

# **PROJECT PROPOSAL**

## **Topic: Taxi Trip Fare Prediction**

### **Teammates:**

- Nitin Somashekhar ([somashekhar.n@northeastern.edu](mailto:somashekhar.n@northeastern.edu))
- Deep Kotadiya ([kotadiya.d@northeastern.edu](mailto:kotadiya.d@northeastern.edu))
- Girish Raut ([raut.g@northeastern.edu](mailto:raut.g@northeastern.edu))

### **Project Description:**

We are planning to develop a Taxi Fare Prediction project that uses advanced machine learning algorithms like SVM (Support Vector Machine) and Random Forest etc. to accurately estimate taxi fares. By collecting and preprocessing historical data on taxi rides, we aim to train models that consider key factors such as travel distance, time of day, and passenger count. These models will be thoroughly tested and validated for accuracy and reliability. Our project is designed to identify critical fare-influencing factors, apply various predictive algorithms, and continuously refine these to enhance precision. This effort will provide a reliable fare estimation tool for taxi companies, ride-sharing services, and passengers, aiming to make transportation more predictable.

### **Project Outcome:**

Upon successfully completing the Taxi Fare Prediction project, we will have a model that can accurately estimate taxi fares. This model will be valuable for ride-sharing services, enabling dynamic pricing strategies to enhance profitability. Passengers will also benefit from more reliable fare estimates, assisting them in budget planning. Throughout this project, we will also deepen our understanding of machine learning techniques, data preprocessing, and the practical application of predictive analytics in real-world scenarios. This hands-on experience will enhance our skills in handling large datasets, optimizing algorithms, and implementing solutions that address complex problems.

### **Dataset:**

We have selected the NYC Yellow Taxi Trip Records dataset because it provides information about yellow taxi rides in New York City. This dataset includes details such as pickup and drop-off locations, times of trips, distances traveled, fare amounts, and passenger counts. It is widely used in studies related to transportation analysis, urban planning, and machine learning applications, providing valuable insights for these fields.

- Link: [Click Here](#)