

ARTIFACT APPENDIX

1. Project Overview

CS5296 Cloud Computing— Theory and Practice, Spring 2024 Group 10 Project

The main code of this project is in: `Project.ipynb`

The data set storage path is: `cal_housing\CaliforniaHousing\cal_housing.data`

2. Github Repository

The link to the Github repository of the project is as follow:

https://github.com/Ysenniko/CityU_CS5296_CloudComputing_Project

3. Detailed Installation Steps

1) Clone the Repository: Users need to clone the project from GitHub to local. Specific command:

```
git clone https://github.com/Ysenniko/CityU_CS5296_CloudComputing_Project.git
```

2) Install Dependencies: The project requires the following dependencies to be installed:

- Recommended openjdk version: v17.0
- Recommended python version: python==3.9
- pyspark==3.5.1
- numpy
- pandas
- pillow
- seaborn
- torch
- torchvision
- torchaudio
- scikit-learn

All dependencies in the `requirements.txt` file can be installed using the following command:

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

Please make sure you have changed to the directory containing the requirements.txt file on the command line and then run the above command.

4. How to Run Our Project

After the code running environment is set up, enter the `Project.ipynb` file. This file contains the main code implementation of the project. The project can be reproduced and verified by running the code in this Jupyter Notebook.