Predicting Divvy BikeShare Demand Using LSTM and Bidirectional LSTM

A Time Series Forecasting Project

Project Overview

Goal: Forecast hourly total bike rentals for Divvy in Chicago

Models: Standard LSTM and Bidirectional LSTM (BiLSTM)

Data: Historical Divvy trips (2024-2025), weather station data, holidays

Benefits: Optimize bike availability, urban planning, reduce shortages

Problem Statement

Bike-sharing demand varies with seasons, weather, and events

Inaccurate predictions can lead to bike shortages or overstock

Challenge: Capture long-term dependencies in time-series data

Objective: Predict hourly total demand with high accuracy using deep learning

Data Sources and Description

Primary Data: Divvy trip records

Features: Trip start/end times, user types, stations, geographical location

External data: Weather (temperature and precipitation from NOAA), holidays

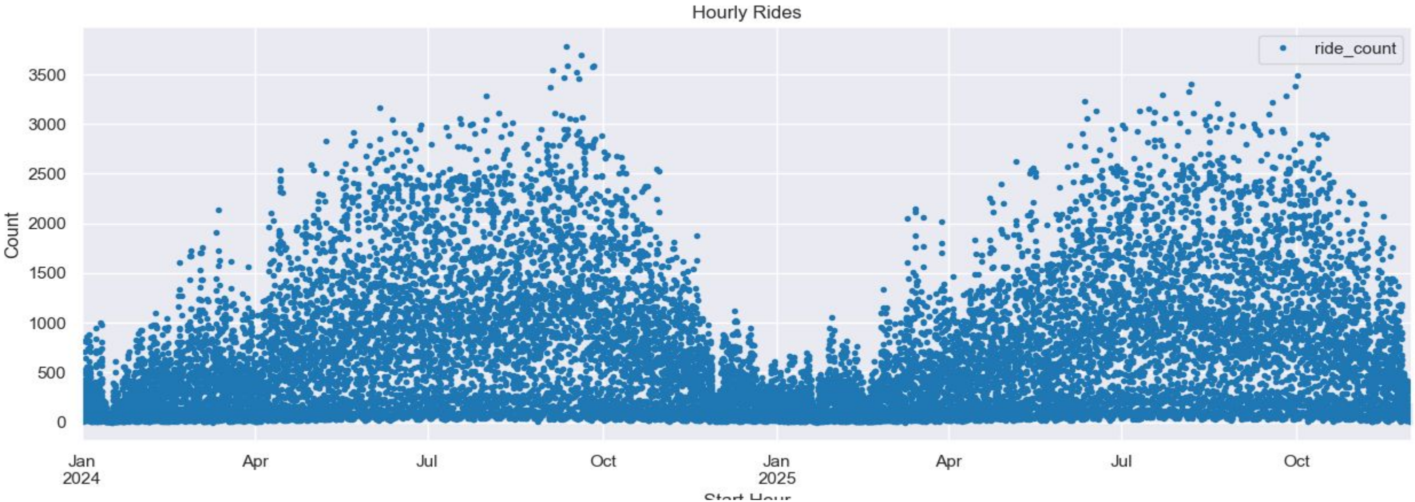
Aggregated: Hourly total rentals as target variable

Dataset Size: 16350 records with 10 features

Data preprocessing and EDA

Observed trends in data first

EDA insights: Peak in summer, correlation with temperature



Data preprocessing and EDA

Next steps: Aggregation to hourly totals, feature engineering, missing values, scaling

Train-test split: 80/20 (tried 70/30 first)

Sequence prep for LSTM input: 12 (or 24) hour sliding windows

LSTM Model Architecture

Type: Recurrent Neural Network for sequential data

Parameters: Hidden size: 50, Layers: 1, Output size: 1

Activation: ReLU; Optimizer: Adam; Loss: MSE

Strengths: Handles long-term dependencies in demand patterns

Bidirectional LSTM Model Architecture

Enhancement: Processes sequences forward and backward

Layers: Similar to LSTM, but Bidirectional wrapper on LSTM layer

Improves: Context understanding for better forecasting

Hyperparameters: Learning rate: 0.001

Training and Evaluation

Training: 100 epochs, batch size 32, early stopping

Metrics used to evaluate after training: MAE, RMSE, R^2

Baseline Comparison: Outperforms naive forecast (RMSE 300)

Challenges: Handling non-stationarity and outliers

Results and Analysis

LSTM: MAE 69.1964, RMSE 108.9451, R^2 0.9652

BiLSTM: MAE 66.4501, RMSE 105.6006, R^2 0.9673 (only slight improvement)

Key Insights: BiLSTM better at seasonal trends; underpredicts holidays

Features importance: Temperature most influential

Conclusion and Future Work

Summary: BiLSTM provides accurate hourly total demand forecasts for Divvy

Impact: Potential reduction in shortages (could do simulations to determine how much)

Limitations: No real-time data; sensitive to extreme events

Future: Station-level predictions, ensemble models, web app deployment

GitHub: https://github.com/ashleyjtoth/bikeshare-demand-prediction-Divvy