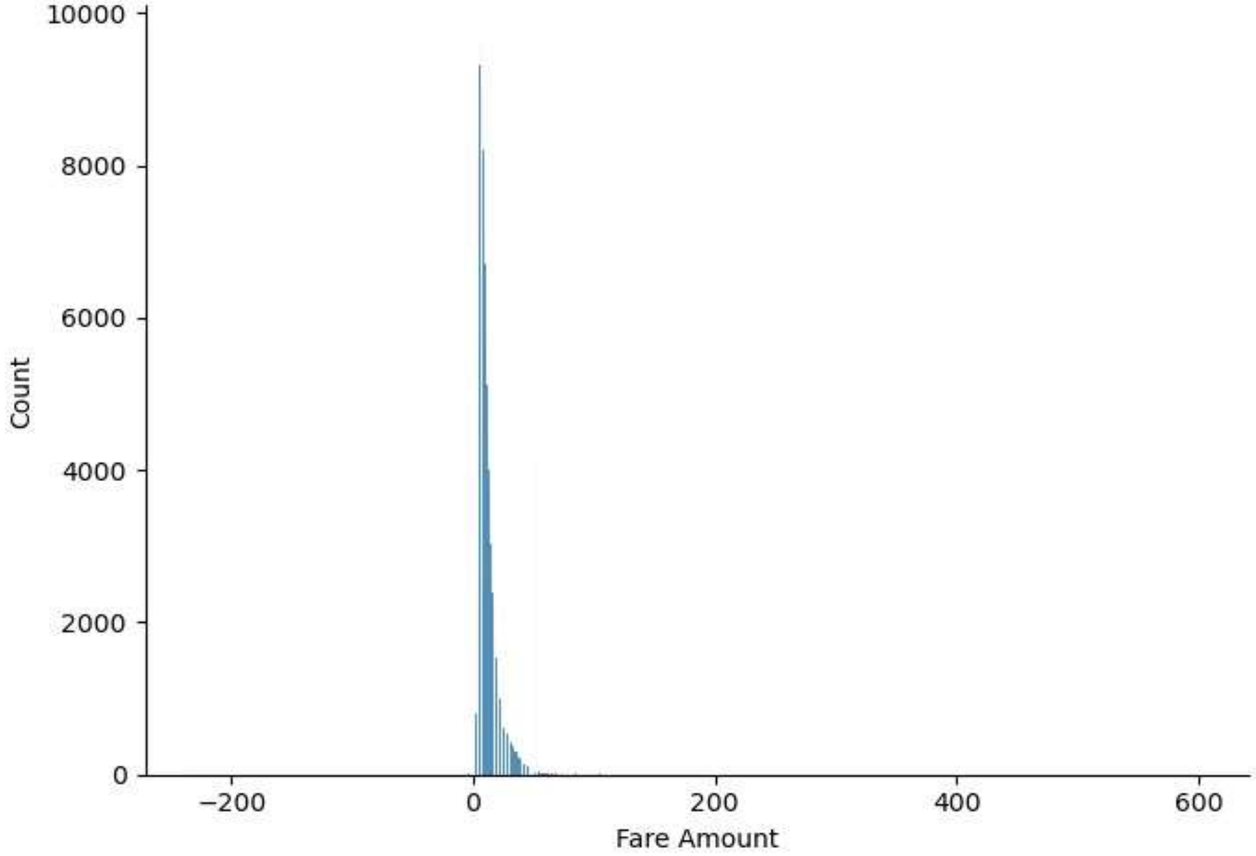
Land Transport AI-ML Case Study

The taxi fares is a critical component of the land transport industry, providing an essential service for commuters and travelers. One of the key challenges for the land transport industry is accurately predicting the fair amount for a given trip, based on a variety of factors such as distance, traffic, and time of day. This is where machine learning can play a vital role in providing a more accurate and efficient fare prediction system.

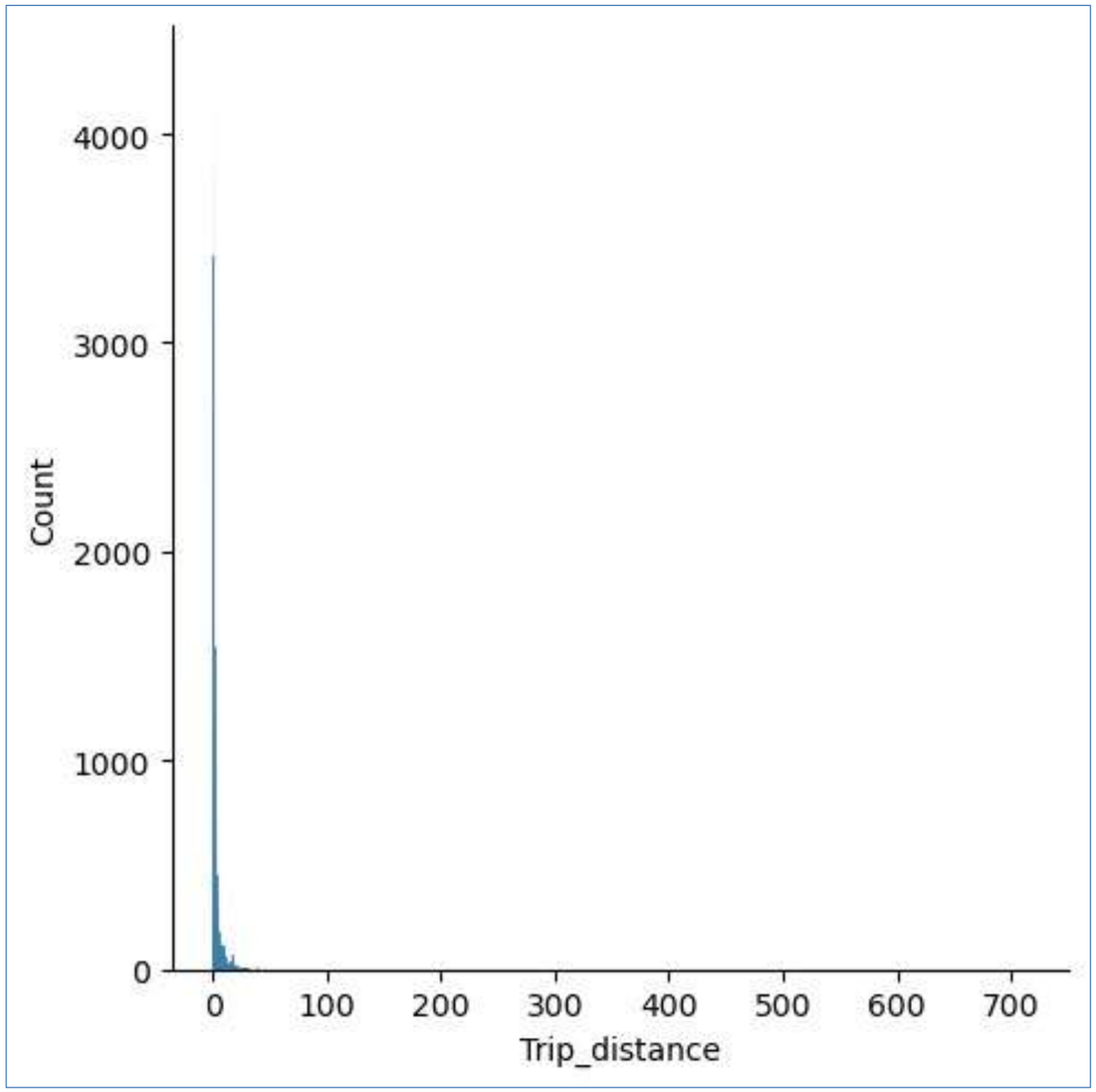
The problem statement for predicting taxi fare amounts in the land transport industry is to develop a machine learning model that accurately predicts the fare amount for a taxi ride based on various factors such as distance, time of day, traffic conditions, and other relevant variables. The goal is to provide accurate fare estimates to passengers and enable taxi companies and drivers to make more informed decisions about pricing strategies, which can ultimately lead to increased revenue and customer satisfaction.

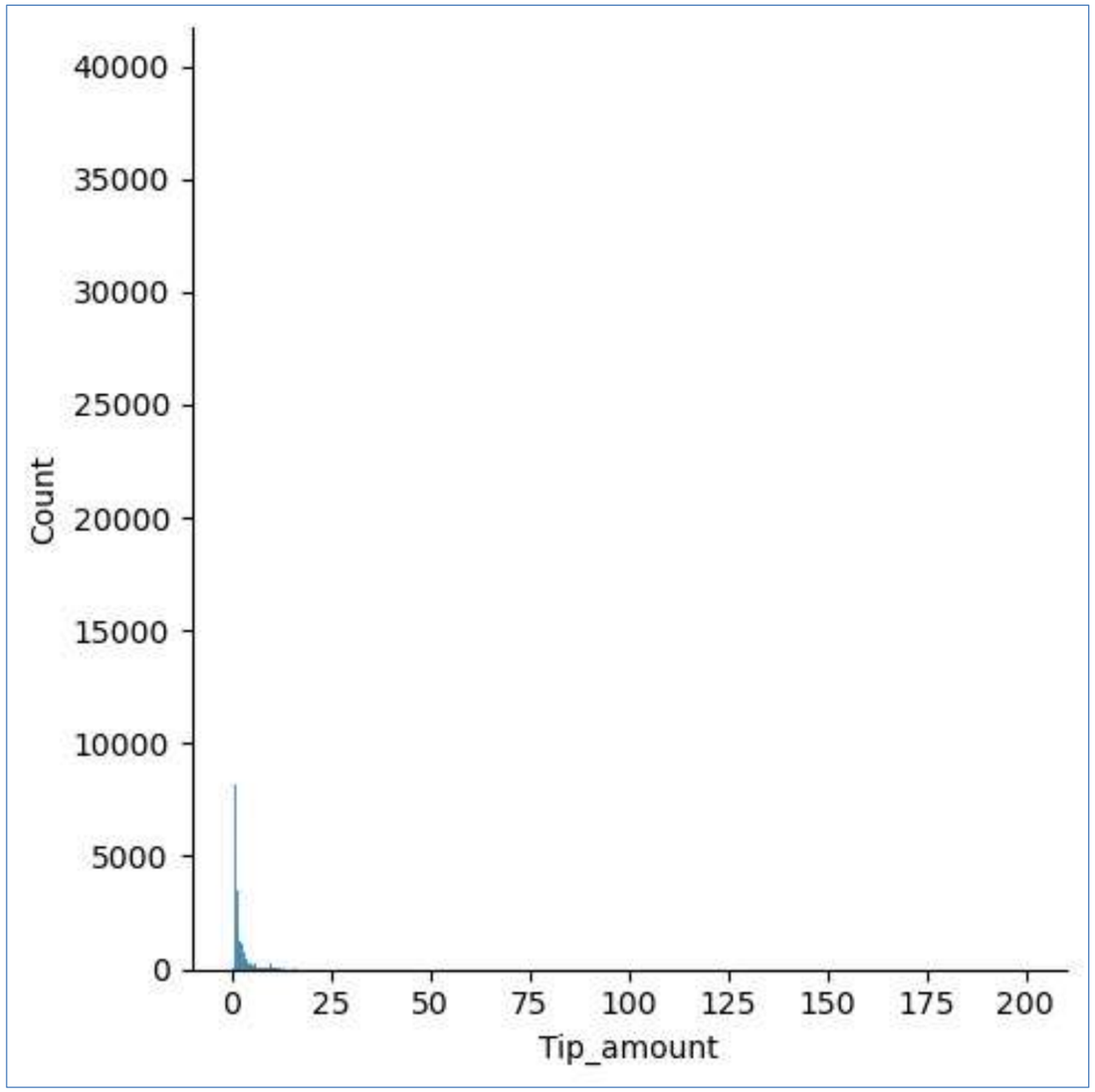


The fare amount in land transport can have a significant impact on both the transportation service providers and the passengers who use those services.

For transportation service providers such as bus or taxi companies, the fare amount directly impacts their revenue and profitability. If the fare amount is too low, the company may not generate enough revenue to cover their operating costs, and they may struggle to stay in business. On the other hand, if the fare amount is too high, it may discourage passengers from using their services, and the company may lose market share to competitors.

Features Responsible





- **<u>Trip Distance</u>**: For lesser trip distances the fare amount is less. The trip distance is a key factor that can have a significant impact on the fare amount in land transport services

- **<u>Tip Amount</u>:** Higher the tip amount, higher the fare amount. The fare amount typically includes the base fare, and additional charges such as distance or time-based fees, service fees or other surcharges.

Auto-ML Methodology Results

Case	Percentile	No. of Features	Random Forest	XGBoost	RNN	MLP	Lasso	Avg. Accuracy
Case 1	25	5	99.5	99.4	98.8	52.5	98.5	**89.74**
Case 2	50	9	99.6	99.8	88.06	99.7	99.9	**97.41**
Case 3	75	13	99.9	99.9	75	74.9	100	**89.94**
Case 4	90	16	99.7	99.9	75.1	65.2	100	**87.98**

Based on our observation , Lasso was the best performing algorithm with 100% accuracy in 75th percentile and 50th percentile is the best percentile with an average accuracy of 97.41%.

Conclusion

In conclusion, the accurate prediction of taxi fares is a critical challenge facing the land transport industry. By leveraging Auto-ML algorithms trained on historical data, it is possible to predict the fare amount for a given trip accurately. This has numerous benefits for both taxi drivers and passengers. The dataset has 1446517 records with 3 Categorical Features and 15 Numerical Features. Out of 1446517 records a random sample of 200000 records is selected for modelling.

For regression, models were created with algorithms using Auto-ML techniques like Lasso, Recurrent Neural Network, Multilayer Perceptron, Random forest and XGBoost . With these models, performance measurement values were obtained for feature sets of 5, 9, 13 and 16. The Auto-ML algorithms were able to predict fair amount with an average accuracy between 89% – 97% and helped to identify factors that determine the fair amount. The major factors include Trip Distance and Tip amount. The Random forest with 99 % accuracy in 75th percentile where tree showed a threshold of total amount >= 29.78 units which will leads to higher fair amount.

Overall, the application of Auto-ML in predicting taxi fares is a crucial tool for the land transport industry. By using Auto-ML algorithms to accurately predict fares, the industry can become more efficient, provide better service to customers, and ultimately, increase profitability.