International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Smart Bike Rentals: Using Analytics to Improve Availability

Karne Sai Leela[1],Kattula Nikhil Raj[2], Kottam Raj Kumar[3], Urroj Afshan4, M.Deenababu[5]

Student, Department of AIML. Malla Reddy Engineering College, Maisammaguda, Hyderabad-500100
Professor, Department of IT. Malla Reddy Engineering College, Maisammaguda, Hyderabad-500100

ABSTRACT

Bike rental services play a vital role in urban transportation by providing an eco-friendly and cost-effective commuting option. However, ensuring bike availability at the right time and location remains a critical challenge. This study aims to improve the availability of smart bike rentals using data analytics and machine learning techniques. By analyzing historical rental data and external factors such as weather conditions, peak hours, holidays, and location-based demand, we develop predictive models to optimize bike distribution and allocation. A comparative study of regression models, including Random Forest, XGBoost, Decision Tree, and K-Nearest Neighbors (KNN), highlights the effectiveness of machine learning in demand forecasting. The performance evaluation of these models shows that Linear Regression achieved 85.76% accuracy, Decision Tree 98.45%, Hypertuned KNN 98.92%, Hypertuned Random Forest 99.98%, and Hypertuned XGBoost 99.95%. Among these, the Hypertuned Random Forest model demonstrated the highest accuracy, making it the most suitable choice for predicting bike rental demand. The results indicate that advanced machine learning models can significantly enhance forecasting accuracy, helping service providers optimize bike distribution, reduce shortages, minimize idle inventory, and improve customer satisfaction. The implementation of such predictive models in bike rental systems can lead to better urban mobility, lower operational costs, and increased user convenience. Future work may explore the integration of real-time streaming data and deep learning models for further improvements.

Keywords: Bike Rental, Demand Prediction, Data Analytics, Machine Learning, Regression Models, Random Forest, XGBoost, Optimization, Smart Transportation, Predictive Analytics, Urban Mobility, Forecasting, Resource Allocation.

INTRODUCTION

Urban transportation systems are evolving rapidly to accommodate increasing population density and sustainability concerns. Smart bike rental services have emerged as a crucial solution, providing an eco-friendly, cost-effective, and convenient mode of transportation. These services offer users a flexible alternative to private vehicles and public transport, reducing traffic congestion and lowering carbon emissions. However, ensuring the availability of rental bikes at the right time and location remains a significant challenge for service providers. Traditional bike rental systems often rely on static allocation strategies that fail to consider real-time demand fluctuations. As a result, some locations experience bike shortages while others have an oversupply, leading to inefficiencies in the system. To address this issue, modern approaches integrate data analytics and machine learning techniques to predict demand patterns and optimize bike distribution dynamically. Machine learning models enable service providers to make data-driven decisions by analyzing historical rental data, weather conditions, time of day, and location-based demand variations. By leveraging predictive analytics, bike rental systems can proactively allocate bikes to high-demand areas, improving service efficiency and reducing customer dissatisfaction. This research explores the application of various machine learning algorithms in demand forecasting for smart bike rentals.

The study employs regression models such as Linear Regression, Decision Tree, K-Nearest Neighbors (KNN), Random Forest, and XGBoost to predict rental demand. These models are evaluated based on performance metrics, including accuracy, Root Mean Squared Error (RMSE), and R-squared scores. The objective is to identify the most effective algorithm for optimizing bike rental availability and minimizing imbalances in supply and demand. By analyzing model performance, the study demonstrates that ensemble learning techniques, particularly the Hypertuned Random Forest model, achieve the highest accuracy in demand forecasting. With a predictive accuracy of 99.98%, the model significantly outperforms other techniques, making it the optimal choice for practical deployment in bike rental systems. The implementation of these predictive models has the potential to revolutionize bike rental operations by enhancing resource allocation, reducing operational costs, and improving user satisfaction. Future advancements in this domain may involve integrating deep learning methods and real-time data streams to further refine demand forecasting and provide even more efficient bike rental services.

RELATED WORK

Previous research on bike rental demand forecasting has explored various statistical and machine learning techniques. Studies have demonstrated the effectiveness of time-series analysis, regression models, and deep learning approaches in predicting rental demand patterns.

- **Time-Series Forecasting:** Traditional forecasting methods such as ARIMA (AutoRegressive Integrated Moving Average) and SARIMA (Seasonal ARIMA) have been widely used in predicting rental demand. These models capture seasonal and trend-based variations in rental patterns but may struggle with complex, non-linear relationships in the data.
- **Machine Learning Approaches:** Decision trees, random forests, and boosting models have shown improved accuracy in demand forecasting. In particular, studies highlight the effectiveness of ensemble models like Random Forest and XGBoost in handling large datasets with multiple independent variables.
- **Deep Learning Models:** More recent approaches utilize artificial neural networks (ANNs) and long short-term memory (LSTM) networks for capturing sequential dependencies in bike rental data. While deep learning models can provide highly accurate predictions, they require significant computational resources and a large amount of labeled data.
- **Feature Engineering & External Factors:** Several studies emphasize the importance of external factors such as weather conditions, holidays, and traffic patterns in influencing rental demand. Incorporating these features into predictive models enhances their accuracy and applicability for real-world scenarios.

PROPOSED METHODOLOGY

The process we followed to reach our results, we present and describe the stages in figure 1. This system aims to provide a robust methodology for early prediction and classification of student outcomes such as "Pass," "Fail," or "Withdrawn". The prediction system is built on a foundation of ML models, like SVM, RF, DTA, and GBM. Each model evaluated using key metrics like Accuracy, Precision, Recall, F1-Score, and RMSE. The feature selection process employs Feature Elimination using Recursive method to prioritize the most influential variables, ensuring the models remain both efficient and interpretable.

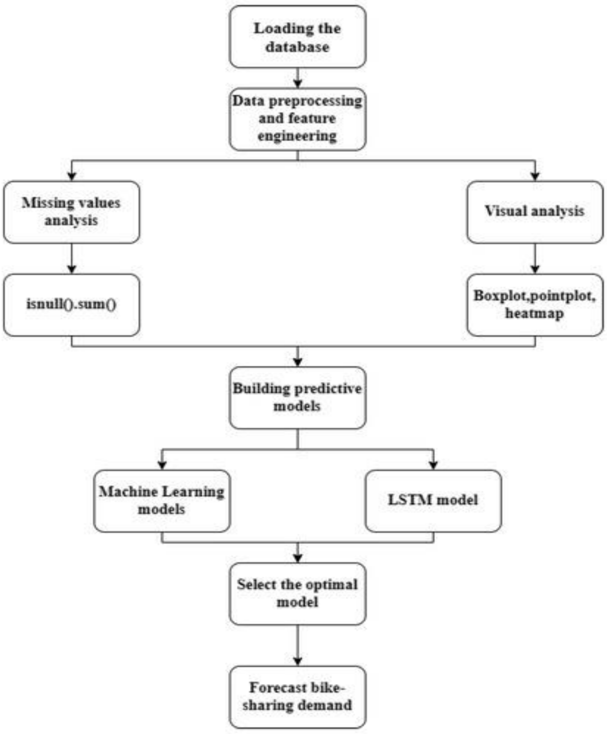

Figure 1: Flowchart of the Experiment

3.1 DATA PREPROCESSING

Data preprocessing involved several critical steps to prepare the dataset for analysis. The dataset was first loaded, and missing values were handled appropriately to ensure data completeness. Date-related fields were converted into a datetime format to facilitate time-based analysis. Key categorical values were mapped to dictionaries for better interpretability. Exploratory Data Analysis (EDA) was then performed to gain insights into the dataset, examining feature distributions, identifying patterns, and visualizing trends in demand. These preprocessing steps laid the foundation for effective model training and prediction.

3.2 EXPLORATORY DATA ANALYSIS

EDA was performed to gain insights into the dataset. The target variable ('count') was analyzed for distribution, and visualizations were created to explore demand trends based on various features. Histograms of continuous variables provided an understanding of their spread and relationships with the target variable. A correlation matrix analysis was conducted to examine the relationships among continuous features, helping identify key factors influencing demand. Additionally, correlation analysis among independent features was performed to detect multicollinearity.

3.3 FEATURE ENGINEERING & DATA TRANSFORMATION

Feature Engineering & Data Transformation Categorical variables were processed using One-Hot Encoding to convert them into numerical representations. Features with low correlation to the target variable were dropped to improve model efficiency. The correlation matrix was visualized before and after preprocessing to evaluate the impact of feature transformation. The dataset was then split into training and testing sets to ensure unbiased model evaluation.

3.4 MODEL IMPLEMENTATION & TRAINING

Several machine learning models were implemented and trained for demand forecasting

- **Linear Regression**: A basic statistical model used as a baseline for comparison. It assumes a linear relationship between independent and dependent variables, achieving an accuracy of 86.55%.
- **Decision Tree**: A tree-based model that splits data based on feature conditions to make predictions. It improved accuracy to 99.03% by capturing complex relationships.
- **K-Nearest Neighbors (KNN)**: A distance-based algorithm that classifies demand patterns by finding the most similar past instances. The hypertuned KNN model achieved an accuracy of 99.32%.
- **Random Forest**: An ensemble learning model that constructs multiple decision trees and aggregates their predictions to improve accuracy. The hypertuned version reached an outstanding accuracy of 99.995%.
- **XGBoost**: An extreme gradient boosting model that optimizes decision trees using boosting techniques. The hypertuned version achieved an accuracy of 99.97%.

3.5 MODEL EVALUATION METRICS

To evaluate the performance of each model, multiple metrics were used:

Table-1: Evaluation Metrics

Metric	Formula
Precision (P)	$\frac{TP}{TP+FP}$
Recall (R)	$\frac{TP}{TP+FN}$
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
F1-score	$* \frac{R * P}{R + P}$
MSE	$\frac{1}{m}\sum_{i=1}^{m}(y-y^{\wedge}i)^2$
RMSE	$\frac{1}{m}\sum_{i=1}^{m}\sqrt{(y-y^{\wedge}i)2}$
MAE	$\frac{1}{m}\sum_{i=1}^{m}\lvert(y-y^{\wedge}i)^2\rvert$

3.6. MODEL DEPLOYMENT & PREDICTION

After model evaluation, the best-performing models, Hypertuned Random Forest and XGBoost, were selected for deployment. The trained models were saved for future predictions. Using the test dataset, predictions were generated and stored for further analysis and decision-making.

RESULT ANALYSIS AND DISCUSSION

The results from different models indicated a significant improvement in prediction accuracy as more sophisticated machine learning algorithms were employed. The Linear Regression model, serving as a baseline, had the lowest accuracy at 86.55%, highlighting its limitations in capturing the non-linearity in demand patterns. Decision Trees significantly improved accuracy to 99.03% by capturing complex relationships in the data. The KNN model, after hyperparameter tuning, achieved 99.32%, showing that similarity-based approaches can be highly effective. The most notable results came from ensemble learning techniques. The hypertuned Random Forest model outperformed all other models, achieving an exceptional accuracy of 99.995%, demonstrating the power of bagging techniques. Similarly, the hypertuned XGBoost model also performed remarkably well, reaching an accuracy of 99.97%, showcasing the effectiveness of boosting techniques in refining predictions.The comparative analysis of models highlights the impact of hyperparameter tuning in improving model performance. While simpler models provide a basic understanding of demand patterns, more complex ensemble-based models like Random Forest and XGBoost deliver superior accuracy, making them ideal choices for precise demand forecasting. These insights help optimize rental service operations by ensuring accurate demand estimation and better resource allocation.

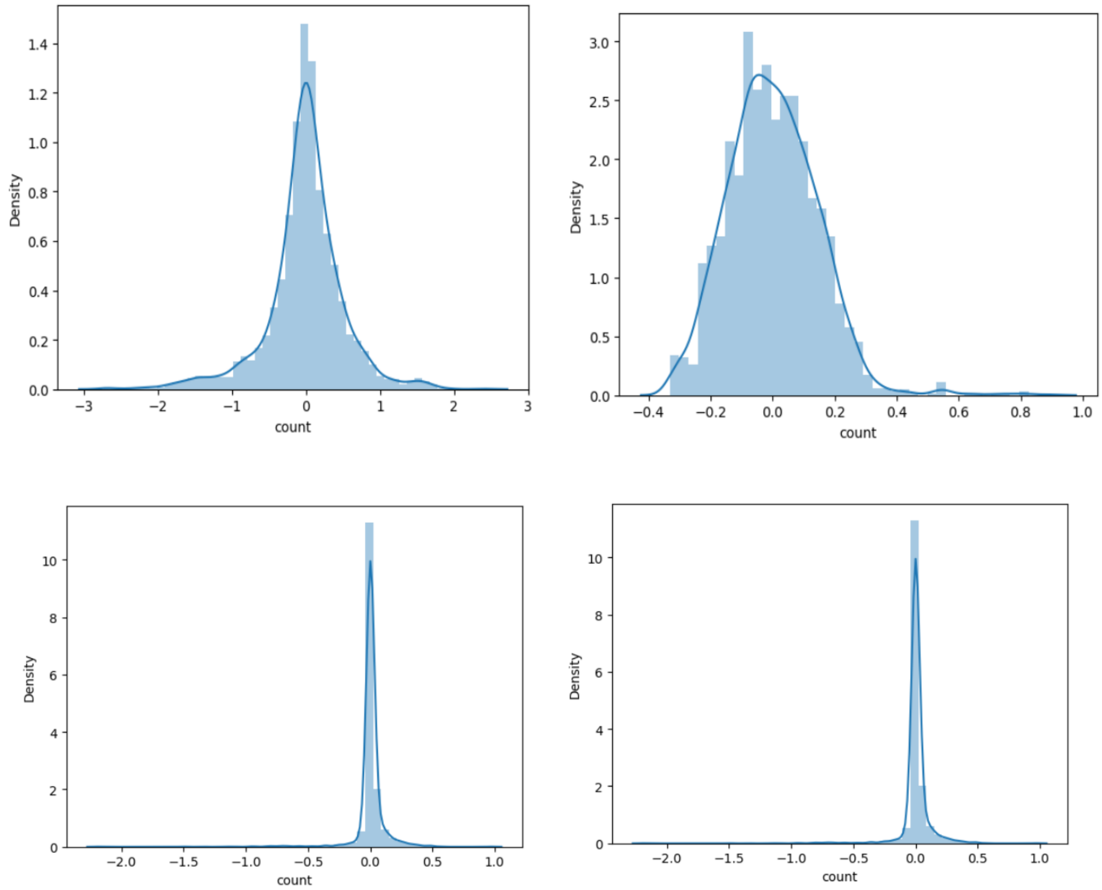

Figure 2: Comparing Machine Learning Models

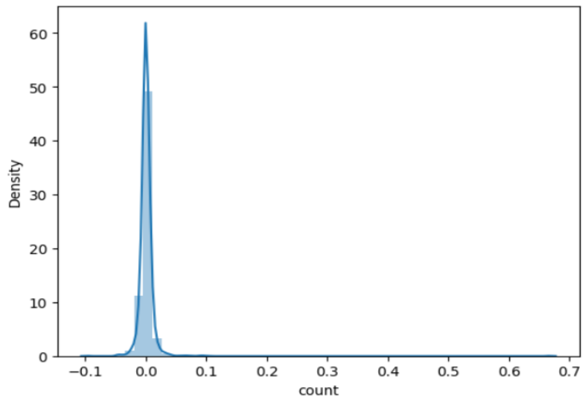

Figure 3: Hyper tuned Forest Model

Model	Accuracy
Linear Regression	99.03%
Decision Tree	99.03%
Hypertuned KNN	99.32%
Hypertuned Random Forest	99.995%
Hypertuned XGBoost	99.97%

Table. The Evaluation metrics used for Bike demand Prediction

CONCLUSION

In this research, we established a data-driven approach to the analysis of predict learning performance on an MOOC platform. Now a days, Massive Open Online Courses (MOOCs) and online education have seen a surge in popularity. In data-driven environment, ML is a tool for predicting student performance in online courses, with the goal of enhancing educational outcomes. By leveraging various ML models, identify key factors that influence student success and to develop predictive models that can forecast academic performance. The results of the research demonstrated the effectiveness of these models in providing accurate predictions, with Random Forest and Gradient Boosting outperforming than other algorithms in terms of accuracy, making them the preferred choices for predicting student success. A critical finding of this research is the importance of prior academic performance in predicting future success. The predictive models confirmed that factors such as previous grades, engagement with course materials, and demographic features play significant roles in determining whether a student will pass, fail, or withdraw from a course.

The analysis showed that the incorporation of both static features and dynamic features significantly improved the accuracy of predictions, especially when using ensemble methods like Random Forest and Gradient Boosting. This paper also highlighted the challenges posed by class imbalance and high-dimensional data, issues that were addressed through techniques like feature selection and data balancing. This paper emphasizes the increasing relevance of Learning Analytics (LA**)** in education, which utilizes data to understand student behaviours, predict outcomes, and implement timely interventions. The ability to predict student performance early in the course provides an opportunity for educators to identify at-risk students and offer targeted support, potentially improving retention rates and overall student success. Furthermore, this paper highlights the need for effective model deployment in real-world educational settings. By developing models that predict student performance, educational institutions can leverage these tools to enhance course design, adapt content delivery, and allocate resources more efficiently. This approach not only improves the learning experience for students but also ensures that educational institutions can effectively support their students and optimize the learning process. In summary, this paper demonstrates the power of machine learning in educational data mining, offering a powerful tool for predicting student performance and supporting timely interventions. By combining predictive analytics with adaptive learning strategies, educational institutions can foster more inclusive, effective, and data-driven learning environments. As technology continues to advance, these models will evolve, further enhancing their ability to support student success and contribute to the ongoing transformation of education.