Predicting Uber Ride Demand Using Heatmaps, VGG16, and LIME

Salini Pradhan

111014877

Introduction

- The growing popularity of ride-hailing services like Uber necessitates real-time demand prediction.
- This project combines heatmaps and VGG16 to classify Uber ride demand into categories: Low, Medium, and High.
- Focus: Data preprocessing, visualization, model training, and interpretability.

Project Objectives

- Implement LIME to explain VGG16 predictions.
- Validate interpretability using diverse datasets.
- Build a framework that prioritizes both accuracy and explainability.
- Provide insights for practical deployment.

Data Collection and Heatmap Generation

- Dataset: Uber pickups in NYC (June 2014).
- Preprocessed data to generate heatmaps showing ride demand density.
- Tools: Python, pandas, folium for interactive map generation.
- Output: High-demand areas like Manhattan visualized.

Data Preprocessing

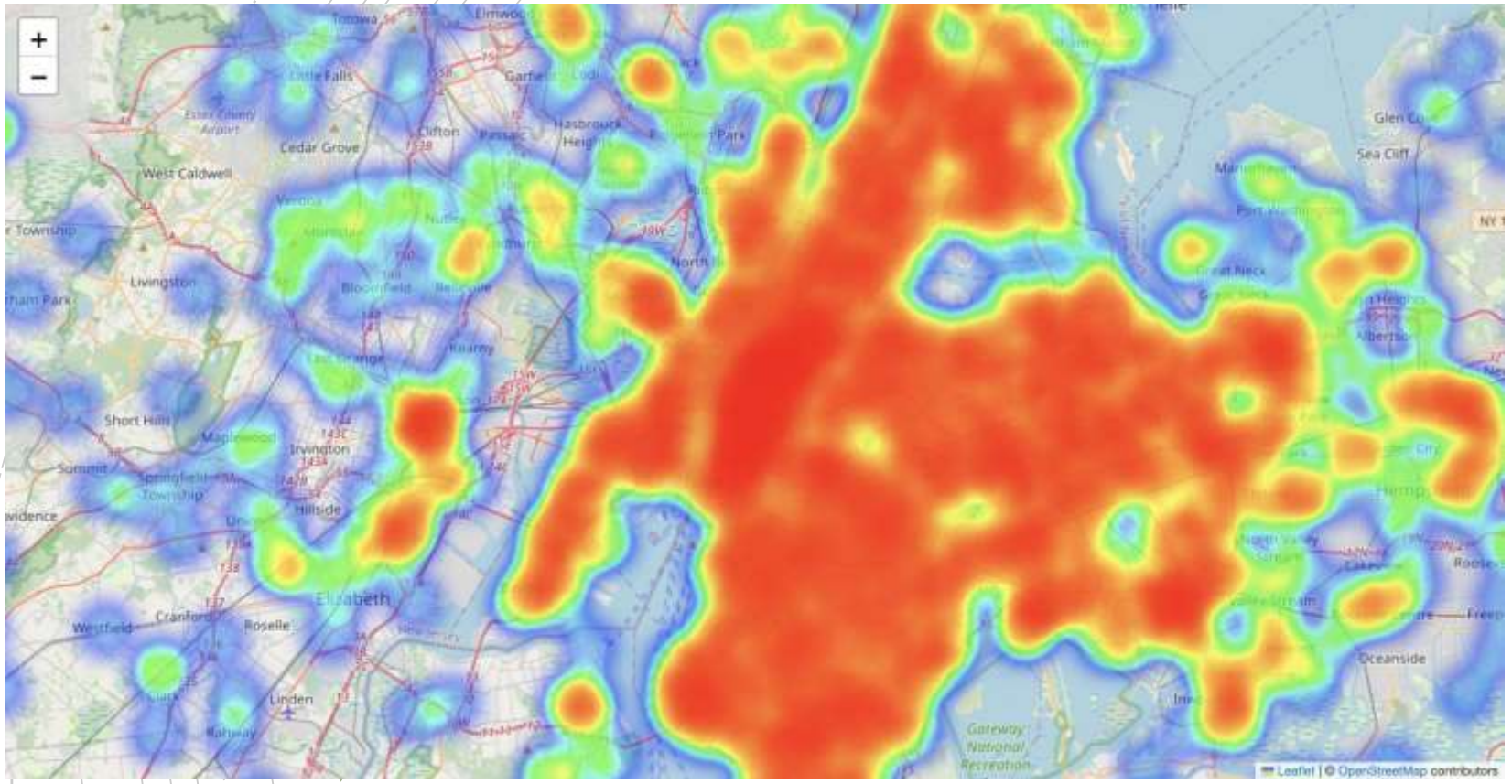
- Converted 'Date/Time' column to datetime for filtering.
- Focused on peak hours (8 AM to 10 AM).
- Filtered dataset visualized using folium heatmaps.
- Challenges: Retaining critical spatial details during preprocessing.

Model Architecture

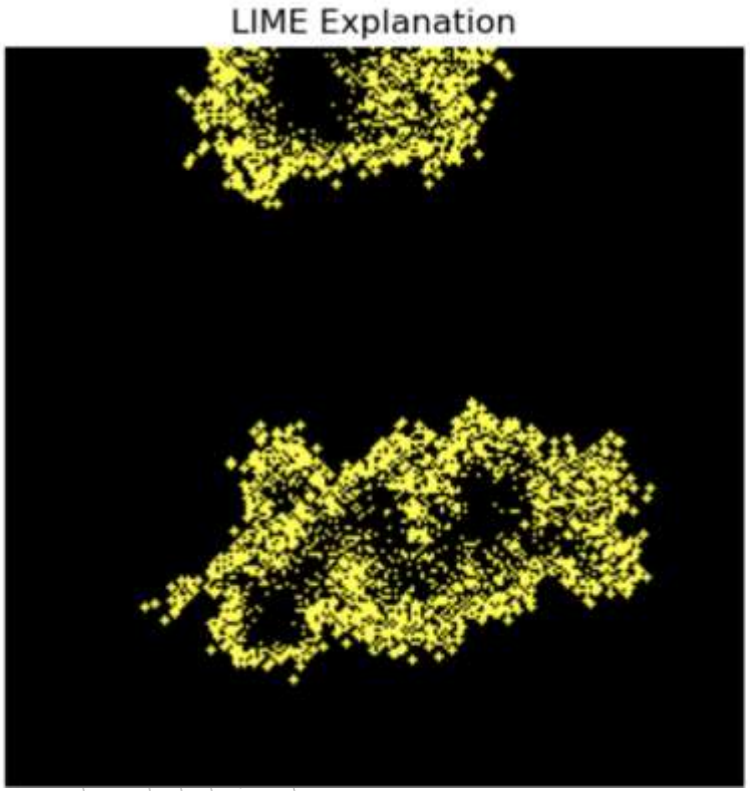
- VGG16 pre-trained on ImageNet used as base.
- Added custom layers:
  - Flatten layer.
  - Dense layer with 256 neurons (ReLU activation).
  - Softmax output layer for classification.
- Optimizer: Adam, Loss Function: Categorical Crossentropy.

Training and Results

- LIME highlights yellow clusters showing areas most important for the model's prediction, likely high-demand regions like Manhattan.
- This helps confirm the model is focusing on relevant data, enhancing trust in its predictions.



- The heatmap shows areas of ride demand, with red zones indicating high demand around New York City and blue/green zones indicating lower demand in surrounding regions.



- LIME highlights yellow clusters showing areas most important for the model's prediction, likely high-demand regions like Manhattan.
- This helps confirm the model is focusing on relevant data, enhancing trust in its predictions.

Interpretability (LIME Explanations)

- LIME highlights regions influencing model predictions.
- Yellow clusters in heatmaps show high-density zones.
- Ensures the model focuses on relevant spatial features.
- Key Insight: Manhattan zones often dominate Medium demand.

Challenges and Solutions

- Handling image data: Ensured consistent resizing and normalization.
- Avoiding overfitting: Focused on regularization and dataset expansion.
- Interpretability: LIME explains predictions, validating model focus.
- Addressed potential distortions in heatmaps using interpolation.

Conclusion and Futurework

- Heatmaps provide intuitive visualization of demand distribution.
- VGG16 effectively classifies demand levels (Low, Medium, High).
- Future Work:
  - Dataset expansion (temporal and geospatial diversity).
  - Enhance model performance with advanced architectures.
  - Deploy real-time demand prediction system.

Thank You!