CS 6140 ML PROJECT PITCH

BLUE BIKE DEMAND FORECASTING

Yashwanth Reddy Vutukoori
Hariharasudan Savithri Anbarasu
Mithun Dineshkumar

INTRODUCTION (PROBLEM STATEMENT)

PROBLEM:

- PREDICTING HOURLY OR DAILY BIKE RENTAL DEMAND AT EACH BLUEBIKES STATION ACROSS BOSTON.

WHY IT MATTERS:

  - HELPS OPTIMIZE BIKE REDISTRIBUTION AND ENSURE AVAILABILITY.
  - REDUCES IDLE INVENTORY AND MISSED TRIPS DUE TO EMPTY/FULL STATIONS.
  - IMPROVES USER SATISFACTION THROUGH BETTER SERVICE RELIABILITY.

GOAL:

- BUILD REGRESSION-BASED MODELS TO FORECAST DEMAND PER STATION USING TIME, LOCATION, AND SIMPLE WEATHER OR HOLIDAY FEATURES.

IMPORTANCE AND CHALLENGES

Importance to city operations:

- Accurate forecasting supports sustainable transportation, reduces congestion, and assists in resource allocation.

Key challenges:

- Temporal variability: Demand peaks differ by hour, weekday, and season.
- Spatial imbalance: Some stations are overused while others remain underutilized.
- External effects: Weather, holidays, and city events strongly influence demand.

DATASET

Dataset: Bluebikes Boston Trip Data for the year 2020 (Kaggle)

Sample Size: Approximately 2 million+ trip records.

Key Features:
- Trip start and end timestamps
- Start and end station names, IDs, and geolocations (latitude, longitude)
- Trip duration and user type (member vs. casual)
- **Target Variable:** Hourly aggregated rental counts per station (created by grouping trip starts)

Preprocessing Steps:
- Extract temporal features (hour, day of week) from timestamps with cyclical encoding
- Aggregate trip data by station and hour to generate demand counts
- Clean station information and handle any missing values
- Optionally merge simple holiday data to account for usage anomalies

REGRESSION METHODS & WORKFLOW

- **Models to be Investigated:**
  - Linear Regression (with polynomial expansions)
  - Random Forest Regressor
  - Deep Neural Network (DNN)
- **Workflow:**
  - Data preprocessing and feature engineering
  - Train/test split
  - Cross-validation for hyperparameter tuning
- **Model Evaluation:**

Use cross-vadation to tune hyperparameters and compare models using RMSE and MAE metrics.

FINAL PRODUCT VISION

Dashboard Prototype:
- Interactive map and timeline views display predicted demand for each station.
- Scenario analysis tools allow users to toggle features like weather and time, instantly updating forecasts to support operational planning.
- Built-in model comparison and performance reporting show results for all regression models, including cross-validation metrics.

Impact:
- Enables data-driven, responsive bike transits to achieve greater reliability and convenience.

THANK YOU

ANY QUESTIONS?