

Machine Learning

Semester Project

Spyridoula Spyrakou (mtn2321) – Christos Georgios Foukanelis (mtn2324) – Panagiotis Tsilimidos (mtn2322)

Execution Details

February 2024



Project File Structure

/:

Demo/ : [Folder with demo code for project]

Demo.ipynb [Notebook containing calls to train/evaluation methods]

Model_Testing.py [.py File containing class and methods for model training/evaluation]

model.sav [pre-trained model]

x_test.csv [training data for demo]

x_train.csv [testing data for demo]

y_test.csv [testing data for demo]

y_train.csv [training data for demo]

src/ : [Folder with full source code and notebooks]

Dataset_Preparation/ : [Folder with data preparation code]

Data.ipynb [Notebook with data curation preprocessing]

Initial_dataset.csv [Initial dataset with complete features]

X_dataset.scv [Combined feature samples]

Y_merged.csv [Combined label samples]

Dataset_Exploration/ : [Folder with data exploration code]

Analysis.ipynb [Notebook with Statistical Analysis of Dataset]

Exploration.ipynb [Notebook with features and ml methods tries]

Final_dataset.csv [Final dataset with relevant features]

Feature_Engineering.ipynb [Notebook with extra domain features]

Modelling/ : [Folder with model training code]

Model.ipynb [Notebook with Model training / tuning / evaluation]

X.csv [Feature Dataset]

Y.csv [Sample Dataset]

environment_droplet.yml [yaml file with environment dependencies]

README.pdf [current file]

REPORT.pdf [project documentation]

README.md [github generated readme file]

requirements.txt [libraries used]

presentation.pptx [project summary presentation]

Setup Instructions

The *requirements.txt* contains all the library dependencies needed to be installed in an existing environment for project execution.

Alternatively, the *environment_droplet.yml* file contains all the dependencies needed for the virtual environment used for the project to be created.

For the source code download, a *git clone* of this repository is all is needed.

GitHub Repo URL: https://github.com/fouk21/ML_PROJECT_MSC_AI

Just for fun, the model is also deployed in flask app on a docker container. Instructions for the docker setup are on the file README.md

Demo Code Execution

For the demo code to be executed, the *Demo.ipynb* should be run (using *Run All*). This notebook contains a pipeline for loading the provided pre-trained model and testing it while also containing a pipeline for training and testing a new model with the datasets provided. The source code used in these two pipelines exists on the *Model_Testing.py* file and is intentionally hidden from the demo code to provide clear distinction between the demo examples to be executed.

Notes

The *object oriented architecture* is used in most of the notebooks because of code repetition and because it provides a robust representation of the end-to-end model/data pipeline.