AUTOMOTIVE COUPON RECOMMENDATION SYSTEM

A PROJECT REPORT

Submitted by:

Yash Bhargava - 221B456

Prathamesh Kale - 221B497

Ritik Prakash - 221B498

Under the guidance of: Dr. Rahul Pachauri

November 2024

Submitted in partial fulfilment of the Degree

of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

Department of Computer Science & Engineering

JAYPEE UNIVERSITY OF ENGINEERING & TECHNOLOGY, A-B ROAD, RAGHOGARH, DT. GUNA - 473226, M.P., INDIA

Declaration by the Student

I hereby declare that the work reported in the B. Tech. project entitled as "**Automotive Coupon Recommendation System**", in partial fulfillment for the award of degree of **B. Tech Computer Science and Engineering** submitted at Jaypee University of Engineering and Technology, Guna, as per best of my knowledge and belief there is no infringement of intellectual property right and copyright. In case of any violation, I will solely be responsible.

Yash Bhargava - 221B456

Prathamesh Kale - 221B497

Ritik Prakash - 221B498

Department of Computer Science and Engineering

Jaypee University of Engineering and Technology

Guna. M.P., India

Date:20/11/2024

विद्या तत्व ज्योतिसम:

JAYPEE UNIVERSITY OF ENGINEERING &TECHNOLOGY

Grade 'A+' Accredited with by NAAC & Approved U/S 2(f) of the UGC Act, 1956

A.B. Road, Raghogarh, Dist: Guna (M.P.) India, Pin-473226

Phone: 07544 267051, 267310-14, Fax: 07544 267011

Website: www.juet.ac.in

CERTIFICATE

This is to certify that the work titled **Automotive Coupon Recommendation System** submitted by **Yash Bhargava**, **Prathamesh Kale, Ritik Prakash,** in partial fulfillment for the award of degree of **B. Tech Computer Science and Engineering** of Jaypee University of Engineering & Technology, Guna has been carried out under my supervision. As per best of my knowledge and belief there is no infringement of intellectual property right and copyright. Also, this work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma. In case of any violation concern student will solely be responsible.

Signature of Supervisor

Dr. Rahul Pachauri

Dept. of CSE (Associate Professor)

Date:

Acknowledgment

We would like to express our gratitude and appreciation to all those who gave us the opportunity to complete this project. Special thanks to our supervisor **Dr. Rahul Pachauri** whose help, stimulating suggestions and encouragement helped us in all the time of development process and in writing this report. We also sincerely thanks for the time spent proofreading and correcting my many mistakes. We would also like to thank our parents and friends who helped us a lot in finalizing this project within the limited time frame.

Thanking you

Yash Bhargava

Prathamesh Kale

Ritik Prakash

Summary

The Automotive Coupon Recommendation System aims to enhance customer engagement and retention rate by delivering personalized coupon recommendations for food outlets. The system leverages data collected from various sources, including customer demographics, purchasing history, location, time of travel, and preferences. This data is further enriched with contextual factors such as traffic conditions, vehicle routes, and customer behavior patterns to create a understanding of user needs.

Our approach involves advanced machine learning algorithms to analyze this data and identify patterns that link customer profiles with their preferred food outlets and offers. The system clusters customers based on similar traits and behaviors, enabling targeted recommendations. Additionally, location-based services ensure real-time coupon suggestions, taking into account the proximity of food outlets and customer travel routes.

The recommendation system also benefits food outlet partners by offering insights into customer preferences, enabling them to tailor offers more effectively. A feedback loop from redeemed coupons and user responses is incorporated to improve the recommendation accuracy over time.

LIST OF FIGURES

Table Of Contents

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

This project is to describe about the customer retention and engagement are critical challenges for businesses, especially in highly competitive industries like food outlets and automotive services. Traditional coupon distribution methods, such as mass emails or generic offers, often fail to target the right audience, leading to wasted marketing efforts and low customer satisfaction. This calls for a more personalized approach that takes into account customer behavior, preferences, and contextual factors to ensure that the right coupons reach the right customers at the right time.

The Automotive Coupon Recommendation System addresses these challenges by utilizing data-driven methods to deliver personalized recommendations. By analyzing customer demographics, purchasing history, and preferences, the system gains insights into user behavior. It further incorporates contextual data such as time of travel, traffic conditions, and vehicle routes to enhance recommendation accuracy. This holistic approach ensures that the coupons suggested are relevant to the customers' needs and situational contexts, thereby increasing the likelihood of redemption and customer satisfaction.

Ultimately, this project aims to revolutionize the coupon recommendation process by combining advanced analytics and contextual awareness.

By offering timely and personalized deals, the system enhances customer engagement, boosts retention rates for food outlets, and improves the overall user experience. Additionally, businesses benefit from optimized marketing strategies, increased customer loyalty, and a stronger competitive edge in the market.

1.2 Project Overview

The **Automotive Coupon Recommendation System** is designed to revolutionize the way businesses engage with customers by delivering personalized and context-aware coupon recommendations. The system focuses on connecting automotive travelers with food outlets and other relevant businesses by offering tailored deals based on their individual preferences, travel patterns, and real-time contextual factors. It leverages a blend of demographic data, purchasing history, and behavioural insights, enriched with contextual elements like traffic conditions, vehicle routes, and time of travel, to create a dynamic recommendation engine that ensures customers receive offers that are timely and relevant to their needs.

The project underscores the growing significance of **personalized marketing** in today's business landscape. Unlike traditional approaches, which often involve generic advertisements and promotions, personalized marketing uses data analytics to understand customer preferences and deliver customized offers. This targeted strategy not only enhances the customer experience but also significantly boosts conversion rates and sales. When customers feel that promotions cater specifically to their needs and lifestyles, they are more likely to engage with the brand, make purchases, and remain loyal over time.

By implementing the Automotive Coupon Recommendation System, businesses stand to gain a competitive edge in the market. Personalized marketing fosters stronger customer relationships, builds trust, and increases brand affinity, which ultimately translates into higher sales and revenue. Furthermore, by offering timely and relevant coupons, businesses can optimize resource allocation, reduce marketing waste, and achieve better return on investment (ROI). This project not only transforms the coupon distribution process but also establishes a new standard for leveraging data-driven insights to enhance customer satisfaction and business growth.

1.2.1 Collaborative Filtering:

- Uses customer purchase history and preferences to recommend coupons based on similar users' behavior.

- Example: If a user frequently redeems coupons for coffee shops, the system might recommend other coffee outlets popular among similar users.

1.2.2 Content-Based Filtering:

•Analyzes the attributes of coupons, such as cuisine type, location, or discount rate, to recommend similar offers based on the user's past preferences.

•Example: A user who redeemed coupons for fast-food outlets might receive recommendations for similar establishments.

1.2.3 Context-Aware Recommendations:

•Integrates contextual factors like time of day, current location, traffic conditions, and vehicle routes.

•Example: During rush hour, the system may recommend nearby outlets with quick service to accommodate the user's travel constraints.

1.2.4 Hybrid Recommendation System:

•Combines collaborative, content-based, and context-aware techniques for improved accuracy and relevance.

•Example: Using both user preferences and current travel context to suggest coupons tailored to the situation.

1.3 Scope Of The Project

1.**Data Preprocessing**: Clean, normalize, and analyze data to uncover relationships between user preferences, travel context, and coupon redemption.

2.**Feature Engineering**: Create and refine key attributes like demographics, travel patterns, and outlet proximity to improve recommendations.

3.**Model Development**: Implement and compare models (e.g., collaborative filtering, content-based, hybrid) to identify the best approach.

4.**Evaluation and Optimization**: Use metrics (e.g., precision, recall) to assess performance and optimize models via hyperparameter tuning.

5.**Deployment**: Design real-time or batch systems for stakeholders, integrating APIs or user interfaces for seamless application.

These tailored points ensure that the Automotive Coupon Recommendation System aligns with best practices in machine learning projects while addressing the specific needs of the application.

1.4 Software and Hardware Requirements

Software Requirements:

1. **Programming Language**:
   - **Python (Version 3.7 or higher)**: Offers extensive libraries for data processing, machine learning, and real-time applications.
2. **Integrated Development Environment (IDE)**:
   - **Jupyter Notebook**: For exploratory analysis and model prototyping.
   - **Google Colab**: For cloud-based development, especially useful for handling larger datasets.
3. **Libraries and Packages**:
   - **Pandas and NumPy**: For efficient data manipulation and analysis.
   - **Scikit-Learn**: For implementing recommendation algorithms and model evaluation.
   - **LightFM/Surprise**: For collaborative filtering and hybrid recommendation system development.
   - **Matplotlib and Seaborn**: To visualize user preferences, coupon usage patterns, and model insights.
   - **APIs (e.g., Google Maps API)**: To integrate real-time location and traffic data.

4. **Version Control**:
   - **Git/GitHub**: For version management and collaboration during system development.

Hardware Requirements:

Processor: Minimum **Intel Core i5** or equivalent; for larger datasets and advanced context-aware models, **Intel Core i7/i9** or **AMD Ryzen 5/7/9** is recommended.

RAM: Minimum **8 GB**, **16 GB or higher** is preferred for handling large-scale user and contextual data efficiently.

Storage: At least **100 GB** of free space to store datasets, recommendation models, and log files.

GPU (Optional): An **NVIDIA GPU with CUDA support** can accelerate model training, especially for large datasets or real-time recommendation systems.

Chapter-2

LITERATURE SURVEY

2.1 Overview of Automotive Coupon Recommendations

In the automotive industry, personalized coupon recommendations are influenced by multiple factors, making it a prime area for machine learning-based systems. Key determinants include user demographics, preferences, location, travel behavior, and contextual data such as traffic and time of day. These attributes collectively shape how recommendations are personalized and delivered to users.

1. **User Preferences**:
   - Historical coupon usage patterns and food preferences help tailor recommendations to individual users.
   - For instance, users frequently redeeming discounts for fast-food chains may prioritize similar options.
2. **Location and Travel Context**:
   - Proximity to outlets and user travel routes are critical factors for recommending relevant coupons.
   - Recommendations can adapt dynamically to current user locations, traffic conditions, and time constraints.
3. **Temporal Factors**:
   - Time of day and seasonal trends impact recommendation strategies. Breakfast coupons are suggested in the morning, while dinner deals are promoted in the evening.

By integrating these attributes with historical data on coupon redemptions, machine learning models can predict and provide highly relevant recommendations in real-time.

2.2 Machine Learning in Coupon Recommendation

Machine learning (ML) offers robust capabilities for analyzing user behavior and contextual factors, enabling accurate and timely coupon recommendations. Its ability to manage multidimensional data and model complex relationships makes it ideal for recommendation systems.

1. **Supervised Learning Techniques in Recommendations**:

   - **Collaborative Filtering**: Predicts user preferences based on similar users' behavior.
   - **Content-Based Filtering**: Recommends coupons by analyzing their attributes (e.g., outlet type, discount rate).
   - **Hybrid Models**: Combines multiple approaches for more accurate and comprehensive recommendations.
2. **Ensemble Methods**:
   - Models like Random Forest and Gradient Boosting (e.g., XGBoost) are well-suited for capturing interactions among user preferences, travel patterns, and contextual data.
   - Ensemble methods improve model accuracy and robustness by integrating insights from diverse data points.
3. **Applications in Coupon Systems**:
   - ML models leverage user preferences, travel behavior, and real-time data to suggest relevant deals.
   - For example, during rush hour, the system may recommend coupons for quick-service restaurants near high-traffic areas.
4. **Advantages and Limitations**:
   - ML models automate and personalize recommendations at scale but may face challenges such as cold-start issues for new users and dependency on high-quality data.

By incorporating machine learning, coupon recommendation systems can deliver precise, user-centric suggestions, enhancing both user experience and business outcomes.

2.3 Key Features Impacting Coupon Recommendations

In the context of automotive coupon systems, several factors influence the relevance and effectiveness of recommendations. Machine learning models utilize these features to optimize predictions and ensure meaningful results.

1. **User Demographics and Preferences**:
   - Factors like age, income, and prior coupon usage patterns play a significant role in shaping recommendations.
   - For example, younger users might favor discounts for trendy coffee shops, while families may prefer deals on casual dining.
2. **Location and Route Data**:
   - Real-time geolocation and travel routes are pivotal for contextual recommendations.
   - Recommendations align with nearby outlets along the user's path, ensuring convenience.
3. **Time of Day and Context**:
   - Temporal features (morning, afternoon, evening) are key for tailoring offers, such as breakfast deals in the morning or dinner promotions in the evening.
4. **Traffic and Environment**:
   - Traffic congestion or weather conditions can guide the type of recommendations provided.
   - For instance, suggesting coupons for restaurants with ample parking during high traffic.
5. **Historical Behavior and Trends**:
   - Patterns in coupon redemption, seasonality, and user activity inform model predictions.
   - For example, increased redemptions for ice cream coupons during summer months.

By understanding and leveraging these features, the system ensures that recommendations are both timely and relevant, fostering user satisfaction and business growth.

2.4 Context and Scenario-Based Coupon Recommendations

Recent studies emphasize the importance of leveraging real-world contextual factors to enhance the effectiveness of recommendation systems. A significant contribution to this area is the use of **context-aware recommendation frameworks** that align marketing efforts with optimal user scenarios.

One such example is the framework proposed for in-vehicle coupon recommendations, which utilizes machine learning to determine the conditions under which users are most likely to redeem a coupon. This framework, supported by data collected through surveys, focuses on understanding factors such as time of day, weather, destination, and passenger presence, and their impact on user decision-making.

1. **Relevance of Contextual Features**:
   - Features like time (e.g., morning vs. evening), weather (sunny vs. rainy), and driving scenarios (with or without passengers) are critical in predicting user behavior.
   - For instance, a user may be more inclined to accept a coffee shop coupon during a morning commute than during an evening outing.
2. **Optimal Marketing Conditions**:
   - The study highlights the role of micro-factors in driving user engagement.
   - For example, insights from such frameworks can inform businesses like Starbucks to determine whether to push coupons during favorable conditions (e.g., sunny mornings) to maximize redemption rates.
3. **Data Collection and Insights**:
   - The dataset, gathered via Amazon Mechanical Turk, showcases realistic driving scenarios to analyze user behavior effectively.
   - By exploring these scenarios, the framework identifies trends and patterns that marketers can use to enhance campaign strategies.
4. **Practical Applications**:
   - This approach is beneficial for food outlets, retail chains, and marketing agencies seeking to improve customer engagement and optimize advertising costs.
   - Businesses can leverage these insights to tailor coupon distribution strategies that align with user preferences and situational factors, leading to increased sales and a competitive edge.

Chapter-3

REQUIREMENT AND ANALYSIS

3.1 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.

2. **Day Attributes**: Afternoons, sunny days and moderate temperatures around 80 degrees work best for acceptance of coupon.

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

4. **Propensity To Eat Out**: Higher likelihood of acceptance when the person generally eats out more often.

5. **Coupon Attributes**: Carry out and budget restaurant coupons have higher chance of acceptance.

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
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%

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

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%

Figure 3.1: EDA on Demographics variables

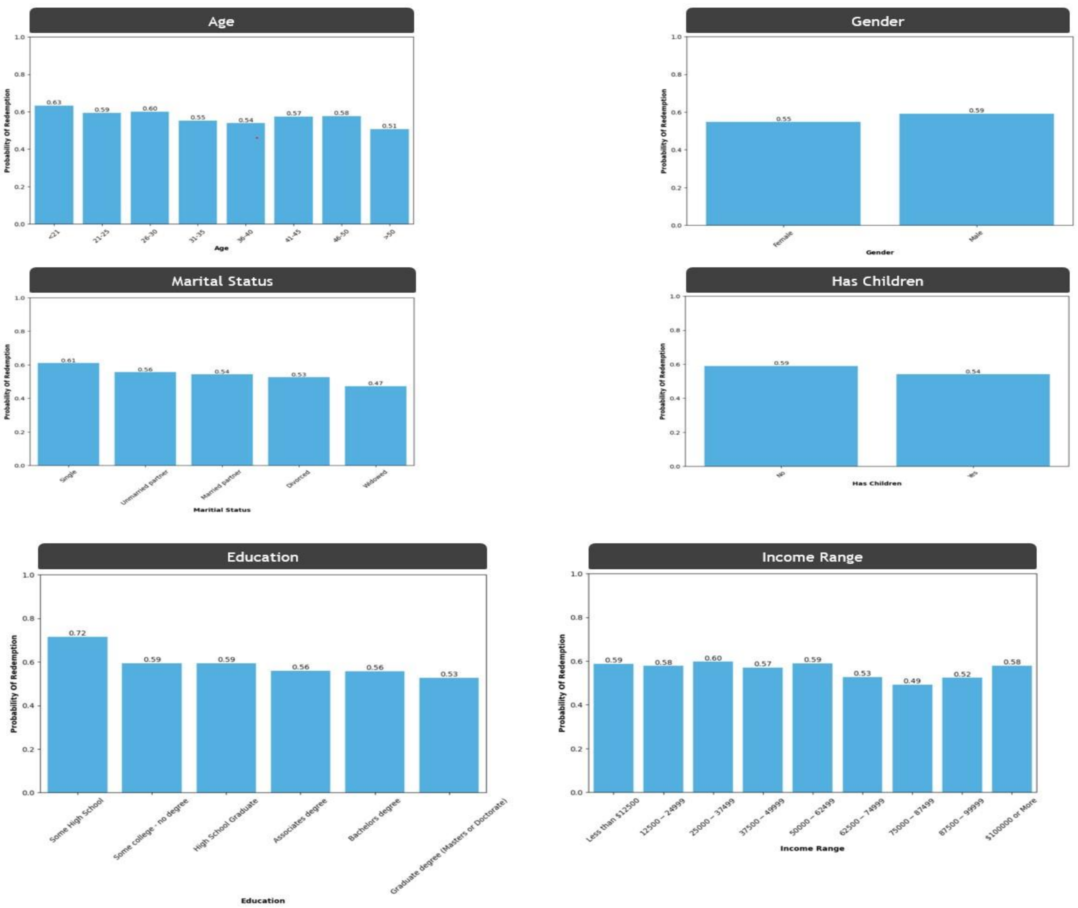


Figure 3.2: EDA on Day Attributes

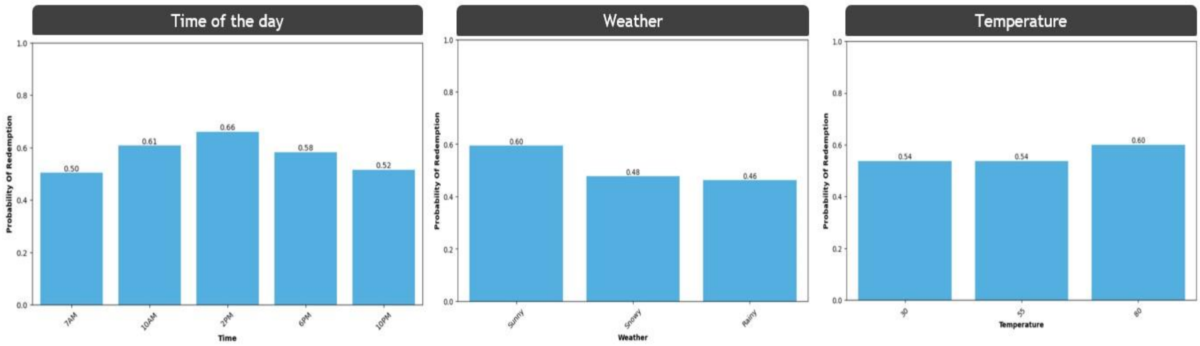


Figure 3.3: EDA on Journey Attributes

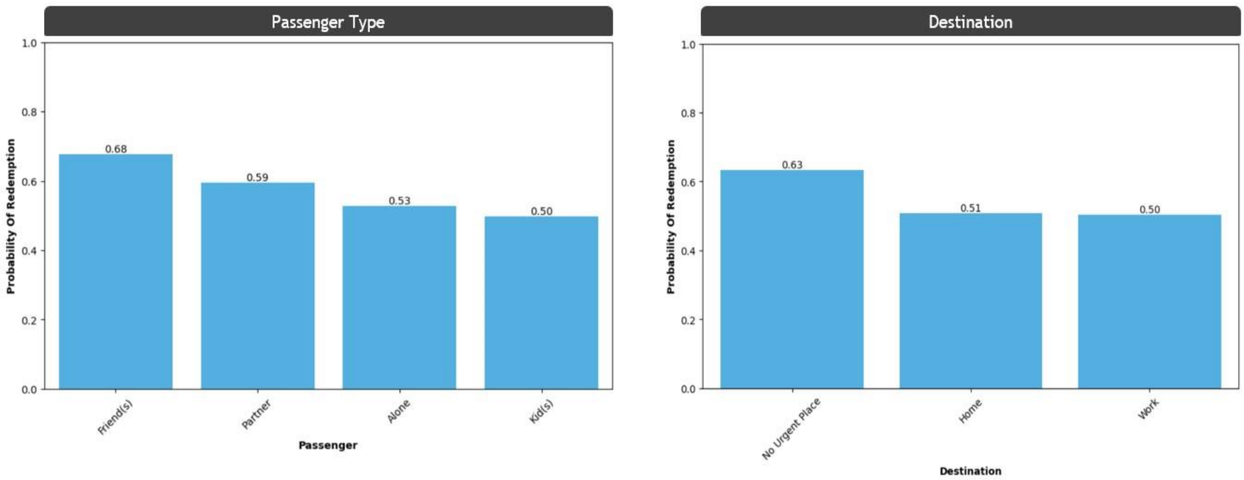


Figure 3.4: EDA on Propensity to eat out

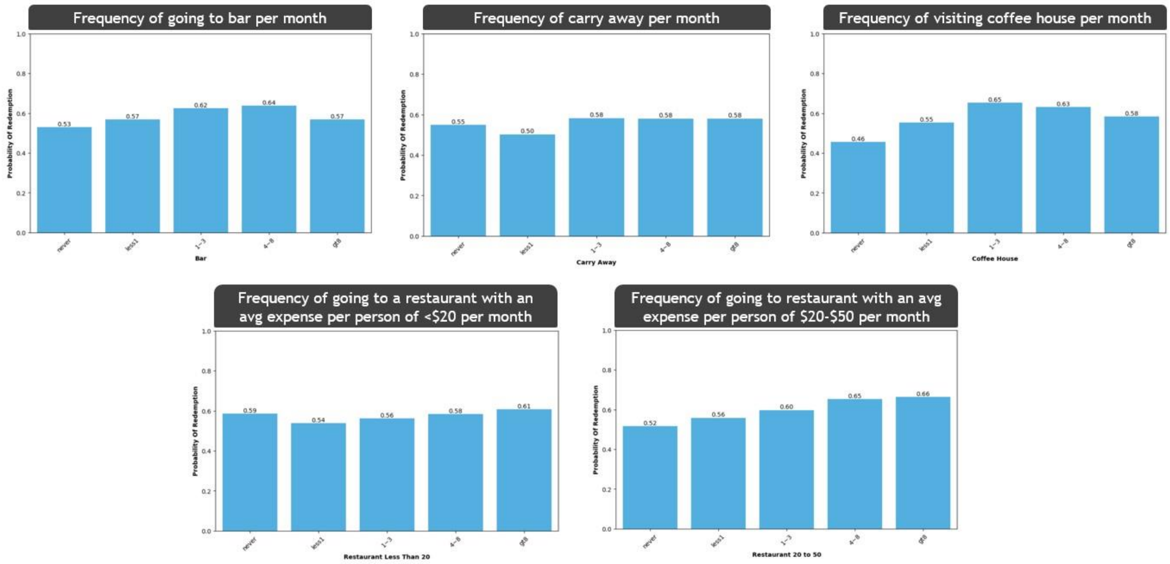
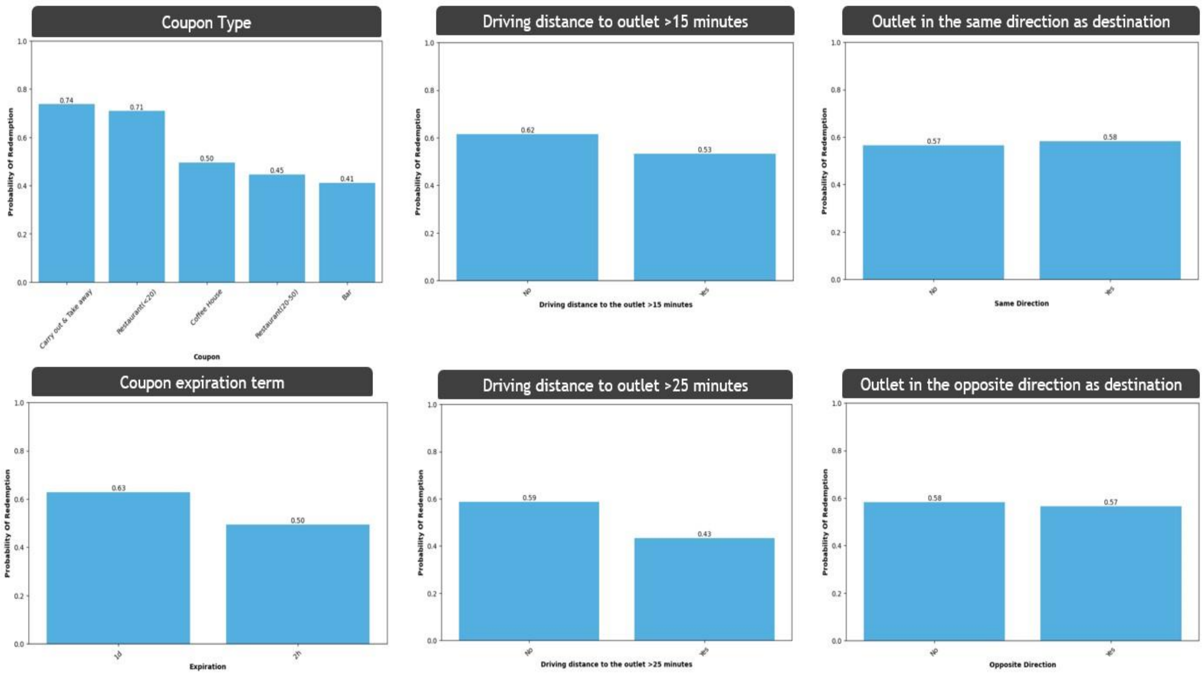


Figure 3.5: EDA on Coupon Attributes



3.2 Data Preprocessing for Automotive Coupon Recommendation

Data preprocessing is a critical step in the development of an effective coupon recommendation system, ensuring the dataset is prepared for analysis and model training. This process involves cleaning, transforming, and organizing data to improve the performance and reliability of machine learning models. Key steps include:

1. **Handling Missing Values**:
    - Missing values in the dataset, such as incomplete user information or location data, are identified and addressed using imputation techniques (e.g., mean, median, or mode). Records with excessive missing information can be excluded to maintain data integrity.
2. **Data Cleaning**:
    - The dataset is checked for anomalies, such as duplicate records, inconsistent entries (e.g., invalid time formats), or unrealistic travel scenarios. These issues are corrected or removed to ensure consistency and accuracy.
3. **Feature Encoding**:
    - Categorical variables like weather conditions, some values of car data, time of day, or outlet types are converted into numerical representations using methods like one-hot encoding or label encoding, enabling the machine learning model to process them effectively.
4. **Feature Scaling**:
    - Continuous variables, such as distance to the nearest outlet or the duration of travel, are standardized or normalized to ensure that all features contribute equally to the recommendation model's performance.
5. **Outlier Detection and Removal**:
    - Outliers in features like user response time or coupon usage frequency are identified using statistical methods (e.g., Z-score, IQR) and treated to avoid skewing model performance.
6. **Dataset Splitting**:
    - The data is divided into training, validation, and testing sets. This ensures that the model is trained on a portion of the data while being evaluated on unseen data, enhancing its ability to generalize to new scenarios.

By following these preprocessing steps, the dataset becomes structured, consistent, and ready for machine learning models to deliver accurate and meaningful coupon recommendations. This ensures that the system is robust, user-centric, and capable of driving better engagement and revenue.

3.3 Feature Selection

Feature selection is a vital step in designing the methodology for an effective coupon recommendation system. This process involves identifying the most relevant attributes that significantly influence coupon acceptance while eliminating redundant or less impactful features. Proper feature selection enhances model accuracy, minimizes overfitting, and improves computational efficiency.

Key Features Considered: The primary features influencing user behavior in accepting coupons include:

- **Time of Day**: The likelihood of coupon usage may vary between morning, afternoon, and evening.
- **Weather Conditions**: Environmental factors (e.g., sunny, rainy) that impact user mood and travel patterns.
- **Destination and Distance**: The purpose of the trip (e.g., work, leisure) and proximity to the coupon-redeeming location.
- **Companions**: Whether the user is traveling alone, with friends, or family, as this can affect decision-making.
- **User Demographics and Preferences**: Age, gender, and prior behavior influencing coupon acceptance patterns.

Techniques Used:

1. **Correlation Analysis**:
    - Identifies features highly correlated with the target variable (coupon acceptance) and removes attributes with little to no impact.
2. **LASSO Regression**:
    - Employs regularization techniques to select only the most impactful features, automatically reducing the effect of irrelevant data.
    -

3. **Feature Importance from Models**:
    - Algorithms like Random Forest and Gradient Boosting compute feature importance scores, highlighting key factors that drive user behaviour.

By focusing on the most significant features, the machine learning models achieve better predictive performance and interpretability, laying the groundwork for an efficient and effective coupon recommendation system. This approach ensures that insights derived from the system can directly inform marketing strategies, driving increased user engagement and revenue generation.

3.4 Model Selection

Model selection is a key step in developing the automotive coupon recommendation system. It involves identifying the most suitable machine learning algorithm or combination of algorithms to predict coupon acceptance accurately while maintaining interpretability and computational efficiency.

Selection of Models: Various supervised learning models are evaluated to determine the most effective one for predicting whether a user will accept a coupon based on contextual features. Commonly considered models include:

1. **Logistic Regression**: A simple yet powerful approach for binary classification, ideal for understanding the relationship between features and coupon acceptance.
2. **Random Forest**: Effective at capturing non-linear relationships and offering insights into feature importance.
3. **Gradient Boosting (e.g., XGBoost)**: Optimized for high accuracy, leveraging regularization techniques to prevent overfitting and improve performance on complex datasets.

Evaluation Metrics: To ensure the best-performing model is selected, key metrics such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC) are employed. These metrics help assess both the correctness of predictions and the balance between false positives and false negatives.

Model Tuning: Once a baseline model is selected, hyperparameter tuning methods like Grid Search or Random Search are used to optimize model parameters for maximum accuracy and robustness.

By choosing and fine-tuning the best-performing models, this step ensures that the system effectively predicts coupon acceptance, driving user engagement and boosting campaign efficiency.

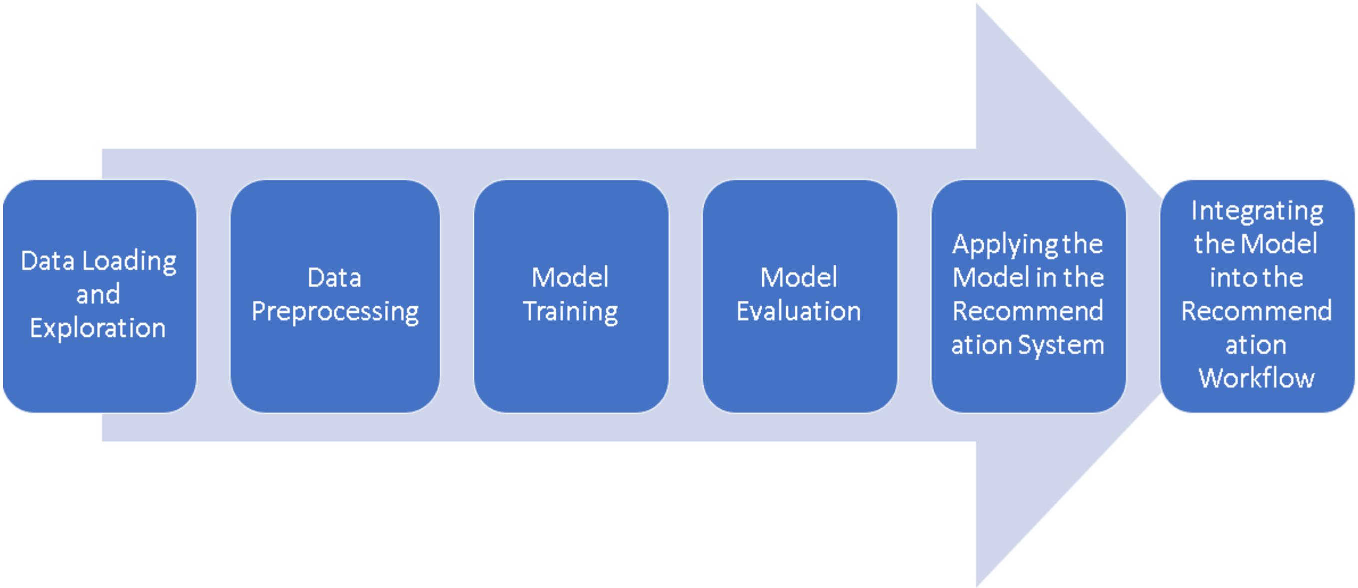
3.5 Model Evaluation Metrics

Model evaluation metrics are essential for assessing the performance and reliability of the machine learning models in predicting coupon acceptance. These metrics provide a comprehensive understanding of how well the model generalizes to new data and handles varying scenarios.

1. **Accuracy**:
    - Measures the overall correctness of the model's predictions by calculating the percentage of correctly predicted outcomes.
2. **Precision**:
    - Evaluates how many of the coupons predicted to be accepted were actually accepted, crucial for minimizing irrelevant recommendations.
3. **Recall**:
    - Assesses the model's ability to correctly identify all instances where a coupon is accepted, ensuring relevant recommendations are captured.
4. **F1-Score**:
    - A harmonic mean of precision and recall, offering a balanced measure when the dataset is imbalanced.

These metrics collectively ensure that the chosen model not only delivers accurate predictions but also provides meaningful and actionable insights for optimizing marketing strategies. Balancing these metrics helps create a robust system capable of recommending coupons effectively in real-world scenarios.

3.6 Recommendation System Workflow



Chapter 4

MODEL IMPLEMENTATION

4.1 Algorithm Selection for Automotive Coupon Recommendation

The selection of machine learning algorithms is a crucial step in designing a reliable model for predicting coupon acceptance in the automotive coupon recommendation system. Various algorithms are explored to handle complex relationships between user behavior, contextual factors, and coupon utilization. Below are the selected algorithms tailored to this system

1. Logistic Regression

Acts as a baseline for binary classification problems, predicting whether a user will accept or reject a coupon.

Offers simplicity and interpretability, making it ideal for understanding the impact of features such as time, weather, or user demographics on coupon acceptance.

2. Random Forest

An ensemble method that effectively handles non-linear relationships between features (e.g., travel route, passenger type, and destination) and coupon acceptance.

Provides feature importance scores, helping to identify which factors most influence user decisions.

3. Gradient Boosting Machines (GBM)

Includes models like XGBoost and LightGBM, which excel in handling interactions among features like time of travel and weather conditions.

Optimized for high predictive accuracy, with regularization techniques to minimize overfitting in diverse datasets.

4. Support Vector Machines (SVM)

Utilizes kernel functions to map complex, non-linear relationships, making it suitable for scenarios with overlapping decision boundaries in user behaviour patterns.

5. K-Nearest Neighbours (KNN)

A simple algorithm that predicts coupon acceptance based on the behaviour of similar users in comparable contexts, such as shared travel routes or weather conditions.

6. Decision Trees

Useful for building interpretable models that explain decision paths based on contextual factors like time, passenger type, or location.

4.2 Using Ensemble Machine Learning Algorithm

Ensemble machine learning algorithms combine the strengths of multiple models to achieve better predictive performance compared to individual models. In the context of an automotive coupon recommendation system, these algorithms are particularly useful because they can effectively handle complex and non-linear relationships between diverse factors, such as user behavior, location, time, and weather conditions.

By integrating predictions from multiple models, ensemble methods reduce overfitting, improve accuracy, and increase the robustness of the recommendation system. These qualities make ensemble algorithms a natural choice for a project that involves dynamic and context-dependent decision-making, as is the case with predicting coupon acceptance.

1. Logistic Regression

- Logistic Regression is a simple and interpretable algorithm for binary classification problems. It models the probability of coupon acceptance as a function of input features.
- In our system, it provides a baseline for predicting whether a user accepts a coupon based on linear relationships between factors such as passenger type, weather, or destination.
- Pros: Computationally efficient; interpretable.
- Cons: Limited in capturing non-linear relationships.

2. Decision Trees

- A decision tree creates a flowchart-like structure where each node represents a feature, each branch represents a decision rule, and each leaf node represents an outcome.
- In our project, a decision tree might split data based on factors such as weather (e.g., sunny or rainy), time of day (e.g., morning or evening), and travel destination to determine coupon acceptance.
- Pros: Easy to interpret and visualize; handles both categorical and numerical data.

- Cons: Prone to overfitting, especially with complex datasets

3. Random Forest

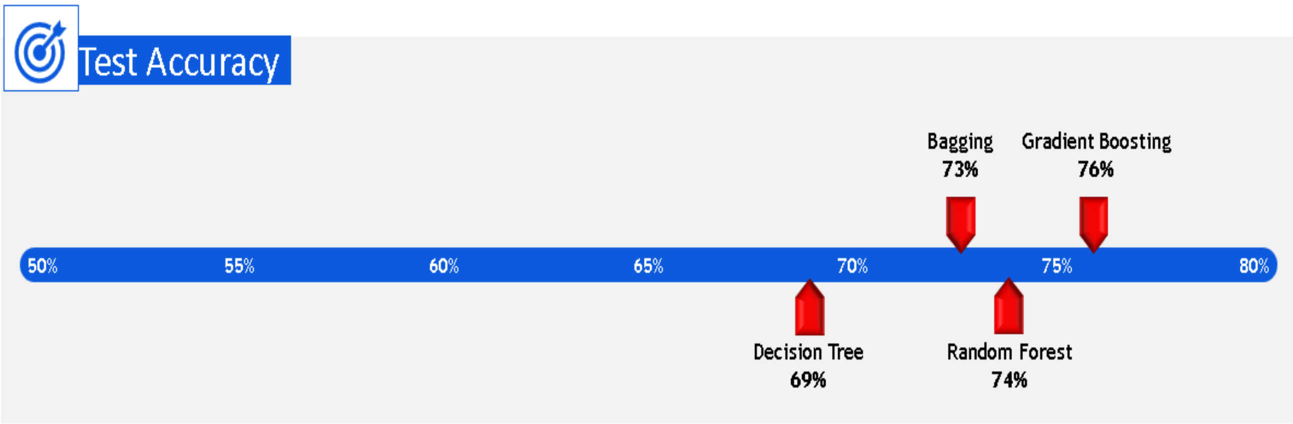
- Random Forest is an ensemble method that builds multiple decision trees during training and averages their predictions to improve accuracy. Each tree is trained on a random subset of features and data points, reducing overfitting.
- In our system, Random Forest can capture interactions among features like user demographics, location, and travel patterns, providing feature importance scores for interpretability.
- Pros: Robust to overfitting; handles high-dimensional data well.
- Cons: Can be computationally intensive for large datasets.

4. Gradient Boosting

- Gradient Boosting is an advanced ensemble method that builds decision trees sequentially, where each tree tries to correct the errors made by its predecessor. Algorithms like XGBoost are commonly used versions.
- In our project, Gradient Boosting captures subtle patterns and interactions between features, such as how time and user preferences combine to affect coupon acceptance.
- Pros: High accuracy; excellent at handling imbalanced datasets.
- Cons: Sensitive to hyperparameters; computationally expensive.
- **Performance:** Gradient Boosting achieved the highest test accuracy of 76% in our project, indicating its superior ability to model the complex relationships inherent in coupon acceptance behavior.

4.3 Modelling Results

Figure 4.1: Test Accuracy Across Models



4.4 Why Gradient Boosting is the Best Fit for the Project?

The Gradient Boosting algorithm achieved the highest test accuracy (76%) in our project due to its ability to:

- Handle complex interactions between multiple features.
- Address imbalanced datasets by focusing more on misclassified instances during training.
- Offer regularization techniques that prevent overfitting while maintaining high accuracy.

Its performance highlights its suitability for dynamic recommendation tasks where user behavior is influenced by multiple, interdependent factors.

Figure 4.2: Selection of number of trees for Gradient Boosting model

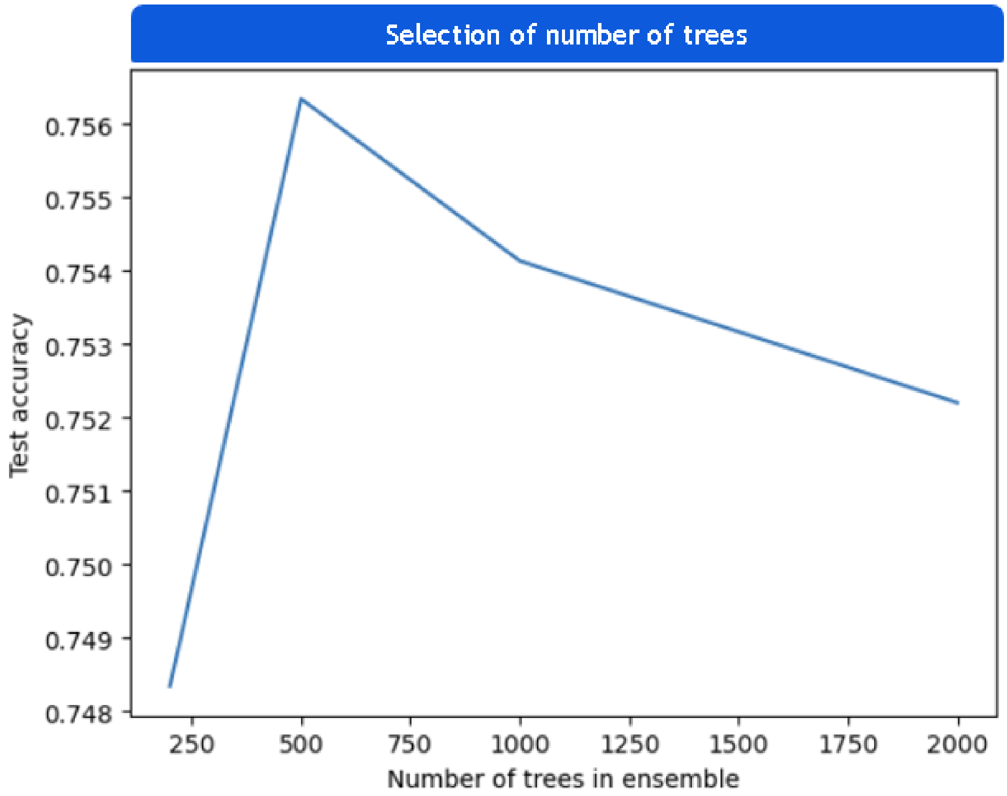


Figure 4.3: Selection of depth of trees for Gradient Boosting model

Figure 4.4: Feature Importance (Top 20) for Gradient Boosting model

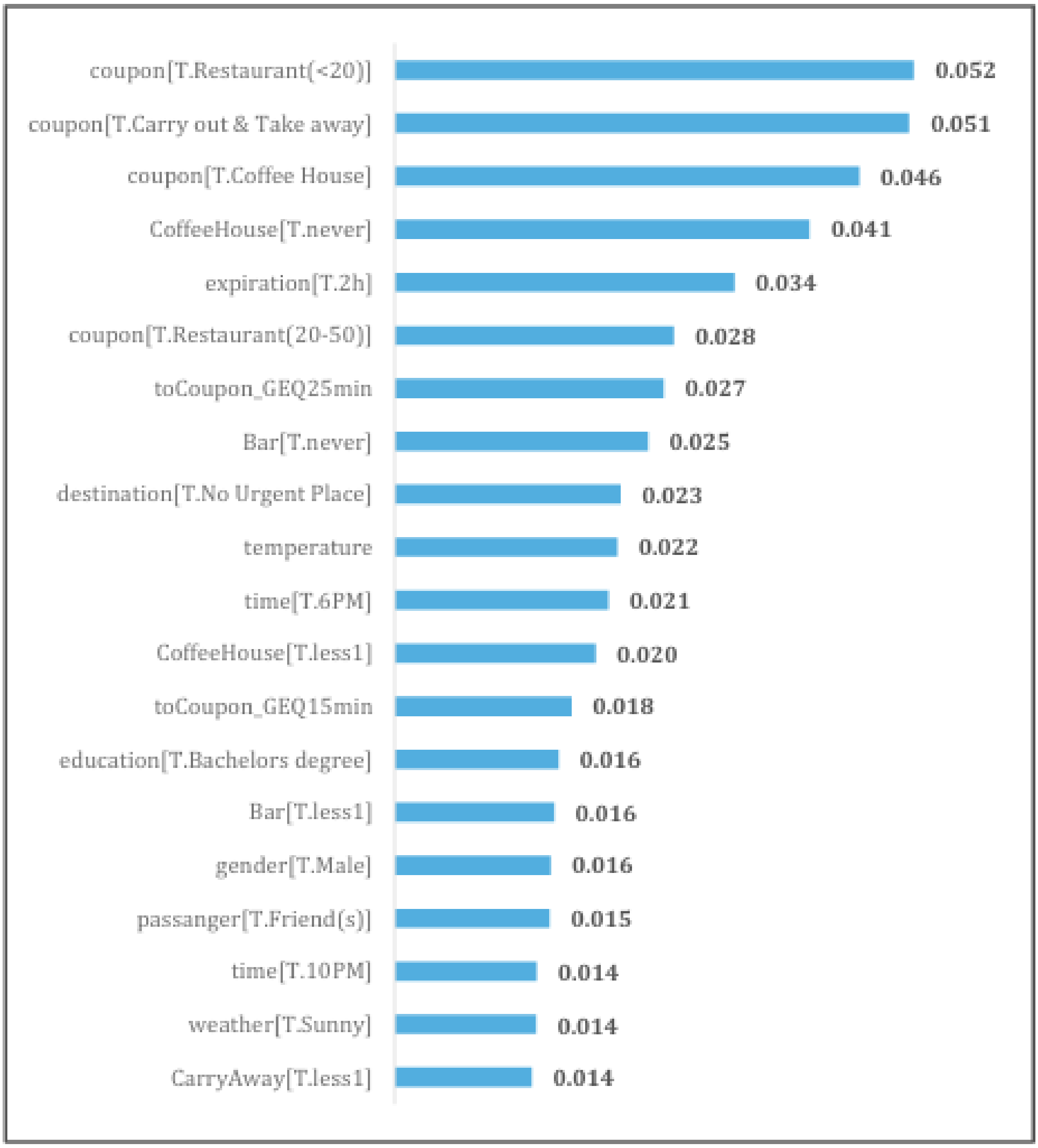


Figure 4.5: Feature Importance

Feature Importance Simplified	
Factors more likely to influence <u>acceptance</u>	• Coupon pertains to take away or a restaurant where per person cost is < $20 • Driving distance to the outlet is <15 minutes • Not travelling to any urgent place • Temperature is moderate • Time is in the afternoon • Weather is sunny • Gender is male • Travelling with friends
Factors more likely to influence <u>denial</u>	• Coupon pertains to coffee house or a restaurant where per person cost is $20-$50 • Person never or rarely goes to a coffee house or bar or gets carry away food • Coupon expires within 2 hours • Driving distance to the outlet is >15 minutes • Early morning or late night

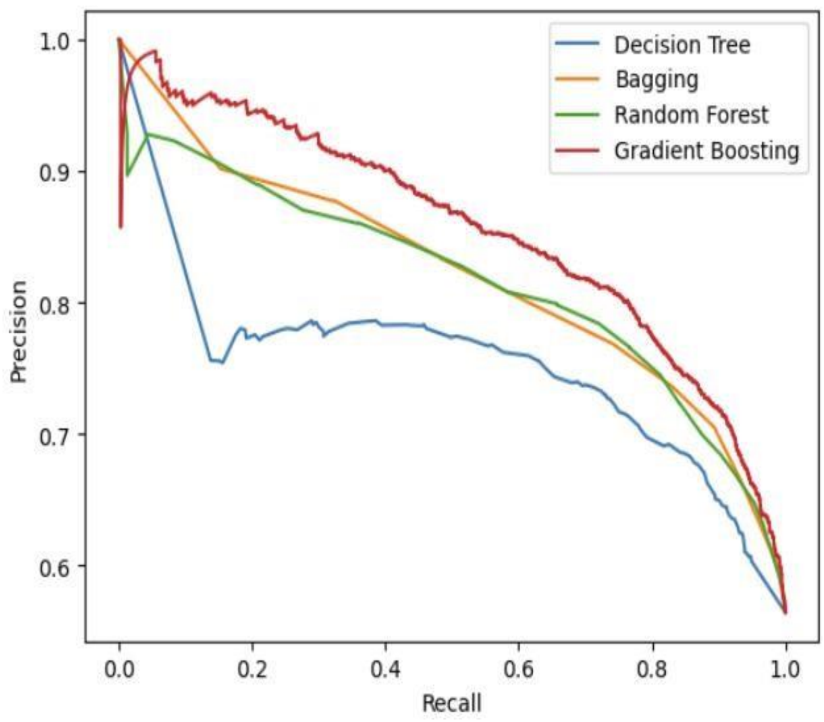
4.5 Recall Factor in the Project

Recall is a critical metric in this project because it measures the system's ability to correctly identify users who will accept a coupon (true positives) out of all users who would have accepted it (true positives + false negatives).

- **Importance of Recall:**
  - High recall ensures the system minimizes missed opportunities by correctly targeting users who are likely to redeem coupons.
  - For businesses, this means reaching potential customers effectively, maximizing the likelihood of coupon redemption, and driving revenue.
- **Example from the Project:**
  If a user is likely to accept a coupon based on contextual factors (e.g., they are traveling with passengers and heading to a food outlet), but the system fails to predict this correctly, it results in a **false negative**. A high recall helps avoid such missed predictions, ensuring better customer engagement and satisfaction.

By combining Gradient Boosting's predictive strength with a focus on optimizing recall, our system ensures it not only achieves high accuracy but also fulfills its goal of effectively driving coupon utilization.

Figure 4.6: Precision Vs Recall



Chapter 5
RESULT AND CONCLUSION

5.1 Performance Parameter Result

Figure 5.1: Precision, Recall, F1 Score

Model Type	Precision	Recall	F1 Score
Decision Tree	71%	76%	74%
Bagging	77%	74%	75%
Random Forest	75%	81%	78%
Gradient Boosting	78%	80%	79%

5.2 Conclusion

- **Increased Customer Engagement:** Customer retention and acquisition rates increase as personalized recommendation increases the chances of new users as well as for present customers to revisit.

- **Increased Sales and Revenue:** By recommending coupons for complementary products or services, the system encourages customers to purchase more, leading to higher average order values.

- **Customer Behavior Analytics:** Businesses gain insights into customer preferences, purchasing patterns, and coupon usage behavior, which can be used to refine marketing strategies.

- **Reduced Coupon Wastage:** Personalized coupon systems minimize the chances of offering discounts to customers who don't need or want them, reducing overall promotional waste.

Chapter 6

REFERENCE

6.1 References

Material Type	Works Cited From
Dataset	• **https://www.kaggle.com/datasets/mathurinache/invehicle-coupon-recommendation?resource=download**
Tools and libraries	• **https://towardsdatascience.com/top-5-machine-learning-libraries-in-python-e36e3e0e02af**
Coupon Marketing	• **https://www.vouchery.io/post/coupon-marketing-strategy-what-is-coupon-marketing**
Research Paper	• **https://ieeexplore.ieee.org/document/10442306**
Infosys course	• **https://infyspringboard.onwingspan.com/web/en/app/toc/lex_auth_013720165067218944334_shared/overview**

	**Name: Yash Bhargava** **Enrollment No: 221B456** **Course: Bachelor In Technology** **Branch: Computer Science and Engineering** **Mother's Name : Mrs. Parul Bhargava** **Father's Name : Mr. Yogesh Bhargava** **Email id : 221b456@juetguna.in** **Contact : +91-9301965586**
	**Name: Prathamesh Kale** **Enrollment No: 221B497** **Course: Bachelor In Technology** **Branch: Computer Science and Engineering** **Mother's Name : Mrs. Sarika Kale** **Father's Name : Mr. Nitin Kale** **Email id : 221b497@juetguna.in** **Contact : +91-8788745895**
	**Name: Ritik Prakash** **Enrollment No.: 221B498** **Course: Bachelor In Technology** **Branch: Computer Science And Engineering** **Mother's Name: Mrs. Rani Singh** **Father's Name: Mr. Vinay Kumar Singh** **Email ID: 221b498@juetguna.in** **Contact: +91-7061833722**