**Data Mining Project - ISM6136.004F23**  
Prof. Abdullah Musamar

**TEAM MEMBERS:**

* Poojitha Nadendla - U08878472
* Harsha Sai Jagu - U10209745
* Saumya Uttam - U74234095
* Himanshu Shrivastava - U79675891

**Predicting Coupon Acceptance in Various Driving Scenarios**

**Introduction**:

The project aims to predict "coupon acceptance" in various driving scenarios. Specifically, it seeks to determine the likelihood that individuals will accept a coupon when presented with different driving-related conditions, such as their destination, current time, weather conditions, passenger presence, and other relevant factors. The predictive model's goal is to provide insights into the factors that influence a driver's decision regarding coupon acceptance while on the road.

**About the Dataset:**

* Reference: [Dataset Link](https://archive.ics.uci.edu/dataset/603/in+vehicle+coupon+recommendation)
* DOI: 10.24432/C5GS4P

**Dataset Characteristics:**

* Type: Multivariate
* Subject Area: Social Science
* Associated Task: Classification
* Feature Types: Categorical and Integer
* Number of Instances: 12,684
* Number of Features: 23

**Features:**

1. Destination: The destination of the individual while driving, categorized as either "No Urgent Place," "Home," or "Work."
2. Passenger: The presence and type of passengers in the car, including "Alone," "Friend(s)," "Kid(s)," or "Partner."
3. Weather: The weather conditions at the time, classified as "Sunny," "Rainy," or "Snowy."
4. Temperature: The temperature conditions, with values such as "55," "80," and "30" representing different temperature ranges.
5. Time: The time of day when the driving scenario occurs, specified as "2PM," "10AM," "6PM," "7AM," or "10PM."
6. Coupon: The type of coupon being offered, with categories like "Restaurant(<$20)," "Coffee House," "Carry out & Take away," "Bar," and "Restaurant($20-$50)."
7. Expiration: The time duration until the coupon expires, either "1 day" or "2 hours."
8. Gender: The gender of the individual, categorized as "Female" or "Male."
9. Age: The age group of the individual, including values such as "21," "46," "26," and others.
10. Marital Status: The marital status of the individual, with categories like "Unmarried partner," "Single," "Married partner," "Divorced," and "Widowed."
11. Has Children: A binary indicator, "1" if the individual has children and "0" if they do not.
12. Education: The highest level of education attained by the individual, with categories including "Some college - no degree," "Bachelors degree," "Associates degree," and more.
13. Occupation: The occupation of the individual, with various job categories such as "Unemployed," "Architecture & Engineering," "Student," and many others.
14. Income: The income range of the individual, such as "$37500 - $49999," "$62500 - $74999," and other income categories.
15. Bar: The frequency of visits to a bar every month, with values like "never," "less1," "18."
16. CoffeeHouse: The frequency of visits to a coffeehouse every month, categorized as "never," "less1," "43," "gt8," and "nan."
17. CarryAway: The frequency of getting take-away food every month, with values like "n43," "gt8," "less1," and "never."
18. RestaurantLessThan20: The frequency of visits to a restaurant with an average expense per person of less than $20 every month, with categories including "43," "less1," "gt8," and "never."
19. Restaurant20To50: The frequency of visits to a restaurant with an average expense per person of $20 - $50 every month, with categories like "18," and "nan."
20. toCoupon\_GEQ15min: A binary indicator, "0" if the driving distance to the restaurant/bar for using the coupon is not greater than 15 minutes and "1" if it is greater.
21. toCoupon\_GEQ25min: A binary indicator, "0" if the driving distance to the restaurant/bar for using the coupon is not greater than 25 minutes and "1" if it is greater.
22. direction\_same: A binary indicator, "0" if the restaurant/bar is not in the same direction as the current destination and "1" if it is in the same direction.
23. direction\_opp: A binary indicator, "1" if the restaurant/bar is in the opposite direction as the current destination and "0" if it is not.
24. Y: A binary indicator, "1" if the coupon is accepted and "0" if it is not.

Bottom of Form

**Scope:**  
The project will encompass data preprocessing, model development, feature analysis, and the creation of an interpretable classification model. It will serve as a tool for informing evidence-based policy decisions and socioeconomic research.

**Approach:**

1. **Data Collection**: The first step is to collect and load the dataset into the project environment. The dataset can be obtained from a reliable source, such as the UCI Machine Learning Repository.
2. **Data Preprocessing**: Data preprocessing is crucial to handle missing values, encode categorical features, and scale/normalize numerical features. Exploratory data analysis (EDA) will also be conducted to understand the dataset's distribution and relationships between variables.
3. **Feature Selection**: Feature selection techniques will be applied to identify the most relevant features that have the most significant impact on an individual's income.
4. **Model Building**: Various classification algorithms will be considered, including logistic regression, decision trees, random forests, support vector machines, and gradient boosting. Multiple models will be built and evaluated for accuracy and robustness.
5. **Model Evaluation**: The models' performance will be assessed using metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve. Cross-validation will be performed to ensure the models generalize well.
6. **Hyperparameter Tuning**: Grid search or randomized search will be used to optimize the model's hyperparameters, further improving predictive performance.
7. **Model Deployment**: The best-performing model can be deployed on a suitable platforms, making it accessible to users for predicting coupon acceptance in driving scenarios.

Top of Form

Bottom of Form