Final Project Description and Summary

MS548 Advanced Programming Concepts and AI

Patrick Hill

April 24, 2024

**Summary**

This project focuses on analyzing and predicting housing prices based on various features of the housing dataset using machine learning models. The code provided loads and preprocesses the data, creates visualizations to understand the distributions and correlations among the features, and then splits the data into training and test sets. Various regression models, including Linear Regression, Ridge Regression, Lasso Regression, and Random Forest Regressor, are employed to predict the median house value. The models are evaluated using metrics like R^2 score, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). Additionally, feature importances are analyzed to understand which factors contribute most to predicting housing prices. Hyperparameter tuning is performed using GridSearchCV to optimize the Random Forest model.

**Project Dependencies and Platforms**

*Python Libraries:*

* Numpy: For numerical operations.
* pandas: For data manipulation and analysis.
* scikit-learn: For implementing machine learning models, model selection, and evaluation metrics.
* bqplot: For creating interactive visualizations in Jupyter notebooks.
* ipywidgets: For creating interactive widgets in Jupyter notebooks.

**Development Environment:**

Jupyter Notebook: An open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text.

**Installation and Running Instructions**

*Environment Setup:*

Ensure Python and Jupyter Notebook are installed on your system. If not, you can install them using the following steps:

*Install Python:*

bash

Copy code

sudo apt-get update

sudo apt-get install python3.6

*Install Jupyter Notebook:*

bash

Copy code

pip install notebook

*Dependency Installation:*

Install the required Python libraries using pip. It is recommended to use a virtual environment to avoid conflicts with other project dependencies.

*Create a virtual environment (optional):*

bash

Copy code

python -m venv myenv

source myenv/bin/activate # On Windows use `myenv\Scripts\activate`

*Install required libraries:*

bash

Copy code

pip install numpy pandas scikit-learn bqplot ipywidgets

*Running the Project:*

Start Jupyter Notebook:

bash

Copy code

jupyter notebook

Open the notebook file (e.g., mlm\_house\_price\_prediction\_v5.ipynb) in the Jupyter interface that opens in your web browser.

Run the cells sequentially to load the data, visualize, train the models, and evaluate their performance.

This setup will allow you to experiment with the dataset and models, make modifications, and see the effects in real-time through interactive visualizations.