iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.81785

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Ola Data Analysis and Ride Comparision

Rudra Pratap[1], Harsh Kumar[2], Avnish Kumar Singh[3], Kunal Solanki[4], Dr. Lalit Kumar[5]
Department of Computer Science, Meerut Institute of Engineering & Technology, Meerut, India

Abstract: ***With the rise of ride sharing we see that accurate fare prediction has become a key issue. In this study we focus on development of a fare prediction system via machine learning methods. We use and compare Two algorithms, Multiple Linear Regression and Random Forest, out of which we try to determine which is the superior model. We evaluate the models with the use of R-squared, Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and RMSLE. What we found is that the Random Forest model does in fact perform better in terms of accuracy and also does a better job with real world data than Linear Regression. This work also puts into light how machine learning may be used to improve pricing in transport services.***
Keywords: Artificial Intelligence, Cab Fare Prediction, Linear Regression, Random Forest, Data Analysis.

I. INTRODUCTION

In recent times online cab services like Uber and Ola have become a very much a part of our daily routine. What we see is that these platforms have to deal with the issue of accurate fare estimation which is based on a variety of factors which include distance, time, traffic, and location. Traditional price models do a poor job at that they put together the relationship between these variables. That is when we see the value of machine learning. What it does is that it looks at large sets of data and brings out what is not at all times easy to see. In this study we are out to develop a fare prediction system using machine learning algorithms. Also we will look at what are the performance of two very common techniques Multiple Linear Regression and Random Forest in the context of cab fare prediction.

II. RESEARCH OBJECTIVES

This project is to put together and bring to life a system which in turn will help users compare ride fares across many cab services in a simple and efficient way. In today's fast paced world users have hard time to jump from one platform to another to check prices before they book a ride. We aim to solve that issue with a single unified platform for fare comparison. Also we have as a key goal to study what factors which play into ride price distance, time, demand etc. By looking at these we try to put forth accurate fare estimates and in that way help users to make informed decisions. Also we are put focused on to improve user convenience by 0 which we mean to reduce the time and effort required to switch between many apps. It is to present to you all relevant info which includes estimated fare and best available choice in one interface. Also we looked at the use of data analysis and basic machine learning in improving fare prediction accuracy which is the focus of this research. By looking at different approaches the system's aim is to identify the best methods for fare prediction and analysis. Also we looked at building a scalable and user friendly solution which in the future we aim to extend to include real time data, advanced analytics and mobile app support.

III. LTERATURE SURVEY

In recent past we have seen an up rise in the area of cab fare prediction and ride hailing data analysis which has been brought about by the large scale growth of services like Uber and Ola. Many studies have reported on the application of machine learning in studying ride patterns and in improving pricing strategies. For instance Srinivas et al. [1] did a study on Uber data which they analyzed via machine learning tools to put forth trends which in turn improved decision making. Their work reports that data driven approaches do in fact improve operational efficiency. Also Kumar et al. [2] put forth a new model for the analysis of Uber data which they put forward as a way to get out of large scale transport data what is valuable. Also it is put forth in their study that machine learning may be very effectively used to find out what is going on in the background in terms of ride data. Another key contribution is by Taruna and Sriramyab [3] who put forth a fare prediction model which they developed using Multiple Linear Regression and Random Forest algorithms. They report that while linear regression gives a fundamental perspective, Random Forest does better in terms of accuracy which it achieves via its ability to deal with non linear relationships. Also it is brought to light that Exploratory Data Analysis (EDA) is of great importance in getting to know the data structure before we apply models.
Camizuli and Carranza [4] report how EDA tools are used to identify trends, outlying values and relationships between variables which in turn improve model performance. Also in related research Goel et al. [5] looked at operational analytics in ride hailing platforms which they looked at from a logistico optimization standpoint. Their study reports that analysis of ride data may be used to improve efficiency, reduce costs.

Overall, the existing literature suggests that combining proper data analysis techniques with advanced machine learning models leads to more accurate and reliable fare prediction systems.

IV. METHODOLOGY

A. Data Collection

The data used in this project consists of cab ride information collected from available datasets and simulated ride scenarios. It includes key attributes such as pickup location, drop location, trip distance, estimated travel time, and fare amount. These parameters are essential for analyzing pricing patterns and building a system capable of comparing fares across different ride service providers.

B. Data Preprocessing

Before applying any machine learning techniques, the dataset was carefully prepared to ensure accuracy and consistency. Missing values were identified and handled appropriately to avoid errors during model training. Irrelevant or duplicate entries were removed to improve data quality. Feature selection was also performed to identify the most important variables influencing fare prediction, such as distance and time. Additionally, data formatting and normalization were applied where necessary to make the dataset suitable for model implementation.

C. Model Implementation

To build an efficient fare prediction and comparison system, two machine learning models were implemented and analyzed:

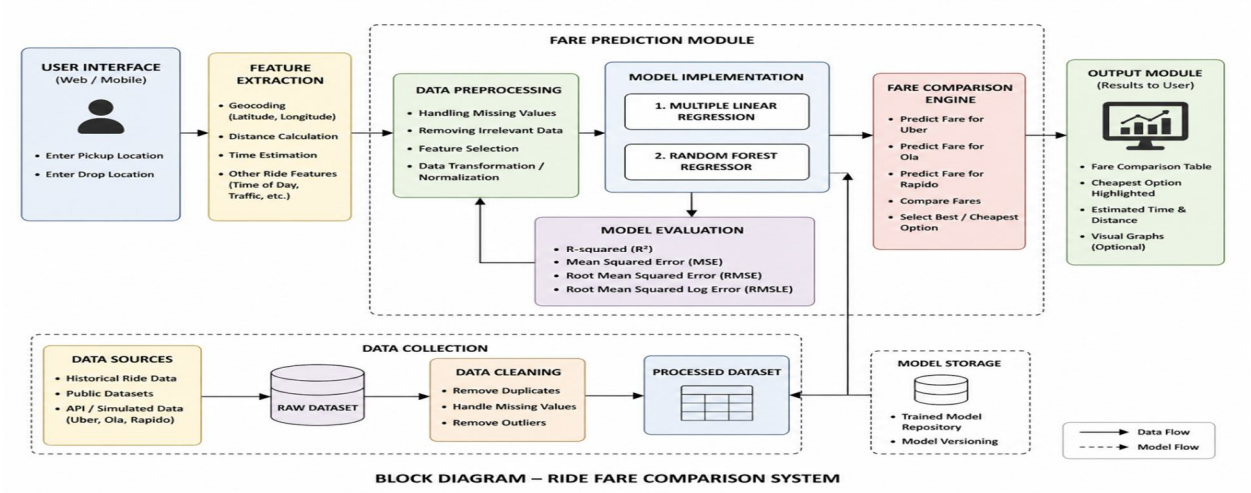
1) Multiple Linear Regression: This model is based on the assumption that there is a linear relationship between input features (such as distance and time) and the target variable (fare). It is simple, interpretable, and useful for understanding how different factors affect ride pricing.
2) Random Forest: Random Forest is an ensemble learning technique that combines multiple decision trees to improve prediction accuracy. It helps in capturing complex relationships in the data and reduces the chances of overfitting, making it more reliable for real-world fare prediction. The outputs from these models are used to estimate ride fares, which are then compared to suggest the most economical option to the user.

D. Evaluation Metrics

To evaluate and compare the performance of the implemented models, several standard metrics were used:

- R-squared (R^2): Measures how well the model explains the variability in the data
- Mean Squared Error (MSE): Calculates the average squared difference between actual and predicted values
- Root Mean Squared Error (RMSE): Provides error in the same unit as the target variable (fare)
- RMSLE (Root Mean Squared Log Error): Useful for handling large variations in fare values and reducing the impact of outliers

These metrics help in determining the accuracy and reliability of each model, allowing the system to choose the most effective approach for fare prediction.

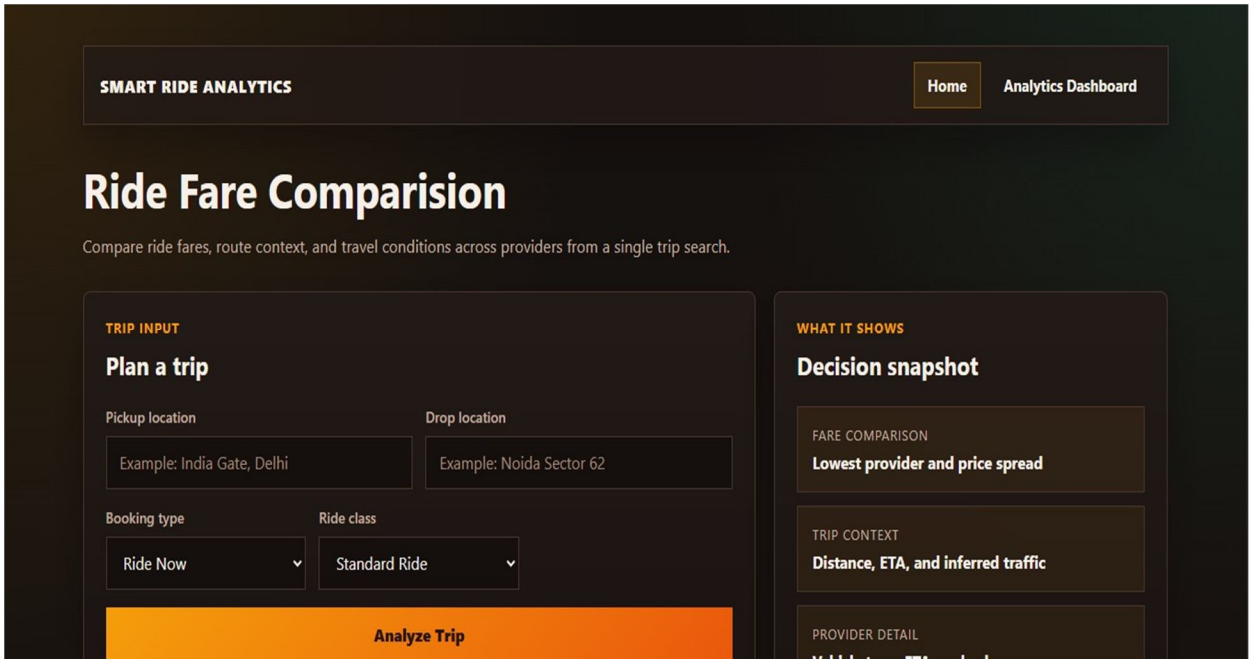


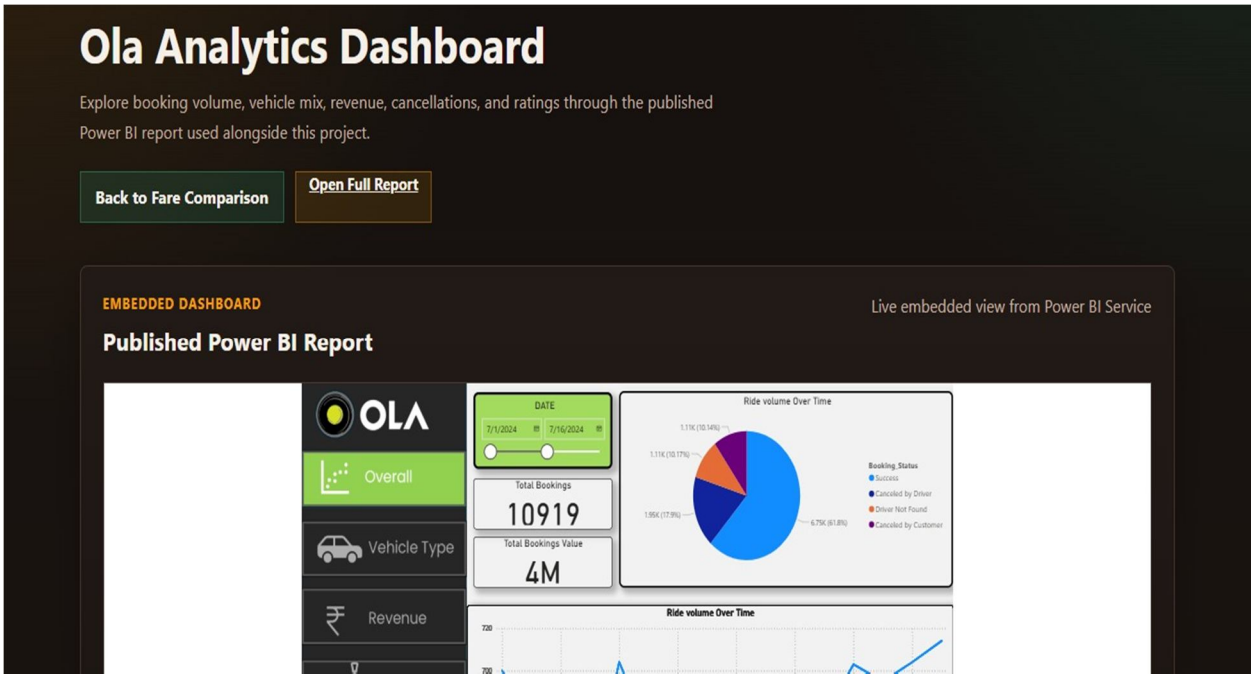
BLOCK DIAGRAM – RIDE FARE COMPARISON SYSTEM

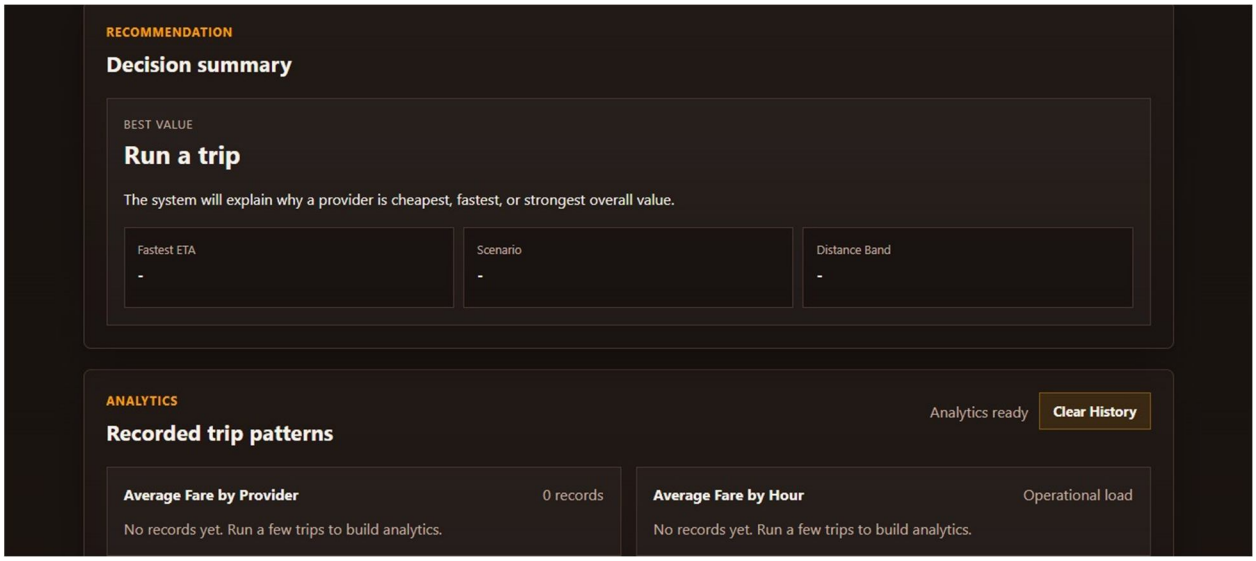
V. RESULTS

After which we put the two models through their paces we saw that the Multiple Linear Regression model did well enough in some areas but had issues with complex patterns in the data. It did very well for simple relationships between variables but had higher error rates in complex settings. The Random Forest model did in fact do much better overall. It gave more accurate results and did a better job at what the model does which is to handle different types of data in which there are variations. Also we saw that the error was lower and the R squared value higher in the case of Random Forest as compared to Linear Regression. That in large part shows that Random Forest is the better choice for this set of problems.

Dashboard Images







VI. CONCLUSION

The in this project we developed the Ride Fare Comparison which addresses a large issue that users have at the time of booking cab services that of comparing fares across many platforms. We integrated data analysis and machine learning into the system which in turn is able to determine ride fares and present a comparison of what different service providers charge in one interface. We used models like Multiple Linear Regression and Random Forest which helped us to study the relationship between key factors of distance, time and fare. Of these the Random Forest model did it's job with greater accuracy**.**

Overall, the project demonstrates how data-driven approaches can improve decision-making for users by saving time, reducing effort, and providing cost-effective ride options. It also highlights the practical application of machine learning in solving real-world problems.

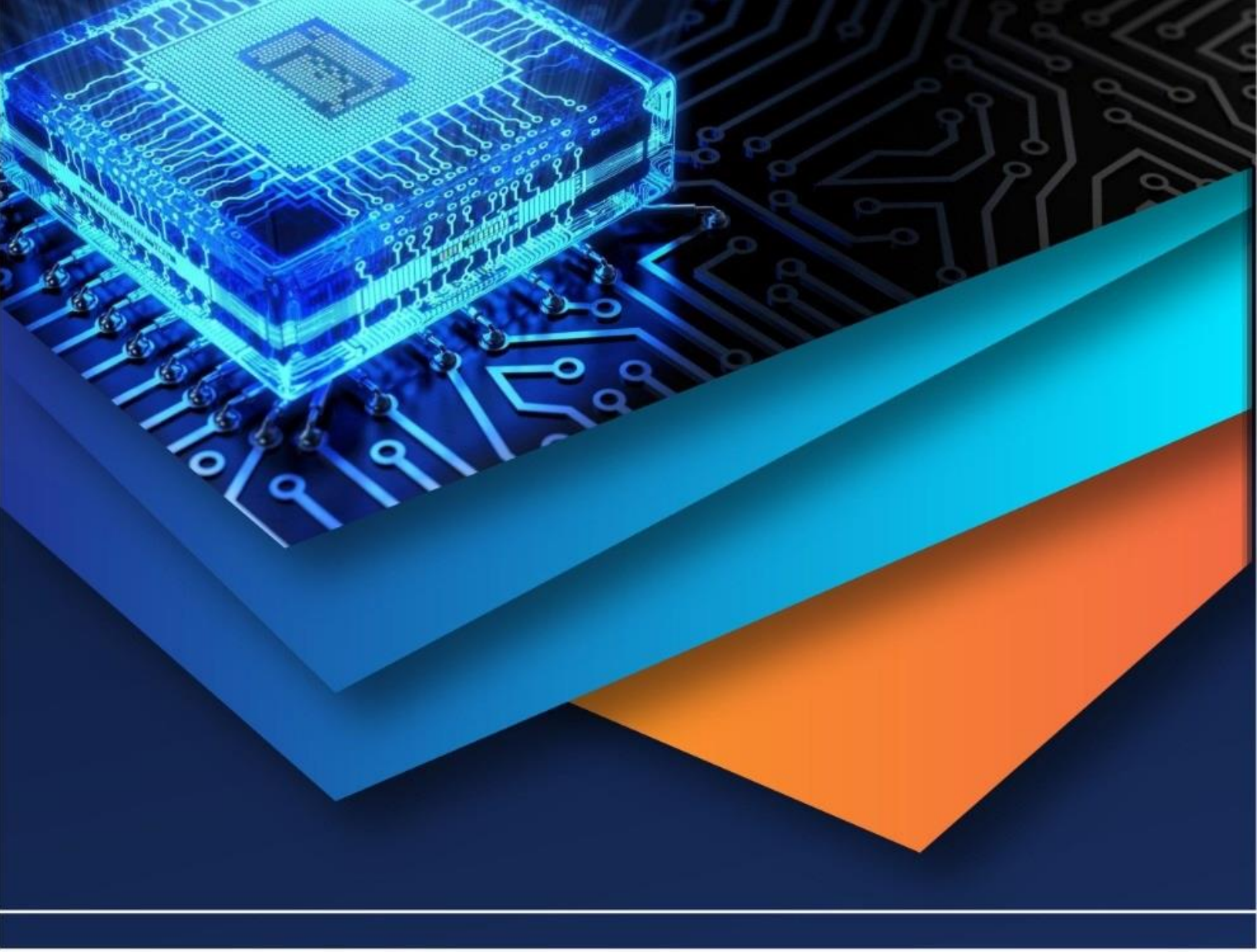
VII. FUTURE SCOPE

Although the current system provides a functional and effective solution, there are several areas where it can be further enhanced. One major improvement would be the integration of real-time APIs from cab service providers like Uber, Ola, and Rapido to provide live fare updates instead of estimated values. This would significantly increase the accuracy and usability of the system. The

project can also be extended by incorporating advanced machine learning models and deep learning techniques to improve prediction performance, especially under dynamic pricing conditions such as surge hours or high-demand periods.Another potential enhancement is the integration of map services like Google Maps to provide route visualization, traffic conditions, and estimated time of arrival (ETA). This would allow users to compare not only fares but also travel time and convenience.In addition, the system can be developed into a full-fledged mobile application with a user-friendly interface, making it more accessible to a wider audience. Features like personalized recommendations, ride history analysis, and smart notifications for price drops can also be added.Finally, expanding the system to include other transportation options such as bike taxis, auto-rickshaws, and public transport could make it a comprehensive travel assistant for users.

REFERENCES

[1] R. Srinivas et al., "Uber Related Data Analysis Using Machine Learning," *Proc. ICICCS*, 2021.
[2] A. P. Kumar et al., "A Novel Approach to Analyze Uber Data Using Machine Learning."
[3] G. Venkat Sai Taruna and P. Sriramyab, "Ola Data Analysis for Dynamic Price Prediction Using Multiple Linear Regression and Random Forest Regression."
[4] E. Camizuli and E. J. Carranza, "Exploratory Data Analysis (EDA)."
[5] R. Goel et al., "Operation Analytics: Uber and Ola Logistics Optimization."



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY