

# Passenger Flow Prediction Report

## AI Model Performance Summary

Generated on August 03, 2025 09:51 AM EAT

Prepared for General Review

### 1. Purpose of This Report

This report evaluates an AI model designed to predict passenger numbers for public transport. The model uses 500,000 trip records, analyzing factors such as time of day, weather, and weekends. It predicts exact passenger counts (e.g., 44 or 80 passengers) and categorizes them as Low (0–50), Medium (51–100), or High (over 100). Training took 2.5 seconds, and predictions are made in 0.004 seconds, enabling real-time applications.

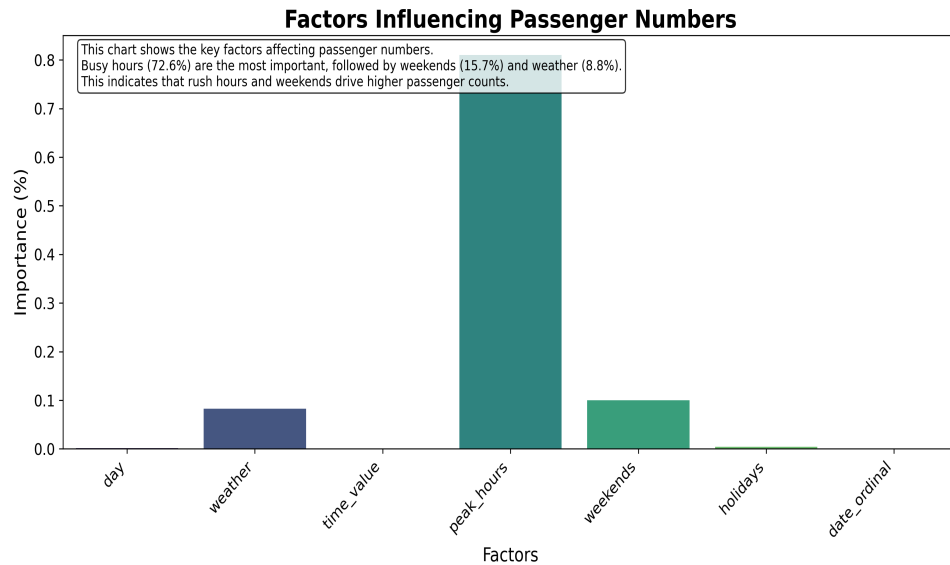
### 2. Prediction Accuracy

The model was tested on 100,000 trips. The following table summarizes its performance, including the R<sup>2</sup> score, which shows how well the model explains changes in passenger numbers:

Measure	Result	Explanation
Average Error (MAE)	18 passengers	Predictions are typically off by about 18 passengers
R <sup>2</sup> Score	60% (0.6034)	The model explains 60% of passenger number variations
Root Mean Squared Error (RMSE)	21.73 passengers	Larger errors are around 22 passengers
Consistency	High	Performance is stable across different data sets

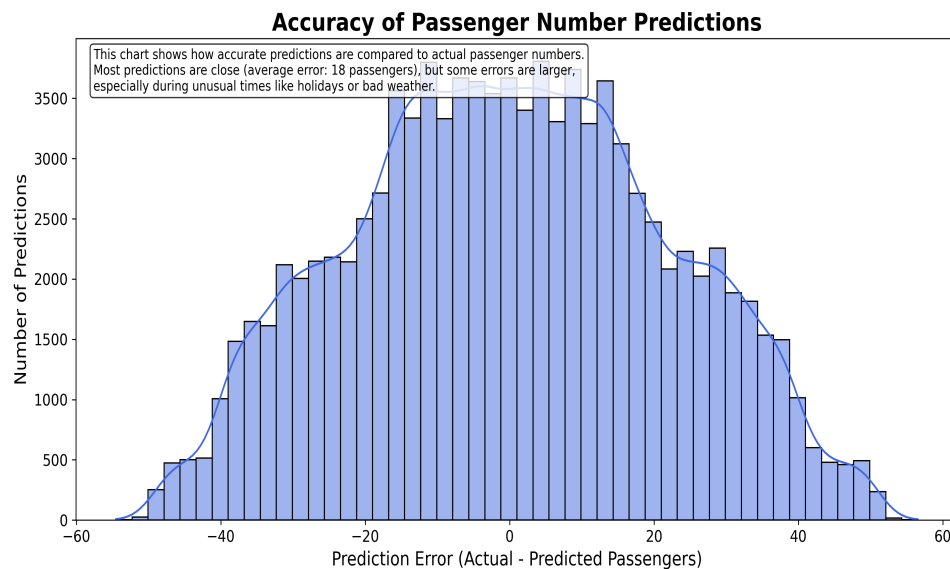
### 3. Factors Influencing Passenger Numbers

The chart below identifies the key factors affecting passenger numbers. **Key Insights:** Busy hours (rush times) have the largest impact (72.6%), followed by weekends (15.7%) and weather (8.8%). This suggests that peak travel times and weekend patterns are critical for predicting passenger counts, while specific dates have less influence.



## 4. Accuracy of Predictions

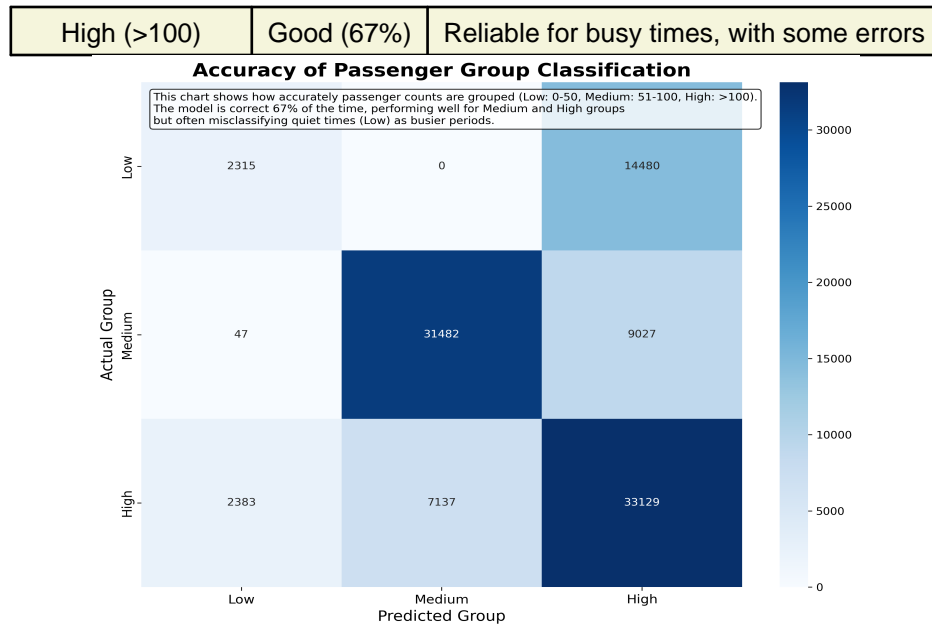
The chart below shows how close predictions are to actual passenger numbers. **Key Insights:** Most predictions are accurate (within 18 passengers), but larger errors occur during unusual times, such as holidays or bad weather, indicating a need for better data in these scenarios.



## 5. Passenger Group Classification

The model categorizes passenger counts as Low (0–50), Medium (51–100), or High (over 100), with an overall accuracy of 67%. **Key Insights:** Predictions are reliable for Medium and High groups (busy and typical times) but less accurate for Low groups (quiet times), often predicting quiet periods as busier. This could lead to scheduling more buses than needed during quiet times.

Group	Accuracy	Explanation
Low (0–50)	Low (21%)	Quiet times are often predicted as busier
Medium (51–100)	High (80%)	Very accurate for typical trips



## 6. Example Predictions

The following examples illustrate the model's predictions:

- Wednesday, 8:00 AM, Sunny, Non-Busy: 44 passengers
- Friday, 5:00 PM, Rainy, Busy: 80 passengers
- Sunday, 12:00 PM, Sunny, Weekend: 59 passengers

## 7. Strengths and Areas for Improvement

### Strengths:

- **Fast Predictions:** Predictions are generated in 0.004 seconds, enabling real-time scheduling and resource allocation.
- **Reliable for Busy and Typical Times:** The model accurately predicts passenger numbers during peak hours (Medium and High groups, with 80% and 67% accuracy, respectively), ensuring effective planning for busy periods.
- **Consistent Performance:** The model maintains stable accuracy across different data sets (train RMSE: 21.71, test RMSE: 21.76), indicating robust generalization.

### Areas for Improvement:

- **Poor Accuracy for Quiet Times:** The model struggles to predict low passenger counts (Low group, 21% accuracy), often misclassifying quiet periods as busier, which may lead to over-scheduling buses.
- **Limited Overall Accuracy:** The  $R^2$  score of 60% (0.6034) indicates that the model explains only 60% of passenger number variations, leaving room for improvement in capturing all trends.
- **Limited Data for Unusual Conditions:** Larger errors occur during holidays or bad weather, suggesting that the current data lacks sufficient information for these scenarios.

## 8. Future Improvements

The following steps can enhance the model's performance:

### Improved Data Collection:

- **Gather Real-World Data:** Collect ticket sales, bus GPS data, or station passenger counts to reflect actual travel patterns, particularly during holidays or events.
- **Incorporate Additional Factors:** Include data on special events (e.g., concerts), road closures, or public holidays to better predict unusual travel patterns.
- **Ensure Data Accuracy:** Verify data quality by cross-checking multiple sources and correcting errors, such as missing or incorrect times.

### Advanced Computing:

- **Utilize NVIDIA GPUs:** Train the model on a computer with NVIDIA GPUs (e.g., RTX 3080 or A100) using XGBoost's GPU support (set `tree_method='hist'`, `device='cuda'`) to process larger datasets and optimize settings faster.
- **Optimize Model Settings:** Test various model configurations (e.g., number of trees, learning rate) through a systematic grid search to improve accuracy.

### Enhanced Model Intelligence:

- **Add New Factors:** Incorporate time-of-day patterns (e.g., morning vs. evening) or traffic data to capture additional travel trends.
- **Improve Quiet Time Predictions:** Apply techniques like SMOTE to balance data for quiet periods, enhancing predictions for low passenger counts.
- **Explore Advanced Models:** Experiment with neural networks or combined models (e.g., XGBoost with LightGBM) to potentially achieve better performance.