PREDICTING AD CLICK

PROBLEM

In the recent time, internet and social media users have grown tremendously. As a result of this, many companies prefer to advertise their products on websites and social media platforms. However, targeting the right audience is still a challenge in online marketing.

Companies spend lot of money to display the advertisement of their products and if the consumers are not likely to buy the product, the advertisement cost can be a burden or costly for the company.

This comes as a challenge for the company to analyze whether a particular consumer click on an advertisement displayed.

DATA

https://www.kaggle.com/fayomi/advertising/version/1#

The data consists of 10 variables:
- Daily Time Spent on Site
- Age
- Area Income
- Daily Internet Usage
- Ad Topic Line
- City
- Male
- Country
- Timestamp
- Clicked on Ad

OBJECTIVE
- To develop a machine learning algorithm that predicts if a particular user will click on an advertisement
- I am interested on the feature "Clicked on Ad" as this feature can have two possible outcomes
  1) 0 → Indicates a case where the customer did not click on the advertisement displayed
  2) 1 → Indicates a case where the customer clicked on advertisement displayed
- My objective is to use the other nine features to accurately predict the value of "Clicked on Ad" feature
- To perform analysis to see how "Daily Time Spent on Site" in combination with "Ad Topic Line" affects the user's decision to click on the add

TECHNIQUES
- Data Preprocessing → Feature Selection is a very important step in data pre-processing that determines the accuracy of a classifier. It helps remove unnecessary correlation in the data that might decrease accuracy. Planning to consider the below preprocessing techniques:
  1) Feature Subset Selection → Data consists of 10 features and all these features won't be useful to predict the Customer clicked on ad
  2) Feature Creation → Since we have a feature "Timestamp" it can be further expanded into day and hour so that we will get new variables that an ML model will be able to process and find possible dependencies
- Model Development
  Proposing to develop two Machine Learning models
  1) Decision Tree → As I need to predict the class label of unknown records
  2) Logistic Regression → As I need to relationship between multiple features

EVALUATION METHODS
- Confusion Matrix → As it provides the information needed to determine how well a classification model performs.
- Summarizing the information with a single number would make it more convenient to compare the performance of different models
- Success is measured using performance metric such as "Accuracy" and "Error Rate"

EXPECTATIONS
- Learn to preprocess the data
- Learn to visualize the data
- To predict the Machine Learning model with high accuracy