Prediction of Street Marketing Success Rate: A Coupon-Recommendation Case

1st Linsey Liu
Department of Psychology
University of Manitoba
Winnipeg, Canada
liuy49@myumanitoba.ca

2nd Dinh Thien Pham
Department of Computer Science
University of Manitoba
Winnipeg, Canada
phamd@myumanitoba.ca

3rd Carson Leung
Faculty of Science
University of Manitoba
Winnipeg, Canada
Carson.Leung@umanitoba.ca

***Abstract*—Based on a large-scale social experiment on in-vehicle coupon recommendation conducted via Amazon Mechanical Turk, this study aimed to predict the acceptance rate of drivers' coupons across various situations. The analysis covered multiple influencing factors, including objective circumstances, accessibility to target shops, demographic information, and coupon-related attributes. After generating general patterns through Monte Carlo simulations and performing comprehensive feature engineering, this study developed a classification approach using the Random Forest Model—recognized for its strong performance in previous studies—alongside the Artificial Neural Network (ANN), to achieve higher prediction accuracy within a controllable time frame. The results revealed the following key findings: a. Coupon information had the greatest impact on drivers' decisions regarding recommendations, while demographic factors were found to be insignificant; b. The proposed model outperformed previous approaches by achieving an accuracy of 0.77, with significantly fewer features after feature engineering. Psychological interpretations were also provided to explain these findings, offering deeper insights into the observed patterns.**

***Index Terms*—feature engineering, Monte Carlo simulation, Random Forest, ANN, customer psychology.**

I. INTRODUCTION

Street marketing is widely recognized as one of the most direct and effective strategies for promoting products or services. Unlike traditional marketing methods such as billboards or TV ads, street marketing involves engaging techniques that actively capture the attention of consumers in real time. These techniques often require careful consideration of various factors to ensure their success. Take, for example, a common coupon recommendation scenario: when drivers park their cars in a parking lot near a bus station, a promoter stands nearby offering coupons, with the goal of encouraging drivers to accept them. What determines the success of this interaction? Should the promoter focus on families or individual drivers? Is the morning or afternoon more suitable for promotion? How does the number of coupons distributed affect acceptance rates? Will coupons with shorter expiration dates prompt faster action? Each of these questions influences the binary outcome: accept or not accept.

This study analyzes data from 12,799 participants who took part in the survey. The dataset includes 25 features, along with an independent variable Y, representing the binary decision (0 for not accept and 1 for accept). These features are grouped into five categories: a. Demographics: gender, age, marital status, income, occupation, education, and whether the participant has children. b. Environment: destination, passenger information, time of day, weather, and temperature. c. Coupon Information: type of coupon (e.g., discount, free item) and expiration time. d. Customer Habits: frequency of visits to bars, restaurants, coffeehouses, or take-out services. e. Accessibility: distance or direction between the parking lot and the target shop.

By segmenting the data in this way, this study aims to not only build prediction models but also to uncover general patterns across features and explore the behavior of consumers from a psychological perspective. Understanding consumer behavior is crucial in street marketing because decisions to accept or reject coupons are influenced by more than just rational factors; emotions, convenience, urgency, and even social influences all play a role. For example, consumers may be more likely to accept a coupon if they feel a sense of urgency (e.g., limited-time offers) or if the offer aligns with their usual habits, such as frequenting a particular restaurant or store. These psychological drivers of consumer behavior need to be considered when designing marketing strategies.

The practical applications of this study are twofold. First, for this specific case of street marketing, the machine learning models developed can predict the likelihood of coupon acceptance under various conditions. This provides useful guidance for promoters, helping them focus on the right demographic, time, and offer type to maximize the effectiveness of their campaigns. For instance, if the study finds that consumers are more likely to accept coupons in the evening, promoters can adjust their schedules accordingly, ensuring that resources are used most efficiently. Second, this research can provide broader insights into consumer decision-making patterns. By identifying which factors most influence whether a consumer accepts a coupon, businesses can optimize their marketing strategies, focusing on high-impact elements and reducing unnecessary effort on less influential factors.

In addition to offering practical applications for street-marketing, the study also emphasizes improving marketing strategies at a broader scale. By understanding which factors—such as time of day, weather, or coupon expiration—are most influential in driving consumer behavior, businesses can better allocate resources. For example, if a specific demographic or time of day shows a higher likelihood of coupon acceptance, businesses can adjust their marketing efforts accordingly, focusing on the most effective strategies and saving on resources. Moreover, these insights can help businesses design more targeted promotions, avoiding one-size-fits-all campaigns and making marketing more efficient.

Several previous studies have explored this dataset, applying models such as decision trees, Bayesian inference, and Random Forest to predict coupon acceptance. In comparison to earlier work, the contributions of this study are as follows: a. Feature Engineering: We conducted comprehensive feature engineering, including feature encoding, interaction terms, dimensionality reduction, and binning. These steps enhance the model's ability to handle complex relationships and reduce unnecessary noise, thereby improving prediction accuracy. b. Statistical Corrections: We corrected errors in statistical methods from prior studies, ensuring that patterns derived from the dataset are more reliable and valid. c. Artificial Neural Networks (ANN): We introduced ANN techniques, comparing them with the top-performing models from previous studies. We adjusted ANN layers and evaluation metrics to improve the prediction results. Through these contributions, this study aims to provide more accurate predictions and a deeper understanding of the factors influencing consumer decisions in street marketing.

II. Related Work

Several studies have previously explored this dataset to develop predictive models (Wang, Rudin, Doshi-Velez, et al., 2017; Hermawan, Fatihahab, Kurniawatibc, Helen, 2021; Depari, Shu, Fachriza, Chow, Wijaya, Witana, 2022). In the following section, we will review each of these studies, provide detailed descriptions, and summarize their limitations. This dataset originates from a large-scale social experiment conducted by Amazon Mechanical Turk, surveying over 12,000 American drivers to uncover potential factors influencing their coupon acceptance decisions.

A. Generating Bayesian Rule Sets Model (2017)

Wang et al. (2017) were the first to collect this dataset and developed a new rule-mining method, "Bayesian Rule Sets (BRS)," comparing its performance with lasso regression, CART, C4.5, RIPPER, and Top*K*. They used 5-fold cross-validation to evaluate accuracy and complexity. The BRS classifier achieved better accuracy than decision trees and RIPPER, with an average of 0.731, though it slightly lagged behind lasso regression (0.745). In terms of complexity, BRS performed similarly to CART, C4.5, and Top*K*, while maintaining the best overall balance among these models. However, lasso regression demonstrated consistently better ROC performance.

Despite these contributions, several issues were noted. First, splitting the dataset into five coupon types and then into cross-validation folds reduced the training data for each fold, leading to lower accuracy. Additionally, the target variable Y was unbalanced within each coupon type, with the highest class ratio reaching 2.9:1, further impairing model performance. Consequently, most model accuracies were around 0.7. Second, BRS was significantly slower than other methods, requiring over 200 seconds to process UCI datasets, while decision trees and regression models took under 1 second. Finally, as data science techniques evolve rapidly, some methods used in this study, though innovative at the time, are now somewhat outdated.

B. Comparing Classifiers (2021)

Hermawan et al. (2021) compared three models—J48, Random Tree, and Random Forest—using 23 features from the dataset. Accuracy and computation time were the evaluation metrics, with models trained via 5-, 7-, and 10-fold cross-validation. Random Forest achieved the highest accuracy (77.09%), whereas Random Tree showed the lowest (67.38%). However, Random Tree had the fastest training time (0.14 seconds), while Random Forest required the longest (10.89 seconds).

This study employed commendable practices, such as multiple cross-validation strategies to minimize data loss and feature selection to exclude less meaningful features. However, limitations persisted. For instance, missing values in customer habit features (e.g., "How many times do you go to the coffee house a month?")—less than 1% of the data—were imputed using the most frequent value. This approach neglected the nature of Missing Completely at Random (MCAR), where missing values often occurred together due to participants skipping pages. As a result, noise was introduced, potentially biasing models that are naturally robust to missing values, like Random Forest. Additionally, although feature selection was conducted, further feature pruning could have reduced redundancy and improved the tree model's robustness and interpretability.

C. Introducing Business Analytics Method (2022)

This study introduced business analytics techniques and compared their performance with established machine learning models such as Random Forest, naive Bayes, and Decision Tree. Random Forest once again demonstrated the best performance, aligning with conclusions from previous studies.

The researchers applied statistical methods, including correlation and Chi-Square tests, to identify relationships between features and the target variable Y. While most tests yielded significant results, the large sample size likely amplified minor relationships, reducing their practical validity. Furthermore, variables with extreme skewness or homogeneity were identified but not excluded from the models, potentially introducing redundancy and increasing the risk of overfitting. Despite these issues, the descriptive analytics section provided valuable insights, such as identifying variables with homogeneous levels. However, more comprehensive feature engineering, such as transforming skewed variables or eliminating redundant ones, could have strengthened model reliability and reduced overfitting risks.

D. Our Consideration

To summarize the shortcomings of previous studies, we conclude the following:

a. Insufficient data processing in previous studies.

For instance, in the 2017 study, Wang's team failed to address columns with significantly more missing values, such as "Car," and did not apply proper handling for other types of missing or skewed values. Additionally, without data detection to manage these issues, the binary target variable Y was unbalanced. Although tree models are relatively insensitive to a small number of missing values, this became problematic for the Bayesian model and lasso regression. The 2021 study made improvements by cleaning the data and removing meaningless features. However, some issues remained in their handling of missing values based on their procedure. The 2022 study also included basic data preprocessing but did so in a less comprehensive manner.

b. Lack of attention to feature engineering in previous studies.

Both the 2017 and 2021 studies did not perform any feature trimming, while the 2022 study conducted limited feature engineering. Though the 2022 study attempted feature engineering, their steps were not exhaustive. Additionally, previous studies gave little consideration to the characteristics of all features in the dataset. More extensive work could be done, such as feature encoding, feature interaction, and dimensionality reduction, to improve model performance. For example, features like "occupation" and "education" contain numerous low-frequency levels that could be consolidated or classified. Similarly, "age" could be binned to reduce the impact of minor observational errors.

c. Use of outdated models and lack of parameter tuning.

In prior research, decision trees were repeatedly built and compared against other models. While decision trees are classical models, a single decision tree is still prone to overfitting and has poor generalization capability. Moreover, when ensemble learning methods like Random Forest were applied, there was rarely any parameter tuning. For improved performance, more recent models, such as deep learning techniques, should be employed, along with a parameter tuning process to optimize results.

d. Misapplication of statistical methods for pattern identification.

In the 2022 study, statistical methods were used to generate broader patterns from the dataset, including Chi-Square tests. However, the researchers concluded that all features were significantly related to coupon acceptance, a result influenced by the over-significance problem arising from the large sample size. Chi-Square tests are used to detect relationships between categorical variables, and a significant result leads to rejecting the null hypothesis. However, when the sample size is large—such as the dataset with over 12 thousand observations—this test may produce too many significant results, even for trivial effects. This misinterpretation of statistical significance in the 2022 study led to the conclusion that all features were critical for predicting coupon acceptance. To address this issue, new statistical techniques should be employed to ensure more reliable conclusions.

Our main contributions are as follows:

- In the data processing stage, we will focus on specific types of data (e.g., skewed data, missing values) and perform thorough data cleaning.
- We will apply corrections to the Chi-Square test in order to identify more accurate patterns within the dataset.
- In the feature engineering stage, we will ensure that variables are properly prepared for tree models. We will scan each feature, reduce dimensionality, and avoid collinearity issues to enhance model performance.
- Given the advancements in deep learning models in data science, we will introduce Artificial Neural Networks (ANN) as a new method for building the prediction model.

III. Main Body

A. Data Preparation

First of all, we need to figure out what features the original dataset contains, as well as their levels and meanings.

- destination: No Urgent Place, Home, Work
- passanger: Alone, Friend(s), Kid(s), Partner (feature meaning: who are the passengers in the car)
- weather: Sunny, Rainy, Snowy
- temperature:55, 80, 30
- time: 2PM, 10AM, 6PM, 7AM, 10PM
- coupon: Restaurant(<$20), Coffee House, Carry out & Take away, Bar, Restaurant($20-$50)
- expiration: 1d, 2h (level meaning: the coupon expires in 1 day or in 2 hours)
- gender: Female, Male
- age: 21, 46, 26, 31, 41, 50plus, 36, below21
- maritalStatus: Unmarried partner, Single, Married partner, Divorced, Widowed
- has_Children:1, 0 (level meaning: 1-yes; 0-no)
- education: Some college - no degree, Bachelors degree, Associates degree, High School Graduate, Graduate degree (Masters or Doctorate), Some High School
- occupation: Unemployed, Architecture & Engineering, Student, Education & Training & Library, Healthcare Support, Healthcare Practitioners & Technical, Sales & Related, Management, Arts Design Entertainment Sports & Media, Computer & Mathematical, Life Physical Social Science, Personal Care & Service, Community & Social Services, Office & Administrative Support, Construction & Extraction, Legal, Retired, Installation Maintenance & Repair, Transportation & Material Moving, Business & Financial, Protective Service, Food Preparation & Serving Related, Production Occupations, Building &

Grounds Cleaning & Maintenance, Farming Fishing & Forestry

- income: $37500 - $49999, $62500 - $74999, $12500 - $24999, $75000 - $87499, $50000 - $62499, $25000 - $37499, $100000 or More, $87500 - $99999, Less than $12500
- car: car that is too old to install Onstar, crossover, do not drive, Mazda5, scooter and motorcycle (feature meaning: the type of your vehicle)
- Bar: never, less1, 1∼3, gt8, 4∼8, never (feature meaning: how many times do you go to a bar every month?; level meaning: 'gt8' stands for 'greater than 8', 'nan' stands for missing value)
- CoffeeHouse: never, less1, 4∼8, 1∼3, gt8, nan (feature meaning: how many times do you go to a coffeehouse every month?)
- CarryAway:4∼8, 1∼3, gt8, less1, never (feature meaning: how many times do you get takeaway food every month?)
- RestaurantLessThan20: 4∼8, 1∼3, less1, gt8, never (feature meaning: how many times do you go to a restaurant with an average expense per person of less than $20 every month?)
- Restaurant20To50: 1∼3, less1, never, gt8, 4∼8, nan (feature meaning: how many times do you go to a restaurant with average expense per person of $20 - $50 every month?)
- toCoupon_GEQ5min:0,1 (feature meaning: driving distance to the restaurant/bar for using the coupon is greater than 5 minutes; level meaning: 0-no, 1-yes)
- toCoupon_GEQ15min:0,1 (feature meaning: driving distance to the restaurant/bar for using the coupon is greater than 15 minutes; level meaning: 0-no, 1-yes)
- toCoupon_GEQ25min:0, 1 (feature meaning: driving distance to the restaurant/bar for using the coupon is greater than 25 minutes)
- direction_same:0, 1 (feature meaning: whether the restaurant/bar is in the same direction as your current destination)
- direction_opp:1, 0 (feature meaning: whether the restaurant/bar is in the same direction as your current destination)
- Y:1, 0 (whether the coupon is accepted)

When examining these features, we identified several key points that require attention:

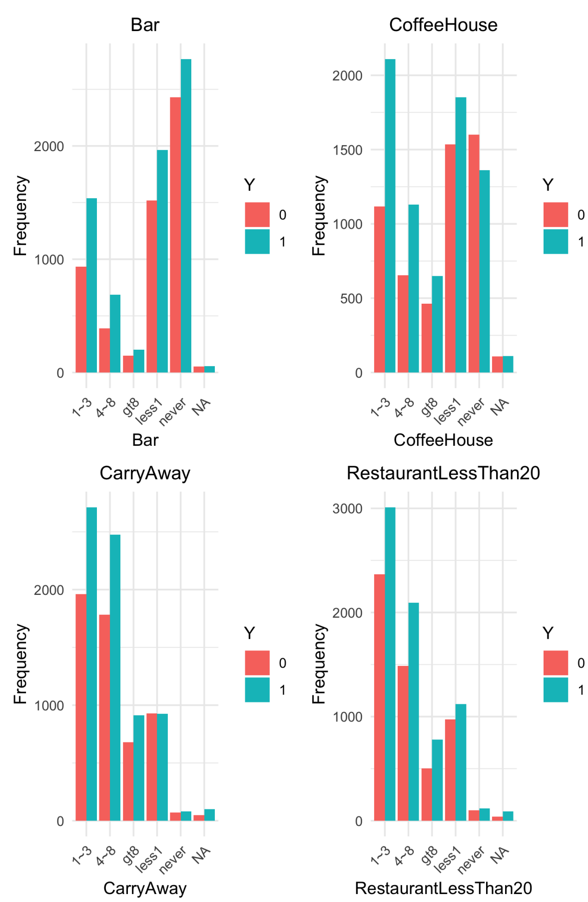
As mentioned earlier, these features can be divided into five segments. The first five features are related to the environment in which the survey and coupon recommendations took place. Next, the coupon type and expiration date are critical pieces of coupon-related information. Following that, the survey asked about demographics, including age, gender, marital status, etc. Since there are five types of coupons, the researchers also inquired about the drivers' habits related to visits to these types of restaurants. Finally, the distance and direction between the drivers and the target restaurants were recorded, which fall under the accessibility segment.

Upon reviewing the demographics segment, we observed that the age attribute contains discrete nodes (e.g., 21, 26) rather than continuous values. This could be due to the convenience of marking age ranges close to the participants' actual ages during the survey. However, such discrete levels are not ideal for analysis. Therefore, binning the age values would make the data more suitable for modeling. Additionally, the levels of the car attribute are confusing and not comparable (e.g., "the car is too old" and "do not drive"). It is advisable to remove this feature from the analysis.

In the habit segment, we found that each variable contained one level of missing values. These missing values need to be evaluated and processed before proceeding. Specifically, we will recode the "NaN" level to ensure it is recognized by R Studio. Before determining how to handle these missing values, we will first present the percentage of missing values (**Table 1**) and explore their relationship with the two levels of Y (**Figure 1**).

Table 1: percentage of missing values

feature	missing value	total	percentage
Bar	109	12684	0.86%
Coffee House	217	12684	1.71%
Carry Away	151	12684	1.19%
RestaurantLessThan20	130	12684	1.02%
Restaurant20To50	189	12684	1.49%



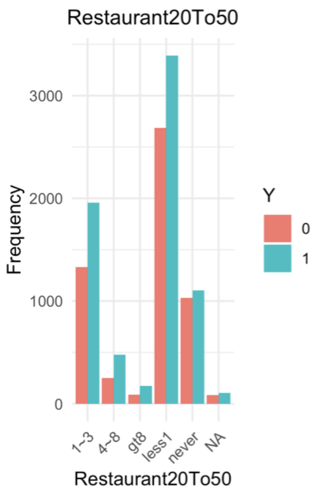


Figure 1: the frequencies of NAs (based on Y levels)

Considering the missing values account for only around 1%, their impact is minimal and can be nearly ignored. As mentioned, these values are missing completely at random (MCAR). For MCAR data, deleting the missing values is generally preferred over imputing them, as imputation might introduce bias. With a large sample size of over 10,000 records, removing a small number of missing values has little effect on model performance. Therefore, we chose to delete these missing values rather than fill them with the most frequent values.

In the accessibility segment, we observed that some features were derived from a single factor. Specifically, 'toCoupon_GEQ5min', 'toCoupon_GEQ15min', and 'toCoupon_GEQ25min' all describe distance. While it is common to convert categorical variables into dummy variables to avoid multicollinearity and capture independent effects, distance is actually an ordinal feature. Dummy coding loses the ordinal nature of the data, which is important for prediction. Therefore, we decided to combine these variables into a single feature. Additionally, the direction-related features ('direction_same' and 'direction_opp') showed redundancy, so we kept only 'direction_same'.

To summarize, in the data preparation phase, we handled missing values in the habit segment, removed confusing and redundant variables (car and direction_opp), and combined the three distance-related variables into one. The age feature, which requires further transformation, was carried forward to the next data engineering step.

B. Feature Engineering

It is obvious that all the features in this dataset are categorical. For a dataset like this, feature engineering, which means the process of selecting, manipulating and transforming raw data into features that can be used in supervised learning, is an important part of machine learning to improve the performance of models. In the first step, we present frequencies of all features by the two levels of Y.

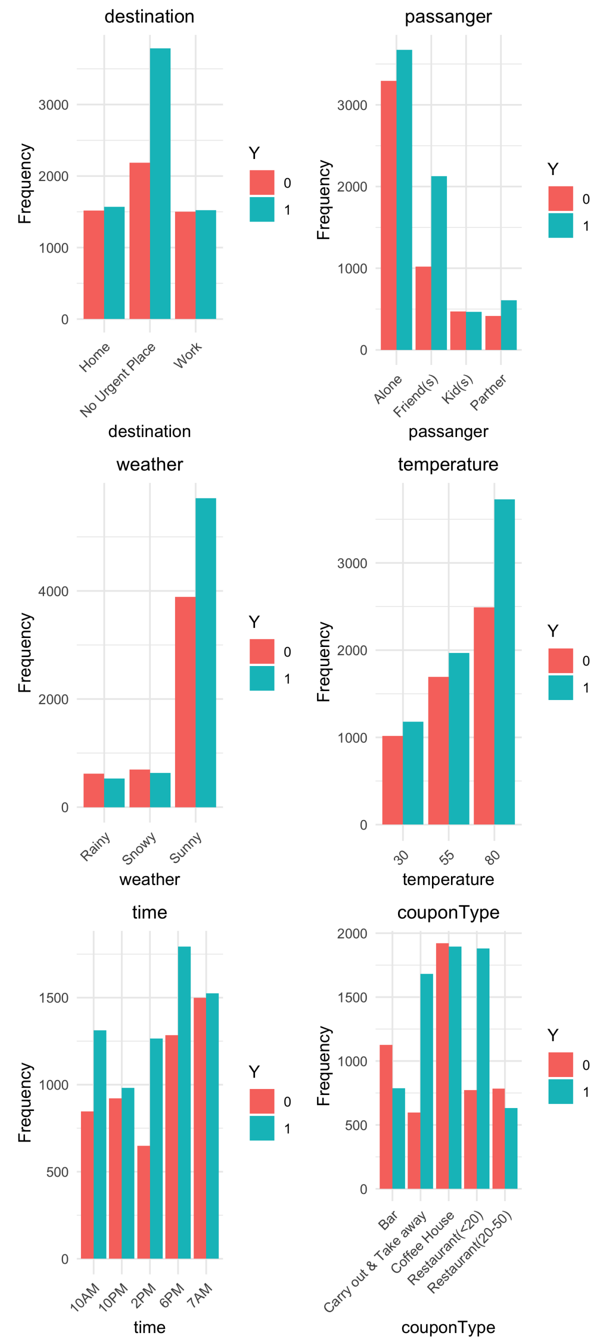


Figure 2-a: the frequencies of levels in features (based on Y)

From **Figure 2-a**, it needs to be mentioned that the 'weather' and 'time' are ordinal. Using label encoding, we converted categorical features to integers to capture the effect of ordinal information by calling the 'Forcats' package.

From **Figure 2-b**, we can see frequencies of different nodes by Y levels in the age feature. To make the result more practical, we divided continuous categories into intervals of 10, known as 'binning'. Here, the ages of '21' and '26' were put into a new interval of '21-30', '31' and '36' were put into '31-40', and '41' and '46' were put into '41-50'. Thus,

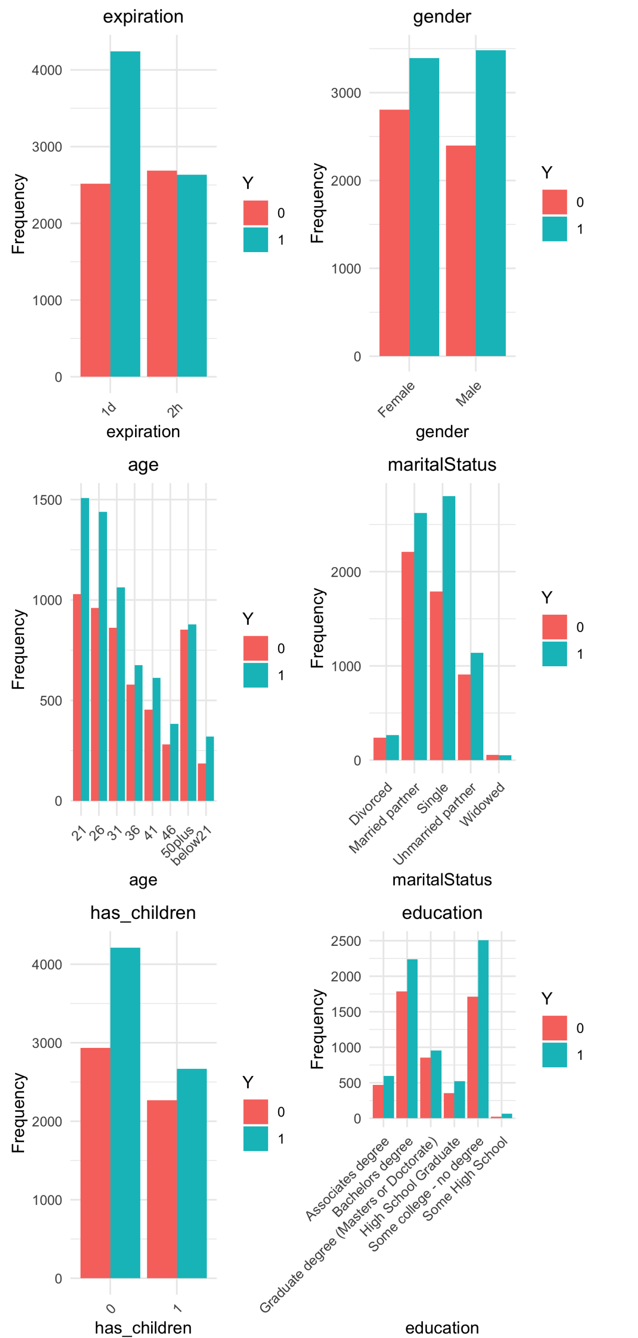


Figure 2-b: the frequencies of levels in features (based on Y)

the new age attribute contains 5 levels: 'below 21', '21-30', '31-40', '41-50', and '50 plus'.

From **Figure 2-c**, we can see the occupation feature is complex. Listing all the main categories for different occupations, this feature contained more than 20 levels, some of which only have unbalanced low frequencies (e.g. some less than 1% whereas some more than 8%). This will induce noise and reduce the performance of models, thus needing dimensionality reduction of feature engineering. One type of High cardinality category processing was considered. instead

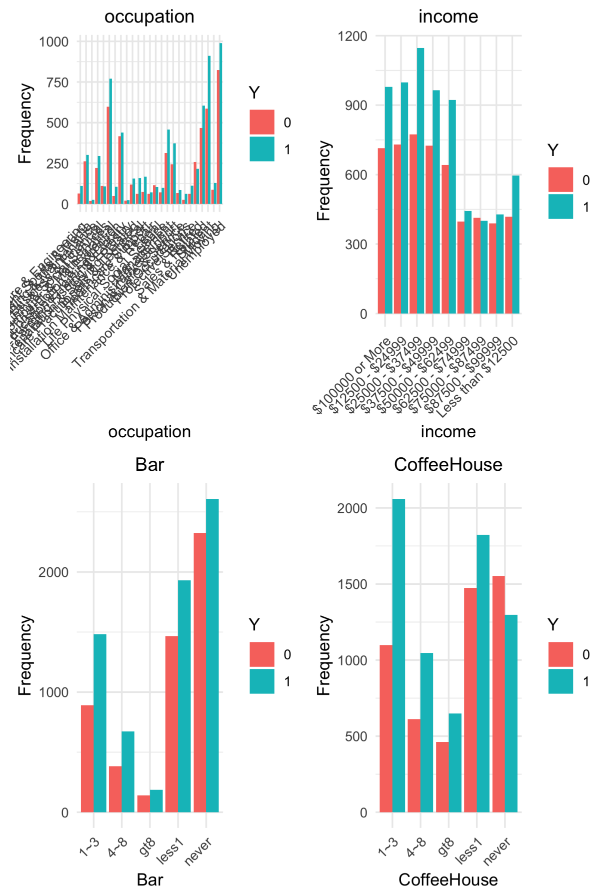


Figure 2-c: the frequencies of levels in features (based on Y)

of showing all the occupations here, we integrated all the occupations into one new level of 'employed'. As a result, the occupation attribute then only contained 4 levels: 'unemployed', 'employed', 'student', and 'retired'. The frequency of each level is shown below.

Table 2: frequencies of levels after dimensionality reduction

level	frequency
unemployed	1814
employed	8295
student	1497
retired	473

Besides, there are 9 levels in the income feature. To reduce model complexity and avoid over-fitting, we referred to the 2024 Current Population Survey Annual Social and Economic Supplements to learn the medium of income in the USA and binned the income into three levels: low (starting from the lowest to %24999), medium (between $25000 and $62499), and high (higher than $62500).

From **Figure 2-d**, it should be highlighted that the features in the habit segment and 'toCoupon' are ordinal. To preserve the ordinal information, we applied label encoding, converting categorical features to integers by using the 'Forcats' package. This step allows us to maintain the inherent order of the data,

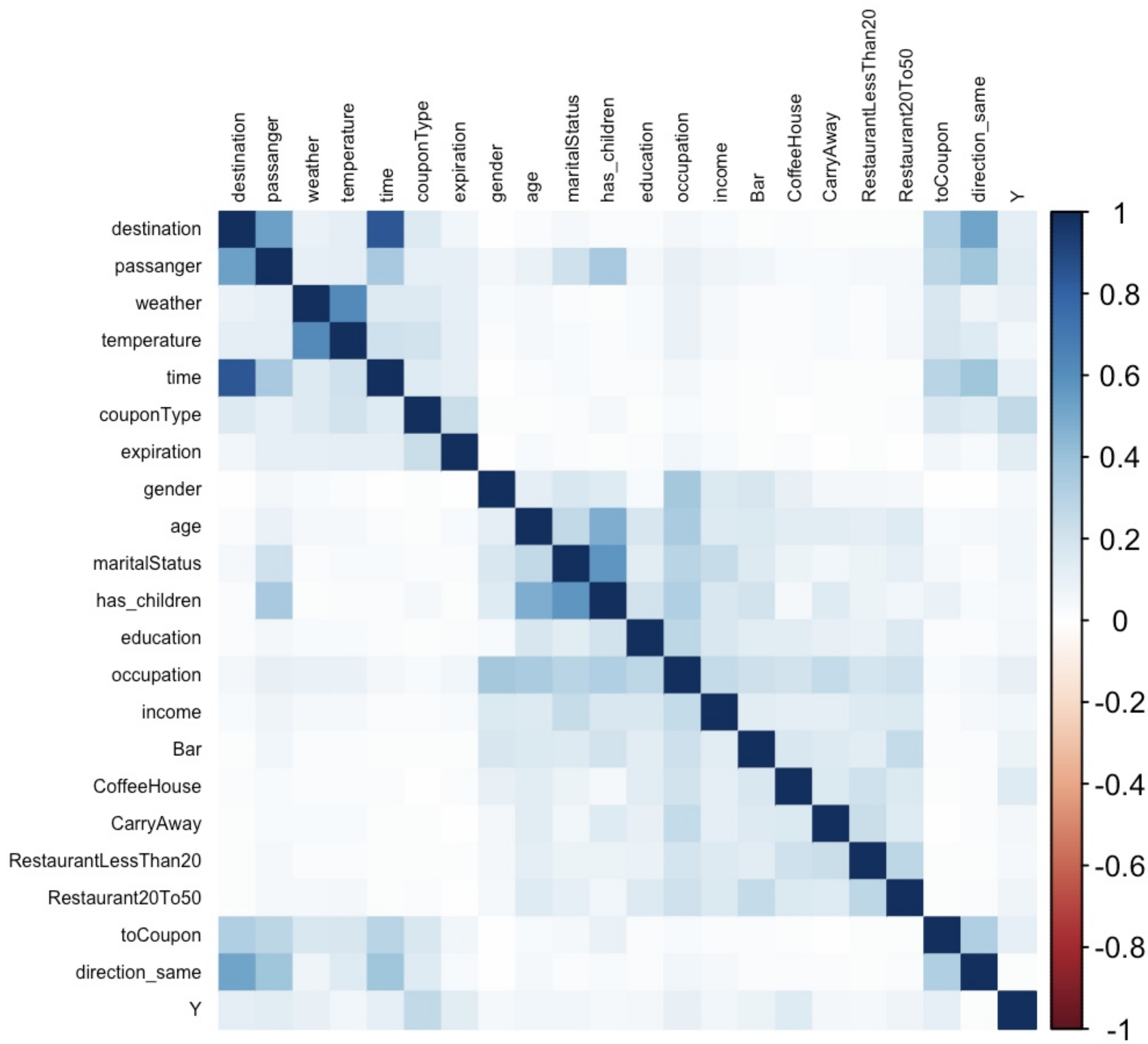


Figure 3: the matrix for Cramer's V Test

which could be significant for the analysis. Additionally, it is worth mentioning that the target binary Y is slightly imbalanced. A moderate imbalance in Y can affect the performance of many predictive models. We will build models and evaluate them to determine whether corrections for this imbalance are needed to enhance performance.

Following a comprehensive review of the data, we performed a Cramer's V test matrix (**Figure 3**) to assess the correlations between all categorical variables. Conducting this test before establishing models has several advantages. First, it helps identify strong correlations among variables. In feature engineering, highly correlated variables can cause redundancy, which in turn increases model complexity and can slow down training, ultimately affecting generalization performance. By identifying correlations, we can filter out redundant or highly related features, thereby retaining the features that best represent the data. Second, the test helps identify and address multi-collinearity, which can have a negative impact on model stability and predictive accuracy. Additionally, it allows for better feature selection by removing meaningless or redundant features, which can improve the overall performance of the model.

The main goals of the Cramer's V test are: a. To identify highly correlated features that might lead to multi-collinearity issues, which can degrade model performance; b. To eliminate redundant features that show little to no relevance to the target

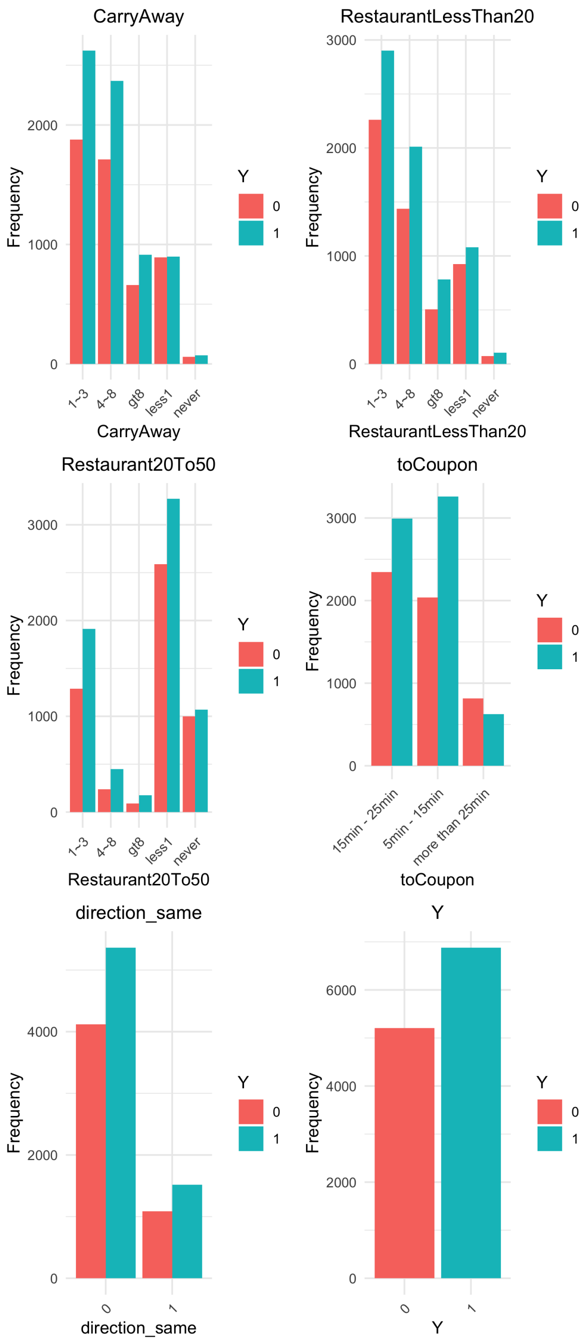


Figure 2-d: the frequencies of levels in features (based on Y)

Y, ensuring the model remains focused on the most informative variables.

To achieve these objectives, we set two specific criteria for selecting features: a. If the correlation between two features exceeds 0.5, they are considered highly correlated and may contribute to multi-collinearity issues. b. If the correlation between a feature and the target Y is less than 0.01, the feature is considered to have minimal relevance to the target and can be removed from the dataset.

Combinations with correlations over 0.5 are:

- destination & time
- destination & passenger
- marital status & has_children
- weather & temperature

Features nearly irrelevant (<0.01) to Y are:

- direction_same
- gender
- Restaurant_Less20

We deleted all three features that were nearly irrelevant to the target Y. For the collinear features with high correlations, we considered several factors when deciding whether to keep or remove them. For example, the destination feature was highly correlated with both time and passenger. In the context of coupon recommendations, promoters can easily observe the customer's companions and know the time, but they do not need to know the customer's potential destination. Based on this logical consideration, we decided to keep time and passenger, and removed the destination feature. Similarly, the combination of marital status and has_children was considered. In practical terms, we opted to keep has_children, as it provides more useful information than marital status. Regarding weather and temperature, the former showed a slightly higher correlation with Y (as indicated by the block color), so we chose to keep weather and remove temperature.

To summarize, in this section we performed data binning, feature integration, feature encoding, and dimensionality reduction. We also conducted a Cramer's V test for feature selection. After completing the feature engineering process, 16 features were selected for the model establishment step III-D.

C. General Pattern Exploration

Previous studies failed to identify general patterns for features and segments due to an over-significance problem. To address this issue, we will perform a Monte-Carlo Simulation Chi-square test in this section. The Monte-Carlo simulation, a numerical method that uses random numbers and statistical models to solve computational or analytical problems, improves the accuracy and stability of results as the number of samples increases, converging toward a theoretical value. Thus, it serves as a robust estimation tool to manage uncertainty.

The core issue causing over-significance in previous studies is the excessively large sample size for conducting a Chi-square test. Statistically, an appropriate sample size should be larger than 40, with all levels having an expected frequency greater than 5. The Monte-Carlo simulation correction works by conducting random sampling and simulating the Chi-square test multiple times. This method calculates the frequencies of results that are significant (i.e., those that influence Y) and reflects the probability that a given feature affects Y.

For this correction, we used R to investigate the effect of each feature on the target Y. Since this simulation aimed to provide a general picture before model building, we included all features without biased data in the analysis. Two features were excluded: 'car', which had 99% missing values, and 'toCoupon_5min', which had only one level. We set the simulation time to 5000 iterations, with a sample size of 100 for each simulation. In each iteration, R randomly selected 100 samples for the Chi-square test and repeated the process 5000 times, calculating the percentage of significant results out of the total.

D. Model Establishing

Previous studies have compared several machine learning models and Bayesian Inference methods. One common finding across two studies was that the Random Forest (RF) model achieved the highest accuracy (0.76) in predictions (Hermawan, Fatihahab, Kurniawatibc, Helen, 2021; Depari, Shu, Fachriza, Chow, Wijaya, Witana, 2022). In the 2017 study, Lasso regression demonstrated the best accuracy (0.74), outperforming the Bayesian Rule Sets model and other tree-based models. The accuracies reported in prior research ranged from 0.67 to 0.76, with computation times varying between less than 1 second and 3 seconds. Given that the dataset is not particularly large, the relatively longer computation time required by the Random Forest model remains acceptable in this context.

Building upon these studies, we aim to select the most effective model and optimize its performance further through feature engineering and parameter tuning. Given its higher accuracy and robustness in handling non-linear datasets, we chose to implement the Random Forest model for optimization in our study. Compared to standard decision tree models, Random Forest reduces the risk of overfitting and sensitivity to data noise by training and integrating multiple trees. Additionally, it can provide feature importance scores, which facilitate the selection and interpretation of the most relevant features.

While leveraging these lessons, we also seek to explore new approaches that are well-suited for classification tasks and may offer even better performance. In datasets with predominantly categorical features, Artificial Neural Networks (ANN) have become a popular choice in recent years. ANN models are capable of approximating complex nonlinear functions through non-linear activation functions and multi-layer neurons, making them particularly well-suited for handling highly nonlinear problems or datasets with intricate relationships between features—relationships that the Random Forest model may not fully capture.

As a result, this section plans to utilize two models to predict the likelihood of accepting a coupon recommendation: a Random Forest model and an Artificial Neural Network model. To evaluate the models' performance, we will split the dataset into 70% training data and 30% testing data.

The algorithm we used for the RF model was developed by Xi and Hemant (2012). There are 4 steps for this algorithm: a. Draw ntree bootstrap samples from the initial data set. b. Grow a tree for each bootstrapped data set, then randomly mtry variables at each node of the tree to split. Grow into a tree until the terminal nodes do not hold fewer than nodesize cases. c. Aggregate the information from other ntree trees to predict the new data the majority vote for classifying the data. d. Calculate an out-of-the-bag error rate using the data not contained in the bootstrap sample. When we started training the random forest model, we used some evaluations to decide how many trees would be needed in terms of error rates corresponding to a varying number of trees in the ensemble (**Figure 4**). The corresponding error rates-Out-of-Bag (OOB) error and class-specific errors-help evaluate the generalization ability, stability, and behavior of the model on the multiple class predictions.

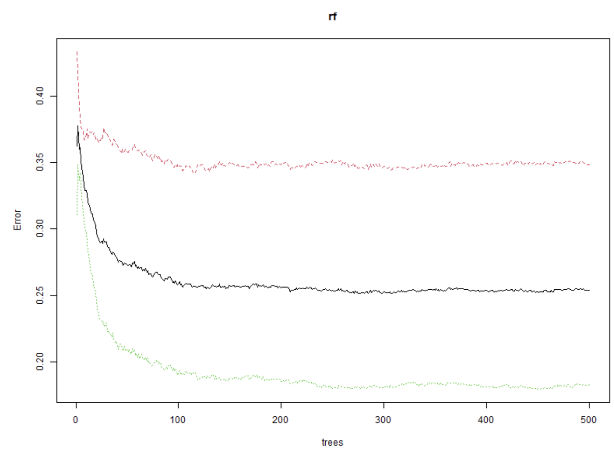


Figure 4: Error rates at different numbers of trees. (Description: Black line: Represents the Overall Out-of-Bag (OOB) Error Rate. Red Line: Represents the Error Rate for Class 0. Green Line: Represents the Error Rate for Class 1.)

Figure 4 shows that all error rates are high since it has only few trees and makes the predictions inefficient. The inclusion of more trees continues to reduce error rates until a point at which error rates always remain relatively constant, thus establishing a level of prediction performance. At approximately 200 – 300 trees error rates appear stabilized, inferring that increasing the number of trees beyond this threshold does not greatly improve performance. Then, we decided to use 300 trees (ntree=300) and a maximum of 5 features (mtry=5) considered at each split, and the model's training, prediction, and evaluation were conducted using a split dataset.

The algorithm we used for the ANN model, Multi-Layer Perceptron Classifier, was developed and maintained by the open source community of scikit-learn. For this algorithm: a. Initialize network parameters and build the number of layers and the number of neurons in each layer of the multi-layer perception. b. Compute linear combinations in each layer then activate functions, while generating predicted probabilities in the last layer. c. calculate the cross-entropy loss. d. Calculate the gradient of each layer through the chain rule and update weights and biases using an optimizer. e. Repeatedly perform forward propagation, loss calculation, back propagation and parameter update until maximum number of iterations is reached or the loss function changes less than the set tolerance. When we started training the ANN model, we used only 1

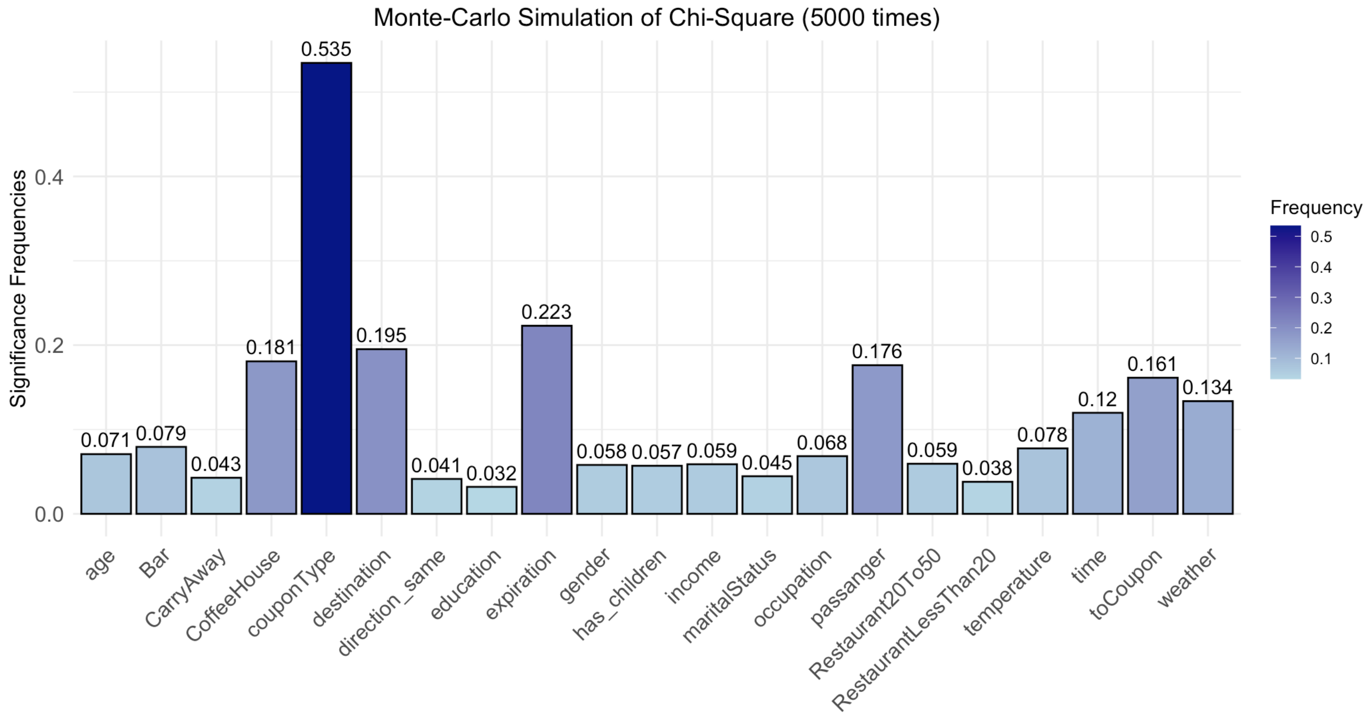


Figure 5-a: the effect of features on the target Y

layer (considering the data size) with 2500 nodes, a maximum of 1000 iteration and a tolerance of 10^{-6}.

IV. Evaluation

A. MonteCarlo Simulation

Figure 5-a shows the heat map result, where higher frequencies show heavy blue and lower frequencies show light blue. Among the features, the most influential factor on the dependent variable is Coupon Type, followed by expiration and Coffee House, while the least one is education.

Following suggestions from peer review, **Figure 5-b** categorizes the features into 5 main segments, with features in the same segment sharing the same color, to assess whether the segment itself plays a significant role. It is evident that the coupon information segment plays the most significant role in influencing a customer's if accept decision. This is likely because, from a consumer psychology perspective, promotional offers, especially those tied to time-sensitive benefits like coupons, create a sense of urgency and scarcity, which can spur immediate action. Customers are more likely to make decisions impulsively when they feel they might miss out on an opportunity, a principle rooted in the "fear of missing out" (FOMO) theory. This theory is particularly relevant in this context, as coupons with clear expiration dates trigger an immediate need to act.

Following coupon information, environmental factors such as the customer's destination, passengers, weather, time of day, and the distance to the restaurant are also influential. This can be explained through the concept of situational influence in consumer behavior, where external factors, such as time constraints or environmental comfort, directly impact decision-making. For instance, if a customer is already near the restaurant, they may be more inclined to accept the coupon due to convenience, especially when weather conditions are favorable, or they are accompanied by passengers who might increase the likelihood of participation.

Only one feature in the habit segment is influential, indicating that drinking coffee is a preference for a specific demographic group. This tendency might explain why, among coffee drinkers, accepting a coupon is almost a reflex action—offering them the opportunity to grab a cup, regardless of other factors. Besides, surprisingly, customer demographics did not significantly influence coupon acceptance in this case, nor did their habits of visiting other types of restaurants beyond coffeehouses. This highlights a crucial aspect of consumer psychology: the increasing irrelevance of demographic-based profiling in certain marketing strategies. As modern consumers are more driven by immediate benefits and contextual relevance, targeting specific traits like age or marital status may not be as effective as targeting situational factors that directly influence decision-making at the moment of interaction.

In summary, this section provided a broader overview, identifying key factors that influence a customer's decision to accept a coupon. We observed that customer demographics have little relevance in this promotional context. Therefore, promoters should focus on offering appealing coupons and timing their promotions effectively, rather than investing excessive effort in user profiling.

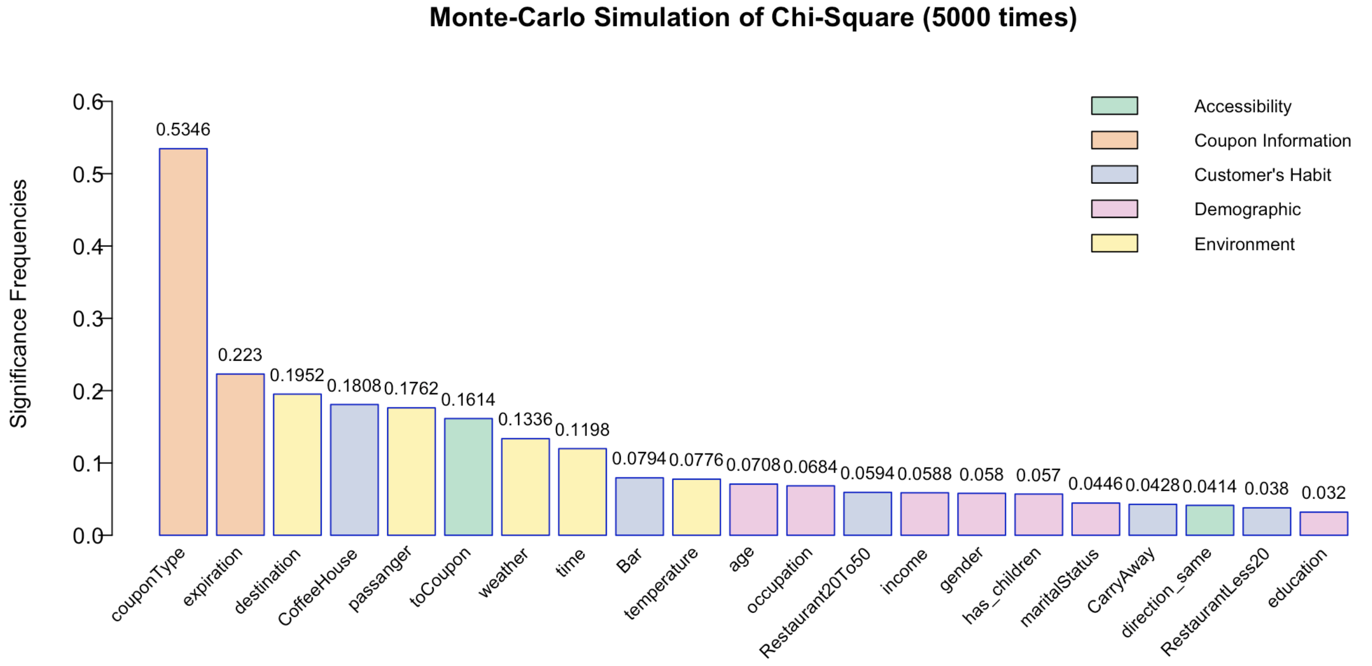


Figure 5-b: the effect of segments on the target Y

B. the Random Forest model

The Random Forest model's performance is evaluated on a binary classification task using metrics such as accuracy, precision, recall, and F1 score. The results are summarized in **Table 3**.

Table 3: Random Forest Performance Metrics

Metric	0	1
Prediction	0.76	0.78
Recall	0.68	0.84
F1 score	0.72	0.81
Overall Accuracy		0.77

In this table, we can see that predictions for Class 1 are more reliable in terms of correctness from precision results. From recall results, only 68% of the true Class 0 instances were correctly identified while 84% of true Class 1 instances were successfully detected. The F1 scores also demonstrated stronger overall performance for Class 1 predictions.

The confusion matrix is shown in Table 4, providing a breakdown of the model's predictions compared to the actual labels. The model performs well overall but shows some room for improvement in reducing false negatives for 0 and false positives for 1.

Table 4: Confusion matrix for Random Forest Performance

Reference	Actual 0	Actual 1
Prediction 0	1060	334
Prediction 1	500	1729

The feature importance plot (**Figure 6**) reveals which variables contributed most to the model's predictive power. The top features are: couponType, CoffeeHouse, Bar, and expiration. Coupon type is the most influential feature, as indicated by its highest mean decrease in accuracy. This suggests that the type of coupon significantly impacts the model's predictions. Other features such as Coffee House and Bar also play substantial roles, reflecting potential habit patterns in the data.

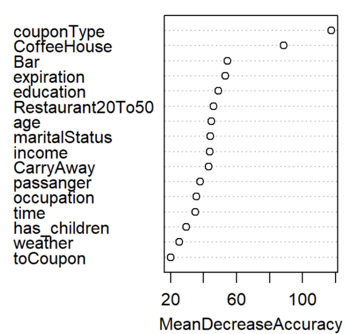


Figure 6: Important features in the random forest model

To summarize, the Random Forest model performs particularly well on Class 1 ('not accept') predictions, while performance on Class 0 ('accept') is comparatively weaker.

C. the ANN model

Under the configuration with maximum iterations of 1,000 and tolerance of 10^{-6}, the ANN training model stopped after 129 iterations because the loss did not improve for 10 consecutive epochs, indicating convergence. The final loss value at iteration 129 was approximately 0.0319. After all, the ANN achieved an accuracy of 76%. Precision and recall values suggest that the model performs better on Class 1, where recall reaches 0.81 and the F1 score is 0.79. For Class 0, the recall is lower at 0.68, highlighting that some true negatives were misclassified as positives. The performance metrics are shown in **Table 5**.

Table 5: ANN Performance Metrics

Metric	0	1
Prediction	0.73	0.78
Recall	0.68	0.81
F1 score	0.70	0.79
Support	1021	1395

When comparing its results with that of random forest model, we found that both models achieved similar overall accuracy (76%-77%). Random Forest demonstrated slightly better recall for Class 0, while ANN performed comparably on Class 1. Besides, ANN required significantly more computational resources due to the large number of nodes (2,500) in the hidden layer and the iterative training process.

To summarize, the ANN model achieved an overall accuracy of 76%. However, similar to Random Forest, improving recall for Class 0 remains a potential area for enhancement.

V. Conclusion

To sum up, this study establishes a general and practical pattern of the coupon recommendation case, achieving both efficiency and improved model performance. For the customer psychology, we learn that the most important factor influencing if they will accept a coupon is the type of the coupon and its expiration time. This indicates that, before starting to do the promotion in different forms, it is better to find an appropriate way to promote it. As to coupon types for five different restaurants, coupons for some restaurants are more acceptable than others. For example, here in this study, coupons for bar, coffee house and expensive restaurant will be more acceptable and inspire customers to take action, whereas coupons for carry-away or low-price restaurant relatively don't work. Results also show that some segments can be considered less when doing this type of promoting, especially the demographics of the customers. Almost all the factors in demographic section have nothing to do with coupon acceptance, indicating that promoters can stop asking information from potential consumers, which seems to be irritating and useless. We also made great contributions to the model establishing. By reducing the number of input features from 25 to 16, the model maintained strong predictive performance with a slight increase in accuracy from 0.76 to 0.77. This result highlights the importance of effective feature selection in removing redundant or irrelevant data while maintaining accuracy. A comparison is made between the proven-best model (random forest) in previous studies and a new ANN model, and verifies both ability for prediction.

However, the model's handling of missing values relied on basic strategies such as filling in or deleting records, which may not fully optimize data quality. Advanced imputation methods, including machine learning-based techniques or multiple imputations, should be explored in future work to address this limitation. Additionally, a direct comparison of the reduced feature set with the original dataset would provide a more complete understanding of the trade-offs between feature reduction and accuracy. The study also reveals limitations in handling the moderately imbalanced target Y, where the "accept" class (5,203 instances) is underrepresented compared to the "not accept" class (6,877 instances). Both Random Forest and Artificial Neural Networks (ANN) exhibited better predictive performance for Y=1, likely due to this imbalance. To address this, future work should incorporate techniques such as SMOTE (Synthetic Minority Over-sampling Technique) to balance the dataset or adjust decision thresholds using ROC curve analysis. Such methods can improve the sensitivity for the minority class and ensure a more balanced prediction across both categories. Addressing class imbalance is particularly critical in applications where false negatives can carry significant consequences.

Based on these findings, we propose two directions for future improvement. a. Comparing model performance using the reduced 16 features versus the full 25 features would clarify the benefits of feature selection and computational efficiency. b. Refining the missing value strategy and integrating threshold adjustment techniques could further optimize performance. By addressing these limitations, the proposed framework can become a stronger and more generalizable approach to solving binary classification problems.

References

[1] Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of pharmaceutical and biomedical analysis*, 22(5), 717-727.

[2] Aggarwal, C. (2015). Data Mining The Text Book.

[3] Barnard. (1963). Discussion on the spectral analysis of point processes.

[4] D. R. Hermawan, M. Fahrio Ghanial Fatihah, L. Kurniawati and A. Helen. (2021). Comparative Study of J48 Decision Tree Classification Algorithm, Random Tree, and Random Forest on In-Vehicle Coupon Recommendation Data, *2021 International Conference on Artificial Intelligence and Big Data Analytics*, Bandung, Indonesia, pp. 1-6.

[5] Genesis Sembiring Depari, Efin Shu, Cut Alya Fachriza, Joceline Chow, Jerica Wijaya, & Ryanto Winata. (2022). Customer's Responses Towards In-Vehicle Coupon Recommendation: An Implemetation of Business Analytics. *Jurnal Ekonomi*, 11(02), 1157–1167.

[6] Ingre, B., Yadav, A. (2015, January). Performance analysis of NSL-KDD dataset using ANN. *In 2015 international conference on signal processing and communication engineering systems* (pp. 92-96). IEEE.

[7] Lin, M., Lucas, H. C., and Schmueli, G. (2013). Too Big to Fail: Large Samples and the P-Value Problem, *Information Systems Research*(24:4), pp. 906-917. *Journal of the Royal Statistical Society: Series B (Methodological)*, 25(3), 294-296.

[8] Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507-514.

[9] T. Wang, C. Rudin, F. Velez-Doshi, Y. Liu, E. Klampfl and P. MacNeille. (2017). Bayesian Rule Sets for Interpretable Classification, *2016 IEEE 16th International Conference on Data Mining (ICDM)*, Barcelona, Spain, pp. 1269-1274.

[10] Wu, B., Zhang, L., Zhao, Y. (2013). Feature selection via Cramer's V-test discretization for remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 52(5), 2593-2606.

[11] Xi Chen, Hemant Ishwaran. (2012). Random forests for genomic data analysis. *Genomics*, 99(6), 323-329.