

PREDICT THE PRICE OF CAB TRIP USING CLASSIFIERS AND REGRESSION

Dr.S.Krishnaveni¹

Assistant Professor

Department of B.Com Business Analytics

PSGR Krishnammal College for Women, Coimbatore, India.

krishnavenis@psgrkcw.ac.in

A.Anjana²,

UG Scholar,

Department of B.Com (Business Analytics),

PSGR Krishnammal College for Women, Coimbatore, India.

anjanaananth2001@gmail.com

ABSTRACT

The main intention of the objective is to layout a set of rules that facilitates to predict the fare of Uber rides for future rides. Machine learning knowledge of algorithms is used to expand regression fashions. Uber supplies carrier to a huge wide variety of clients every day. Now it becomes simply crucial to arrange their records well, to come up with new commercial enterprise thoughts to get the first-class effects. Eventually, it will become honestly vital to estimate the fare costs correctly.

Keywords: Predictive analysis, fare prediction, pricing version

I.INTRODUCTION

Uber has been a first-rate source of journey for people dwelling in urban regions. Some people don't have their cars at the same time as a few don't drive their cars intentionally because of their busy schedule. So special types of people are the usage of the offerings of Uber and other taxi offerings. In this text, it's going to take you thru the consequences that Uber faces and the solution for the problem found with the usage of Python.

The dataset consists of facts of approximately more than thousand Uber pickups in Coimbatore from 2010 to 2021. We can do extra with this dataset in preference to just analysing it. In this segment, it's going to take you thru Uber trips Problems and solution using Python. Uber has emerged as leading enterprise within the provision of recent transportation alternatives within the coeval international. Uber is particularly within the commercial enterprise of networking and all the organization's emerging operations may be gestate in phrases of in reality presenting a median thru which the relevant call for can meet up with the applicable supply. Uber is the best potency company to evaluate and publish real global supportable data. So fare amount is our target variable and rest of the variables are our predictor variables.

II.OBJECTIVE

This Objective is framed in the goal as to achieve the increase in the trips being booked in the day to day period. The datasets collected from kaggle and modified according to the objective. The dataset contains attributes like START_DATE, END_DATE, REQUEST ID, DRIVER ID, DISTANCE, and PRICE. The main motto is to achieve the price increase to increase the revenue of the organization. The machine learning algorithms are used to predict the fare according to distance and time. The ideas are suggested to increase not only the Price but also applicable for other factors also.

III.RELATED WORK

This research will be useful to those, who're concerned in fare forecasting. In the previous generation, the fare changed into handiest depending on distance, however with the enhancement, in technologies, the cab's fare is dependent on lots of things like time, area, number of passengers, traffic, quantity of hours, base fare, and so forth. The look at is based on Supervised mastering whose one application is prediction, in device getting to know. This research aims to observe predictive evaluation, which is a method of analysis in Machine Learning. Many corporations like Ola, Uber, and many others makes use of Artificial Intelligence and system learning technology to find the answer to correct fare prediction hassle. [2]

A survey suggests that the Flight and Cab fares vary in step with different factors like place, time of the day, and so forth. Cabs as well, wherein the fare depends upon the wide variety of passengers, visitors, and

so on. The vendor has facts about all of the factors, but the buyers can get admission to the records that is constrained and we cannot expect the price lists. Uber and Ola use factors like traffic in a specific vicinity, and call for and supply elements motive of the paper is to investigate the factors that have an impact on the deviation in the tariffs and the way they're associated with the trade inside the prices. [5]

The patterns and functions of the transportation system, including traditional method of tour which includes taxis and subways as well as revolutionary gear like ride-hailing structures (Uber, Lyft, etc.), are critical studies subjects in economics, transportation, and operational studies fields. By calculating and analysing the effect of those elements on Uber riders' fee amounts, we achieve conclusions which might be instructive and useful in practice. [3]

Utilizing huge-scale urban facts sets to are expecting taxi and Uber passenger's demand in cities is valuable for designing higher taxi dispatch structures and enhancing taxi services. In this paper, we are expecting taxi and Uber demand using real-global statistics sets. Our technique includes two key steps. First, we use temporal-correlated entropy to measure the call for regularity and achieve the maximum predictability. Second, we implement and verify 5 famous representative predictors (Markov, LZW, ARIMA, MLP, and LSTM) in accomplishing most predictability. [8]

Using Spatio-temporal time series fashions can assist us to better recognize the demand for e-hailing services and to expect it greater accurately. This paper analyses the prediction overall performance of 1 temporal model (vector autoregressive (VAR)) and spatio-temporal fashions (Spatial-temporal autoregressive (STAR); least absolute shrinkage and selection operator applied on STAR (LASSO-STAR)) and for distinctive scenarios (primarily based on the number of time and space lags), and carried out to each rush hours and non-rush hours periods. The effects show the want of considering spatial fashions for taxi demand. [4]

This model is able to predict Uber surge multipliers, the overall mean and the historical average in all but 3 of the 49 locations in Pittsburgh and outperforming three nonlinear methods in 28 of the 49 locations. Cross-correlation of Uber and Lyft surge multipliers is also explored. [7]

The data is multidimensional in form and gives an acute insight into many different features of a taxi trip like the Pick Up point, Drop Off point, Trip Distance, fare Amount, Taxable Amount, etc. Here the goal is develop the best model that predicts accurately for this given dataset while also giving an overview into the algorithms used with some details. Using this, users can benefit by forecasting the fare amount that is charged for each trip and which location will be cheaper for them to hail a taxi from. [1]

Ride-on-demand services are becoming more popular in Uber and OLA cabs. To help both drivers and customers. It uses dynamic pricing to balance supply and demand in an attempt to increase service quality. Dynamic prices changes generate problems for passengers. The price is "unpredictable". In order to give a solution to this problem, it is therefore important to give passengers more detail, and forecasting dynamic prices. This Acts as a feasible solution [6]

IV. METHODOLOGY

A.LINEAR REGRESSION

Linear regression is a primary and typically used type version of predictive evaluation. These regression estimates are used to explain the relationship between one structured variable (Feature) and one or more unbiased variables (Features). The architecture of the proposed device includes 3 predominant components: information garage, information processing, and data analytics.

B.DESICION TREE CLASSIFIER

The decision tree classifier creates the type version via building a decision tree. Decision tree classifiers provide a readable type version this is potentially correct in lots of one-of-a-kind application contexts, which include power-based programs. The classifier and test are being performed on the basis of the attributes given in the dataset.

This supervised learning method gives the probability of a given value. Linear regression uses fairly common machine learning algorithms that predict accurate outcomes. Linear regression uses a classification algorithm, used when the target variable is in nature.

STEP 1: Imported the dataset from kaggle and modified the dataset and saved in excel .csv format.

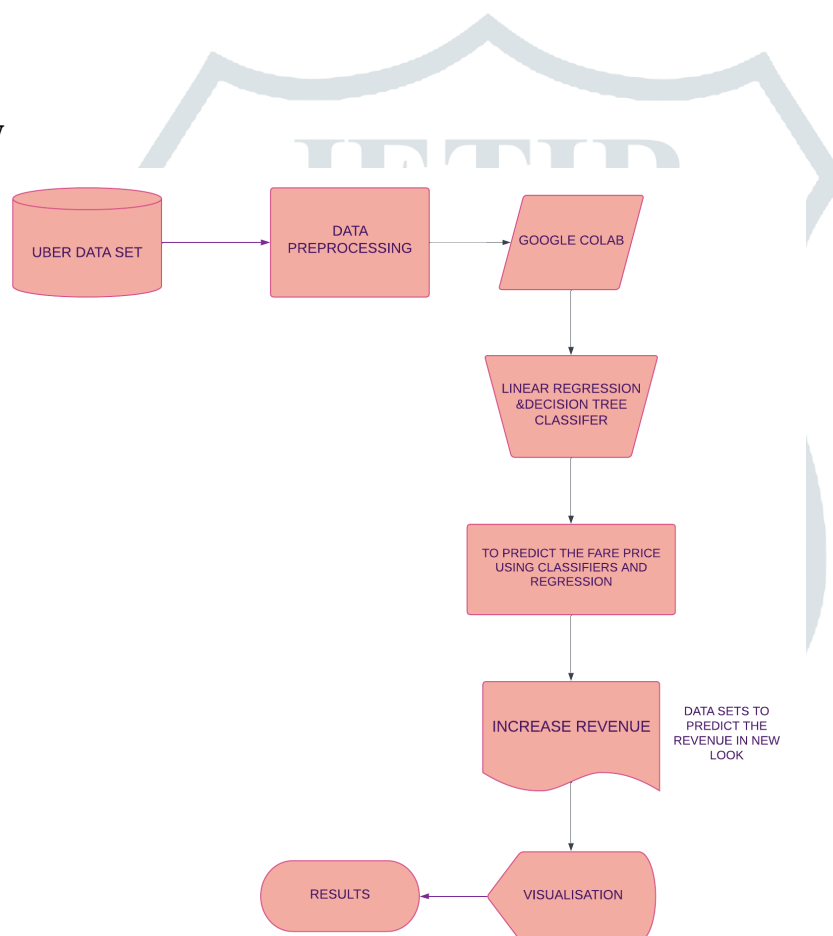
STEP 2: Used Google colab for executing python code

STEP 3: Pre-processing is done by removing all unwanted data from data set.

STEP 4: Linear regression is used to predict the price of the cab being booked.

STEP 5: Decision tree classifier is used to predict the output and positive accuracy is calculated.

WORK FLOW

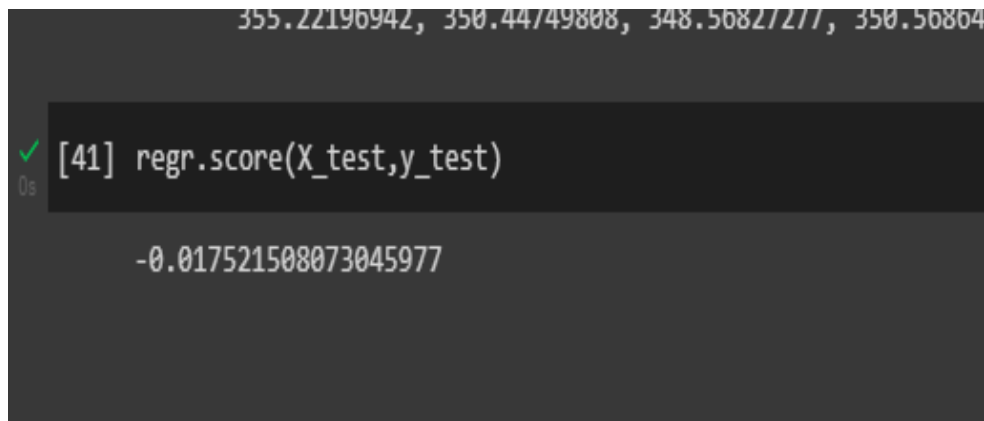


IMPLEMENTATION

Linear regression model is used to predict the Fare of the Uber cab being booked in the future times. But it gives a negative accuracy. So by using the decision tree classifier algorithm we are able to predict the price of the uber cab being booked and it gives a positive accuracy value. This helps to predict that the fare increase and it also contributes to the revenue increase of the organisation.

V.RESULTS

LINEAR REGRESSION

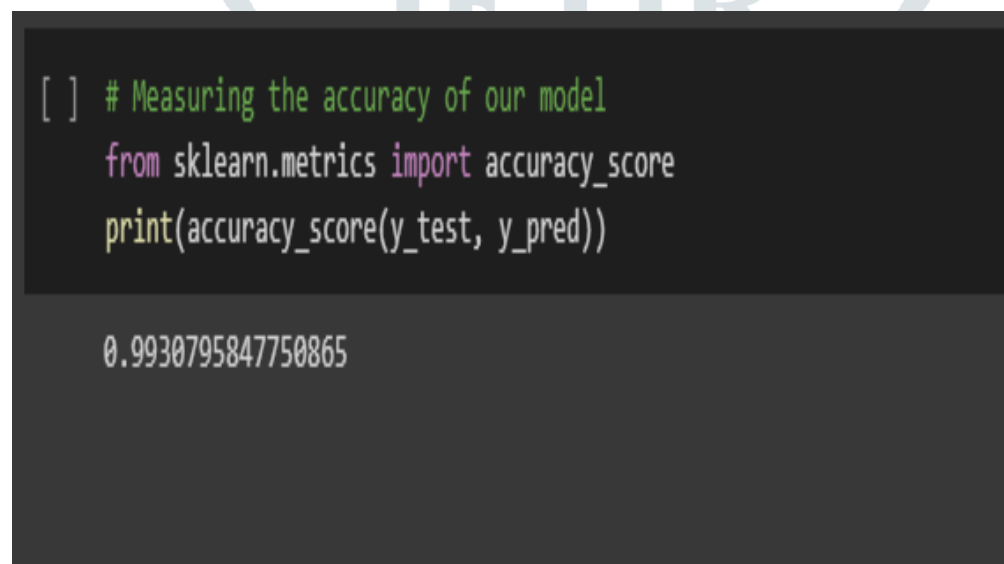
A screenshot of a Jupyter Notebook cell. At the top, there is a line of text: "355.22196942, 350.44749808, 348.56827277, 350.56864". Below this, the cell contains a code block starting with a green checkmark icon and the text "[41] regr.score(X_test,y_test)". The output of the code is "-0.017521508073045977".

```
355.22196942, 350.44749808, 348.56827277, 350.56864  
✓ [41] regr.score(X_test,y_test)  
-0.017521508073045977
```

Fig5.1 Accuracy prediction of linear regression

The Linear regression predicts the price of the trip but the score is predicted in the negative value.

DECISION TREE CLASSIFIER

A screenshot of a Jupyter Notebook cell. The cell contains a code block with the following text: "[] # Measuring the accuracy of our model", "from sklearn.metrics import accuracy_score", and "print(accuracy_score(y_test, y_pred))". The output of the code is "0.9930795847750865".

```
[ ] # Measuring the accuracy of our model  
from sklearn.metrics import accuracy_score  
print(accuracy_score(y_test, y_pred))  
0.9930795847750865
```

Fig5.2 Accuracy prediction of Decision tree classifier

To compare both the prediction accuracy, the decision tree classifier algorithm is used. In this Algorithm the accuracy score is predicted upto 99%. This Denotes that decision tree classifier predicts the better accuracy.

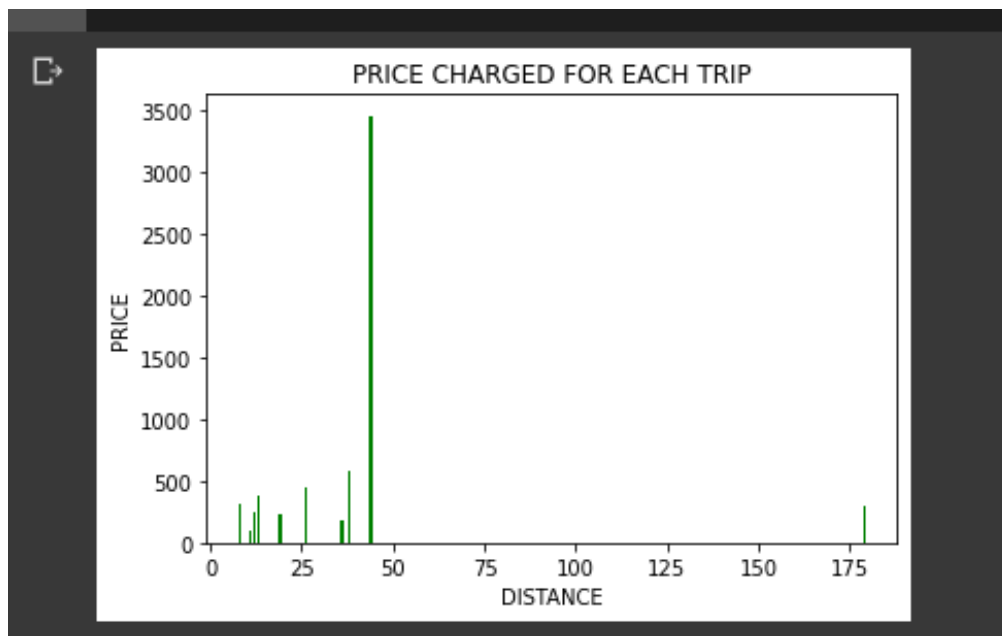


Fig 5.3 Visualisation in Increase of the price

Python Google colab is used to predict the Uber trip price and it helps to predict the price for future trips. There is a 99% of chance for increase in uber fare trip in the future transactional trips.

VI. CONCLUSION

In this paper, Python Google colab is used to predict the Uber trip price and it helps to predict the price for future trips. This is the major part of the revenue and it also helps to increase the revenue. The further work can be proceeded with the different machine learning algorithms and we can also predict the fare price of different cities and countries.

REFERENCES

1. A retrospective on Taxi Hailing methodology Sanskar Shah Student, Department of Information Technology, K J Somaiya College of Engineering, Maharashtra, India. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 09 | Sep 2020 www.irjet.net p-ISSN: 2395-0072
2. Banerjee, Pallab & Kumar, Biresh & Singh, Amarnath & Ranjan, Priyeta & Soni, Kunal. (2020). Predictive Analysis of Taxi Fare using Machine Learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 373-378. 10.32628/CSEIT2062108.
3. Chao, Junzhi. (2019). Modeling and Analysis of Uber's Rider Pricing. 10.2991/aebmr.k.191217.127.
4. Faghih, Sabihev & Safikhani, Abolfazl & Moghimi, Bahman & Kamga, Camille. (2017). Predicting Short-Term Uber Demand Using Spatio-Temporal Modeling: A New York City Case Study.
5. Khandelwal, K., Sawarkar, A. ., & Hira, S. (2021). A Novel Approach for Fare Prediction Using Machine Learning Techniques. *International Journal of Next-Generation Computing*, 12(5). <https://doi.org/10.47164/ijngc.v12i5.451>
6. Kunal, Arora & Kaur, Sharanjit & Sharma, Vinod. (2021). Prediction of Dynamic Price of Ride-On-Demand Services Using Linear Regression. International Journal of Computer Applications & Information Technology. 13. 376-389.

7. Predicting real-time surge pricing of ride-sourcing companies Matthew Battifaranoa , Zhen (Sean) Qiana,b, A Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, United States b Heinz College, Carnegie Mellon University, Pittsburgh, PA 15213, United States
8. Zhao, Kai & Khryashchev, Denis & Huy, Vo. (2019). Predicting Taxi and Uber Demand in Cities: Approaching the Limit of Predictability. IEEE Transactions on Knowledge and Data Engineering. PP. 1-1. 10.1109/TKDE.2019.2955686.

