

Team 6 Code Sample doc

This document provides detailed instructions on setting up the environment, running the code, and interpreting the results for the METCS777 Term Project. The project focuses on bike rebalancing using predictive analytics and optimization techniques.

Environment Setup

To ensure that all dependencies are met, follow these steps to set up your project environment:

Prerequisites

Python 3.8 or higher
pip (Python package installer)

Installation Steps

1. Clone the Repository: Open your terminal and run the following command to clone the project repository:

```
git clone
```

```
https://github.com/dreamfireyu/METCS777-Term-Project.git
```

```
cd METCS777-Term-Project
```

2. Install Required Libraries: Use pip to install the required Python packages listed in requirements.txt:

```
pip install -r requirements.txt
```

3. Spark installation:

MAC:

<https://github.com/charlie-ph/BigDataAnalytics/blob/master/Installations-HowTos/How-To-Install-Spark-On-MACOS.md>

Windows:

<https://github.com/charlie-ph/BigDataAnalytics/blob/master/Installations-HowTos/How-To-Install-Spark-On-Windows.md>

How to Run the Code

Navigate to the src/ directory where the source code files are located. Here are the steps to execute each script:

1. Data Preparation: These scripts clean and prepare the data for analysis.

 Data_Preparation.ipynb: Raw data ETL and net_flow data preprocessing

 Station_Data_Pre.ipynb: Station data ETL process.

2. Predictive Modeling: Executes the machine learning models to predict net_flow data for every station.

models/: Different models trained for net_flow prediction.

4. Optimization Model: Solves the bike rebalancing optimization problem.

 Model_Optimization.ipynb

5. WebApp Demo:

 python webapp.py

Results

After running the code, the outputs will be stored in the same directory.

Predictive Results: These include CSV files with predicted bike demands per station in the next 24 hours.

Optimization Results: Outputs from Model_Optimization.ipynb show the optimal number of bikes to be rebalanced at each station.

Dataset and Results Explanation

Dataset

Input Data: trip_data/

Contains historical data on bike rentals and returns.

Columns: "tripduration", "starttime", "stoptime", "start station id", "start station name", "start station latitude", "start station longitude", "end station id", "end station name", "end station latitude", "end station longitude", "bikeid", "usertype", "postal code".

Processed Data:

Merged_Data_202101_202303_v3.csv

Merged_Station.csv

Cleaned and transformed data ready for modeling.

Understanding the Results

Predictive Modeling:

The machine learning models predict the net-flow at each station based on historical data and environmental factors like weather.

Optimization Model:

Using the predictions, the optimization model calculates how many bikes should be moved to or from each station to meet expected demand and minimize costs.

Results include lists of each station with the number of bikes to add or remove.

Example Output

```
Station 101: +5 bikes
Station 102: -3 bikes
Station 103: +8 bikes
```

This indicates that 5 bikes should be moved to Station 101, 3 bikes should be removed from Station 102, and 8 bikes should be added to Station 103 to optimize for the next operational period.

Conclusion

This documentation should assist in effectively setting up, running, and understanding the code and results of the METCS777 Term Project. For further details or assistance, refer to the project's GitHub repository or contact the project maintainers.