International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Uber Price Prediction System

Bharti Vidhury[1], Ritwik Kandwal[2], Harshita Aggarwal[3], S Danishwaran[4], Aditya Pande[5]

[1]Project Guide
New Delhi

Abstract: ***Uber is one of the most popular ride-sharing services in the world. To provide the best service possible to the customers, this Uber price prediction system will use a combination of machine learning algorithms and historical data to accurately predict the cost of a ride. Nowadays, Taxi services are the primary methods of transportation. The businesses have gone under the rapid transformation and moved towards the digital innovation. In the past, product developers at software businesses have created a variety of ways, but they did not take into account the need for a customer's transportation in a certain area. In order to develop a precise model for forecasting future pricing, the Uber price prediction system will examine historical trip data, traffic, weather, time of day, and other pertinent information. To produce its predictions, it will employ a range of methods including linear regression, decision trees, random forests, and gradient boosting.***

Keywords: Uber, supervised machine Learning, business, price prediction, forecasting future pricing and ride-sharing

1. Introduction

In this chapter, we've provided a quick introduction to machine learning, the various algorithms utilised, and how it can be used to forecast the cost of an Uber journey using various factors like the weather, time of day, distance travelled, etc. We trained the system using a variety of machine learning methods, including random forest, linear regression, decision trees, etc., and then identified the algorithm that makes the best accurate pricing predictions. This algorithm is then used to forecast the cost of an Uber journey. The cost of an Uber ride relies on factors such as the distance travelled between the source and destination, the time of day, the weather, the wind speed, the dew point, visibility, latitude, and longitude, among others. Additionally, we look for relationships between the dataset's various columns.

Introduction to Machine learning

Binary and distributive issues are two subtypes of classification problems in machine learning. Machine learning, a branch of computer science and artificial intelligence (AI), focuses on utilising data and algorithms to replicate how individuals learn, gradually improving the system's accuracy. The three primary types of machine learning approaches are supervised learning, autonomous learning, and reinforcement learning. We'll use the supervised learning approach. We employ supervised learning and a range of methods, such as linear regression, random forests, decision trees, gradient-enhancing regression coefficients, and the k-fold crossover validation approach, to find which method offers the most accurate price prediction. For price, the surge factor is a crucial variable. This setting is particular to this place. Uber Heart, I say. It's a mystery box. It is dependent on other factors.

1.1. Benefits of Uber

Uber is a transportation company with an app that allows passengers to hail a ride and drivers to charge fares and get paid. More specifically, Uber is a ride sharing company that hires independent contractors as drivers.

The benefits of Uber:

- **Convenient & Cashless:** Instead of chasing down a taxi on a street, or calling and waiting for a car service, e-hail app users can hail a car from any location and have it arrived in minutes. Uber doesn't even need to ask you for an address.
- **Professional Service:** The drivers of Uber use their own cars so the cars are kept clean as well as well maintained.
- **Competitive Pricing:** It is difficult for Uber to keep one and definite price for the ride. The price of the Uber ride changes according to surge multiplier.
- **Safety & Flexibility:** The Uber services are even available at night for the customers.

2. Literature Review and Related Work

Dr. Balika J. Chelia, Jai Singh, Devansh Chaturvedi, and Avinash Kumar Singh published a research paper on a taxi fare prediction system using key feature extraction of artificial intelligence [1]. By matching passengers, the service known as "Journey Sharing" lowers the demand for vehicles on the road as a whole. Yet, there is a significant drawback. Very pricey and unsuitable for long distances or high band width. The issues raised above are scientifically defined in this article. To properly index requests for taxis, for instance, the system also offers information on trip demands. When picking the best building, she pays particular consideration to the taxi score based on many criteria including geographic area, origin, and destination. Fill with io and travel. A sizable amount of ratings show how accurate the suggested system is. Requests can be handled by the system in a matter of milliseconds. It makes the deployment simpler than with other providers.

Jeremy P Toner presented the research paper on the demand for taxis and the value of time [5]– welfare analysis. This study evaluated the value of taxi customers' walking, waiting, and in-vehicle times, as well as the elasticities of customer demand for taxis in terms of price and waiting time. The study's findings are presented in this document. Using expressed preference and transfer pricing approaches, the data were collected. Thus, of the three main sections of

the report, the first deals with the conception and analysis of the expressed preference trial into the value of time, and the second employs transfer price information to recalibrate demand models. The social welfare study of the town's current pricing and quantity restrictions is presented in the third main part.

Ahmed. G. Kvazim published a study titled The article "Taxi Fare Prediction Utilising Multilayer a perception system and Radial Basis" [7]. To forecast taxi fares, several prediction algorithms have been created. Some of these algorithms merely take into account the distance between the pickup and drop-off locations, while others also take the number of people and trip duration into account. The standard errors that resulted from the poor accuracy of these models' predictions varied from $2 to $4. Additionally, this problem has not been modelled using artificial neural networks. This study aims to develop a neural network-based system that takes into consideration the trip distance and other pertinent information in order to accurately predict the amount of fare for taxis in New York City. Both the radial based function network (RBFN) and the multi-layer perceptron are constructed; the former is done purely on the trip's journey distance, while the one that follows is accomplished by considering the trip's journey distance, the total number of passengers, the moment of day, the period of the week, and the month of the journey (temporal factors). The results show that when the journey distance as well as additional influencing factors were included in the models, neural networks were effective in correctly modelling the taxi fare with a mean-square error < 0.005.

Ankit Kumar and Vishal Shah presented their research on A method using unsupervised machine learning to predict the cost of taxi rides [4]. In this research, a machine learning-based model that can predict the demand for taxis in various geographic locations of a city by optimising the within-cluster sum of squared distances has been suggested. In order to anticipate the cost of transportation from one fixed place to another fixed location depending on the time and location of booking, a reliable and accurate pricing prediction model has been built. Keywords K-means clustering for unsupervised learning Price prediction for a taxi journey. Alberto Rossi, Gianni Barlacchi, M. Bianchini, B. Lepri of Computer Science has proposed the research paper on Modelling Taxi Drivers' Behavior for the Next Destination Prediction [3].

A Recurrent Neural Network (RNN) approach is presented that models the taxi drivers' behavior and encodes the semantics of visited locations by using geographical information from Location – Based Social Networks (LBSNs).

3. System Analysis

3.1. We will cover the numerous trials conducted to identify the most precise model for estimating the cost of an Uber journey in this chapter. We will talk about the issue and the system we are developing to address it. We'll work with a variety of machine learning techniques, including decision trees, linear regression, random forests, and others. We will examine the model to see how various factors, such as the weather, time of day, destination, surge multiplier, icon, etc., affect the cost of a ride.

3.2. Existing Problem

Uber is a transportation company with an app that allows passengers to hail a ride and drivers to charge fares and get paid. The existing problem of our project is how to calculate the most accurate price of a ride in Uber depending on the different parameters like traffic, weather, time, surge multiplier, destination etc.

3.3. The Suggested Repair

In order to develop a precise model for forecasting future pricing, the Uber price prediction system will examine historical trip data, congestion, the weather, time of day, and other pertinent information. To evaluate the model's accuracy, a number of methods will be used, including linear regression, decision trees, the Gradient Boost Regressor, random forests, and the k-fold cross validation approach. In the future, the price will be determined using the most precise algorithm.

4. Methodology

4.1. Defining the Problem:

In order to develop a precise model for forecasting future pricing, the Uber price prediction system will examine historical trip data, traffic, weather, the moment in the day, and other pertinent information. To create its predictions, it will combine a number of methods, including Gradient Boost, decision trees, and linear regression.

4.2. Data Collection:

With the aid of Kaggle, we are gathering the data for our research. Kaggle is a platform that is open-source that may be used for free to gather data for machine learning.

4.3. Exploratory Data Analysis:

EDA is a useful technique for comprehending the data. It makes an effort to gather as much data as possible.

4.4. Feature Engineering

It is described as a process for transforming the initial data set into characteristics. It is used to enhance the performance and accuracy of machine learning models.

4.5. Modelling

Different models, including linear regression, decision trees, random forests, gradient boosting, regressors, and K fold cross validation, will be used in our research.

4.6. Testing

The model created in the previous step will testing using historical test data to calculate precision.

5. Conclusion and Future Scope

Uber is top-most Famous ride-sharing services in the entire world. To provide the best service possible to the customers, this Uber price prediction system will use a combination of machine learning algorithms and historical data to accurately predict the cost of a ride.

In order to develop a precise model for forecasting future pricing, the Uber price forecasting system will examine historical trip data, congestion, the weather, the moment in the day, and other pertinent information. To create its predictions, it will combine a number of methods, including random forests, regression-based decision trees, gradient boosting regressors, and k fold cross validation. Because the random forest model's accuracy is the greatest for our project, we are employing it.

In this we are concluding the following features like the trips that are maximum are of which category like business category. Maximum round trips are in which month whether the trip is meeting its requirement. Then booking of flight is least in which month like September these all the features we are concluding with the help of feature engineering. The future scope of this project is that we have used the supervised machine learning algorithms which we were not previously used with the help of which we can predict the price more accurately on the various factors. The additional feature has enhanced our project. We can tell where we require large number of ride at a particular time so that we can provide the customers with the better experience.

References

[1] Devansh Chaturvedi, Jai Singh, Avinash Kumar Singh, and Dr. Balika J. Chelliah (2021) are the first. System for Predicting Taxi Fare Using Artificial Intelligence's Key Feature Extraction.
[2] Gaurav Hajela, Pawel Pratyush, and AkshataGangrade (2021). using dynamic spatial and temporal analysis to anticipate demand for taxis. Alberto Rossi, Gianni Barlacchi, M. Bianchini, and B. Lepri (2018) are cited in
[3] Modelling the Behaviour of Taxi Drivers to Predict the Next Destination. Vishal Shah and Ankit Kumar (2022)
[4] An unsupervised artificial intelligence method for tax price prediction. Jeremy P. Toner (1991)
[5] The Value Of Time And The Demand For Taxis.
[6] MajidKhedmati and Mohammad Fili (2020). Forecasting town visits using data mining methods. Ahmed. G. Quasim (2020),
[7] Using Multilayer Perceptrons and Radial Basis, one can predict taxi fares.