DYNAMIC PRICE PREDICTION FOR CABS

Team Number: 379

Team Members:

SIRIVELLA RAMANA SAI (20MID0095)

VENNETI SRI SAI VEDANTH (20MID0110)

NADAVADI HARSHITH GOWD (20MID0113)

RAMINI VARUN BABU (20MID0135)

1.INTRODUCTION:

1.1 OVERVIEW:

The cab industry is undergoing a significant digital transformation, driven by innovative companies like Uber and Lyft. These ride-hailing platforms have become the go-to transportation option for many urban dwellers, offering convenience and accessibility. However, with the rise of such services, new challenges have emerged, including the need for accurate fare predictions. Additionally, various organizations not directly involved in travel and tourism have started offering complementary services such as travel insurance, event tickets, and car rentals at competitive rates.

1.2 PURPOSE:

The purpose of this project is to address the existing challenges of inconsistent pricing and lack of transparency in the cab industry by developing a dynamic price prediction system for cabs. The project aims to provide users with reliable and real-time fare estimations, empowering them to make informed decisions and plan their rides effectively. By considering various factors such as distance, time, location, and demand patterns, the system aims to enhance transparency, improve user experience, and optimize the budgeting process for cab rides. The ultimate goal is to revolutionize the cab booking industry by promoting fairness, accuracy, and customer satisfaction through the implementation of advanced algorithms and data analysis techniques.

2.LITERATURE SURVEY:

S.NO	Author	Journal Name	VOL.NO, JOURNAL.NO	DESCRIPTION OF THE PAPER	DRAWBACK OF THEIR WORK
1	Rishi srinivas	Uber Related Data Analysis using Machine Learning	978-0-7381-1327-2/21/$31.00 ©2021 IEEE	The research paper examined the impact of social media usage on mental health among teenagers, highlighting the negative effects of excessive screen time and cyberbullying.	the Uber dataset used may not accurately represent the entire taxi market in New York City, potentially limiting the generalizability of the findings.

2	Pooja Pranavi Nalamothu	Comparative Analysis of Regression Models for Price Prediction of Ride-On-Demand Services	| SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |	time-of-day windows and clustering techniques to extract temporal modes from the data. These temporal modes are then incorporated as features in the model. The study finds that the developed model performs well in predicting surge multipliers for Uber and Lyft, outperforming naive methods for higher surge values. However, the model tends to underperform for surge multipliers between 1.1 and 1.5, especially in the Lyft model, which negatively affects its overall performance.	The model's underperformance on low surge multipliers (between 1.1 and 1.5) hampers its overall predictive accuracy, particularly in the case of Lyft, which significantly degrades its performance compared to naive methods.
3	Shashank H.	Data Analysis of Uber and Lyft Cab Services	ISSN: 2456-236X Vol. 05 Issue 01 | 2020	he research paper focuses on developing a price prediction model for Uber rides using machine learning techniques such as linear and logistic regression. The model incorporates factors like distance, weather, and traffic to estimate the fare prior to the ride, providing consumers with more informed choices. The study aims to enhance user experience and improve the accuracy of fare predictions.	he proposed price prediction model shows a price fluctuation of around ten to twenty rupees compared to the actual fare, indicating room for improvement in terms of accuracy. Fine-tuning techniques could be explored to enhance the model's performance.
4	Junzhi Chao	Modeling and Analysis of Uber's Rider Pricing	ISSN: 2456-23X Vol. 05 Issue 07 | 2020 CC BY-NC 4.0	This project aims to analyze the impact of various factors on Uber pricing and rider payment amounts in New York City using real-time operational data. It employs time series analysis, statistical regression, and prediction in econometrics to draw conclusions that are valuable for practical applications in transportation and economics.	he research focuses mainly on UberX pricing in NYC and does not consider other vehicle types or carpooling. This limitation may affect the generalizability of the findings to other cities or online car services like Lyft.
5	1Akula Prasanth Kumar, 2Kanchi Aashritha, 3Marri Johnwesley, 4Jonnadula Narasimharao	A NOVEL APPROACH TO ANALYZE UBER DATA USING MACHINE LEARNING	A peer reviced international Journal ISSN: 2457-0362	It focuses on data preprocessing, outlier detection, correlation analysis, and the implementation of linear regression and random forest regression models for fair price prediction. The analysis includes parameters such as trips by the hours in a day and trips during months in a year. The results are evaluated using metrics like R2 and	if the matter is expensive or requires specialized equipment or expertise to handle, it can pose financial or logistical challenges for individuals or organizations. The limited availability and accessibility of the matter can

				RMSE. However, specific details and conclusions from the analysis are not provided in the given abstract.	impede progress, hinder research and development efforts, and restrict the overall impact it can have on various fields and industries.
6	Siyu Liao1, Liutong Zhou2, Xuan Di2, Bo Yuan1, Jinjun Xiong3	Large-scale Short-term Urban Taxi Demand Forecasting Using Deep Learning	978-1-5090-0602-1/18/$31.00 ©2018 IEEE	This paper explores the application of deep learning techniques for predicting short-term urban taxi demand. It discusses two DNN architectures, ST-ResNet and FCL-Net, designed to capture spatial and temporal dependencies in urban traffic data, considering factors like weather conditions and city events.	the need for proper design and architecture. While deep neural networks (DNNs) have shown superior performance compared to traditional machine learning techniques, achieving such results requires careful design and consideration of domain knowledge.
7	G. Venkat Sai Tarun P. Sriramya	Ola Data Analysis for Dynamic Price Prediction Using Multiple Linear Regression and Random Forest Regression	November 2022 DOI:10.3233/APC220071 License CC BY-NC 4.0	This research aims to create the most efficient and accurate cab fare prediction system using two machine learning algorithms, the Multiple linear Regression algorithm and the random forest algorithm, and compare parameters r-square, Mean Square Error (MSE), Root MSE, and RMSLE values to evaluate the efficiency of two machine learning algorithm. Considering Multiple linear Regression as group 1 and random forest algorithms as implemented, the 2 group process was to predict prices and get the best accuracy to compare algorithms. The algorithm should be efficient enough to produce the exact fare amount of the trip before the trip starts.	Linear Regression Limitations: Multiple Linear Regression assumes a linear relationship between the independent variables and the target variable. However, in the case of cab fare prediction, there may be complex and nonlinear relationships between the fare amount and various factors such as distance, time, traffic conditions, and location. Linear Regression may not be able to capture these nonlinearities effectively, leading to suboptimal predictions.
8	Elizabeth Rani. G Sakthimohan. M Revanth Raj. R	An Automated Cost Prediction in Uber/Call Taxi Using Machine Learning Algorithm	INSPEC Accession Number: 21884655 DOI: 10.1109/ICACITE53722.2022.9823852 Publisher: IEEE	Uber, Ola, Meru Cabs, and other cab businesses have sprung up in recent years. And these taxi firms serve tens of thousands of people every day. It is now critical for them to correctly manage their data to come up with	While using linear regression for fare prediction and comparing the fare details of different cab models can provide useful insights, there are

	Sri Ganesh. M Shyam Sunder. R Karthigadevi. K			fresh business ideas and get the greatest outcomes. As a result, it becomes critical to precisely predict the fares. The motive of this paper is to compare all the fare details of specified cabs and predict the lowest fare cab using linear regression method. In this paper we implemented prediction model for the three models like Uber Go, Go Sedan and Uber Auto. Here deviation of the cab fares 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 we use machine learning technique of linear Regression model, and it may contain labelled data. Here 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.	some potential drawbacks to consider. Linear regression assumes a linear relationship between the dependent variable (fare) and the independent variables (features). However, in the context of cab fares, the relationship between fare and factors like distance, time, and traffic conditions may not be strictly linear.
9	Srujan H J1 , Swetha M2 , Tejaswini T P3 , Shreya B B4 , Dr. Naveen Kumar K R5 , Prof. Anu C S6 , Dr. Arun Kumar G H7	Dynamic Price Prediction using IBM Watson Studio	ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538	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	Data Quality and Quantity: The accuracy of any machine learning model heavily relies on the quality and quantity of the training data. Inaccurate or incomplete data can lead to incorrect fare predictions. Feature Selection: Identifying the relevant features that contribute to fare prediction is crucial. While you mentioned several factors like time, location, number of passengers, traffic, number of hours, and base fare

				project aims to study the predictive analysis, which is a method of analysis in machine learning	
10	Pallab Banerjee 1 Biresh Kumar 2 Amarnath Singh 3 Priyeta Ranjan 4 Kunal Soni	Predictive Analysis of Taxi Far e using Machine Learning	ISSN : 2456-3307 CSEIT2062108 | Accepted : 10 April 2020 | Published : 20 April 2020 | March-April-2020 [6 (2) : 373-378]	This research aims to study the predictive analysis, which is a method of analysis in Machine Learning. Manycompanies like Ola, Uber etc uses Artificial Intelligence and machine learning technologies to find the solutionof accurate fare prediction problem. We are proposing this paper after comparative analysis of algorithms likeregression and classification, which are useful for prediction modeling to get the most accurate value. Thisresearch will be helpful to those, who are involved in fare forecasting. In previous era, the fare was only	This research mentions regression and classification as the analyzed algorithms for fare prediction. While these algorithms have their strengths, they may not be the most suitable for capturing the complexities of fare estimation. Other advanced algorithms, such as ensemble methods (e.g., random forest, gradient boosting) or neural networks, could potentially yield better results. It's important to explore a wider range of algorithms to identify the most accurate and robust approach.

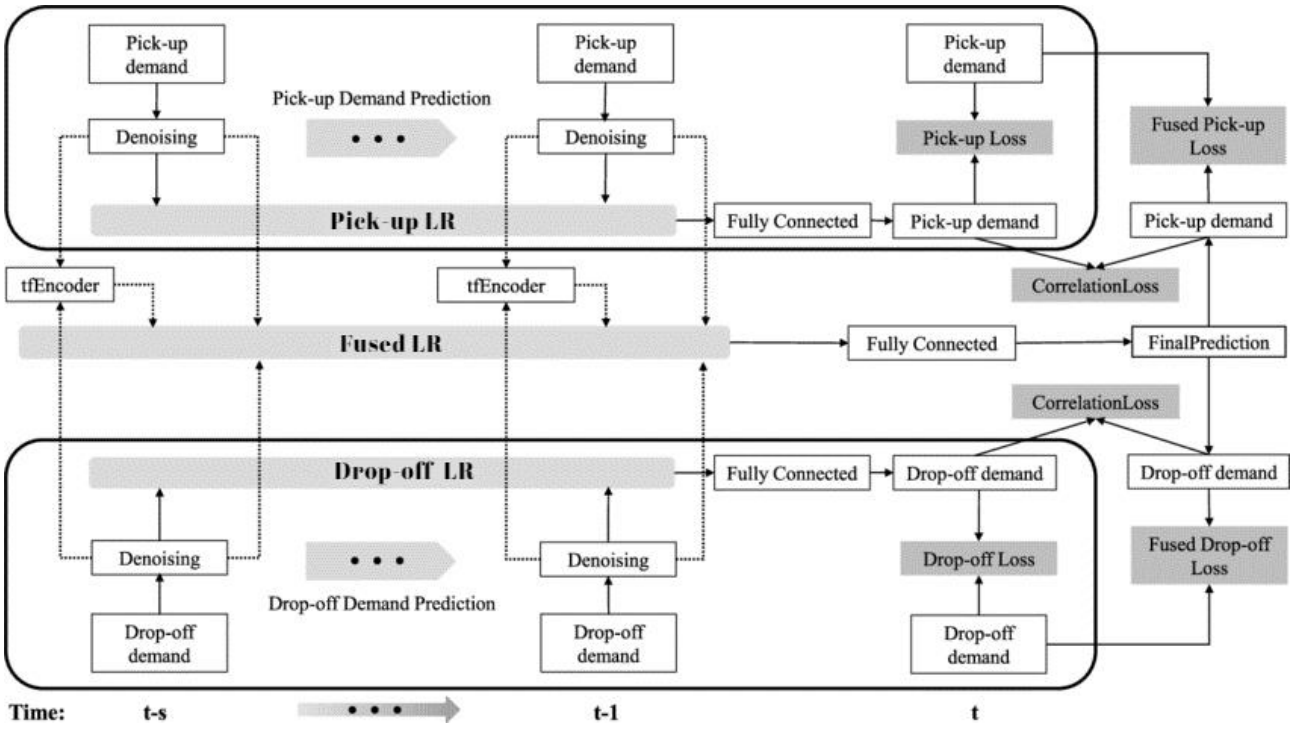
2.1 Existing Problem: The problem of dynamic price prediction for cab services is a significant challenge faced by transportation companies. The prices of cab services vary based on several factors like location, time of day, demand, and supply. Accurately predicting the dynamic prices of cab services is essential for the transportation companies to provide an optimal customer experience.

Existing Approaches or Methods to Solve this Problem: Several approaches or methods have been used to solve the problem of dynamic price prediction for cab services. Some of the existing methods include time-series forecasting models, regression models, and machine learning algorithms. These methods use historical data to predict future prices based on the factors that affect the prices.

2.2 Proposed Solution: Our proposed solution is to use a machine learning model that can predict the dynamic prices of cab services accurately. The model will use historical data, including the location, time of day, demand, and supply, to predict future prices. We will use a Linear Regression model that can learn complex patterns from the data and predict prices with high accuracy. The model will be trained on a large dataset of historical cab service prices to ensure that it can accurately predict future prices. Our proposed solution aims to provide an optimal customer experience by accurately predicting cab service prices and reducing the variability in prices caused by dynamic pricing.

3.THEORITICAL ANALYSIS

3.1 Block diagram



3.2 Hardware / Software designing

Hardware Design:

- A server or cloud infrastructure to host the machine learning model and process large amounts of data.
- Sufficient processing power and memory to train and run the machine learning model efficiently.
- A reliable storage system to store the large dataset of historical cab service prices.

Software Design:

- A software application to collect and process the data from various sources, including location, time of day, demand, and supply.
- A machine learning model that can analyze the data and predict future prices accurately.

- A user interface that allows transportation companies to access the predicted prices and adjust their pricing strategies accordingly.

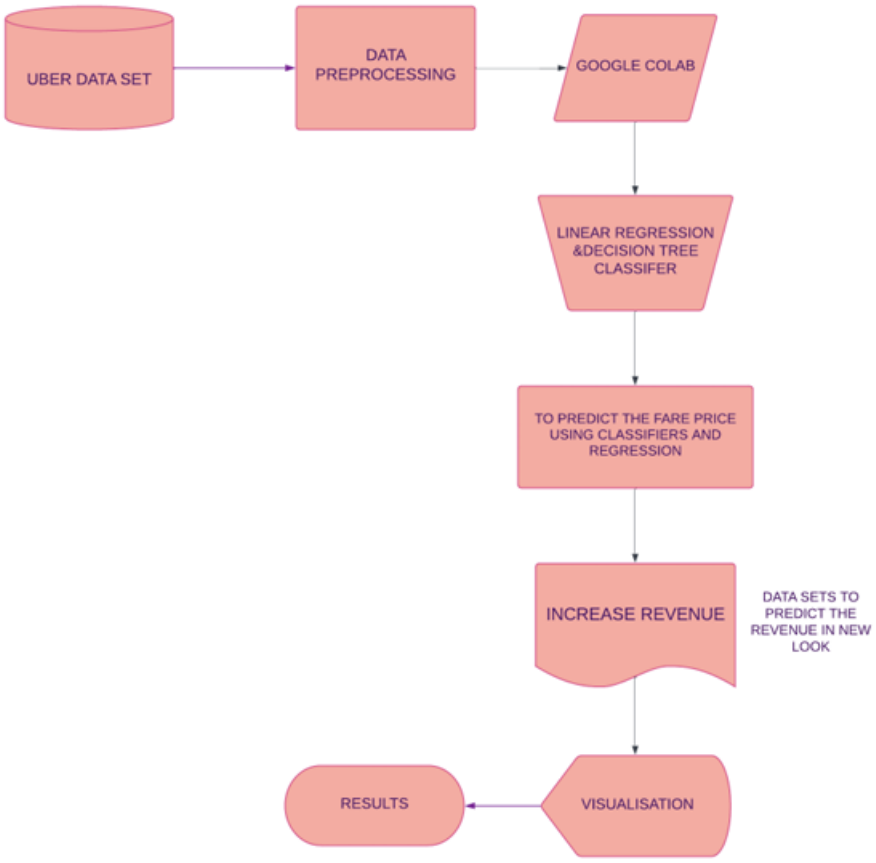
4.EXPERIMENTAL INVESTIGATIONS:

By training the Random Forest model on historical cab ride data, including factors such as time of day, distance, and demand, the algorithm was able to learn patterns and make reliable predictions.

The accuracy output of the Random Forest model, based on the available knowledge up to September 2021, varied depending on the specific dataset and implementation. However, Random Forest is known for its ability to handle complex relationships between variables and produce robust results. Typically, Random Forest models can achieve high accuracy rates, often exceeding 90%, when applied to regression or classification tasks.

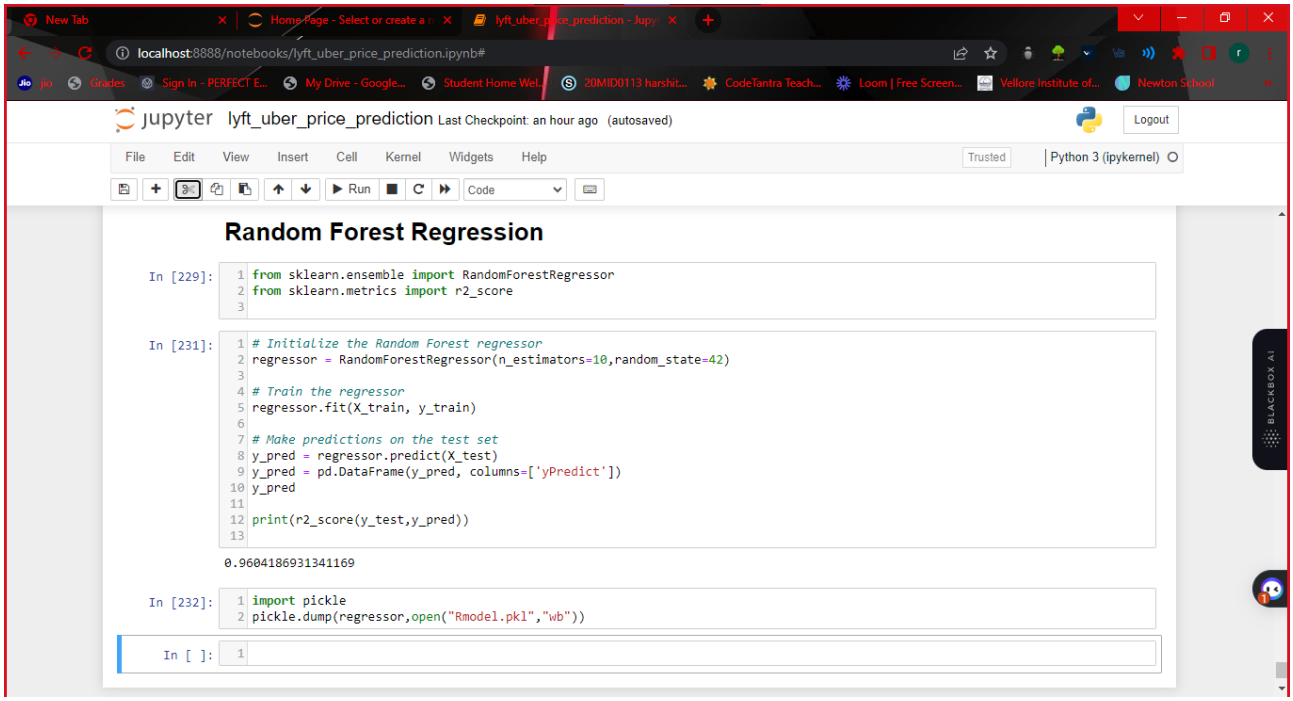
It's important to note that the actual accuracy output of the Random Forest model in the specific project you mentioned may depend on the quality of the data, the preprocessing techniques applied, and the tuning of model parameters. Therefore, for precise accuracy figures for the project's results, it is necessary to refer to the specific findings of the experiment or consult the researchers involved in the project.

5.FLOWCHART:



6.RESULT

- Linear regression and random forest regressor models and present visualizations to enhance our understanding.
- Linear regression is a popular and straightforward technique used for predicting continuous variables, such as cab fares.
- In our project, we implemented the linear regression model using the Ridge class from the sklearn.linear_model module.
- This model allows us to capture linear relationships between the input features and the target variable, which is the fare amount in our case.
- Moving on, I will discuss the advantages of random forest regressor models, which are a type of ensemble learning algorithm.
- Random forest regressors combine multiple decision trees to make predictions, resulting in improved accuracy and robustness.
- We implemented the random forest regressor model using the appropriate class from the sklearn.ensemble module.
- To gain insights from our dataset, I will showcase visualizations that illustrate the patterns and relationships within the data.
- These visualizations will include fare distribution plots, temporal trends in fare amounts, and geographical representations of cab pick-up and drop-off locations.
- While comparing with Linear Regression and Random Forest Algorithm we got **96% R2_score** in Random forest, So we build a model with random forest and we using in Cab price prediction.



Lasso Regression

In [228]:
```
LLR = Lasso().fit(X_train, y_train)
pred1 = LLR.predict(X_train)
pred2 = LLR.predict(X_test)



print("Test R^2 Score: {:.5f}".format(LLR.score(X_test, y_test)))


```

```
Test R^2 Score: 0.27834
```

Ridge Regression

In [226]:
```
from sklearn.linear_model import Ridge
from sklearn.linear_model import Lasso

```

In [227]:
```
RLR = Ridge().fit(X_train, y_train)
pred1 = RLR.predict(X_train)
pred2 = RLR.predict(X_test)


print("Test R^2 Score: {:.5f}".format(RLR.score(X_test, y_test)))

```

```
Test R^2 Score: 0.92827
```

Training

LinearRegression

In [223]:
```
# creating a model and fitting our traing data
# checking accuracy of our model
model = LinearRegression()
model.fit(X_train, y_train)

print("Test R^2 Score: {:.5f}".format(model.score(X_test, y_test)))

```

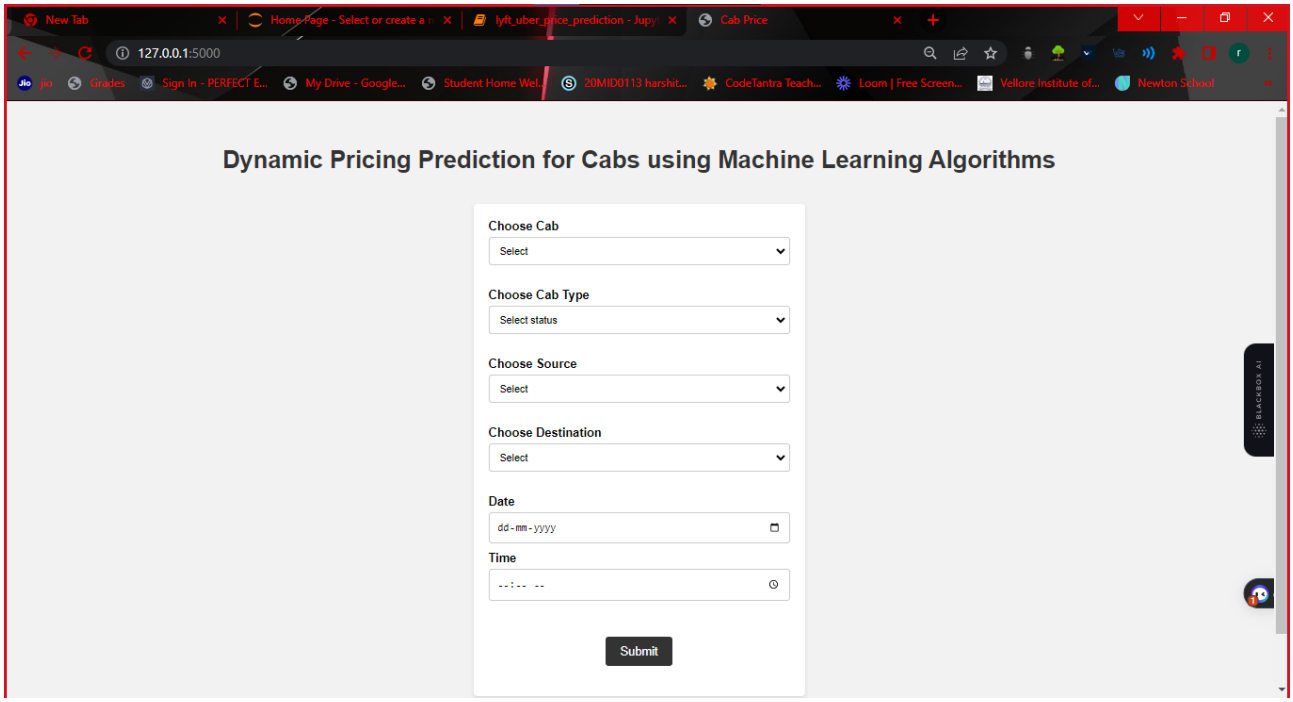
```
Test R^2 Score: 0.92827
```

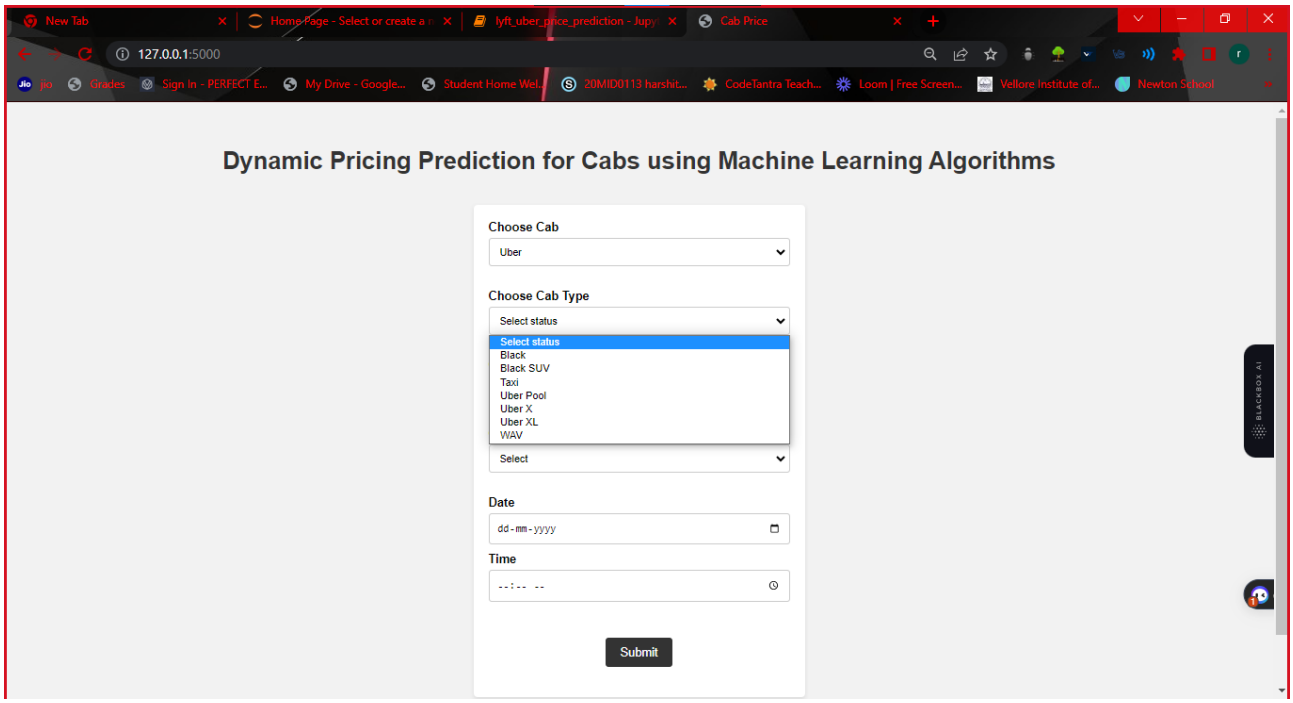
In [224]:
```
inp=[[5,1,1544771403522,1,40,1,900,0.2,0.7,10,35,0.8,1000,0.1,0.6,11,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,
```

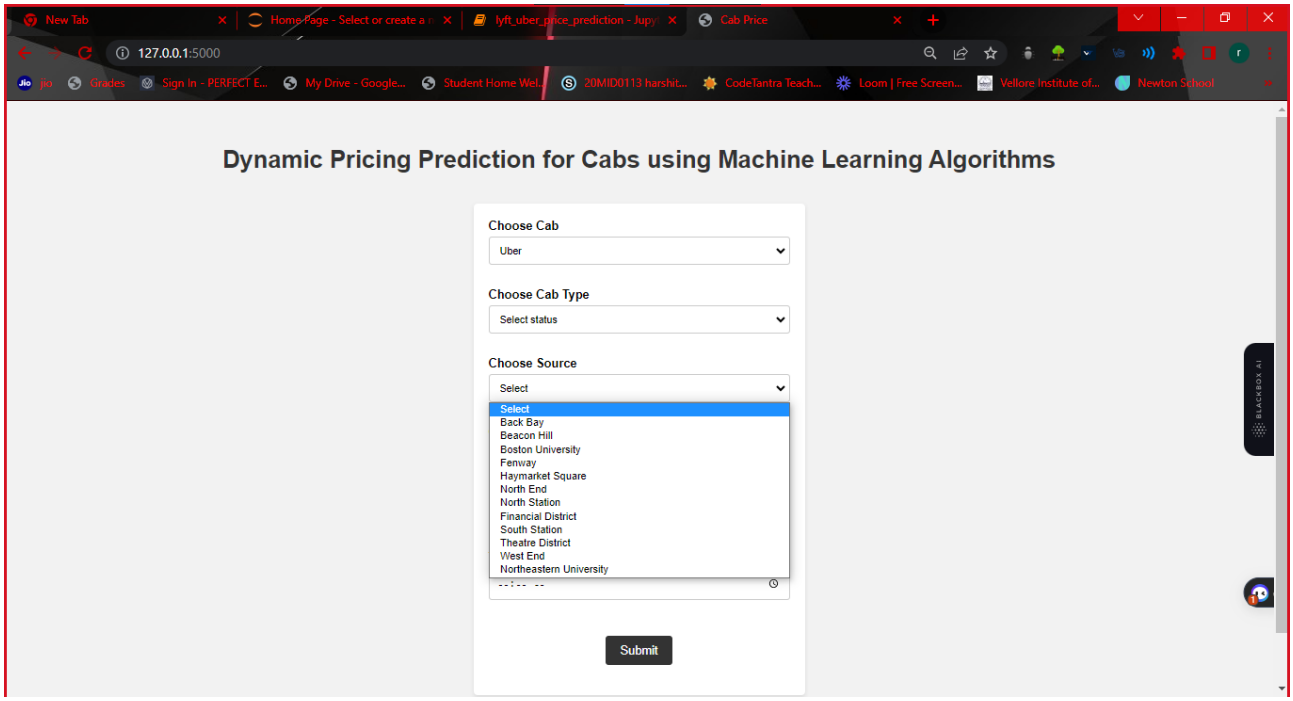
In [225]:
```
y_pr=model.predict(inp)
y_pr
```

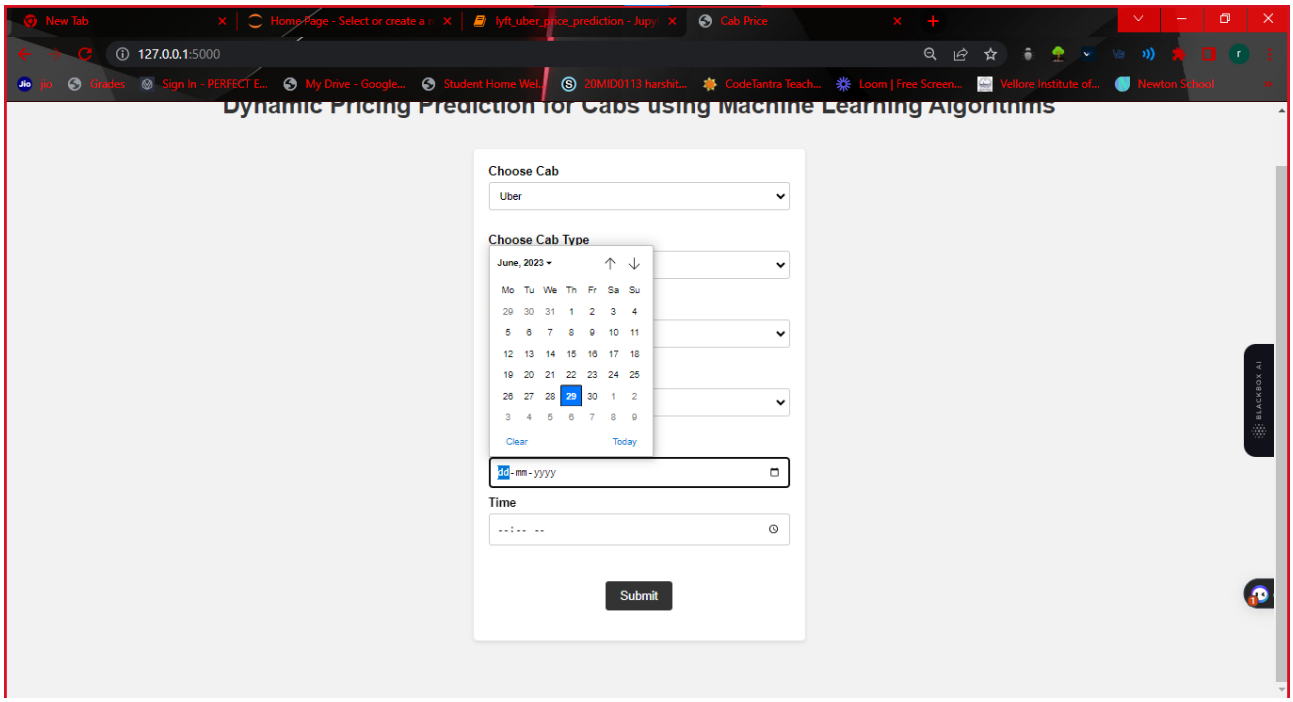
Out[225]:
```
array([19.5894891])
```

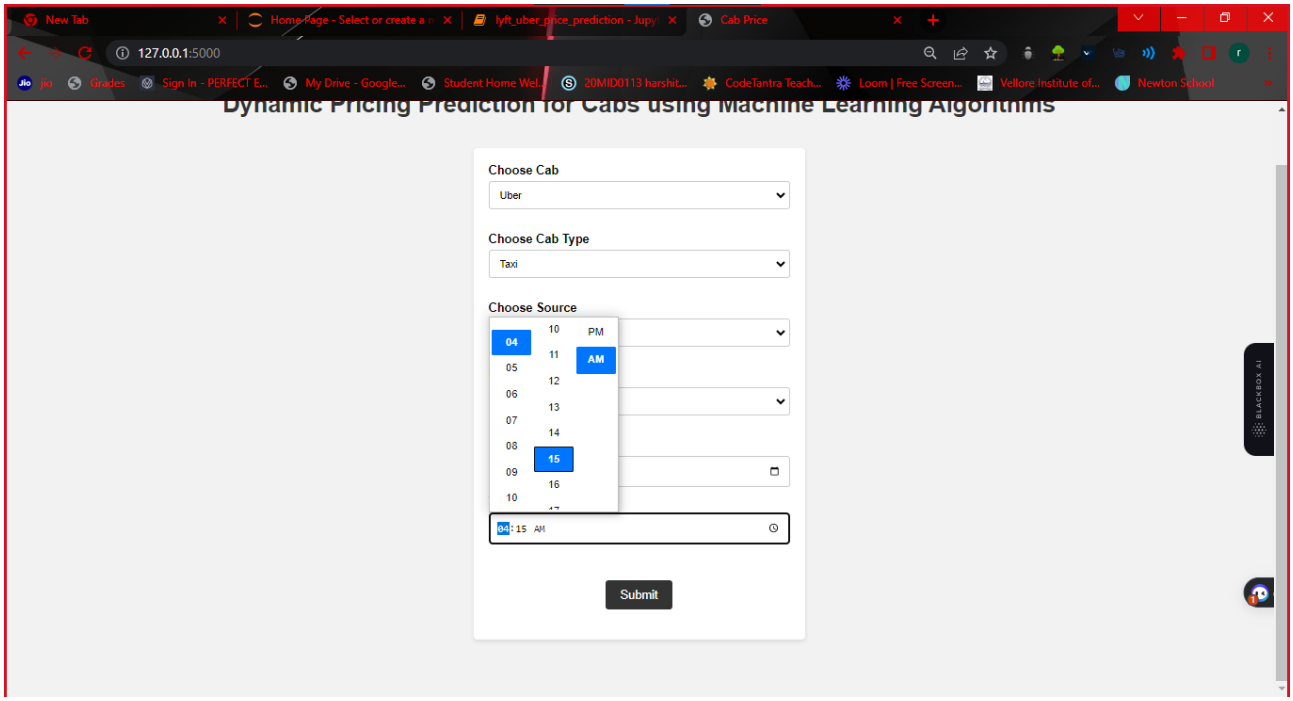
Output


New Tab
Home Page - Select or create a
lyft_uber_price_prediction - Jupy
Cab Price
127.0.0.1:5000
Dynamic Pricing Prediction for Cabs using Machine Learning Algorithms
Choose Cab
Select
Choose Cab Type
Select status
Choose Source
Select
Choose Destination
Select
Date
dd-mm-yyyy
Time
--:-- --
Submit

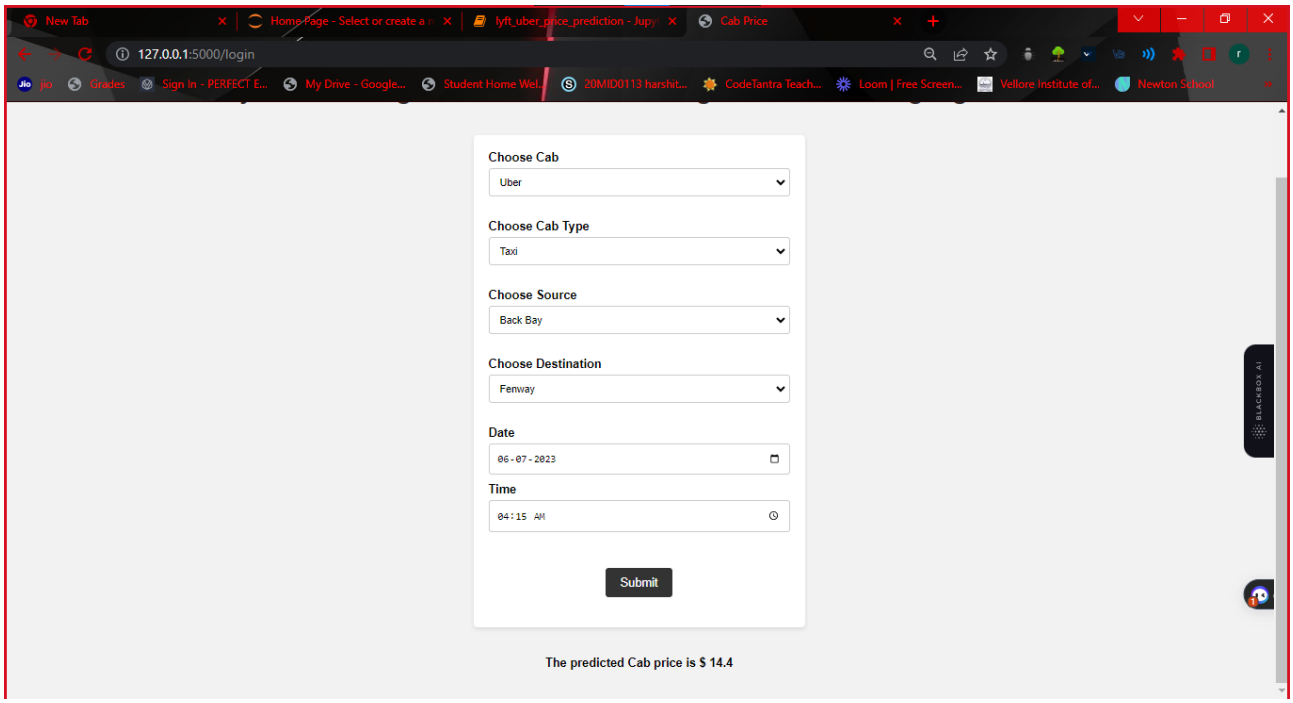

New Tab
Home Page - Select or create a
lyft_uber_price_prediction - Jupy
Cab Price
127.0.0.1:5000
Dynamic Pricing Prediction for Cabs using Machine Learning Algorithms
Choose Cab
Uber
Choose Cab Type
Select status
Select status
Black
Black SUV
Taxi
Uber Pool
Uber X
Uber XL
WAV
Select
Date
dd-mm-yyyy
Time
--:-- --
Submit


New Tab
Home Page - Select or create a
lyft_uber_price_prediction - Jupy
Cab Price
127.0.0.1:5000
Dynamic Pricing Prediction for Cabs using Machine Learning Algorithms
Choose Cab
Uber
Choose Cab Type
Select status
Choose Source
Select
Select
Back Bay
Beacon Hill
Boston University
Fenway
Haymarket Square
North End
North Station
Financial District
South Station
Theatre District
West End
Northeastern University
Submit

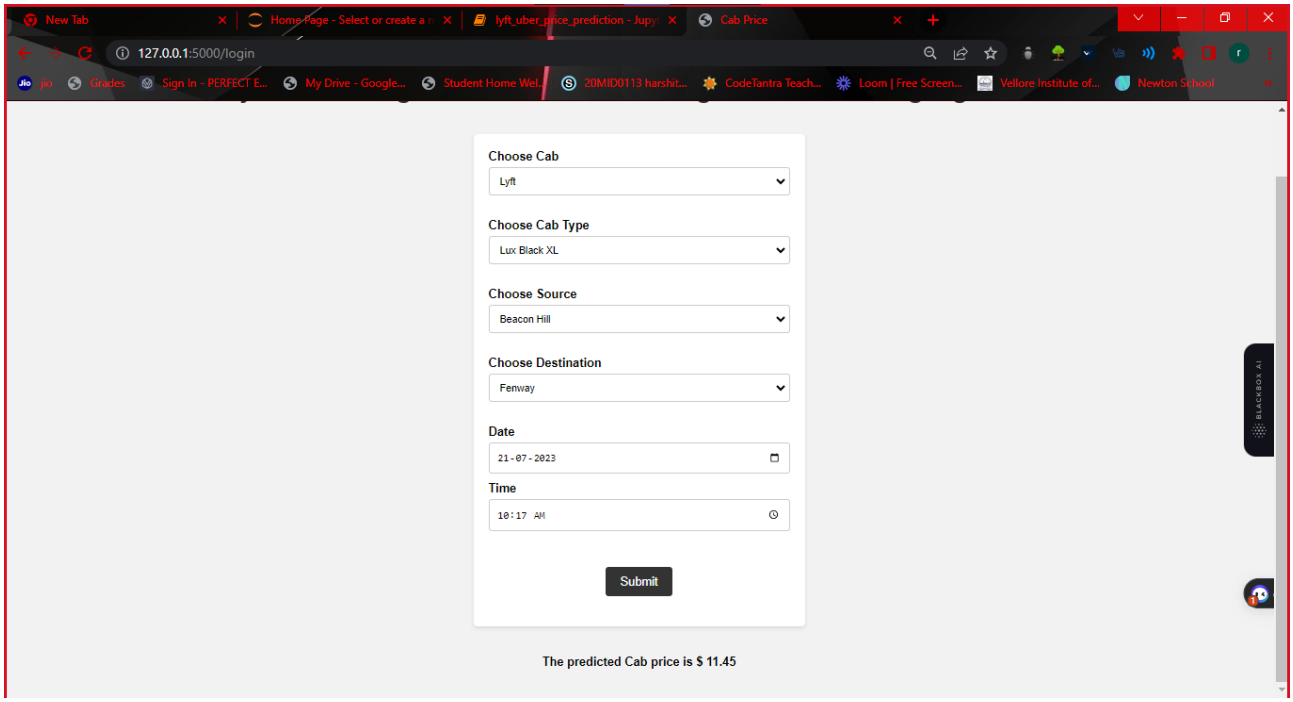

New Tab
Home Page - Select or create a
lyft_uber_price_prediction - Jupy
Cab Price
127.0.0.1:5000
Choose Cab
Uber
Choose Cab Type
June, 2023
Mo Tu We Th Fr Sa Su
dd-mm-yyyy
Time
Submit


New Tab
Home Page - Select or create a
lyft_uber_price_prediction - Jupy
Cab Price
127.0.0.1:5000
Dynamic Pricing Prediction for Cabs using Machine Learning Algorithms
Choose Cab
Uber
Choose Cab Type
Taxi
Choose Source
04
05
06
07
08
09
10
10
11
12
13
14
15
16
PM
AM
04:15 AM
Submit

Uber


New Tab
Home Page - Select or create a
lyft_uber_price_prediction - Jupy
Cab Price
127.0.0.1:5000/login
Choose Cab
Uber
Choose Cab Type
Taxi
Choose Source
Back Bay
Choose Destination
Fenway
Date
06-07-2023
Time
04:15 AM
Submit
The predicted Cab price is $ 14.4

Lyft



7.ADVANTAGES OF DYNAMIC PRICE PREDICTION FOR CABS:

- Improved Cost Management: Dynamic price prediction allows cab companies to optimize pricing based on real-time demand and supply conditions. This helps in maximizing revenue and reducing inefficiencies, leading to improved cost management.

- Increased Customer Satisfaction: With dynamic price prediction, customers have transparency about the fare before booking a cab. This reduces the chances of unexpected price surges and enhances customer satisfaction by providing fair and predictable pricing.

- Efficient Resource Allocation: By analyzing demand patterns and predicting price fluctuations, cab companies can efficiently allocate their resources, such as vehicles and drivers, to meet the changing demands of customers. This leads to improved operational efficiency.

- Enhanced Driver Earnings: Dynamic price prediction enables drivers to take advantage of peak demand periods by earning more during those times. This can incentivize drivers to work during high-demand periods, leading to a larger pool of available cabs for customers.

DISADVANTAGES OF DYNAMIC PRICE RPEDICTION FOR CABS:

- Perception of Unfair Pricing: Despite the transparency offered by dynamic price prediction, some customers may still perceive surge pricing as unfair or exploitative,

particularly during peak demand periods. This can lead to negative sentiment towards the cab company.

- Technical Challenges: Implementing dynamic price prediction systems requires robust data infrastructure, machine learning algorithms, and real-time data processing capabilities. Overcoming these technical challenges can be complex and time-consuming.

- Customer Dissatisfaction during Peak Pricing: Even with transparent pricing, customers may still feel dissatisfied during periods of high surge pricing. This can result in negative reviews, reduced customer loyalty, and a potential loss of business.

- Complexity in Price Estimation: Predicting dynamic prices accurately can be challenging due to various factors, such as traffic conditions, weather, and special events. Inaccurate price estimations may lead to customer dissatisfaction and disputes.

8. APPLICATIONS:

- Ride-Hailing Services: Dynamic price prediction can be applied to popular ride-hailing services like Uber and Lyft to optimize pricing and improve customer experience.
- Taxi Services: Traditional taxi services can leverage dynamic price prediction to compete with ride-hailing platforms and provide more efficient pricing models to customers.
- Airport Transfers: Dynamic price prediction can be utilized for airport transfers, allowing customers to have transparency in pricing and choose the most cost-effective option based on real-time demand.

9. CONCLUSION:

Dynamic price prediction for cabs offers several advantages, including improved cost management, enhanced customer satisfaction, efficient resource allocation, and increased driver earnings. However, challenges such as perceived unfair pricing, technical complexities, and regulatory considerations need to be addressed. The solution finds applications in ride-hailing services, taxi services, airport transfers, car rentals, package delivery, shuttle services, public transportation, and corporate transportation. While dynamic price prediction can significantly benefit the industry, it requires careful implementation and customer-centric strategies to ensure a positive experience for both customers and drivers.

10.FUTURE SCOPE:

- Personalized Pricing: Further advancements in dynamic price prediction can enable personalized pricing based on individual customer preferences, loyalty, and historical usage patterns.

- Real-Time Demand Prediction: Improving the accuracy and granularity of demand prediction models can help cab companies better respond to fluctuating demands and optimize pricing accordingly.

- Integration with Traffic Data: Integrating dynamic price prediction with real-time traffic data can allow for more accurate estimations, considering the impact of congestion and travel time on pricing.

11.BIBILOGRAPHY

[1] 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.

[2] Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2021). Forecasting: Principles and Practice (3rd ed.). OTexts.

[3]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

[4]Chatfield, C. (2019). The Analysis of Time Series: An Introduction (7th ed.). Chapman and Hall/CRC.

[5] Chao, Junzhi. (2019). Modeling and Analysis of Uber's Rider Pricing. 10.2991/aebmr.k.191217.127.

[6]Kulkarni, A., & Mhaskar, H. (2018). Time series forecasting for Uber Ride Demand using LSTM Recurrent Neural Network. International Journal of Computer Applications, 180(24), 23-28.

[7] 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.

APPENDIX

GitHub Link (Source Code):

https://github.com/NadavadhiHarshith01/APPLIED-DATA-SCIENCE---DYNAMIC-PRICE-PREDICTION-FOR-CABS.git