# Waiter's Tip Prediction using Machine Learning

## Project Report

Soumyadeep Poddar

## Introduction:

In this machine learning project, we explore a Waiter's tip prediction model using Python. The project involves analyzing a dataset containing information about restaurant bills and tips. The initial data preprocessing includes data loading, visualization of distributions, and correlation analysis. Categorical variables are encoded, and the data is split into training and validation sets. Standardization is applied to the features, followed by training and evaluation of various regression models including Linear Regression, XGBoost, RandomForest, and AdaBoost.

## Objective:

The main objective of this project is to develop a predictive model that can accurately estimate the tips received by waiters in a restaurant based on various factors such as the total bill amount, day of the week, time of day, and group size. The project involves data preprocessing, feature scaling, and training multiple regression models. By comparing the performance of different models on both training and validation data, the project aims to identify the most suitable model for tip prediction.

## Methodology:

1) Data Collection & Loading:
    - ✓ The dataset is taken from geeksforgeeks. It contains various columns such as the total bill, tip, time, day among other variables.
2) Exploratory Data Analysis:
    - ✓ Checked for missing values and outliers.
    - ✓ Analysed the distribution of the data.
    - ✓ Analysed some patterns of offering tips on various factors.
3) Encoding categorical columns:
    - ✓ Converted all the categorical variables to numerical ones for correlation analysis through heatmap.
4) Model Building and Evaluation:
    - ✓ Splitted the dataset into training and test sets and scaled the training set.
    - ✓ Fitted Linear Regression, XGBRegressor, RandomForestRegressor and AdaBoostRegressor models and evaluated them on the basis of training and validation error through Mean Absolute Error.
    - ✓ We observe that XGBoost Regressor performs better in comparison to others.

## Conclusion:

In conclusion, this project successfully addressed the task of predicting waiter's tips in a restaurant setting using machine learning techniques. Through data preprocessing, feature scaling, and model training, we were able to build and evaluate regression models for tip prediction. The comparison of model performances revealed that the XGBoost Regressor exhibited the lowest Mean Absolute Error on the validation data, demonstrating its effectiveness in predicting tip amounts. This predictive capability can be leveraged by restaurants to optimize staff earnings and enhance customer service. The project highlights the value of data-driven insights in decision-making within the hospitality industry.

## Limitations:

One limitation of this project is the reliance on a specific dataset, which might not encompass all possible factors influencing tips. Also, the size of the dataset is small, which can affect the models' accuracy. The dataset's scope is limited to a set of attributes like total bill amount, day, time, and group size, potentially overlooking other important variables that could affect tipping behavior. Additionally, while the project evaluated different regression models, it did not explore more complex techniques such as neural networks or ensemble methods. As a result, the predictive performance might be further enhanced with more sophisticated modeling approaches. Furthermore, the project assumed that the relationships observed in the data would hold in various restaurant settings, which might not always be the case due to varying customer behaviors and cultural differences.