Rutgers Bus Arrival Time Prediction Enhancement

Authors: Daneliz Urena & Rohan Sharma

Introduction

Efficient and reliable campus transportation is essential for Rutgers students. Current arrival-time predictions (e.g., Passio GO) often lack accuracy, causing frequent delays and frustration. Our project aims to enhance the accuracy of these predictions by integrating multiple data sources—including real-time bus geolocation, weather data, and information on campus events—to generate a more reliable and context-aware predictive model.

Objectives

- Improve arrival-time accuracy using supervised learning.
- Integrate multiple data sources (websocket, weather APIs, event schedules).
- Implement a model that adapts to special campus schedules and events.

Impact

- **Significance**: Timely and accurate bus arrivals can save thousands of students precious time, improving their daily academic and social routines.
- **Excitement**: This challenge tackles a real-world, campus-specific issue, making data science more tangible and beneficial to the broader student community.
- **Practicality**: More precise arrival predictions may reduce road congestion, shorten wait times, and enhance overall satisfaction with public transportation.

Data Sources

1. **Bus Geolocation (Passio GO)**: Real-time location and estimated arrival times.
2. **Weather Data**: Temperature, precipitation, and other meteorological factors.
3. **Campus Events (Rutgers GetInvolved)**: Large events, holiday schedules, and key academic periods (first week, midterms, finals).

Data Preparation

- **Collection**: Stream data from Passio GO, weather APIs, and campus events.

- **Feature Engineering**: Use timestamps, event tags, and weather variables. Mark academic and holiday schedules.
- **Storage**: Store records in a structured database, aligning timestamps across data sources.

Methodology

- **Neural Network Model**: Capture non-linear relationships among route data, events, and weather.
- **Daily Updates**: Incorporate each day's real outcomes to refine predictions.

Implementation Steps

1. **Database Setup**: Create tables for bus logs, weather data, and event schedules.
2. **Model Training**: Train and validate using historical data.
3. **Prediction Loop**: Generate daily predictions, compare to actuals, then retrain.

Evaluation

- **Metrics**: MAE (Mean Absolute Error) and RMSE (Root Mean Squared Error) against ground truth.
- **Comparison**: Benchmark against existing Passio GO predictions.
- **Feedback**: Gather student opinions and measure sustained accuracy over time.

Conclusion

By combining real-time data with contextual factors, this project aims to reduce waiting times and enhance the overall student experience. Such an approach underscores the role of data science in solving practical, everyday problems in academic communities.