

Configuration Manual

MSc Research Project
Data Analytics

Tushar Patil
Student ID: x18182020

School of Computing
National College of Ireland

Supervisor: Manaz Kaleel

National College of Ireland

MSc Project Submission Sheet

Student Name: Tushar Santosh Patil

Student ID: X18182020

Programme: Data Analytics

Year: 2020

Module: MSc Research Project

Supervisor: Manaz Kaleel

Submission

Due Date: 17/08/2020

Project Title: Food Authentication Using Dimensionality Reduction techniques and Ensemble Algorithms on Spectroscopic Datasets

Word Count: 305 **Page Count:** 5

School of Computing

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:

Date: 14/08/2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	<input type="checkbox"/>
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Tushar Patil
Student ID: x18182020

1. Project Directory details:

The screenshot shows the JupyterLab interface for a project directory named 'ResearchProject_Python_Latest'. The directory contains the following files and folders:

- ..
- execute_program.ipynb
- adaboost_dct_honey.pkl
- adaboost_dct_meat.pkl
- adaboost_dctumap_meat
- adaboostdct__orig_olive.pkl
- adaboostdct_lda_honey.pkl
- adaboostdct_lda_olive.pkl
- adaboostdct_orig_honey.pkl
- adaboostdct_pca_honey.pkl
- adaboostdct_pca_meat.pkl
- adaboostdct_pca_olive.pkl
- adaboostdct_umap_honey.pkl
- adaboostdct_umap_olive.pkl
- adaboostsvm_lda_meat.pkl
- createEmbeddings.py
- datasetsSplitTrainTestVal.py
- datasetsVisualizations.py
- honeydata.csv
- hyperparam.csv
- Implementation.py
- meats.txt
- oliveoils.txt
- Options.py
- randomforest_honey.pkl
- randomforest_olive
- rf_meat.pkl
- trainClassifiers.py

Project Directory name : ResearchProject_Python_Latest (in the screen print above)

The Project comprises of 6 python files, 3 datasets files, 1 file containing hyperparameters and 15 files containing saved pre-trained model objects. All these files are present in the project directory name mentioned above as shown in the screen print. Below are the file details.

Python Code files (6 files):

1. Implementation.py
2. datasetsSplitTrainTestVal.py
3. createEmbeddings.py
4. trainClassifiers.py
5. datasetsVisualizations.py
6. Options.py

Datasets (3 files):

1. meats.txt
2. oliveoils.txt
3. honeydata.csv

Hyperparameter (1 file):

1. hyperparam

Pre-trained model objects (15 files):

1. adaboost_dct_meat.pkl, adaboostdct_pca_meat.pkl, adaboostsvm_lda_meat.pkl, adaboost_dctumap_meat and rf_meat.pkl.
2. adaboost_dct_honey.pkl, adaboostdct_lda_honey.pkl, adaboostdct_orig_honey.pkl, adaboostdct_pca_honey.pkl and adaboostdct_umap_honey.pkl.
3. adaboostdct__orig_olive.pkl, adaboostdct_lda_olive.pkl, adaboostdct_pca_olive.pkl, adaboostdct_umap_olive and randomforest_olive

2. Instructions to execute the program :

1. Launch the Jupyter and navigate to the project directory - **ResearchProject_Python_Latest** containing all the python files and dataset files.
2. Click on new button and create a new notebook with Python 3. This will open a new Jupyter notebook. Name this notebook as **execution_program** as shown in below screen shot highlighted in yellow.

Files Running Clusters

Select items to perform actions on them.

0 / ResearchProject_Python_Latest

- ..
- execute_program.ipynb**
- adaboost_dct_honey.pkl
- adaboost_dct_meat.pkl
- adaboost_dctumap_meat
- adaboostdct_orig_olive.pkl
- adaboostdct_lda_honey.pkl
- adaboostdct_lda_olive.pkl
- adaboostdct_orig_honey.pkl
- adaboostdct_pca_honey.pkl
- adaboostdct_pca_meat.pkl
- adaboostdct_pca_olive.pkl
- adaboostdct_umap_honey.pkl
- adaboostdct_umap_olive.pkl
- adaboostsvm_lda_meat.pkl
- createEmbeddings.py
- datasetsSplitTrainTestVal.py

- In an execution_program notebook page, in a cell, enter the command- **%load Implementation.py** and click on Run button and execute the cell containing: **%load Implementation.py**
- All the code in the file Implementation.py is loaded in the cell as shown in below screen print.

jupyter execute_program Last Checkpoint: Last Monday at 17:38 (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Run

```
In [5]: # %load implementation.py
#!/usr/bin/env python

# In[20]:

import numpy as np
import pandas as pd
from sklearn.ensemble import AdaBoostClassifier, RandomForestClassifier

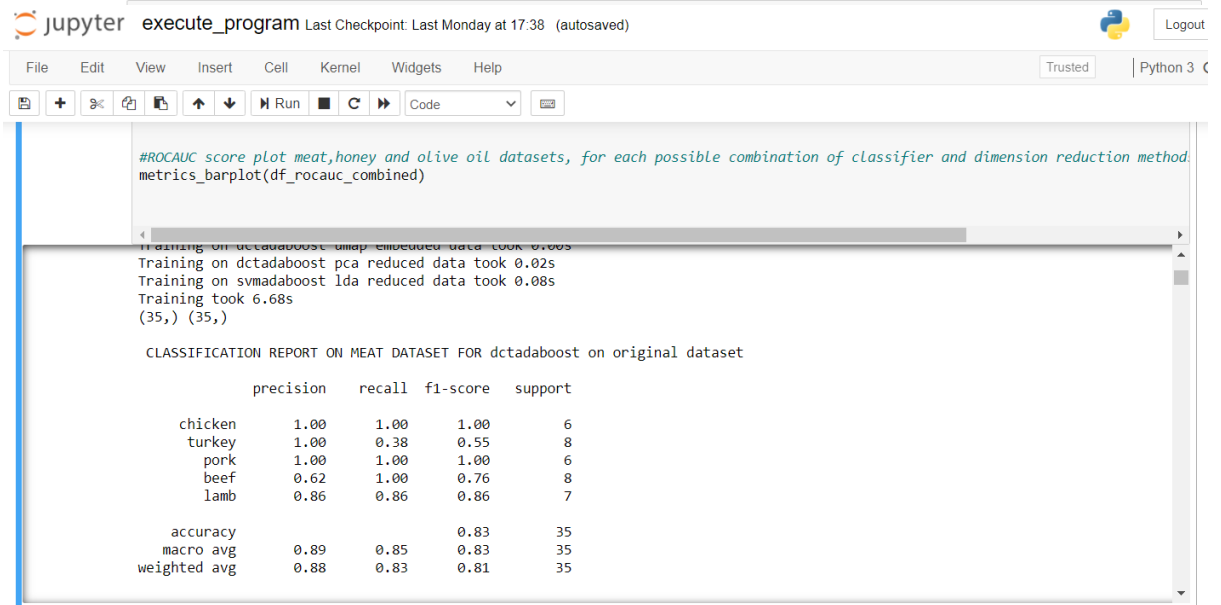
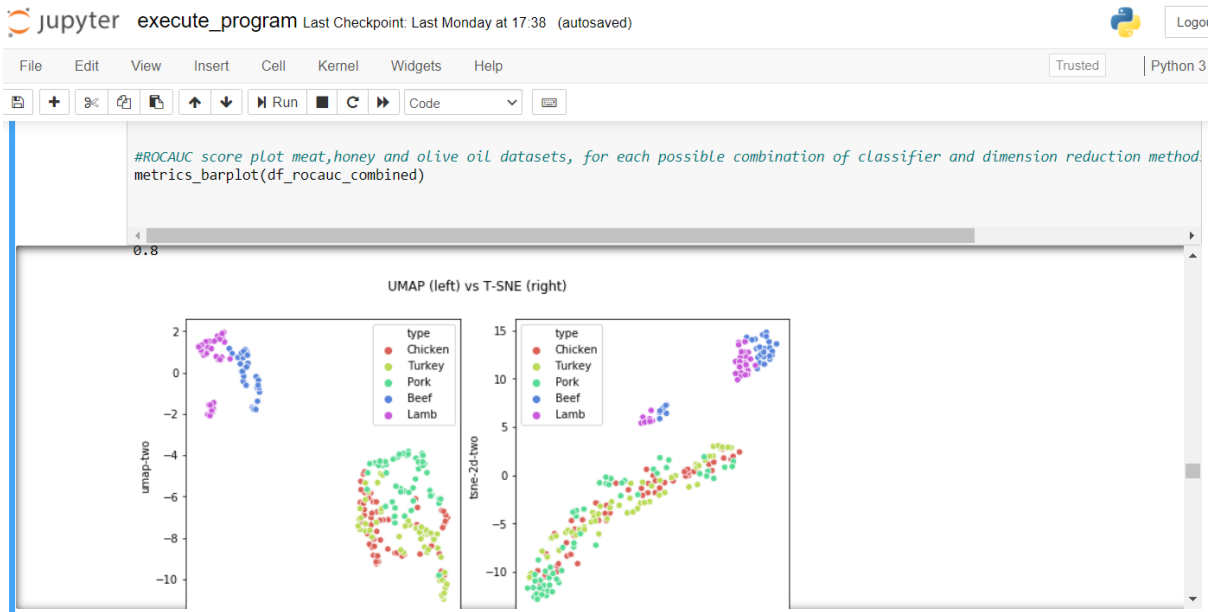
get_ipython().run_line_magic('matplotlib', 'inline')
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import seaborn as sns
from sklearn import preprocessing

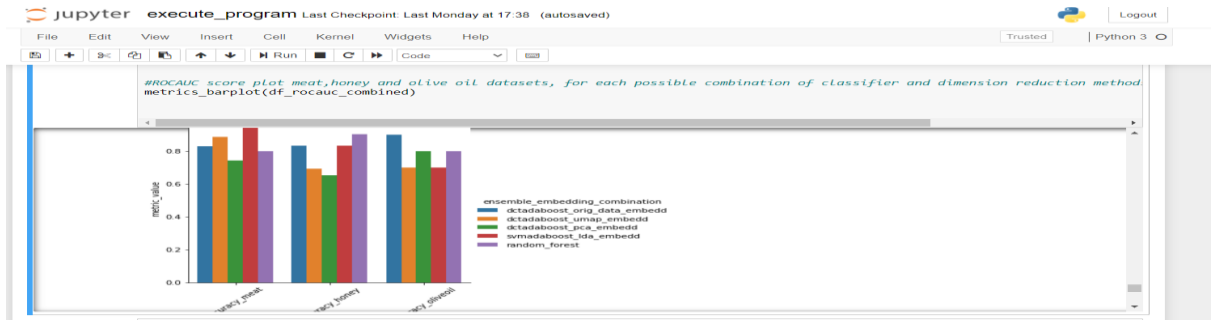
import import_ipynb
import datasetsSplitTrainTestVal
from datasetsSplitTrainTestVal import *
import trainClassifiers
from trainClassifiers import *
import datasetsVisualizations
from datasetsVisualizations import *

from sklearn.metrics import confusion_matrix, precision_score, recall_score, f1_score, classification_report
from sklearn.metrics import accuracy_score
from sklearn.metrics import roc_auc_score

get_ipython().run_line_magic('load_ext', 'autoreload')
get_ipython().run_line_magic('reload_ext', 'autoreload')
```

5. Click on the run button and execute the cell containing the loaded code from Implementation.py.
6. All the execution results are visible at the bottom of Jupyter notebook
 - Visualization graph
 - Classification Report
 - Accuracy bar graph
 - ROC-AUC score bar graph





7. Run the entire code again to again train the classifiers to get the expected results :

- accuracy of 97 % for ADABOOST-SVM-LDA classifier on meat dataset.
- accuracy of 90 % for Random Forest classifier on Honey dataset.
- accuracy of 90% for ADABOOST-SVM-LDA classifier on olive oil dataset.