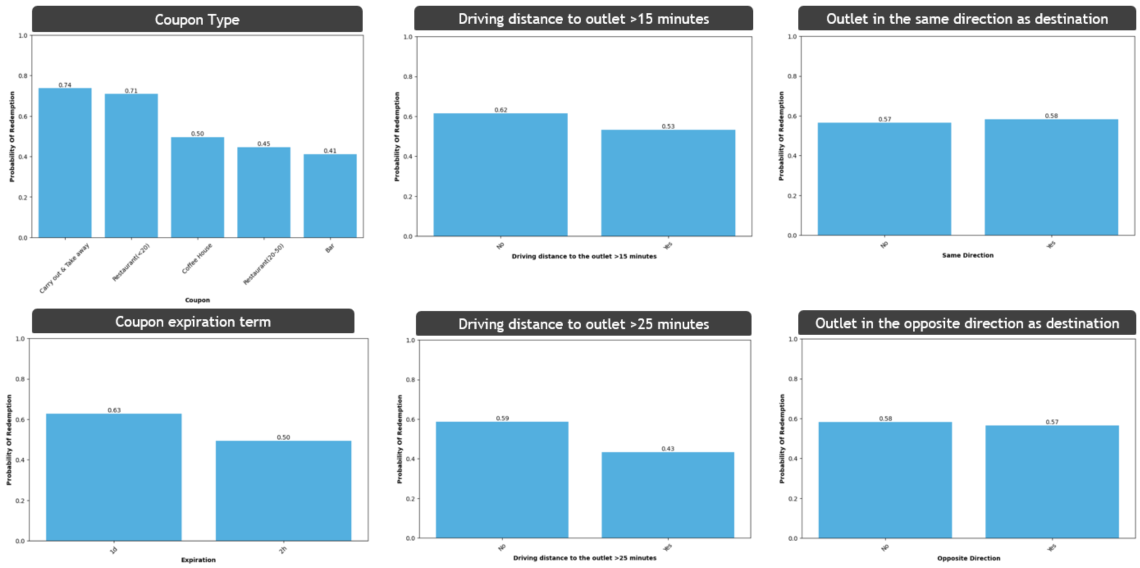

Figure 2: Test Accuracy across models

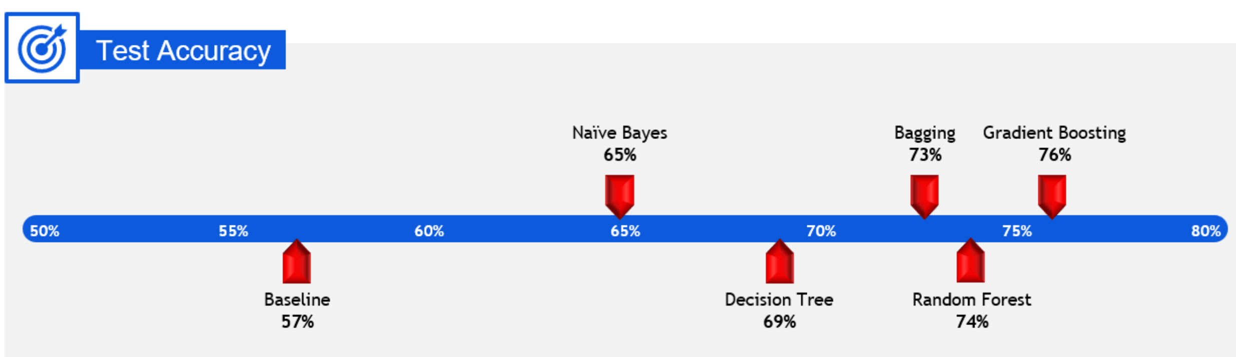

Figure 3: Selection of number of trees for Gradient Boosting model

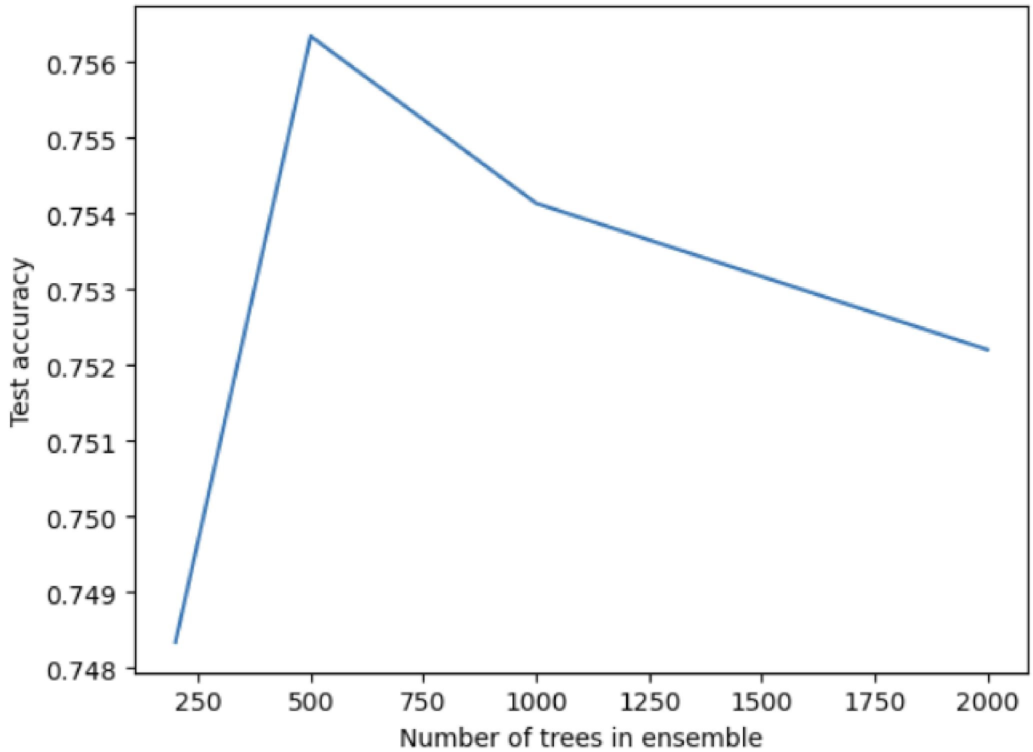

Figure 4: Selection of max depth for Gradient Boosting model

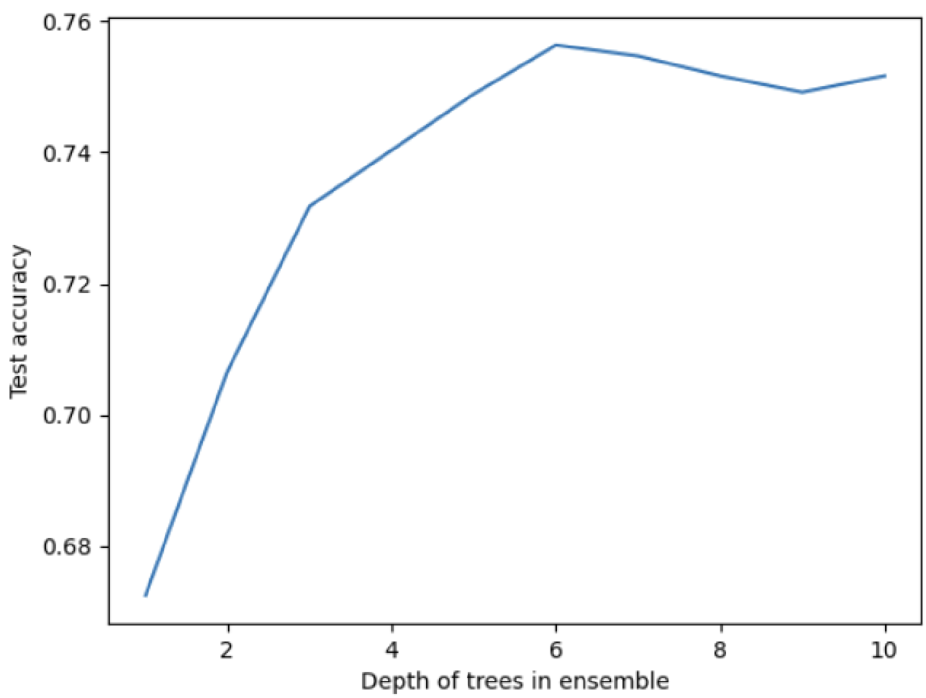

Figure 5: Selection of number of estimators for Random Forest model

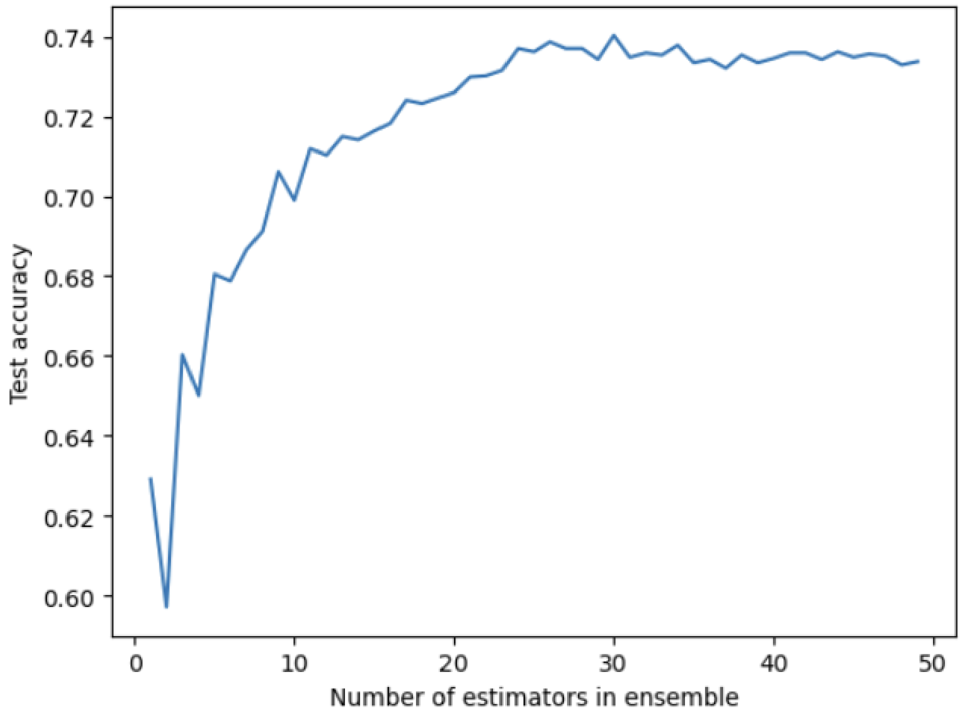

Figure 6: Selection of max depth for Decision Trees model

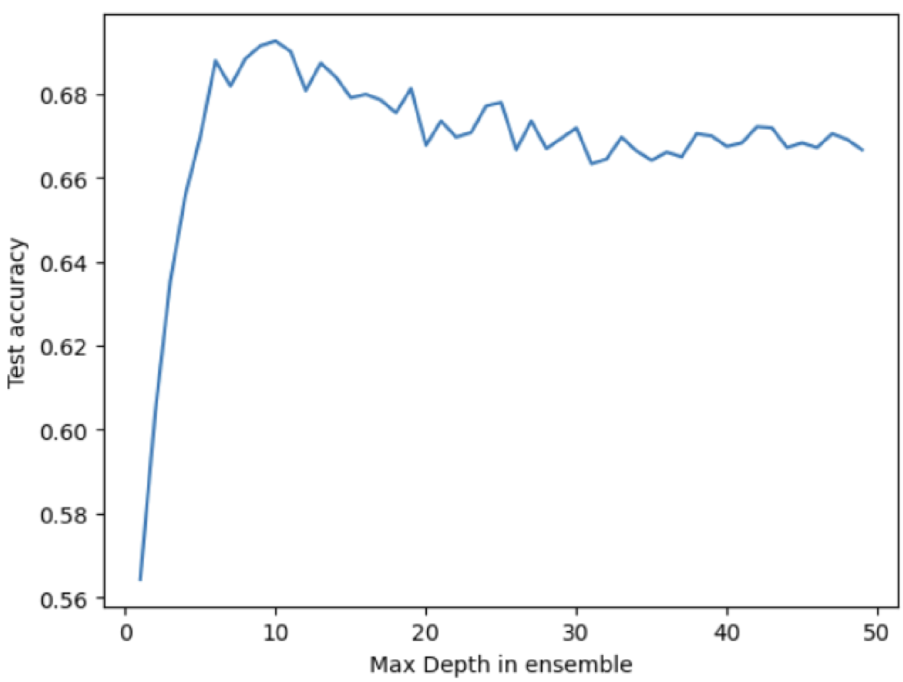

Figure 7: Precision vs Recall curve for all the ensemble methods

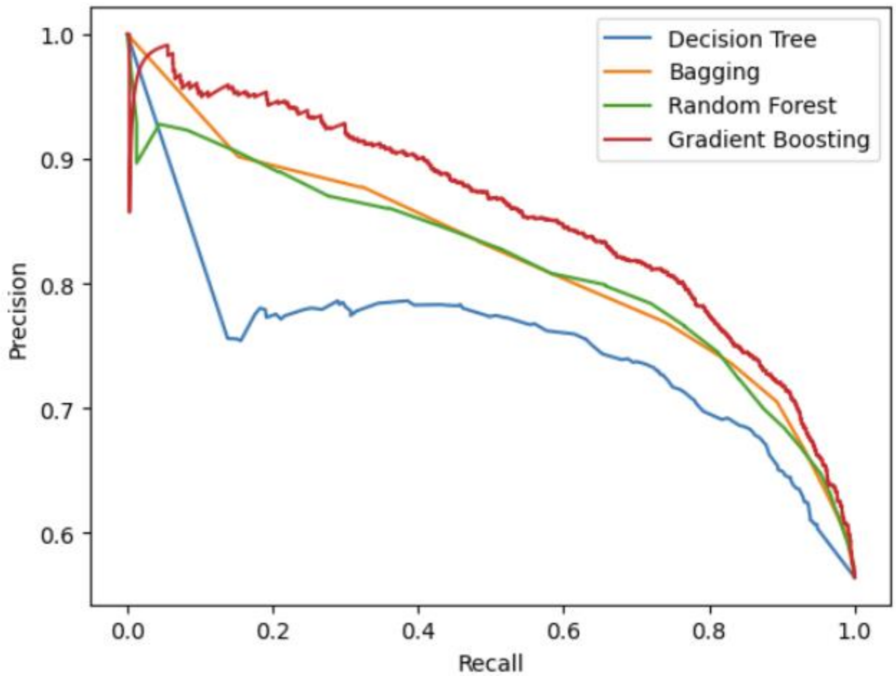

Figure 8: Feature Importance (Top 20) for Naïve Bayes Model

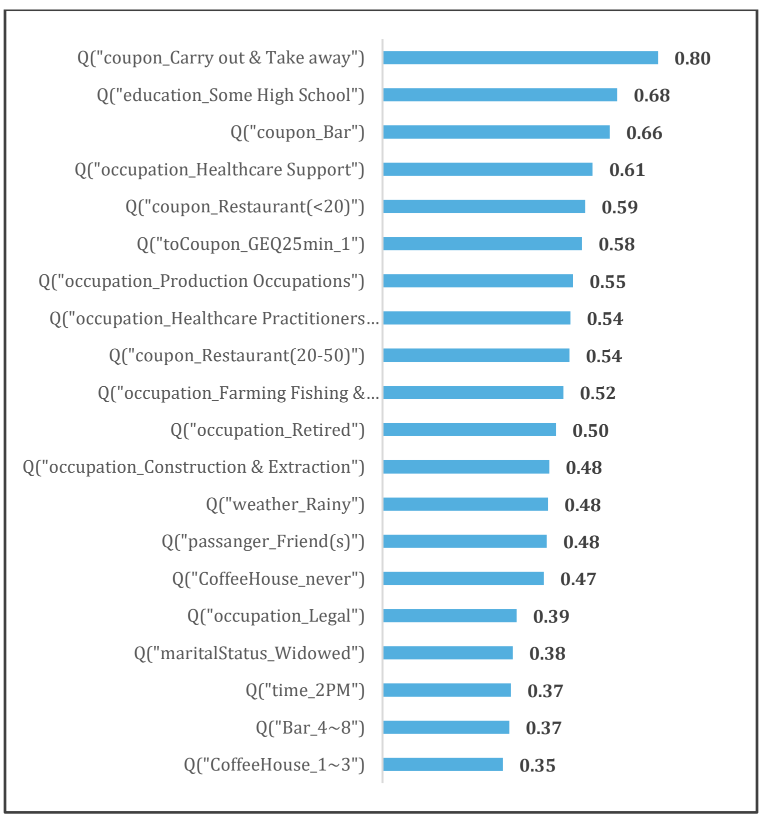

Figure 9: Feature Importance (Top 20) for Decision Trees model

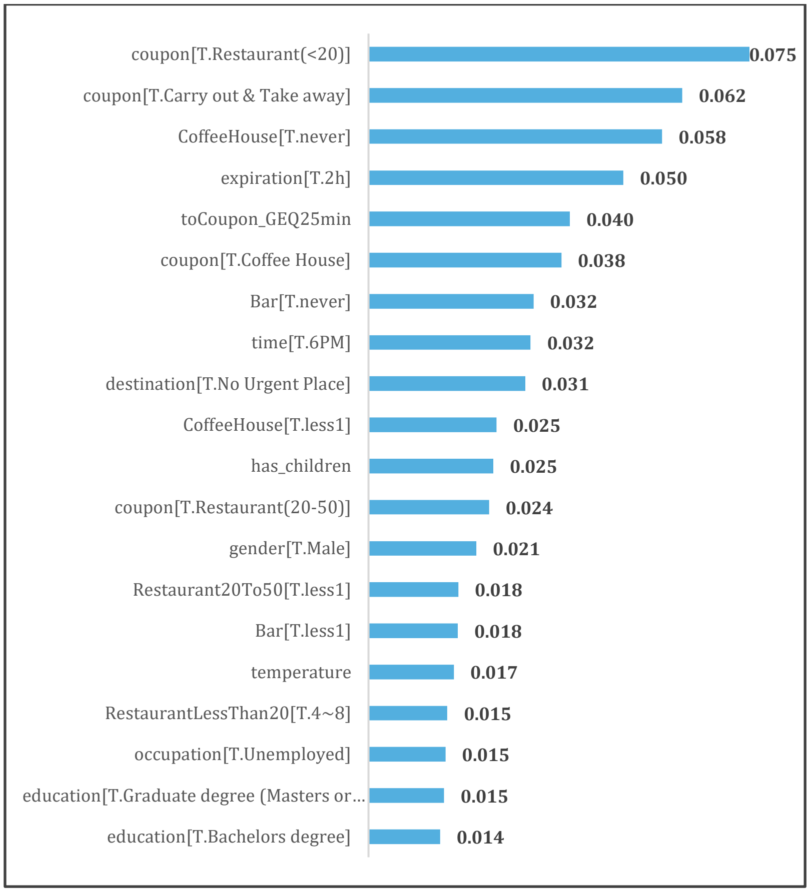

Figure 10: Feature Importance (Top 20) for Random Forest model

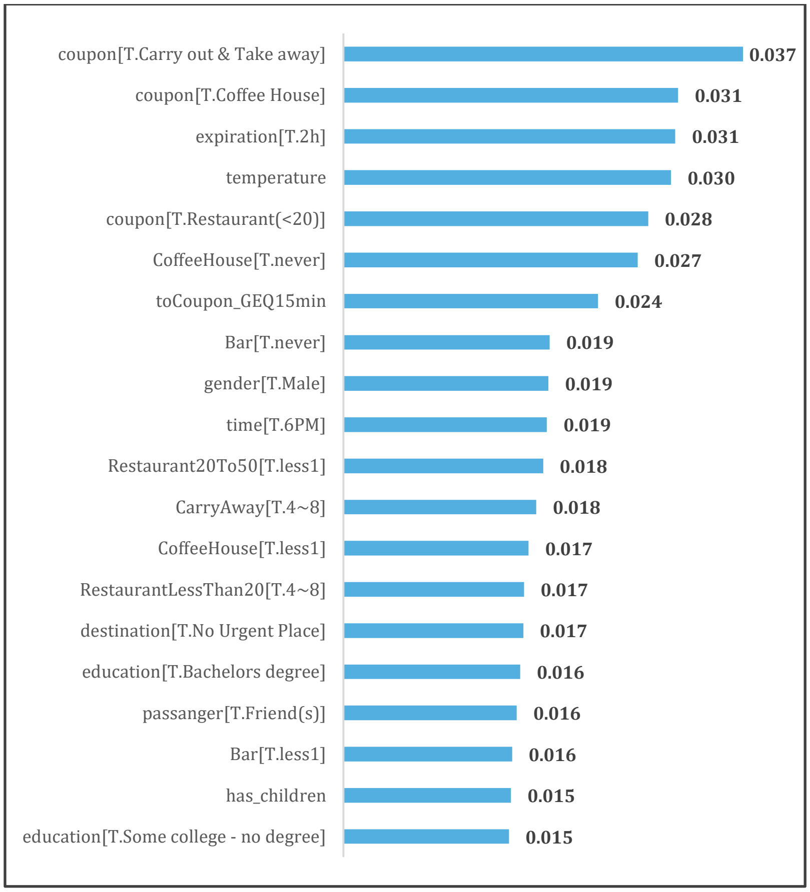

Figure 11: Feature Importance (Top 20) for Gradient Boosting model