Traffic Prediction

Adhwaith Natarajan, Ibrahim Mohsin, James Dankwah, Krish Vazirani

Accurate Traffic Prediction Matters

Enhanced Passenger Experience

Real-time predictions help improve reliability and satisfaction

Optimized Operations

For dynamic dispatch and information systems

Our Approach: Random Forest Regressor for Bus Arrival Delays

Data Ingestion

GTFStables & vehicle position logs

Data Alignment

Real-time observations matched to scheduled stops

Feature Extraction

Temporal context, route identity, vehicle motion, elapsed delay

Model Training

Random Forest predicts arrival delay in minutes

Data & Modeling

Data Sources

GTFS Static Schedule Data

Scheduled arrival/departure times (Lametro)

Real-time Vehicle-Position Logs

574,123 records over 3 hours (Lametro).

Feature Construction

1

Temporal & Positional

Hour, day-of-week, rush-hour flags, location

2

Route/Trip

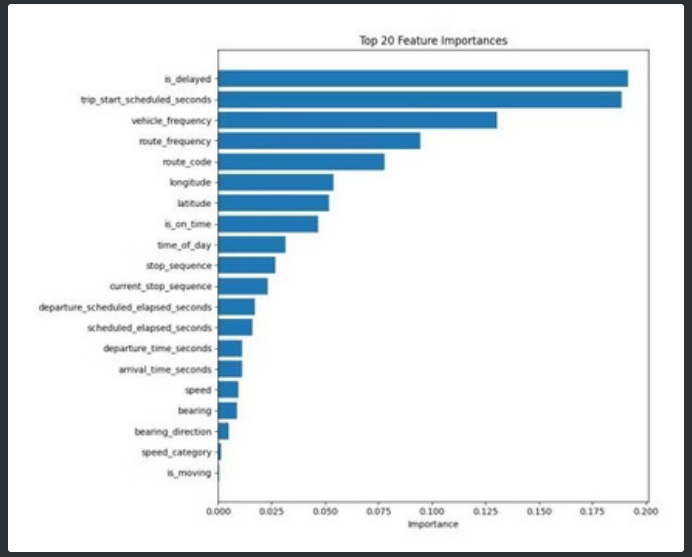
Parsed identifiers, occurrence frequencies.

3

Vehicle Motion

Speed, bearing, location validity.

Modeling Approach



Trained on 80/20 split (40,188 training, 10,047 test examples)

Evaluation Metrics

$$MAE = \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{y}_i$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Results & Conclusion

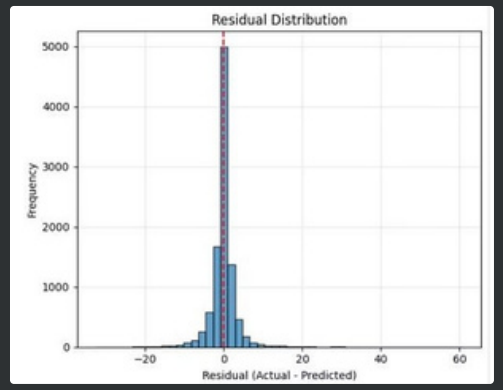
Performance Highlights

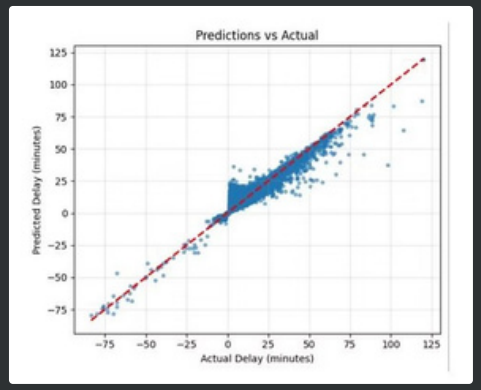
1.86
MAE (minutes)

3.42
RMSE (minutes)

Near-operational accuracy for short-term predictions

Occasional larger errors for extreme delays





Key Takeaways

1. **Random Forest Efficacy**

   Extracts meaningful structure from noisy data

2. **Data Pre-Processing Critical**

   Staged matching of real-time observations is key to results

3. **Future Enhancements**

   Incorporate external signals for extra accuracy in extreme scenarios