Gyalpozhing College of Information Technology
Royal University of Bhutan

ITS307 Data Analytics

Bachelors of Science in Information Technology

Gyalpozhing College of Information Technology

Taxi Fare Prediction

Researcher(s)

Nima (12200067)

Purna Kumar Limbu (12200075)

Sonam Wangchuk (12200084)

Tshering Deker (12200091)

Guided by: Mrs. Nima Dema

Table of contents

Literature Review

Fare and Duration Prediction: A Study of New York City Taxi Rides

According to this study, predicting fare and duration of a ride help passengers decide when is the optimal time to start their journey or help drivers decide which of the given two potential rides will be more profitable. In order to predict duration and fare, features such as pickup and dropoff coordinates, trip distance, start time, number of passengers, and a rate code detailing whether the standard rate or airport rate was applied (Antoniades et al., 2017).

In this study, the data used are all the subsets of New York City Taxi and Limousine Commission's trip data which contains observations of around 1 billion taxi rides in New York City. The total data is split between yellow taxi (operates in Manhattan) and green taxi (operates in the outer areas of the city). The original dataset contains features as pickup and dropoff locations, longitude and latitude coordinates, time and date of pickup and dropoff, ride fare, tip amount, payment type, trip distance, and passenger count.

The study was done to estimate ride duration without real time data, by analyzing data collected from taxis. Through this estimation, it would help in making future predictions. With the objective to model and account for traffic in predictions, two additional features were calculated from the data: rides in an hour and average speed during the hour. Rides in an hour represent the number of started rides within the hour of each observation. The average speed represents the average speed of all those rides.

It was found that the prediction result was fairly accurate. It was also found that more variables needed to be considered and modeled to further improve the accuracy. For example, the effect of location between pickup and dropoff points should be considered as well as the differences in driver's speed.

Similarly, for our project, we would like to use two features which are discussed in the above study: pickup and drop off points and distance traveled. However, we would like to use the location names of pickup and drop off point instead of using its longitude and latitude coordinates. Features such as payment type, tip amount, and average speed are not included in our project dataset. Instead of the number of passengers, we would use the number of seats as one of many features to predict the fare.

Taxi Fare Rate Classification Using Deep Network

As stated in this journal, it is difficult for individuals and organizations to estimate taxi trip fare using conditions such as time and day, which affects the traffic condition and starting location in a big city. The paper presents a comprehensive trip fare rate prediction based on time and location.

The taxi fare rate prediction was done using the following features:

- id: unique identifier for each trip
- Taxi id: unique identifier for each taxi
- Timestamp: julian timestamp which identifies the trip start time
- Starting latitude: latitude coordinate of the pick up location
- Starting longitude: longitude coordinate of pick up location

The target label is Renenue_class, which is categorical and reflects the taxi fare rate using five different classes: low, normal, medium, high and very high.

As for the features, more derived features were introduced to improve the performance of the system: time (morning, afternoon, evening), weekends, holiday, distance from airport, distance

from city center, and distance from one of the tourist spots in the given place. It was found that the time of day affects the taxi fare. For instance, the fare is lower in the afternoon than in the evening - the demand for taxis is greater in the evening. Weekends and holidays were also taken into consideration after observing that it also affects the fare rate.

Because the prediction system was based on deep learning with fewer features, it was concluded that the result was not optimal but was acceptable (Upadhyay & Lui, 2017).

As for our project, it will be based on structured machine learning with features such as pick up and drop off point, number of passengers, fuel price at that particular time, and distance traveled. Unlike in the above paper, our project's target will be regression in nature. Irrespective of the time, the taxi fare remains same in our country, therefore, time is not considered in our case.

The Research on Planning of Taxi Sharing Route and Sharing Expenses

This journal article focuses on the issues of poor carrying rates, arbitrary route planning, and taxi rates. The focus of this work is on taxi sharing routes and the sharing expenditure model, where the target functions are the maximum carrying rate, the shortest driving distance, and the sharing expense of drivers. They take into account the issues with taxi capacity limitations, driving distance restrictions, passenger load limits, and fee issues. They utilize the passenger's pool to categorize passengers into distinct directions and different beginning locations, then they apply the price algorithm, station-supervised mutation, station fragment cross design, and championship selection technique to solve the model. They examine the taxi data in Lanzhou City through this research and experiment with different ways to present this information. The study's findings also demonstrate how the taxi sharing mode has clearly improved in terms of carrying rate, driving distance, and driving advantages when compared to daily nonshared

practice. They draw the conclusion that the sharing mode can be used in taxi sharing routes and that the sharing costs are fair and beneficial to both drivers and passengers.

Real Time Prediction of cab fare using machine learning

According to this study, for predicting the longer-term events predictive analysis uses data which is an archive. For capturing the trends which are important mathematical models are used from past data. The model then uses present data to predict the longer-term or to derive actions to require optical outcomes, tones of appreciation in recent time for predictive analytics thanks to development in support technology within areas of massive data in machine learning. Many industries use predictive analytics for making an accurate forecast like giving the amount of fare for the ride within the city. These resource planning are enabled by the forecast as an example, cab fare can be predicted more accurately. A lot of factors are taken into consideration for a taxi start-up company. This research work tries to know the patterns and use different methods for fare prediction. This research work is developed for predicting the cab fare amount within a certain city. The research work involves different steps like training, testing by using different variables like pickup, drop-off location for predicting cab fare.

An Automated Cost Prediction in Uber/Call Taxi Using Machine Learning Algorithm

According to the study, cab businesses such as Uber, Ola, Meru Cabs and others have sprung up in recent years and serve thousands of people across the world every day. It is now critical for them to correctly manage their data in order to come up with fresh business ideas and get the greatest outcomes. As a result, it becomes critical to precisely predict the fares. This paper

compared all the fare details of specified cabs and predicted the lowest fare cab using linear regression method. In this paper, they implemented prediction models for the three models like Uber Go, Go Sedan and Uber Auto. Here deviation of the cab fares is also compared and using these data, build an application that can assist the users to select the cab with the determined benefits and lowest fare. In this model they use the machine learning technique of linear Regression model. The methodology and outcomes of this work can contribute to a more real-world demand. This application can improve the transport accessibility, reduce waiting time and reduce the transportation fare etc.

References

Antonaides, C., Fadavi, D., & Foba Amon Jr, A. (2017). *Fare and Duration Prediction: A Study of New York City Taxi Rides*. 6.

G, E. Rani., M, Sakthimohan., R, R. Raj., M, S. Ganesh., R, S. Sunder., & K, Karthigadevi. (2022). An Automated Cost Prediction in Uber/Call Taxi Using Machine Learning Algorithm. *IEEE Xplore*, 764–767. https://doi.org/10.1109/ICACITE53722.2022.9823852

Prem Jacob, T., Pravin, A., Mohana Prasad, K., Judgi, G. T., & Rajakumar, R. (2022, March 1). *Real Time Prediction of Cab Fare Using Machine Learning*. IEEE Xplore. https://doi.org/10.1109/ICEARS53579.2022.9752315

Upadhyay, R., & Lui, S. (2017, September). *Taxi Fare Rate Classification Using Deep Networks*. Encontro Português de Inteligência Artificial, Portugal.

Zhang, X., Zhang, Q., Yuan, Z., Wang, C., & Zhang, L. (2020). The Research on Planning of Taxi Sharing Route and Sharing Expenses. *Mathematical Problems in Engineering*, *2020*, e6348578. https://doi.org/10.1155/2020/6348578