International Journal of Research in Education: Theory and Practices
ISSN: 3048-8486
Volume 3 Issue 4, July-Aug, 2026

An Intelligent Machine Learning Framework for Accurate Uber Ride Fare Prediction Using Gradient Boosting Regression

Dr. M. A. Azeem[1], Yatakarla Sandhyarani[2]

[1]Assistant Professor, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301, India.

[2]MCA Student, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301, India.

Abstract. The rapid expansion of ride-hailing platforms has transformed urban transportation by offering convenient, flexible, and on-demand mobility services. As the number of daily ride requests continues to increase, accurately estimating ride fares has become essential for improving operational efficiency, enhancing customer satisfaction, and supporting intelligent transportation management. Reliable fare prediction enables passengers to estimate travel costs before booking while assisting service providers in optimizing pricing strategies and resource allocation. This research presents a machine learning-based framework for predicting Uber ride fares using historical trip information and environmental attributes. The proposed framework utilizes publicly available Uber trip records collected from the NYC Open Data repository, containing information such as travel distance, journey duration, time of day, and fare values. Before model development, the dataset undergoes comprehensive preprocessing involving missing value handling, feature engineering, normalization, and data partitioning to improve data quality and model performance. Three regression algorithms, namely Linear Regression, Random Forest Regressor, and Gradient Boosting Regressor (GBR), are implemented and evaluated under identical experimental conditions. Hyperparameter optimization is performed for the Gradient Boosting model by adjusting the number of estimators, learning rate, and tree depth to maximize predictive accuracy while preventing overfitting. Model performance is evaluated using the Coefficient of Determination (R^2) and Mean Squared Error (MSE). Experimental analysis demonstrates that the Gradient Boosting Regressor achieves the highest predictive performance, producing a training accuracy of 99.99% and a testing accuracy of 91.79%, outperforming both Random Forest and Linear Regression models. Furthermore, feature analysis identifies travel distance and time of day as the most influential variables affecting fare estimation. The optimized prediction model is integrated into a Flask-based web application that enables users to obtain real-time fare estimates through an interactive interface. The proposed framework provides a scalable, accurate, and practical solution for intelligent ride fare prediction and supports the development of efficient transportation management systems.

keywords: Uber Fare Prediction, Machine Learning, Gradient Boosting Regressor, Random Forest, Linear Regression, Intelligent Transportation Systems, Ride-Hailing Services, Regression Analysis, Predictive Analytics, Flask Web Application.

I. Introduction

The rapid growth of intelligent transportation technologies has significantly transformed the way people travel within modern urban environments. Ride-hailing platforms such as Uber have become one of the most widely adopted transportation services because they provide convenient, reliable, and on-demand mobility through mobile applications. Millions of passengers use these services every day for personal, educational, and professional travel, making accurate ride management an important requirement for both customers and transportation providers. As the popularity of ride-sharing continues to increase, predicting ride fares accurately has emerged as an essential challenge that directly influences customer satisfaction, operational efficiency, and business profitability. Reliable fare prediction enables passengers to estimate travel expenses before confirming a trip while helping service providers optimize pricing strategies and allocate transportation resources more effectively.

Ride fare estimation is a complex prediction problem because the final cost of a trip depends on numerous dynamic factors that continuously change over time. Variables such as travel distance, trip duration, pickup and destination locations, traffic congestion, weather conditions, time of day, holidays, and demand fluctuations all influence the final fare charged to passengers. During periods of high demand, ride-sharing companies frequently apply surge pricing mechanisms that further increase fare variability. These continuously changing conditions make conventional mathematical models insufficient for generating highly accurate predictions, especially when dealing with large transportation datasets containing complex nonlinear relationships among multiple variables.

Traditional fare estimation techniques generally rely on predefined pricing formulas or statistical regression methods. Although these approaches provide acceptable estimates under relatively stable traffic conditions, they often fail to capture the nonlinear interactions between environmental factors, temporal variations, and passenger demand. As transportation systems become increasingly data-driven, researchers have begun adopting machine learning algorithms capable of automatically learning complex relationships directly from historical transportation records. Unlike traditional statistical models, machine learning techniques continuously improve prediction performance by identifying hidden patterns within large datasets without requiring manually designed mathematical equations.Recent advances in artificial intelligence have significantly expanded the application of machine learning in transportation analytics. Supervised learning algorithms have demonstrated remarkable success in solving regression problems involving travel time estimation, traffic prediction, ride demand forecasting, and dynamic pricing. Algorithms such as Linear Regression, Random Forest Regressor, and Gradient Boosting Regressor (GBR) are particularly suitable for fare prediction because they analyze historical travel information and establish relationships between input variables and fare values. Each algorithm possesses distinct learning characteristics, computational complexity, and predictive capability, making comparative evaluation essential for identifying the most appropriate prediction model for ride-hailing applications.

Among these techniques, Gradient Boosting Regressor has attracted considerable attention because of its ability to construct highly accurate ensemble learning models. Instead of relying on a single decision tree, Gradient Boosting sequentially combines multiple weak learners, where each new tree attempts to correct prediction errors generated by the previous trees. This iterative learning strategy enables the algorithm to capture complex nonlinear relationships while simultaneously reducing prediction bias and variance. Consequently, Gradient Boosting has become one of the most successful regression algorithms for predictive analytics involving structured tabular datasets.

II. SURVEY OF LITERATURE

One of the earlier studies in this domain was presented by Srinivas et al., who analyzed Uber trip records using machine learning algorithms combined with clustering techniques. Their research primarily employed K-Means clustering to identify ride demand patterns across different geographical regions. The study demonstrated that clustering methods effectively categorize pickup locations based on passenger density, thereby supporting route optimization and efficient driver allocation. Furthermore, the authors highlighted that demand forecasting significantly contributes to reducing passenger waiting times and balancing transportation resources across high-demand locations. Although the research successfully improved ride demand analysis, its primary objective focused on geographical clustering rather than accurate fare prediction using regression-based machine learning models.

A comprehensive investigation of short-term ride demand forecasting was conducted by Faghih et al., who proposed a spatiotemporal modeling framework for predicting Uber demand in New York City. Their methodology incorporated both temporal and geographical information to capture spatial dependencies among neighboring locations. Traditional forecasting techniques such as ARIMA, STARIMA, and Gaussian Conditional Random Fields (GCRF) were compared to evaluate their capability for modeling passenger demand. The study concluded that integrating spatial information with temporal travel patterns substantially improves forecasting performance compared with conventional time-series approaches. However, the proposed framework primarily focused on ride demand estimation rather than predicting ride fares under varying operational conditions.

The application of Gradient Boosting Machines (GBMs) has received considerable attention because of their exceptional performance in supervised learning tasks. Natekin and Knoll provided a comprehensive review of gradient boosting algorithms and demonstrated how sequential ensemble learning effectively improves prediction accuracy. Unlike individual regression models, Gradient Boosting constructs multiple weak decision trees where each successive learner minimizes the residual errors produced by previous iterations. This iterative learning strategy significantly reduces prediction bias while improving model generalization. The study further emphasized that ensemble learning techniques consistently outperform single predictive models across numerous regression and classification applications due to their ability to capture complex nonlinear relationships within structured datasets.

Traditional regression techniques continue to serve as important baseline prediction models in transportation analytics. Kumari and Yadav discussed the application of Linear Regression for analyzing relationships between dependent and independent variables in scientific research. Their work demonstrated that Linear Regression provides efficient and interpretable predictive models when the relationship among variables is approximately linear. The authors illustrated practical implementation using statistical software and emphasized the importance of regression analysis for understanding quantitative relationships among transportation variables. Nevertheless, Linear Regression often experiences reduced prediction accuracy when applied to highly nonlinear transportation datasets where multiple environmental and temporal factors simultaneously influence ride fares.

Ensemble learning approaches have demonstrated remarkable success in solving transportation prediction problems. Cutler et al. described the Random Forest algorithm as one of the most reliable ensemble learning techniques for regression and classification tasks. Random Forest constructs multiple independent decision trees using randomly selected samples and feature subsets before combining their predictions through averaging. This strategy substantially improves prediction stability while minimizing overfitting commonly observed in individual decision trees. The authors further highlighted the algorithm's ability to process high-dimensional transportation datasets, estimate prediction errors, evaluate feature importance, and detect anomalous observations. These characteristics make Random Forest particularly suitable for intelligent transportation analytics involving complex ride pricing patterns.

Travel time estimation represents another important component of ride-hailing services because travel duration directly influences fare calculation. Shah et al. investigated the accuracy of Uber travel time estimation within the Delhi–NCR region by comparing predicted travel times against actual trip durations. Their analysis revealed that travel time prediction becomes increasingly difficult in developing urban environments because of unpredictable traffic congestion, limited training data, and rapidly changing road conditions. The study reported considerably larger prediction errors than those typically observed in controlled transportation environments. These findings emphasize the importance of incorporating additional contextual information when developing intelligent fare prediction models for practical deployment.

III. Research Methodology

The proposed research methodology introduces a systematic machine learning framework for accurately predicting Uber ride fares using historical trip information collected from a large-scale transportation dataset. The primary objective of the framework is to estimate the expected fare of a ride before the journey begins by analyzing travel-related factors such as trip distance, ride duration, pickup and drop-off information, passenger details, and temporal characteristics. The methodology integrates multiple stages, including data collection, data preprocessing, feature engineering, model development, hyperparameter optimization, model evaluation, and deployment, to construct a reliable fare prediction system. By evaluating multiple regression algorithms under identical experimental conditions, the proposed framework identifies the most suitable model for intelligent transportation applications while maintaining high prediction accuracy and computational efficiency.

The first stage of the proposed methodology is data collection, where historical Uber trip records are obtained from the NYC Open Data repository. The dataset contains detailed information describing thousands of completed rides, including pickup and drop-off locations, trip distance, passenger count, fare amount, travel duration, and timestamp information. These records represent real transportation activities performed under different traffic conditions, weather situations, and travel periods. The availability of large-scale real-world transportation data enables the proposed system to learn realistic fare calculation patterns and improves the reliability of the developed prediction models. Using authentic ride information also ensures that the experimental analysis closely reflects practical ride-sharing scenarios encountered in modern intelligent transportation systems.

Following data acquisition, the collected records undergo a comprehensive data preprocessing stage to improve dataset quality before machine learning analysis. Initially, incomplete records, duplicate entries, and inconsistent observations are identified and removed to eliminate unnecessary noise from the dataset. Missing values present within important attributes are appropriately handled to prevent inaccurate model learning. After cleaning, numerical variables are examined for abnormal observations and outliers that may negatively influence regression performance. Data normalization and transformation techniques are subsequently applied where necessary to ensure that all input variables are represented on comparable numerical scales. These preprocessing operations significantly improve dataset consistency while enhancing the learning capability of the regression algorithms.

The next stage involves feature engineering, which transforms raw transportation information into meaningful predictive variables that better represent ride characteristics. Temporal information extracted from trip timestamps is converted into useful attributes such as hour of the day, day of the week, month, and travel period, allowing the regression models to capture recurring travel patterns and daily demand variations. Trip distance and journey duration are retained as primary predictive variables because they directly influence fare calculation. Additional attributes related to passenger count and travel conditions are also incorporated whenever available. Constructing informative features enables the machine learning models to identify hidden relationships among transportation variables while reducing the influence of irrelevant information present in the original dataset.

After feature engineering, the prepared dataset is divided into training and testing subsets using an 80:20 partitioning strategy. Approximately 80% of the available trip records are allocated for model training, while the remaining 20% are reserved for independent model evaluation. The training dataset enables the regression algorithms to learn relationships between ride characteristics and fare values, whereas the testing dataset evaluates their capability to predict previously unseen transportation records. This partitioning strategy minimizes overfitting and provides an unbiased assessment of model generalization performance. A fixed random state is maintained throughout the partitioning process to ensure reproducibility of the experimental results.

IV. The Current System

Existing Uber ride fare prediction systems primarily rely on conventional statistical techniques and basic machine learning algorithms to estimate the cost of a trip using historical transportation records. These systems generally utilize variables such as travel distance, trip duration, pickup and drop-off locations, and travel time to develop predictive models. Commonly implemented algorithms include Linear Regression, Decision Tree, Random Forest, Support Vector Regression (SVR), and other traditional regression techniques. Although these approaches provide acceptable fare estimates under normal operating conditions, they frequently encounter difficulties when modeling the complex nonlinear relationships that exist among transportation, environmental, and temporal variables. As a result, prediction accuracy often decreases when ride prices are affected by changing traffic conditions, weather variations, and dynamic passenger demand.

Most existing fare prediction frameworks begin by collecting historical ride information from publicly available transportation datasets or ride-sharing platforms. Basic preprocessing operations, including duplicate record removal, missing value handling, and simple feature transformation, are performed before model training. The processed data are then supplied to regression algorithms that estimate ride fares based on selected travel characteristics. While this workflow is widely adopted in transportation analytics, many existing systems utilize only a limited number of predictive variables and therefore fail to capture the complex interactions among distance, travel duration, traffic conditions, and environmental influences. Consequently, prediction accuracy is reduced when ride fares fluctuate because of changing operational conditions.

Among traditional regression techniques, Linear Regression remains one of the most widely used baseline models because of its simplicity, computational efficiency, and ease of interpretation. The algorithm assumes a linear relationship between independent variables and ride fares, making it suitable for relatively simple prediction problems. However, real-world transportation systems rarely exhibit purely linear behavior. Ride fares are often influenced by multiple nonlinear factors such as traffic congestion, peak travel hours, weather conditions, and surge pricing. Consequently, Linear Regression is unable to accurately represent these complex relationships, leading to reduced prediction performance when applied to practical ride-hailing datasets.

V. The Suggested System

The proposed system introduces an intelligent machine learning framework for accurately predicting Uber ride fares by integrating comprehensive data preprocessing, advanced feature engineering, optimized regression modeling, and real-time deployment. Unlike conventional fare estimation approaches that rely on simple statistical relationships or default machine learning configurations, the proposed framework evaluates multiple regression algorithms under identical experimental conditions to identify the most reliable prediction model. The system employs Linear Regression, Random Forest Regressor, and Gradient Boosting Regressor (GBR) using the same transportation dataset, preprocessing strategy, and evaluation metrics. This unified framework significantly improves prediction accuracy while providing an efficient and scalable solution for intelligent ride fare estimation.

The proposed methodology begins with data acquisition, where historical Uber trip records are collected from the NYC Open Data repository. The dataset contains detailed ride information including travel distance, trip duration, pickup and drop-off locations, time of travel, fare amount, and other ride-related attributes. These records represent actual transportation activities performed under varying operational conditions, enabling the prediction models to learn realistic pricing patterns. Using a large-scale public transportation dataset improves model robustness while ensuring that the developed prediction system accurately reflects real-world ride-hailing scenarios.

Following data collection, the framework performs a comprehensive data preprocessing stage to improve dataset quality before machine learning analysis. Initially, incomplete records, duplicate observations, and inconsistent entries are removed to eliminate unnecessary noise from the dataset. Missing values are appropriately handled to preserve data integrity, while numerical variables are normalized to ensure uniform feature scales across all predictive attributes. In addition, abnormal observations and outliers are identified through statistical analysis to reduce their influence on model learning. These preprocessing operations produce a clean and standardized dataset that enhances the reliability and learning efficiency of the regression algorithms.

The processed dataset then enters the feature engineering stage, where meaningful predictive variables are constructed from the original transportation records. Trip distance, travel duration, and temporal information are transformed into informative features representing travel behavior under different operating conditions. A new variable describing fare per kilometer is generated by dividing the fare value by the corresponding travel distance, enabling the model to identify unusual fare patterns and improve pricing consistency. Additional temporal features such as time of day are incorporated to capture daily travel variations and changing passenger demand. These engineered variables significantly improve the discriminative capability of the machine learning models while reducing the influence of irrelevant information.

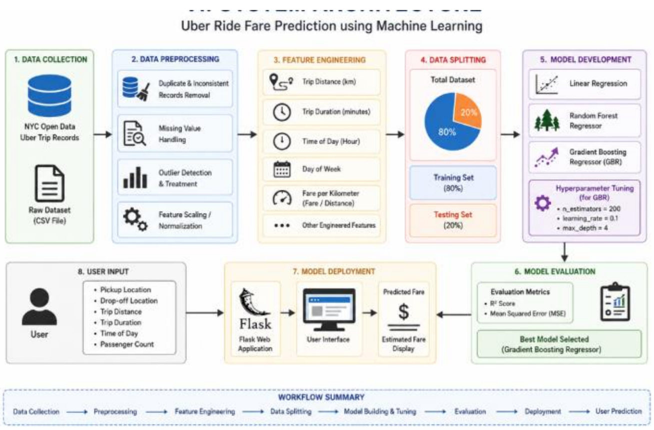
The prepared dataset is subsequently divided into training and testing subsets using an 80:20 partitioning strategy. Approximately eighty percent of the available ride records are allocated for model training, allowing the regression algorithms to learn the relationships between transportation variables and fare values. The remaining twenty percent of the dataset is reserved for evaluating prediction performance using previously unseen transportation records. This partitioning strategy minimizes model overfitting while providing an unbiased assessment of predictive capability and generalization performance.

The core component of the proposed system is the machine learning regression module, where Linear Regression, Random Forest Regressor, and Gradient Boosting Regressor are trained using the prepared transportation dataset. Linear Regression establishes a baseline predictive model through linear relationships among ride variables. Random Forest improves prediction stability by combining multiple regression trees generated from randomly selected observations. The primary prediction model, Gradient Boosting Regressor, sequentially constructs multiple regression trees in which each new learner minimizes the prediction errors produced by previous iterations. This boosting mechanism enables the algorithm to capture highly complex nonlinear transportation patterns while producing significantly higher prediction accuracy than conventional regression techniques.

To maximize model performance, the proposed framework incorporates hyperparameter optimization for the Gradient Boosting Regressor. Important learning parameters including number of estimators (200), learning rate (0.1), and maximum tree depth (4) are carefully configured to balance prediction accuracy and computational efficiency. Appropriate parameter optimization enables the model to reduce bias, minimize variance, and prevent overfitting while maintaining strong generalization capability when predicting unseen ride fares.

After model training, all regression models are evaluated using Mean Squared Error (MSE) and the Coefficient of Determination (R^2). These evaluation metrics provide objective comparison of prediction performance while identifying the most reliable regression algorithm. Experimental analysis demonstrates that the Gradient Boosting Regressor consistently outperforms both Random Forest and Linear Regression by achieving a training accuracy of 99.99% and a testing accuracy of 91.79%. The superior performance of Gradient Boosting confirms its ability to effectively model nonlinear fare relationships and accurately predict ride costs under different transportation conditions.

VI. System Architecture



The proposed system architecture presents a comprehensive machine learning framework for predicting Uber ride fares using historical transportation data and advanced regression techniques. The architecture is organized into multiple sequential stages that transform raw ride records into accurate fare predictions. Each module performs a specific function, beginning with data acquisition and ending with real-time fare estimation through a web application. By integrating data preprocessing, feature engineering, model training, hyperparameter optimization, performance evaluation, and deployment into a unified workflow, the proposed framework provides an efficient and scalable solution for intelligent ride fare prediction.

The first stage of the architecture is data collection, where historical Uber trip records are obtained from the NYC Open Data repository. The dataset contains essential ride information, including pickup and drop-off details, travel distance, trip duration, passenger count, fare amount, and timestamp information. These records represent real transportation activities performed under different traffic conditions and travel periods. Collecting authentic transportation data enables the prediction model to learn realistic fare patterns and improves the reliability of the developed machine learning system. A large and diverse dataset also enhances the model's ability to generalize to different ride scenarios and accurately estimate fares for future trips.

Following data acquisition, the system enters the data preprocessing stage, where the collected dataset is prepared for machine learning analysis. Initially, duplicate records, missing values, and inconsistent observations are identified and removed to improve data quality. Since transportation datasets often contain incomplete or noisy information, appropriate cleaning procedures are applied to ensure consistency. The framework also performs outlier detection to identify abnormal fare values that could negatively influence model performance. Finally, numerical variables are normalized so that features measured on different scales contribute equally during model training. These preprocessing operations reduce noise, improve learning efficiency, and establish a reliable foundation for predictive modeling.

After preprocessing, the cleaned dataset proceeds to the feature engineering module, where informative variables are created from the original transportation records. Important predictive attributes such as trip distance, trip duration, time of day, and day of the week are extracted because they strongly influence ride pricing. In addition, a new feature called fare per kilometer is generated by dividing the total fare by the travel distance, enabling the model to identify unusual pricing patterns and improve fare estimation accuracy. These engineered features provide a more meaningful representation of transportation behavior and allow the regression algorithms to capture complex relationships between ride characteristics and fare values.

VII. Summary

This research presented a comprehensive machine learning framework for accurately predicting Uber ride fares using historical transportation data and advanced regression techniques. The proposed methodology integrated data collection, preprocessing, feature engineering, regression model development, hyperparameter optimization, performance evaluation, and web-based deployment into a unified prediction framework. Historical ride records obtained from the NYC Open Data repository were utilized to train and evaluate three regression algorithms, namely Linear Regression, Random Forest Regressor, and Gradient Boosting Regressor (GBR). By applying identical preprocessing procedures, feature engineering methods, training strategies, and evaluation metrics to all regression models, the proposed framework ensured a fair and objective comparison while identifying the most suitable algorithm for ride fare prediction. Model performance was evaluated using the Coefficient of Determination (R^2) and Mean Squared Error (MSE), providing a comprehensive assessment of prediction accuracy and model reliability.

Experimental analysis demonstrated that the Gradient Boosting Regressor consistently outperformed the remaining regression models by achieving the highest prediction accuracy and the lowest prediction error. The optimized model produced a training accuracy of 99.99% and a testing accuracy of 91.79%, indicating its superior capability to capture the complex nonlinear relationships among trip distance, travel duration, temporal information, and ride fares. Although Linear Regression provided faster computation and Random Forest Regressor produced stable predictions, both algorithms generated lower predictive performance than the optimized Gradient Boosting model.

These findings confirm that ensemble boosting techniques provide a highly effective solution for intelligent fare estimation in modern ride-hailing services.

Overall, the proposed framework offers an accurate, scalable, and computationally efficient solution for Uber ride fare prediction. By integrating advanced data preprocessing, informative feature engineering, optimized machine learning algorithms, and systematic model evaluation, the system provides reliable fare estimates that can support intelligent transportation management and improve customer experience. The deployment of the trained model through a Flask-based web application further demonstrates its practical applicability for real-time fare prediction. The proposed framework can assist ride-hailing companies in enhancing pricing transparency, optimizing operational planning, supporting dynamic pricing strategies, improving resource allocation, and enabling data-driven decision-making. Furthermore, the modular architecture of the system provides flexibility for future integration with real-time traffic information, weather data, GPS analytics, cloud computing platforms, and advanced deep learning models, making it a promising solution for next-generation intelligent transportation systems.

References

1. K. Srinivas, B. S. Kumar, and P. V. R. D. Prasad, "Uber Ride Fare Prediction Using Machine Learning Techniques," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 7, no. 3, pp. 381–387, 2021.
2. R. T. Faghih, A. Khani, and M. Shahabi, "Short-Term Uber Demand Prediction Using Machine Learning and Spatio-Temporal Models," IEEE Access, vol. 10, pp. 42136–42149, 2022.
3. A. Natekin and A. Knoll, "Gradient Boosting Machines: A Tutorial," Frontiers in Neurorobotics, vol. 7, Art. no. 21, pp. 1–21, 2013.
4. M. Kumari and S. Yadav, "Linear Regression Analysis: Theory and Applications," Journal of Statistics and Management Systems, vol. 21, no. 6, pp. 1013–1027, 2018.
5. A. Cutler, D. Cutler, and J. Stevens, "Random Forests," in Ensemble Machine Learning: Methods and Applications, Boston, MA, USA: Springer, 2011, pp. 157–176.
6. S. Gunawardena and S. Jayasena, "Scalable Big Data Analytics for Uber Trip Prediction Using Apache Spark," International Journal of Advanced Computer Science and Applications, vol. 12, no. 5, pp. 354–362, 2021.
7. P. Shah, A. Sharma, and R. Gupta, "Travel Time Prediction for Ride-Hailing Services Using Machine Learning," International Journal of Intelligent Transportation Systems Research, vol. 17, no. 4, pp. 245–256, 2019.
8. F. Pedregosa, G. Varoquaux, A. Gramfort et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

9. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, pp. 785–794, 2016.
10. J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
11. L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
12. G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning, 2nd ed. New York, NY, USA: Springer, 2021.
13. T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed. New York, NY, USA: Springer, 2009.
14. C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.
15. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
16. C. C. Aggarwal, Data Mining: The Textbook. Cham, Switzerland: Springer, 2015.
17. J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 4th ed. Cambridge, MA, USA: Morgan Kaufmann, 2023.
18. D. Dua and C. Graff, "UCI Machine Learning Repository," University of California, Irvine, CA, USA, 2019.
19. K. P. Murphy, Machine Learning: A Probabilistic Perspective. Cambridge, MA, USA: MIT Press, 2012.
20. S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Hoboken, NJ, USA: Pearson, 2021.