

# **Project Title**

## **Uber Fare Prediction**

### **TEAM MEMBERS:**

**Sailesh Pilla(11593815)**

**Yaswanth Sai Satish Sreerama(11601948)**

**Kishore Kumar Paila(11600316)**

### **Idea Description:**

Our Project is about Uber, the largest taxi company in the world, serving thousands of consumers every day. I noticed some open-source data from Kaggle. Due to the service's global reach, it is necessary to precisely compute fare pricing considering several factors, including distance, weather, time, and demand.

### **Goals and Objectives:**

Goals of this project is to predict fare accordingly in different places with respect to different variables by training the model first and then test it on test dataset.

### **Objectives:**

1. Thoroughly understand the dataset and determine whether the data is correct or not, and if not, perform a clean-up.
2. Create a regression model over the training data that has been separated.
3. Evaluate the model and compare R2 and Root mean squared error scores to find the projected model error so that you may understand how accurate your predictions were.

## **Motivation:**

The reason I wanted to know what type of algorithm or model they were using to see what all elements play a role in forecasting price and see which dimension plays a significant function in accurately predicting the fare is that we utilize uber a lot in our daily lives to get to our destinations.

## **Significance:**

This is accomplished by predicting fare quickly whenever a user enters his pickup and destination; it calculates it automatically considering all the factors using a machine learning algorithm. It attracts users when he can see the amount that he needs to pay before booking; it will give user to allocate or plan their budget accordingly and they can choose the type of vehicle accordingly. Here, utilizing ML models increases the likelihood that the prediction will be accurate. If it is incorrect, we can compute the error and improve our model's training by using more accurate data or by omitting irrelevant data. A better prediction can be made with the right training data.

## **Literature Survey:**

We are attempting to comprehend how Uber operates, how the uber fare is computed, and what additional criteria are taken into account to estimate the fare using ML models. Given that there are numerous variables that affect how fare is predicted, we will attempt to use a multi-linear regression.

1. Data Exploration
2. EDA (Explanatory Data Analysis)
3. Data Pre-Processing
4. Feature Selection
5. Splitting of Data
6. Applying Models and Performing Prediction

## Features:

- key - a unique identifier for each trip
- fare\_amount - the cost of each trip in USD
- pickup\_datetime - date and time when the meter was engaged
- passenger\_count - the number of passengers in the vehicle
- pickup\_longitude - the longitude where the meter was engaged
- pickup\_latitude - the latitude where the meter was engaged
- dropoff\_longitude - the longitude where the meter was disengaged
- dropoff\_latitude - the latitude where the meter was disengaged

## Expected Outcome:

Using the data's longitude, latitude, and passenger count, predict the fare amount. To improve our ability to predict the fare amount variable from the given dataset, we can extract new features from the dataset.

## References:

[1][Fares Dataset](#)

[2][Fare Amount Prediction](#)

[3][Comparing Fares](#)