iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.45682

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Dynamic Price Prediction using IBM Watson Studio

Srujan H J[1], Swetha M[2], Tejaswini T P[3], Shreya B B[4], Dr. Naveen Kumar K R[5], Prof. Anu C S[6], Dr. Arun Kumar G H[7]
[1, 2, 3, 4, 5]*Internship Guide, Assistant Professor,* [6, 7]*Internship Coordinator, Assistant Professor, Computer Science and Engineering Department, BIET, VTU University*

Abstract: ***Many companies like Ola, Uber, etc., uses Artificial Intelligence and machine learning technologies to find the solution of accurate fare prediction problem. We are using random forest regression algorithm, which is useful for prediction modeling to get the most accurate value. The project will be helpful to those, who are involved in fare forecasting. In previous era, the fare was only dependent on distance, but with the enhancement in technologies the cab's fare is dependent on a lot of factors like time, location, number of passengers, traffic, number of hours, base fare etc. The project is based on Supervised Learning whose one application is prediction, in machine learning. The project aims to study the predictive analysis, which is a method of analysis in machine learning.***
Keywords: Machine Learning, Supervised Learning, Random Forest, Dynamic Prediction, IBM Watson Studio

I. INTRODUCTION

Cab services are a central transportation method in almost all urban areas. Like so many other fields today the cab business is undergoing a rapid digital transformation with new actors like Ola/Uber taking market shares with innovative digital products.
Many organizations do not have a direct role in travel and tourism but offer related products and services. Some examples would be offering travel insurance, parking facilities at airports, theatre and event tickets, car hire, and travel by rail or coach to airports, etc. at competitive rates.
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[1]. So fare amount is our target variable and rest of the variables are our predictor variables. Here in this project we used two different datasets namely cab-rides and weather dataset to predict dynamic prices of cabs using supervised machine learning model.

II. LITERATURE SURVEY

This literature survey 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.
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, 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.
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.

©IJRASET: All Rights are Reserved

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. 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.

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.

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.

III. METHODOLOGY

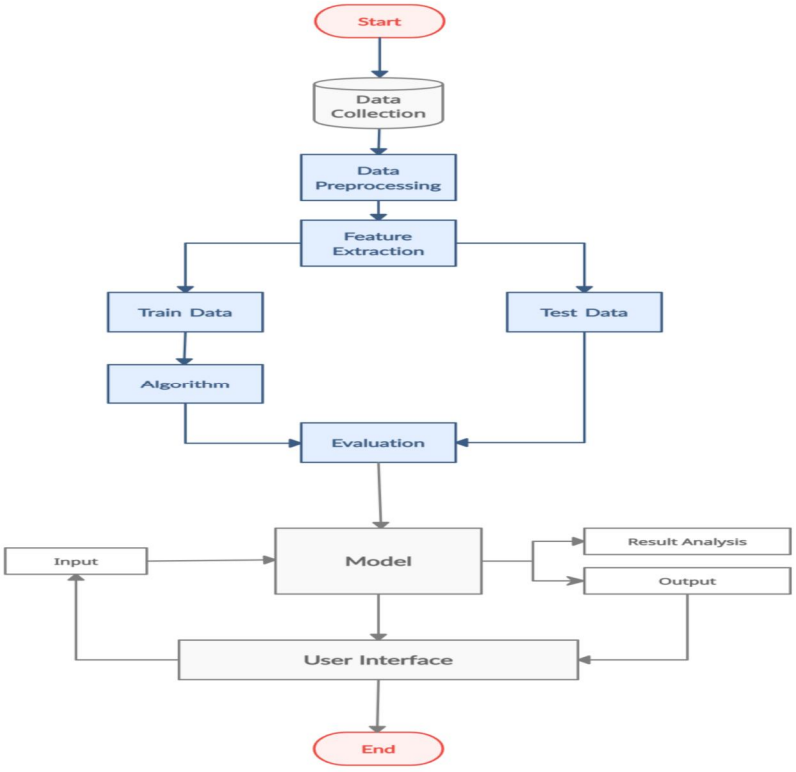


Fig. 1: Methodology Diagram

The above figure depicts the methodology for the dynamic prediction of cab price.

1) *Data Collection:* The process of gathering data depends on the type of project, for an ML project, realtime data is used. The data set can be collected from various sources such as a file, database, sensor and other sources and some free data sets from internet can be used. Kaggle and UCI Machine learning Repository are the repositories that are used the most for data collection for Machine learning models.
2) *Data Pre-processing***:** Data pre-processing is a process of cleaning the raw data i.e. the data is collected in the real world and is converted to a clean data set. There are certain steps executed to convert the data into a small clean data set and make it feasible for analysis, this part of the process is called as data pre-processing., Most of the real-world data is messy, like: Missing Data, Noisy Data, Inconsistent Data

3) *Feature Extraction:* When the input data to an algorithm is too large to be processed and it is suspected to be redundant then it can be transformed into a reduced set of features. Determining a subset of the initial features is called feature selection. The selected features are expected to contain the relevant information from the input data, so that the desired task can be performed by using this reduced representation instead of the complete initial data. Feature extraction involves reducing the number of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power, also it may cause a classification algorithm to overfit to training samples and generalize poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. Many machine learning practitioners believe that properly optimized feature extraction is the key to effective model construction.

4) *Model Selection*

a. Model selection is the process of selecting one final machine learning model from among a collection of candidate machine learning models for a training dataset. Model selection is a process that can be applied both across different types of models and across models of the same type configured with different model hyper parameters. There are several Machine learning algorithms to be used depending on the data you are going to process such as images, sound, text, and numerical values. The algorithms that you can choose according to the objective that you might have it may be Classification algorithms or Regression algorithms.

b. Some of the regression models are: Linear Regression, Logistic Regression, Random Forest Regression/Classification, Decision Tree Regression/Classification.

c. A Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap and Aggregation, commonly known as bagging. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees.

5) *Train and Test Data:* For training a model we initially split the model into 2 sections which are “Training data” and ‘’Testing data”. The classifier is trained using training data set, and then tests the performance of classifier on unseen test data set.

6) *Training set:* The training set is the material through which the computer learns how to process information. Machine learning uses algorithms to perform the training part. Training data set is used for learning and to fit the parameters of the classifier.

7) *Test set:* A set of unseen data used only to assess the performance of a fully-specified classifier.

8) *Evaluation:* Model Evaluation is an integral part of the model development process. It helps to find the best model that represents the data and how well the chosen model will work in the future. To improve the model hyper-parameters of the model can be tuned and the accuracy can be improved. Confusion matrix can be used to improve by increasing the number of true positives and true negatives. The output is predicted by analysing the test data as input along with test data output and then the output is displayed.

9) *Interface:* A web interface is built to take input and display an output. Flask language is used to build a web interface and pickle library is used to integrate both model and web page

IV. MODEL DESIGN

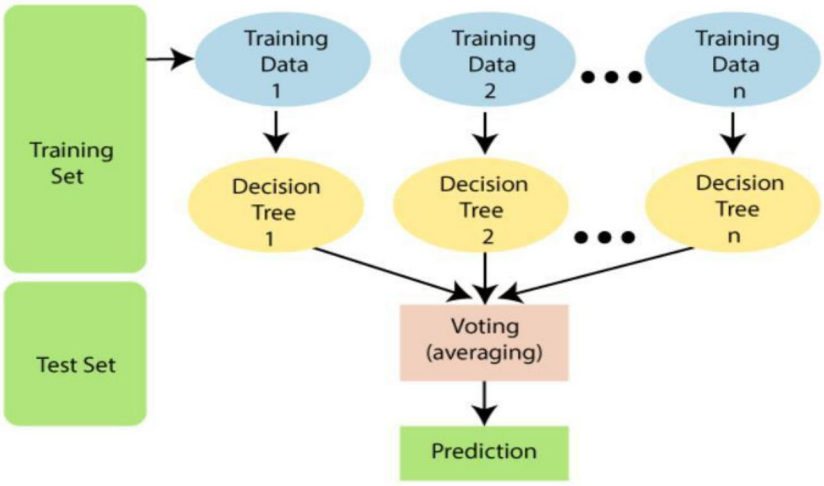


Fig 3: Block Diagram of Random Forest Regressor

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.[6]

1) Step-1: Select random K data points from the training set.
2) Step-2: Build the decision trees associated with the selected data points (Subsets).
3) Step-3: Choose the number N for decision trees that you want to build.
4) Step-4: Repeat Step 1 & 2.
5) Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

V. RESULTS

This segment provides a detailed description of the implementation results as well as the performance of the proposed system.

```
Out[68]: RandomForestRegressor(max_depth=4, n_estimators=20, n_jobs=-1, random_state=52)
```

Predecting the Result

```
In [69]: ypred=rand.predict(x_test)
         print(ypred)

         [33.44544798 19.16381383  9.54753035 ...  6.02421004 26.79738243
          17.55244465]
```

Score of the model

```
In [70]: rand.score(x_train,y_train)
Out[70]: 0.7575275520145969
```

Fig 4: Model Building with random Forest Regressor and Model Evaluation

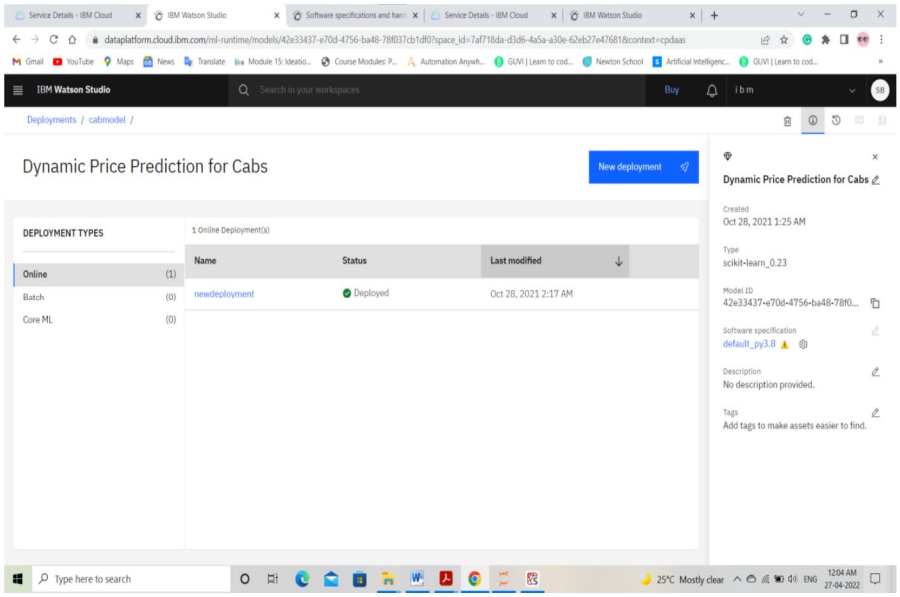


Figure 5: Model Deployment in IBM Watson Studio



Fig. 6: Home Page of the User Interface

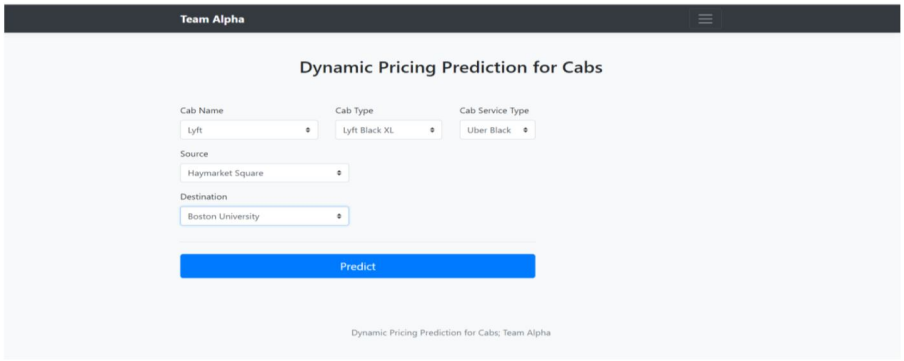


Fig. 7: Prediction Page of the User Interface

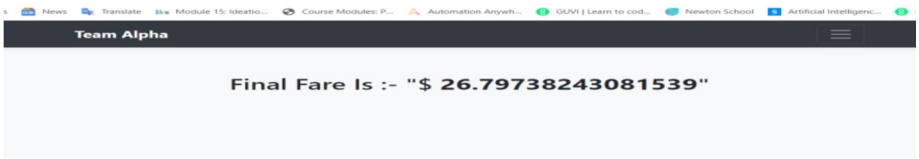


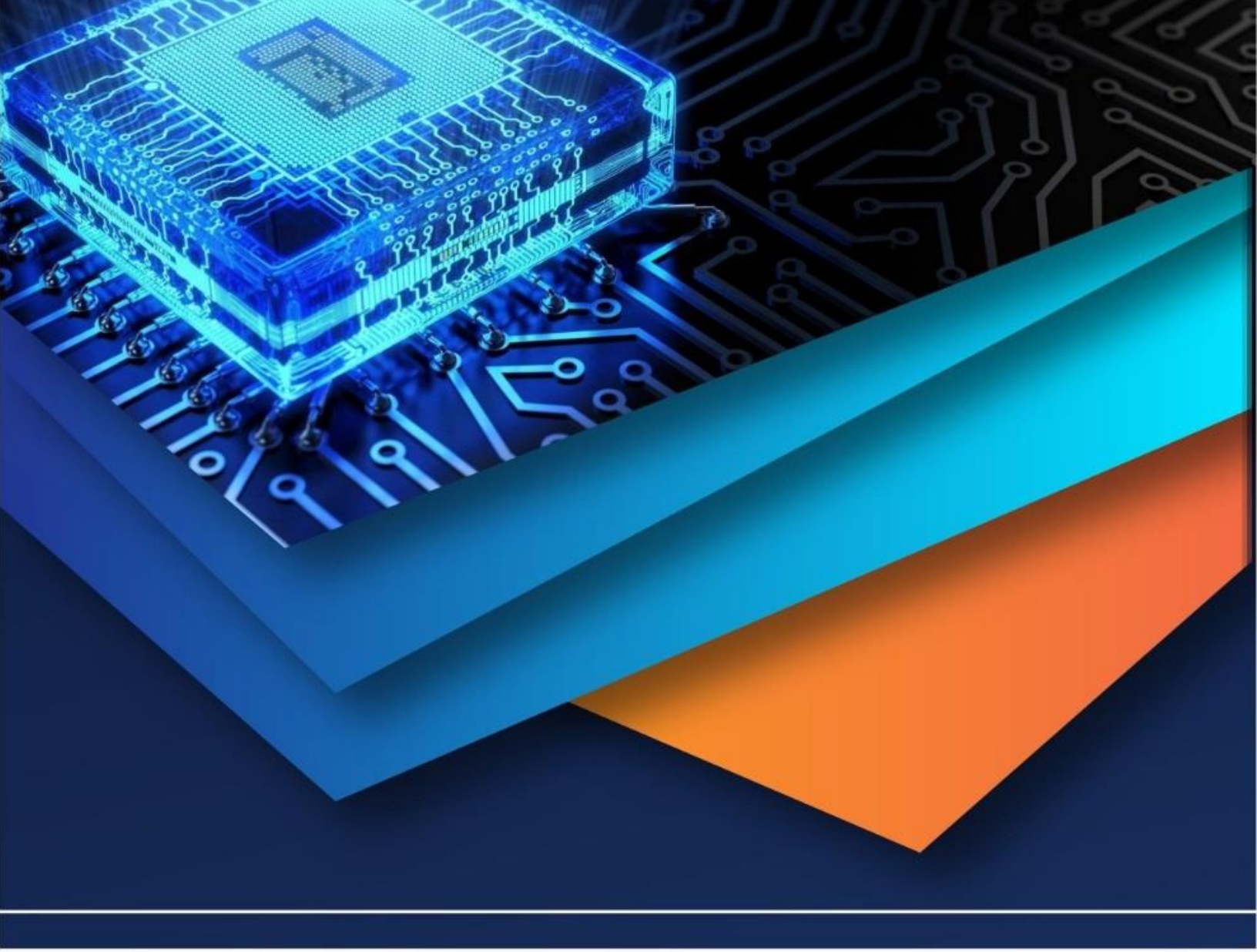
Fig. 8: Result Page of the User Interface

VI. CONCLUSION

Cab price prediction can be a challenging task due to the high number of attributes that should be considered for the accurate prediction. The major step in the prediction process is collection and pre-processing of the data. The proposed system is a sequential learning model with recurrent machine learning for predicting the cab price in different areas in the city. Learning from the past historical data, the demand prediction is done for the location. Cab rides and weather data set is used to train our model. This model gives the prediction of cab price. This work can be extended in the future by adding more input such as holidays, festivals etc. Cabs can be organized and send based on the prediction of the model. In addition, it can save so much time and energy that is currently being spent by cabs to find passengers.

REFERENCES

[1] https://www.thesmartbridge.com/Aboutus
[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, Sabiheh & Safikhani, Abolfazl & Moghimi, Bahman & Kamga, Camille. (2017). Predicting Short-Term Uber Demand Using Spatio-Temporal Modeling: A New York CityCase 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.s
[7] https://machinelearningmastery.com/a-gentle-introduction-to-model-selection-for-machine-learning/
[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.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)