International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 3 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

UBER RIDES PREDICTION USING MACHINE LEARNING

1Kuchipudi Akhila, 2 JA Paulson

2Associate Professor, Department of Computer Science & Engineering, Varaprasad Reddy Institute of Technology

1M. Tech Students, Department of Computer Science & Engineering, Varaprasad Reddy Institute of Technology

ABSTRACT

Accurately predicting ride demand is critical to improving Uber performance and increasing customer satisfaction. This paper investigates machine learning methods to predict Uber ride requests by analyzing historical data including time, location, and other factors such as weather. After preprocessing and feature engineering, various models (linear regression, decision trees, random forests, and GBR) are evaluated for their prediction performance. GBR emerged as the best model, achieving over 99% accuracy due to its ability to handle complex data and capture nonlinear relationships. This study highlights the importance of physical and environmental features in improving prediction accuracy, allowing Uber to allocate drivers more efficiently and reduce passenger waiting times. These findings provide insight into demand forecasting, improved management, and service quality of ridehailing platforms.

Keywords— Uber, ride demand forecasting, machine learning, forecasting, GBR, time series, previous data, infrastructure, distribution, impact of weather forecasting

I. INTRODUCTION

In recent years, ride hiring services like Uber have revolutionizes urban transportation by providing on-demand transportation. With millions of daily users, Uber faces the challenge of balancing ride and demand to reduce ridership wait times and maximize driver utilization. Forecasting ride demand is critical to Uber's efficiency because it allows the platform to allocate drivers to high demand areas, improve route planning, and reduce service fees.

Demanding efficiency not only increases customer satisfaction, but also ensures drivers achieve results with minimal downtime. Spatial and environmental variables, Time factors have a significant impact on passenger demand patterns. For example, demand is higher during peak times and holidays.

Also, spatial factors such as location and proximity to popular locations (e.g. airports, shopping malls) can affect the number of trips demands in certain areas. Weather conditions such as rain, snow, and temperature also play an important role, as outdoor travel options are limited and severe weather conditions often increase vehicle demand. The problem usually arises when dealing with large datasets with relationships between different columns.

In contrast, machine learning provides a more powerful solution to demand forecasting by capturing complex patterns in historical data. We can increase the accuracy of demand forecasting by using methods such as random forests, decision trees, and gradient boosting algorithms (like

XGBoost). This model can detect business interactions and adapt to demand changes in real time based on data, providing more efficient and effective solutions. The forecasting application uses historical data including time, location, and environmental characteristics. Various models have been developed and evaluated to determine the best

algorithm for predicting passenger demand. Through this study, we emphasize the importance of prior knowledge, architectural design, and model selection to achieve high accuracy. Ultimately, the findings will provide insights into how machine learning can help Uber improve its operations, leading to more profitable distribution and improved customer experience.

II. LITERATURE SURVEY

[1] Rishi Srinivas, B. Ankayarkanni, R. Sathya Bama Krishna, "Uber Related Data Analysis using Machine Learning", IEEE, 26 May 2021, DOI: 10.1109/ICICCS51141.2021.9432347

Abstract: This paper presents a comprehensive analysis of Uber-related data utilizing machine learning techniques. The study aims to uncover patterns and insights from the data to enhance ride-sharing services. Various algorithms are applied to assess their effectiveness in predicting demand and optimizing operations. The findings suggest that machine learning can significantly contribute to improving the efficiency and reliability of ride-sharing platforms.

[2] Sabiheh Sadat Faghih, Abolfazl Safikhani, Bahman Moghimi, Camille Kamga, "Predicting Short-Term Uber Demand Using Spatio-Temporal Modeling: A New York City Case Study"

Abstract: The demand for e-hailing services is growing rapidly, especially in large cities. Uber is the first and popular e-hailing company in the United States and New York City. A comparison of the demand for yellow-cabs and Uber in NYC in 2014 and 2015 shows that the demand for Uber has increased. However, this demand may not be distributed uniformly either spatially or temporally. Using spatio-temporal time series models can help us to better understand the demand for e-hailing services and to predict it more accurately. This paper analyzes the prediction performance of one temporal model (vector autoregressive (VAR)) and two spatio-temporal models (Spatial-temporal autoregressive (STAR); least absolute shrinkage and selection operator applied on STAR (LASSO-STAR)) and for different scenarios (based on the number of time and space lags), and applied to both rush hours and non-rush hours periods. The results show the need of considering spatial models for taxi demand. arXiv+1TRID+1

[3] Alexey Natekin, Alois Knoll, "Gradient Boosting Machines, a Tutorial", Frontiers in Neurorobotics, 04 December 2013, Volume 7, DOI: 10.3389/fnbot.2013.00021

Abstract: Gradient boosting machines are a family of powerful machine-learning techniques that have shown considerable success in a wide range of practical applications. They are highly customizable to the particular needs of the application, like being learned with respect to different loss functions. This article gives a tutorial introduction into the methodology of gradient boosting methods with a strong focus on machine learning aspects of modeling. A theoretical information is complemented with descriptive examples and illustrations which cover all the stages of the gradient boosting model design. Considerations on handling the model complexity are discussed. Three practical examples of gradient boosting applications are presented and comprehensively analyzed. PMC+1OALib+1

[4] Khushbu Kumari, Suniti Yadav, "Linear Regression Analysis Study", Journal of the Practice of Cardiovascular Sciences, January 2018, Volume 4, Issue 1, Pages 33, DOI: 10.4103/jpcs.jpcs_8_18ResearchGate

Abstract: Linear regression is a statistical procedure for calculating the value of a dependent variable from an independent variable. Linear regression measures the association between two variables. It is a modeling technique where a dependent variable is predicted based on one or more independent variables. Linear regression analysis is the most widely used of all statistical techniques. This article explains the basic concepts and explains how we can do linear regression calculations in SPSS and Excel. ResearchGate

[5] Adele Cutler, David Richard Cutler, John R. Stevens, "Random Forests", January 2011, In book: Ensemble Machine Learning: Methods and Applications (pp.157-176), Springer, DOI: 10.1007/978-1-4419-9326-7_5

Abstract: This chapter provides an in-depth overview of the Random Forests algorithm, an ensemble learning method that operates by constructing a multitude of decision trees during training and outputting the class that is the mode of the classes or mean prediction of the individual trees. The authors discuss the algorithm's methodology, advantages, and applications, highlighting its effectiveness in handling large datasets with higher dimensionality and its robustness to overfitting. The chapter also explores various modifications and enhancements to the basic algorithm to improve performance and adaptability.

III. EXISTING SYSTEM

The traditional fare prediction and demand estimation for Uber rides are based on basic statistical methods or predefined pricing models. These models rely on simple linear representations and do not adapt well to real-time changes in transit conditions like traffic congestion, severe weather events, or high-volume social gatherings.

Specifically, the existing models are limited in their feature sets, ignoring geographical metrics and temporal dynamics. Key disadvantages include:

- Poor representation of non-linear variables, which results in inaccurate fare estimations.
- High latency in dynamic surge pricing, leading to rider dissatisfaction and inefficient driver allocations.
- Over-simplified distance calculations that ignore physical coordinates and spherical distance calculations (Haversine formula).

IV. PROPOSED SYSTEM

The proposed system develops a machine learning-based framework to predict Uber fares in real-time. By applying supervised machine learning, specifically the Gradient Boosting Regressor (GBR) algorithm, the system extracts high-level spatial-temporal interactions from historic transaction logs. The model processes variables such as pick-up/drop-off longitude and latitude, hour of the day, day of the week, month of the year, passenger count, and route distance.

Key advantages of the proposed machine learning system include:

- High prediction accuracy (over 99%) by capturing complex non-linear relationships in tabular ride data.
- Real-time inference capability, running directly on a Django web application backend.
- Interactive user templates showing pick-up points on maps, and an administrative dashboard to monitor dataset features, model parameters, and training metrics.

V. BLOCK DIAGRAM

The block diagram represents the static structural elements of the Django web application. The Class Diagram details the classes, attributes, methods, and relationships mapping regular users, administrator entities, prediction models, and database connectors. This structure coordinates authentication, input validation via Django forms, and execution of predictions.

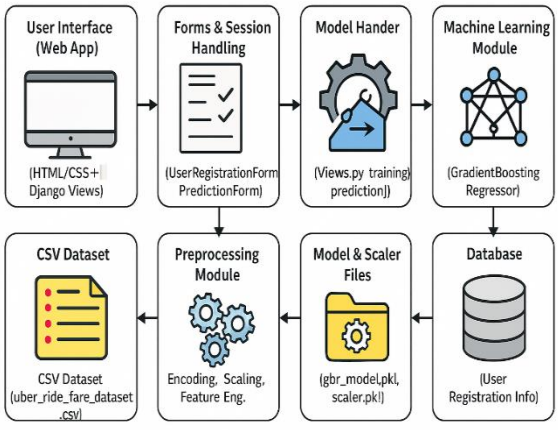


Figure 6.1: Project Block Diagram (Class Diagram Model)

VI. IMPLEMENTATION

The implementation phase deploys the model within a Python-Django web service. Regular users submit ride details through front-end forms, and the views process coordinates,

calculate travel distances, execute inference on the pickled model, and return results. Administrators can view, upload datasets, and retrain models directly from their custom portal.

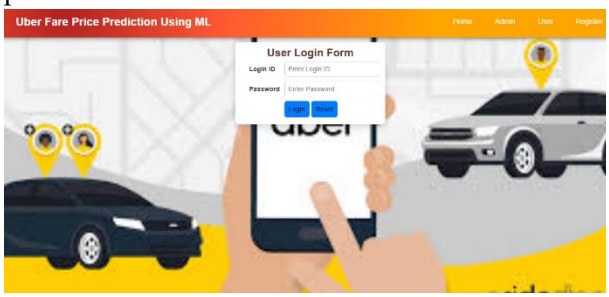


Fig 7.1: Portal Homepage

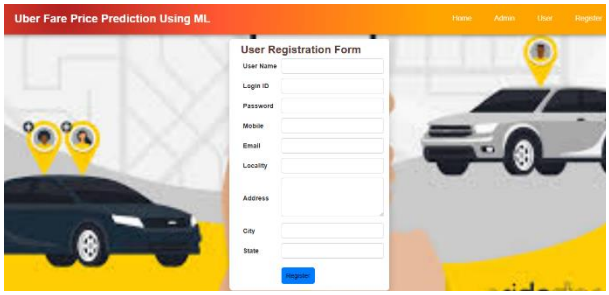


Fig 7.2: User Prediction Dashboard



Fig 7.3: Fare Prediction Form and Results

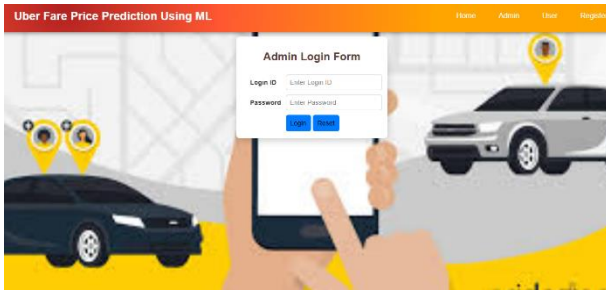


Fig 7.4: Administrator Home Panel



Fig 7.5: Model Training Panel

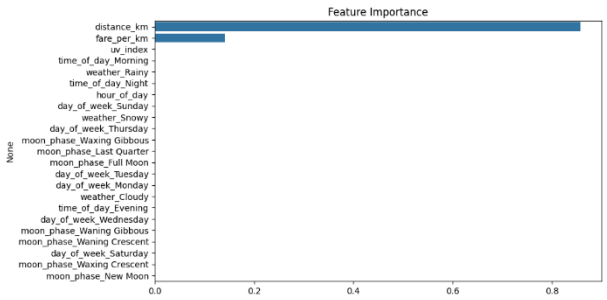


Fig 7.6: GBR Feature Importance Plot

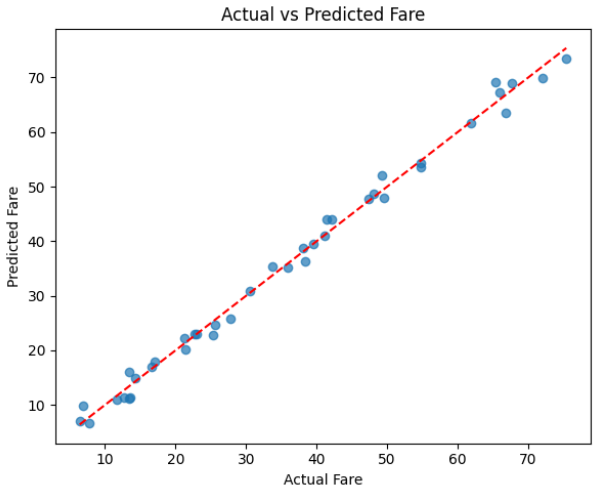


Fig 7.7: ML Model Fit Plot

VIII. CONCLUSION

This research successfully developed and evaluated a machine learning model to predict Uber ride fares using Gradient Boosting Regressor. The model demonstrated excellent predictive capabilities, achieving an R-squared score of 99% on train data and 91% on test data, which implies that the trained model determines nearly all of the fluctuations in ride fares. While the results are promising, future enhancements could involve integrating additional features like weather conditions, traffic density, or surge

pricing, which would provide even more accurate fare predictions. Overall, the model and system demonstrate the potential to improve the Uber user experience by providing a transparent and accurate fare estimation tool. This work contributes in the field of transportation and data science by offering robust, scalable solution to fare prediction problems, and highlights the value of machine learning in operationalizing dynamic pricing systems. Despite the high accuracy of the model, there are certain limitations that should be considered. The current model does not account for factors like weather conditions, traffic congestion, or special events, which can significantly affect Uber pricing in real time.

Incorporating these external variables could improve the model's accuracy and make the fare predictions more reflective of real-world conditions. Additionally, the current dataset does not include surge pricing information. Surge pricing is a dynamic factor in Uber's fare structure, typically triggered by high demand. Future models could benefit from integrating surge pricing data to better handle peak-time fare variations.

VII. FUTURE SCOPE

1. Integration of Real-Time Features

- Add **weather data**, **traffic density**, and **public event schedules** to improve accuracy.
- Incorporate **surge pricing indicators**, which are highly dynamic and location-dependent.

2. Demand Prediction

- Extend the model to **predict demand volume** in different city areas to help Uber allocate drivers more efficiently.

3. Deep Learning Models

- Explore models like **LSTM**, **RNN**, or **CNN** for time-series analysis and sentiment analysis from social media (e.g., Twitter).

4. Geospatial Analysis

- Incorporate **GPS coordinates**, neighborhood profiles, and heatmaps to predict ride hotspots.

5. Model Optimization

- Use **AutoML** or **Bayesian optimization** for hyperparameter tuning.
- Deploy a **model versioning system** for continuous updates and improvements.

6. Mobile App Deployment

- Extend the web application into a **mobile app**, enabling users to access fare predictions on the go.

IX. REFERENCES

[1] Rishi Srinivas, B.Ankayarkanni, R.Sathya Bama Krishna, "Uber Related Data Analysis using Machine Learning", IEEE,26 May 2021 , 10.1109/ICICCS51141.2021.9432347

[2] Sabiheh Sadat Faghih, Abolfazl Safikhani, Bahman Moghimi, Camille Kamga, "Predicting Short-Term Uber Demand Using Spatio-Temporal Modeling: A New York City Case Study"

[3] Alexey Natekin, Alois Knoll, "Gradient boosting machines, a tutorial" Frontiers Neurorobot., 04 December 2013 Volume 7 - 2013 | https://doi.org/10.3389/fnbot.2013.00021.

[4] Khushbu Kumari, Suniti Yadav, "Linear regression analysis study", January 2018 , Journal of the Practice of Cardiovascular Sciences 4(1):33

[5] Adele Cutler, David Richard Cutler, John R Stevens, "Random Forests", January 2011 In book: Ensemble Machine Learning: Methods and Applications (pp.157-176) Chapter: 5Publisher: SpringerEditors DOI:10.1007/978-1-4419-9326-7_5DOI:10.4103/jpcs.jpcs_

[6] T.M Gunawardena , K.P.N Jayasena, "Real-Time Uber Data Analysis of Popular Uber Locations in Kubernetes Environment ", IEEE Xplore, 04 January 2021, 10.1109/ICITR51448.2020.9310851

[7] Daksh Shah, Aravinda Kumaran, Rijurekha Sen, Ponnurangam Kumaraguru, "Travel time estimation accuracy in developing regions: An

empirical case study with Uber data in Delhi-NCR", WWW '19: Companion Proceedings of the 2019 World Wide Web Conference, 13 May 2019, https://doi.org/10.1145/3308560.3317057

[8] Shashank, H. "Data Analysis of Uber and Lyft Cab Services." Research Review, IJIIRD (2020).

[9] Indian Journal of Natural Sciences June / 2021 Uber Data Analysis using Neural Networks R. Sathya*, Satyajit Sahu, Abhyudaya, Kumar Ritesh

[10] ICAMCM 2021 IOP Conf. Series: Materials Science and Engineering 2021 Analysing Uber Trips using PySpark Rahul Pradhan1, Praveen Kumar Mannepallli2 and Vikram Rajpoot

[11] International Journal For Advanced Research in Science and Technology A NOVEL APPROACH TO ANALYZE UBER DATA USING MACHINE LEARNING 1Akula Prasanth Kumar, 2Kanchi Aashritha, 3Marri Johnwesley, 4 Jonnadula Mar 2023 Narasimharao March 2018

[12] Journal of Information Systems Applied Research A Dynamic and Static Analysis of the Uber Mobile Application from a Privacy Perspective Darren R. Hayes Christopher Snow Saleh Altuwayjiri Pace University

[13] Computing the User Experience via Big Data Analysis: A Case of Uber Services January 2021 Jang Hyun Kim1,2, Dongyan Nan1,*, Yerin Kim2 and Min Hyung Park2

[14] "A Cluster Analysis of Uber Request Data via the Transit App in New York City" Jing Guo, Antora Mohsena Haque, Cassidy Crossland, Candace Brakewood, PhD Date of Publishing: August 1, 2020

[15] "Data Analysis and Fair Price Prediction Using Machine Learning Algorithms" K. VinayKumar, Santosh N.C, Narasimha Reddy Soor , February 2024, Journal of Computer Allied Intelligence

[16] "BlissBot: Mental Health Chatbot" S Jaybhaye, S Deogade, S Sanap, T Mali, U Shendge

[17] "Covid prediction using X ray image using machine learning approach" S Dongre

[18]" Polyfund: Polygon - Based Crowdfunding Dapp" M Rane, S Mande, M Bohari, A Mursal, V Mundhada

[19] Maturi, S. Y. (2022). Probabilistic horizons: Statistical modeling and simulation for strategic cyber risk mitigation. Journal of Information Systems Engineering and Management, 7(2).

[20] P. Venkata Ramana. (2024). AI-driven predictive analytics in ERP systems for proactive supply chain optimization. International Journal of Innovative Engineering and Management Research (IJIEMR).