



# CUSTOMER CLICK PREDICTION

PROBLEM STATEMENT, VARIABLE DESCRIPTION & DELIVERABLES

# Customer Click Prediction – Objective & Deliverables

---

## Introduction:

In the fast-paced world of digital advertising, our goal is to empower marketing decisions through data-driven insights. This project centers on predicting customer clicks on advertisements using a robust machine learning model. By harnessing the potential of historical data and relevant features, we aim to enhance advertising strategy and resource allocation for optimal campaign performance.

## Objective:

Develop an effective machine learning model to predict customer clicks on advertisements. By analyzing key features such as Daily Time Spent on Site, Age, Area Income, Daily Internet Usage, and others, the model will provide valuable insights into customer behavior, ultimately improving the efficiency of ad campaigns.

## Background:

As our company navigates the dynamic landscape of digital marketing, understanding customer engagement with advertisements becomes crucial. This initiative is fueled by the need to optimize ad campaigns, ensuring maximum impact and return on investment.

## Expected Outcomes:

The envisioned machine learning model will serve as a predictive tool, offering actionable insights into the factors influencing customer clicks. This goes beyond predicting clicks; it lays the foundation for strategic decision-making, positioning our company at the forefront of targeted and effective advertising.

## Recommended Project Steps & Guidelines:

- **Data Exploration and Cleaning:** Explore the dataset, identify and handle missing values, outliers, and duplicates. Gain an understanding of the distribution of each feature.
- **Data Preprocessing:** Encode categorical variables, scale numerical variables, and handle any class imbalance. Split the data into training and test sets for model development.
- **Exploratory Data Analysis (EDA):** Conduct EDA to identify patterns and relationships between user characteristics and ad clicks.

- **Feature Engineering:** Create new features or transform existing ones to improve the model's predictive performance.
- **Build Classification Model:** Develop a logistic regression model to predict the probability of a user clicking on an ad.
- **Model Evaluation:** Assess the performance of the model using metrics such as accuracy, precision, recall, and F1-score. Iteratively refine the model parameters for optimal predictive accuracy.
- **Reporting and Visualization:** Create a comprehensive report summarizing key findings. Develop visualizations to illustrate model performance and insights gained from the analysis.

### Variable Description:

The dataset provided for this project comprises information related to user interactions with online advertisements. Each row represents a unique user, and the dataset includes the following columns:

**Daily Time Spent on Site (Numeric):** The amount of time, in minutes, that a user spends on the website where the ad is displayed.

**Age (Numeric):** The age of the user who interacted with the advertisement.

**Area Income (Numeric):** The income level of the geographical area where the user is located.

**Daily Internet Usage (Numeric):** The amount of time, in minutes, that a user spends on the internet daily.

**Ad Topic Line (Text):** The topic or content of the advertisement that the user interacted with.

**City (Categorical):** The city of residence of the user.

**Male (Binary):** Binary variable indicating the gender of the user (1 for male, 0 for female).

**Country (Categorical):** The country in which the user is located.

**Timestamp (DateTime):** The timestamp indicating when the user interacted with the advertisement.

**Clicked on Ad (Binary):** The target variable indicating whether the user clicked on the ad (1 for clicked, 0 for not clicked).