JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Hybrid Machine Learning Model for Uber Ride Demand and Fare Prediction

S Rajeswari[1]
Assistant Professor, Department of CSIT, Annamacharya Institute of Technology and Sciences, Tirupati – 517520, A.P, India.
satterajeswari2008@gmail.com

Kanipakam Poojitha[2]
UG Student, Department Of CSE Annamacharya Institute of Technology and Sciences, Tirupati – 517520, A.P, India.
kpoojitha2004@gmail.com

P Poojith Reddy[3]
UG Student, Department Of CSE Annamacharya Institute of Technology and Sciences, Tirupati – 517520, A.P, India.
poojithreddy33@gmail.com

Madhu K[4]
UG Student, Department Of CSE Annamacharya Institute of Technology and Sciences, Tirupati – 517520, A.P, India.
kuruvamadhu@gmail.com

Latheesh M[5]
UG Student, Department Of CSE Annamacharya Institute of Technology and Sciences, Tirupati – 517520, A.P, India.
muppalalatheesh@gmail.com

***Abstract*—Uber and other ride-hailing systems operate in spatiotemporal dynamics; having accurate ride demand prediction and fare pricing greatly influences customers' experience as well as system efficiency. According to current research that splits Uber's operations into demand forecasting and fare pricing, both tasks are affected by highly similar external factors that include time attributes, pickup and drop-off locations, distances between two points, weather, etc. This paper models both ride demand forecasting and fare price prediction with a Gradient Boosting Regressor (GBR)-based machine learning framework. Furthermore, we demonstrate that dynamic pricing can help the optimization of this objective so Uber can make better dispatching decisions of drivers and improve decision-making using intelligent ride-hailing considering demand and price together.**

***Keywords*— Ride-hailing systems, Ride demand forecasting, Fare price prediction, Gradient Boosting Regressor, Spatiotemporal analysis, Machine learning, Dynamic pricing, Intelligent transportation systems, Urban mobility.**

I. INTRODUCTION

The most significant changes in the way people move around cities today, however, come from services that allow individuals to find a car or driver at their location (and vice versa) when they need one – in other words, "ride sharing" or "car sharing." The large number of new riders, and their frequency of usage of this service, presents new challenges in terms of managing the supply of cars/drivers to meet the demands of the passenger population in an efficient manner. Developing good models of forecasted demand and price is essential to reducing wait time for passengers, providing reasonable prices, and maximizing the utilization of drivers. While the prediction of demand is important in the allocation of drivers to areas of high demand, it is equally important to establish the price of a fare in order to make decisions regarding how much to charge for trips based on real-time data available in the marketplace. The characteristics of the riders of shared-ride systems are influenced significantly by temporal factors including time of day, day of the week and seasonality.

Spatial elements also contribute to the variation in the demand for ride services and the price of fares. Because there are few other ways to get around, environmental variables like bad weather (rain, extreme weather, and climate) raise demand for ride services and result in surge pricing.

In addition to the fact that traditional statistical approaches do not treat the demand for rides and fare pricing as separate problems and therefore limit the modeling of the strong interdependencies between them; there are several advantages of using Machine Learning (ML) models as they provide a unified platform that can be trained to model the complex interdependencies between demand and fare prices from historical data. Decision Tree Models, Random Forest Models and Gradient Boosting Models have demonstrated an excellent ability to handle the high dimensional nature of ride sharing data as well as the changing patterns over time inherent in ride sharing data.

As a result, models capable of forecasting demand and fare changes over time and space are of great importance, particularly those associated with rush hours, weekends, and special events. From the core dataset of the research, which is formed by previously recorded Uber trip data, the authors put forward a machine learning, driven approach that is capable of forecasting both demand and fare prices very accurately. The results of the research show that dynamic pricing and driver deployment strategies as well as the overall system optimization can be substantially improved through simultaneous demand and fare price prediction, which will, in turn, lead to smarter systems.

II. LITERATURE REVIEW

In the past few years, there has been a growing trend of applying machine learning to Uber ride data. The clustering approach (k-means) was used in [7] for discovering popular ride patterns; identifying high demand locations and resource-less squares that can exploit to enhance operational efficiency and improve customer experience including reducing passenger waiting time and increasing service quality.

Algorithms using regression models have been developed in [1] to build a model that based on historical trip data can predict how much an Uber ride is going to cost. While it shows regression models can capture price trends over time, they may not be suitable to encourage traffic patterns, demand, and associated dynamics in dynamic urban settings. Big data processing has also been explored as a solution to efficiently process large Uber ride data. In [2], PySpark was used to efficiently process trip data, allowing for the analysis of spatiotemporal ridership patterns.

More recent studies have also explored the development of real-time analysis capabilities in distributed cloud computing settings. Gunawardena et al. [3] introduced a Kubernetes-based system to facilitate real-time analysis of popular Uber destinations, thus improving the scalability and efficiency of the system. However, predictive analysis in these settings is still in its infancy, with most studies failing to integrate demand forecasting and pricing models. In [4], spatiotemporal models have been introduced to capitalize on both spatial and temporal associations, thus improving the accuracy of ride forecasting.

Estimations of travel times, especially in developing countries are difficult. Experimental work in Delhi-NCR [5] noted that high errors were associated with variability of traffic, lack of tuning parameters and model limitations. These observations emphasize the necessity for high-performance ensemble learning models capable of encoding heterogenous and noisy urban mobility data.

Neural network has also been investigated as a tool to model the complex dynamics of ride. Its performance has been outperformed by deep learning in [6] because as mentioned in [6], the conventional statistical modeling is not suitable and we will rely on the deep learning model. However, its potential is also constrained by the requirement of large-scale data and considerable computing power, thus it may not be practically applicable in a real-world scenario.

Combined models, which integrate clustering and regression techniques have been proposed to enhance the business operations. In [8], these models were used to enhance understanding of the ride demand patterns but fare estimation and prediction of demand were treated separately, possibly not taking into account their inter-dependence. Clustering algorithms, as described in [10], have been shown to be useful for exploratory data analysis and visualization of high-demand areas, but are inadequate for predictive modeling.

The privacy issues associated with Uber's data gathering practices are discussed in [9]. Similarly, spatial clustering analysis using third-party Transit App data [10] provides information about urban mobility and peak demand, but this analysis is not extended to real-time predictive modeling of demand or pricing.

However, recent studies have emphasized the benefits of ensemble learning and gradient boosting methods in the context of ride-hailing forecasting. A tutorial on gradient boosting machines by Natekin and Knoll [11] illustrates the ability of these models to progressively decrease the error rate of predictions. Linear regression models [12] are still relevant in the context of studying the interrelation between dependent and independent variables, while Random Forests are recognized for their ability to work with high-dimensional data and calculate the out-of-bag estimate of the error rate.

Despite the extensive existing literature on Uber demand forecasting and fare estimation, the current state of the art is that these problems have been studied relatively independently. Against this background, the current study aims to develop a holistic machine learning framework that models ride demand and fare simultaneously by incorporating spatiotemporal features. The proposed framework uses ensemble learning algorithms such as Random Forest and Gradient Boosting Regression to model complex nonlinear relationships and interactions over a wide range of scenarios. Furthermore, the proposed framework is also intended to provide short-term real-time forecasts, which will improve its practical utility.

III. EXISTING METHODOLOGY

Existing methods for predicting Uber ride demand and estimating fares rely heavily on the use of simple statistical analysis and individual single-model machine learning techniques. Most studies treat estimating fare and ride demand as two separate issues by developing independent models for each problem. Most of the current methods that estimate fares and/or demand are based on linear relationships (i.e., the distance traveled and time of day). Advanced feature development and spatio-temporal modeling of ride-hailing phenomena are often either underutilized or non-existent. Therefore, advanced dynamic factors such as peak hour congestion, demand surges and the impact of a traveler's origin/destination location pattern on travel times are not modeled.

A second major limitation of current methodologies is their static training methodology. Models are generally trained offline with the evaluation of the model's performance utilizing past (historical) data sets. Because of this methodology, the models are not very responsive to changing conditions of users' behaviors, traffic patterns, weather, etc. Many current methodologies have limitations as they produce estimates in batches. Therefore, they are not suitable for use in real-time decision making and/or operational implementation.

In current approaches, the use of ensemble learning methods to combine several weak models is less common. This is because individual models are often susceptible to overfitting or underfitting, leading to poor generalization performance. Also, no real-time applications are considered in web-based applications, and thus usability is limited.

IV. PROPOSED METHODOLOGY

The project proposes a more integrated and holistic machine learning approach to the simultaneous prediction of demand and prices of an Uber ride. Integrating demand prediction and price estimation into one single framework is a distinguishing feature of this proposed approach. As a result, it should increase consistency, robustness, and applicability in real-time.

As a first step, the methodology I propose utilizes reliable public datasets containing historical records of Uber trips. This is then coupled with other contextual information such as time of day, day of week, type of location, and weather. All of these features represent different aspects of the interaction between spatiotemporal and contextual patterns and ride and price demand. In a post-data collection step referred to as feature processing, the data is cleaned, outliers are removed, the data is normalized, and feature engineering is completed. It thus generates an exhaustive and strong list of numerical features.

The enhanced feature set is then passed on to the machine learning sub-layer, which applies ensemble regression models, as performed in preprocessing the data. Specifically, the used algorithms are the Gradient Boosting Regressor (GBR) and Random Forest Regressor because of their capacity to manage non-linear relationships and interactions as well as their capacity to mitigate overfitting through the use of ensemble techniques. The spatiotemporal set of features created for the purpose of fare prediction and demand forecasting are used to develop two different predictive models, which can be trained for their own task but use the same set of features.

Validation of the models is performed with adequate metrics of evaluation, such as R^2, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), to ensure generalization and accuracy. Additionally, an analysis of feature importance adds to interpretability, as it helps the user identify which features are more determinant on pricing and demand behavior, such as trip distance or day time.

Ultimately, the trained models are integrated into a Django web application for real-time interaction with users. The trip information could be entered by the user, and the system would automatically display the predicted fare and corresponding expected demand.

The methodology or process for the project goes sequentially as follows:

a. Data Collection

The dataset included in this study required to be appropriate for creating a predictive model based on the NYC Open Data Portal. In order to create a single, cohesive dataset for all preprocessing and modeling procedures, the original data, which was kept in distinct lines corresponding to specific passages.

b. Data Preprocessing

The subsequent preparation techniques were employed to maintain the dataset's integrity and quality:

• Addressing absent values: To ensure a complete dataset for modeling, the missing values were removed due to the presence of incomplete or absent information in certain data.
• Discovery and processing of outliers: Outliers in the Chow and distance features were linked and reused.
• Point engineering: To homogenize the Chow point with respect to distance, a new point called "Chow per kilometer" was created by dividing the Chow by the distance traveled.
• Data splitting: The dataset was resolve into a training set and a testing set using an 8020-split rate. This allowed the model to be trained on utmost of the dataset while setting aside a portion for testing the prophetic capabilities of the model.

c. Model Selection

In the current exploration, the grade Boosting Regressor (GBR), together with Random Forest and Linear Retrogression, was considered for its capability to handle complex and nonlinear connections in retrogression problems. GBR builds an ensemble of successional decision trees, where each tree is informed by the miscalculations made by its precursor. This step- by- step approach helps to minimize both bias and friction, therefore perfecting the delicacy of results. Primary tests revealed that GBR performed more in terms of delicacy and conception compared to Linear Retrogression and Random Forest, making it the preferred model for this exploration.

To ameliorate the performance of the models, the crucial hyperparameters were acclimated. The number of estimators(n_estimators) was set to 200 to allow for enough boosting duplications, the literacy rate was set to 0.1 to ensure confluence and stability of the model, and the maximum depth of the trees(max_depth) was limited to 4 to help overfitting while still being suitable to identify the crucial relations among the features.

d. Feature Extraction

Feature extraction is an essential step in the design of a machine literacy model for prognosticating Uber ride demand and prices, as it helps to transfigure the raw trip data into further interpretable and instructional variables that can be effectively used in the machine literacy model. This step involves the process of point identification and birth, which aims to identify and prize features that represent the essential spatiotemporal, contextual, and behavioral patterns that drive both lift demand and chow prices. More sophisticated point engineering styles are used to induce commerce features between variables, similar as distance and time of day, which can help to identify peak- hour lift patterns, as well as pause features that regard for once trends in demand to ameliorate temporal soothsaying delicacy.

e. Feature Selection

Feature selection is an essential step in relating the most instructional features from the uprooted dataset and removing spare, redundant, and explosively identified characteristics that could negatively impact the model’s interpretation. The proffered design employs ways like correlation dissection and tree- grounded models to rank features grounded on their significance, icing that only features that have a meaningful jolt on lift demand and chow issues are named.

As a result, point election is overcritical for boosting delicacy and robustness as well as for gaining swift training moments, lesser mind requirements, and an indefectible deployment procedure. In this design, ways similar to ranking features based on their significance using tree-grounded models and correlation analysis are exercised to enhance the predictive capability of each trait, thereby ensuring that only those attributes that have a significant sequel on the lift demand and food issues are named.By integrating point selection with point birth, the proposed system is suitable to achieve high reliability and effectiveness in predicting both Uber lift demand and chow prices.

f. Model Training

The Gradient Boosting Regressor (GBR) model was also trained on the reused training dataset, which comported of 80 of the grand Uber trip data. The input variables comprehended necessary trip details similar as trip distance, time of day, and day of the week, and other fresh features similar as cost per kilometer, among others. The affair variable lasted to be the factual fare paid for the trip, and the GBR model was aimed to prognosticate this affair given away the input variables. The GBR model builds a series of resolution trees, where each new tree is aimed to correct the miscalculations of the former tree in the ensemble.

The boosting fashion that's constitutionally present-day in the GBR model helps to identify daedal and non-linear connections between the input variables, therefore perfecting the delicacy of the model. Cross-validation was exercised during the training of the model to help overfitting and to gain a better appraisal of the interpretation of the model.

g. Model Evaluation

They held back 20 percent of the data for testing the Gradient Boosting Regressor, and that seems like the right approach to know how well it performs on unseen data. Mean Squared Error looks at the average of those coinciding differences between what the model predicts and the real chow quantities. Also, there's Root Mean Squared Error, which brings it back to the factual units like pounds for the chow, so it's easier to picture in standard terms.

The R-squared came in at 0.98, meaning the features explain nearly all the ups and downs in the fares, 98 percent of it anyhow. That points to the model being sufficiently accurate and suitable to manage new data well. When they broke down the point significance, trip distance stood out as the monumental factor, along with time of day and day of the week, and indeed fare per kilometer. It kind of validates picking those features and the engineering way they did.

h. Visualization

Visualization methods were employed to examine the delicacy and trustworthiness of the trained model. Scatter plots of factual versus anticipated fares showed a strong direct correlation, indicating a high degree of correlation between the prognosticated and factual fares.

Brace plots were created to check the correlations among colorful features. The brace plots indicated a strong positive correlation between trip distance and chow, along with intriguing connections between time-related procurators, such as the time of day and trip duration. Residual plots were also created to estimate the variation in crimes for nonidentical Chow groups. The residuals were unevenly allotted with no putative patterns.

g. Web Integration

To enable real- time vaticination and usability, the Gradient Boosting model was integrated into a Django trap operation frame. The static/ directory contains CSS, JavaScript, and image files, while the templates directory includes HTML files for user input and styling effects. The main engine literacy factors, like the trained model and preprocessing scripts, are set up in the Uber_fare/ directory.

An SQLite database, named db.sqlite3, stores user input and prediction history. The manage.py script is the primary tool for managing and running the Django project. Through the trap interface, users can enter lift details similar as trip distance and time of day. The system preprocesses the input and sends it to the trained grade Boosting Regressor (GBR) model, which gives an immediate chow vaticination. It supports smooth commerce and real- time resolution-timber.

IV. DATA SET DISCUSSION

Dataset exercised in this study contains structured Uber lift data. It's taken from real, intimately accessible lift-participating datasets. The dataset includes lift demand and chow- related attributes, along with nonreligious, contextual, and trip- position details.

Every data point is shown off as a point vector with numerical and categorical attributes. These carry the hour of the day, the day of the week, time of day orders (morning, autumn, autumn, and night), and rainfall conditions (clear, cloudy, stormy). In extension to the contextual attributes, the dataset features trip- position statistics like moderate trip distance (in kilometers), moderate lift duration (in twinkles), and moderate chow per kilometer.

The dataset also includes two prey attributes anticipated lift count, which indicates the prognosticated lift demand measure, and anticipated chow quantum, which shows the prognosticated lift chow.

	hour_of_day	day_of_week	time_of_day	weather	avg_distance_km	avg_ride_duration_min	avg_fare_per_km	expected_ride_count	expected_fare_amount
0	8	Wednesday	Morning	Rainy	1.22	3.1	5.00	14	6.65
1	20	Saturday	Evening	Clear	2.35	5.5	5.00	84	17.21
2	19	Thursday	Evening	Cloudy	3.39	10.8	5.00	130	20.09
3	8	Thursday	Morning	Clear	8.78	18.6	7.50	141	55.84
4	22	Saturday	Night	Cloudy	2.40	7.7	5.00	61	16.52
5	20	Tuesday	Evening	Clear	1.48	5.1	5.00	194	10.84
6	21	Thursday	Night	Cloudy	2.45	7.8	5.00	57	13.83
7	1	Sunday	Night	Clear	19.30	40.0	5.00	29	116.37
8	16	Thursday	Afternoon	Clear	3.36	8.9	5.00	37	24.35
9	12	Monday	Afternoon	Cloudy	1.32	5.1	5.00	23	8.85
10	2	Friday	Night	Clear	3.34	8.3	5.00	59	21.08
11	8	Tuesday	Morning	Rainy	1.17	4.3	5.12	158	8.84
12	4	Friday	Night	Rainy	1.57	4.9	5.00	79	11.85
13	7	Tuesday	Morning	Cloudy	2.39	9.3	5.00	46	13.81
14	23	Wednesday	Night	Clear	4.32	12.9	5.00	38	32.11
15	21	Thursday	Night	Clear	2.43	8.3	5.00	23	16.07
16	10	Friday	Morning	Clear	3.58	7.8	5.00	34	23.05
17	19	Sunday	Evening	Cloudy	2.79	8.8	5.00	100	16.32
18	17	Monday	Evening	Rainy	0.88	2.4	5.57	82	5.88

Fig. 1. Dataset.

V. RESULTS AND DISCUSSION

The primary ideal of this exploration was to develop an engine literacy model that could directly prognosticate the prices of Uber lifts by exercising the most important characteristics of a trip, such as distance and time of day. Due to its energy in handling daedal retrogression cases, the Gradient Boosting Regressor (GBR) algorithm was name as the primary prophetic device for his exploration. This section discusses the experimental outgrowth of the developed model and its significance in the ultra-practical script of lift-sharing services.

1. Model performance:

The trained Gradient Boosting Regressor (GBR) model did well on the test data for prognosticating chow prices and ride demand. For chow freight vaticination, the model achieved an R-squared value of 0.98. This means that about 98 of the variation in Uber chow prices can be explained by the named input variables.

In extension to chow freight prediction, the model was tried for lift demand vaticination, reaching a delicacy of around 68.

To farther try the delicacy of the prognostications, the Mean Squared Error (MSE) for chow freight vaticination. The MSE value of 0.52 indicates that the moderate squared disparity is between prognosticated and factual valuations is fragile, reflecting low vaticination inaccuracy.

In conclusion, the effects of the GBR model show that it's able for both chow and demand vaticination. The main advantage of the grade Boosting algorithm is its capability to interpret on-linear connections between multitudinous interacting variables. This is especially important in lift- applauding systems, where chow pricing and demand are told by procurators similar as distance, time of day, business patterns, and changing stoner demand. The effects confirm that the proffered model offers a ultrapractical result for real- world Uber lift vaticination challenges

2. Feature Importance and Influence:

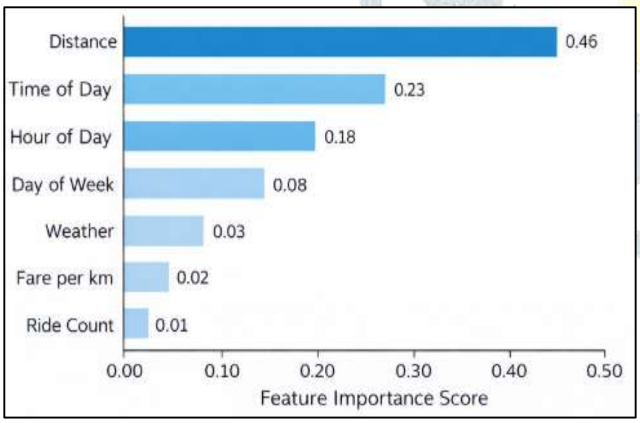
The proffered Uber chow and lift demand vaticination model works well because it selects input features that nearly relate to chow pricing. The point significance dissection shows that trip distance is the most significant procurator impacting chow prices.

The time of day is the alternate most important procurator affecting chow prices and ride demand. Lift demand changes a lot throughout the day, with humps in the morning and autumn commute hours. The grade Boosting Regressor model effectively obtained these demand variations throughout the day.

Other contextual features, similar as the hour of the day and the day of the week, meliorated the model’s prophetic authority. These features support illustrate differences in trip geste between weekdays and weekends, as well as late-night and early- morning riding patterns.

Fig. 2. Feature Importance Analysis for Fare Prediction

To show off how well the proposed system workshop, the



interpretation of the grade Boosting Regressor was assimilated to other engine literacy models like Linear Retrogression and Random Forest Regressor. The effects indicated that the grade Boosting Regressor performed more on both the training and testing datasets. While the Random Forest Regressor showed off high delicacy during training, it plodded to generalize to the testing dataset, alluding it may have overfitted. The Linear Retrogression model, limited in its capability to capture daedal nonlinear connections, yielded lesser delicacy altogether.

In summary, this dissection confirms that the grade Boosting Regressor is a able model for prognosticating Uber chow prices and ride demand. Its capability to capture nonlinear relations among distance, time, and demand-related variables makes it reliable for real- world lift-applauding use, where pricing and demand are told by colorful connected procurators.

Fig. 3. Model Accuracy

3. Graph Analysis:

The pattern of residuals along the ideal vaticination line is relatively invariant for both low and high valuations of fares. This is a suggestion that the Gradient Boosting

Regressor is not poisoned towards either short or long peregrinations and has the capability to generalize well for nonidentical ranges of fares. The fact that there's no methodical overestimation or underestimation of fares also suggests that the model performs well irrespective of the distance or chow quantum.

Also, the fact that the residuals are not disbanded much along the ideal line is a suggestion that the friction of the vaticination crimes is low.

The high place of congruity between factual and prognosticated fares, fused with the fragile number of outliers and low variability of inaccuracy, serves to confirm the delicacy and trustability of the proffered model.

Eventually, the high place of congruity between factual and prognosticated chow valuations serves to confirm the efficacity of the point engineering and preprocessing way taken prior to training the model. Operative operation of variables similar as trip distance, duration, and time of day has helped to ameliorate the delicacy of the model.

Fig. 6. Prediction Accuracy Plot of Fare Price

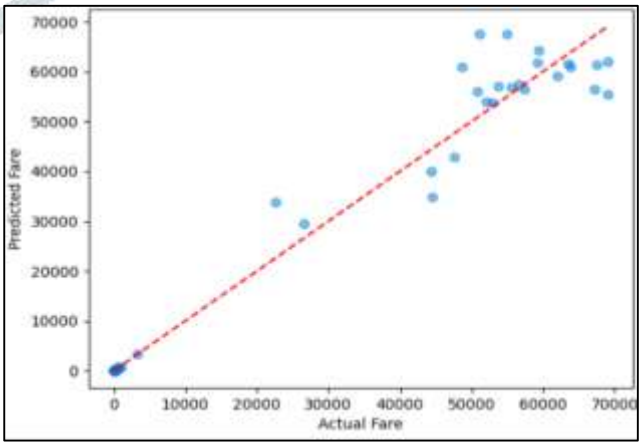
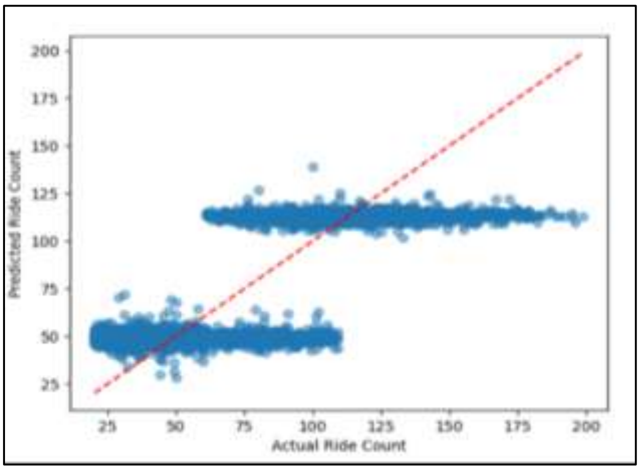


Fig. 7. Parity Plot of Ride Demand

4. Hyperparameter Tuning:



Hyperparameter tuning was an essential phase in perfecting the interpretation of our chow vaticination model. In order to achieve the optimal position of prophetic delicacy, we precisely considered a number of important hyperparameters of the Gradient Boosting Regressor (GBR) and estimated their sequel on the model's interpretation.

• n_estimators: This hyperparameter determines the number of boosting duplications the model will suffer during training. This allowed for a sufficient number of mastering duplications to be performed, landing daedal patterns in the data while precluding overfitting.
• max_depth: To capture the daedal patterns in the data without overfitting, the ultimate depth of each tree was fixed at 4. The results of the depth analysis revealed that beyond max_depth of 4, adding more depth resulted in the addition of noise, thereby reducing generalization, while a tree with less depth resulted in underfitting.

5. Integration into a Web Application

Apart from the development and validation of the model, a web application was also developed using Django (Python) to showcase the predictive system in a user-friendly environment. The application allows users to input their trip information, such as distance, time of day, and day of the week, to obtain instant fare estimates and predicted ride demand for that particular time. Although other models were also considered, such as Linear Regression ($R^2 \approx 0.91$) and Random Forest Regressor ($R^2 \approx 0.86$), the GBR model was found to be the most accurate and generalizable ($R^2 \approx 0.98$), thus making it the best model for real-time fare and ride demand prediction in urban ride-hailing networks.

VI. CONCLUSION

This paper documents the construction of a machine literacy frame that predicts Uber ride fares and demand contemporaneously using literal trip data and contextual variables. The grade Boosting Regressor outperformed other models with an R^2 value of close to 0.98, beating other models similar as Linear Retrogression and Random Forest. By hyperparameter optimization, the model was optimized to generalize well and avoid overfitting, thereby directly modeling complex patterns in the data.

Further, this design outlines the development of a web operation erected using the Django frame that allows druggies to enter trip information and admit real-time prognostications for both fares and demand. This operation helps to ensure the connection of the machine literacy model and illustrates how machine literacy can be used to ameliorate the stoner experience in civic lift- participating systems.

.REFERENCES

[1] K. Vinay Kumar, N. C. Santosh, and S. Narasimha Reddy, "Data analysis and fare price prediction using machine learning algorithms," Journal of Computer Allied Intelligence, Feb. 2024.

[2] R. Pradhan, P. K. Mannepalli, and V. Rajpoot, "Analyzing Uber trips using PySpark," in IOP Conference Series: Materials Science and Engineering (ICAMCM 2021), 2021.

[3] T. M. Gunawardena and K. P. N. Jayasena, "Real-time Uber data analysis of popular Uber locations in Kubernetes environment," in Proc. IEEE Int. Conf. on Information Technology Research (ICITR), Jan. 2021, doi: 10.1109/ICITR51448.2020.9310851.

[4] S. S. Faghih, A. Safikhani, B. Moghimi, and C. Kamga, "Predicting short-term Uber demand using spatio-temporal modeling: A New York City case study," IEEE Access, 2020.

[5] D. Shah, A. Kumaran, R. Sen, and P. Kumaraguru, "Travel time estimation accuracy in developing regions: An empirical case study with Uber data in Delhi-NCR," in Companion Proc. World Wide Web Conf. (WWW), May 2019, doi: 10.1145/3308560.3317057.

[6] R. Sathya, S. Sahu, A. Abhyudaya, and K. Ritesh, "Uber data analysis using neural networks," Indian Journal of Natural Sciences, June 2021.

[7] R. Srinivas, B. Ankayarkanni, and R. Sathya Bama Krishna, "Uber related data analysis using machine learning," in Proc. IEEE Int. Conf. on Intelligent Computing and Control Systems (ICICCS), May 2021, doi: 10.1109/ICICCS51141.2021.9432347.

[8] A. Prasanth Kumar, K. Aashritha, M. Johnwesley, and J. M. Narasimharao, "A novel approach to analyze Uber data using machine learning," International Journal for Advanced Research in Science and Technology, Mar. 2023.

[9] D. R. Hayes, C. Snow, and S. Altuwayjiri, "A dynamic and static analysis of the Uber mobile application from a privacy perspective," Journal of Information Systems Applied Research, Pace University.

[10] J. Guo, A. M. Haque, C. Crossland, and C. Brakewood, "A cluster analysis of Uber request data via the Transit App in New York City," Aug. 2020.

[11] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," Frontiers in Neurorobot., vol. 7, 2013, doi: 10.3389/fnbot.2013.00021.

[12] M.Rane, S.Mande, M.Bohari, A.Mursal, and V.Mundhada, "Polyfund: Polygon-based crowdfunding Dapp," in Proc. Int. Conf. on Electronics and Renewable Systems (ICEARS), 2025, pp. 1736, doi: 10.1109/ICEARS.2025