National College of Ireland

Configuration Manual

MSc Research Project
Cybersecurity

Loveth Ukamaka Diwe
Student ID: x21180211

School of Computing
National College of Ireland

Supervisor: Imran Khan

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: LOVETH UKAMAKA DIWE……

Student ID: X21180211……..……

Programme: Cyber Security…… **Year:** 2023…….

Module: Research Project……..……

Lecturer: Imran Khan……..……

Submission Due Date: 14/08/2023……..……

Project Title: Detection of FTP and SSH Bruteforce attacks using Deep Belief Network Model ……..……

Word Count: 1203…… **Page Count:** 10……..……..……

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: LOVETH UKAMAKA DIWE……

Date: 13/08/2023……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
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.	□

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

Loveth Ukamaka Diwe
Student ID: x21180211

1 Introduction

The configuration manual details the Host systems, applications and the execution of codes used for the project development. It also offers brief explanations and guides on the applications and installation techniques.

Below is the design specification used for the development of this study.

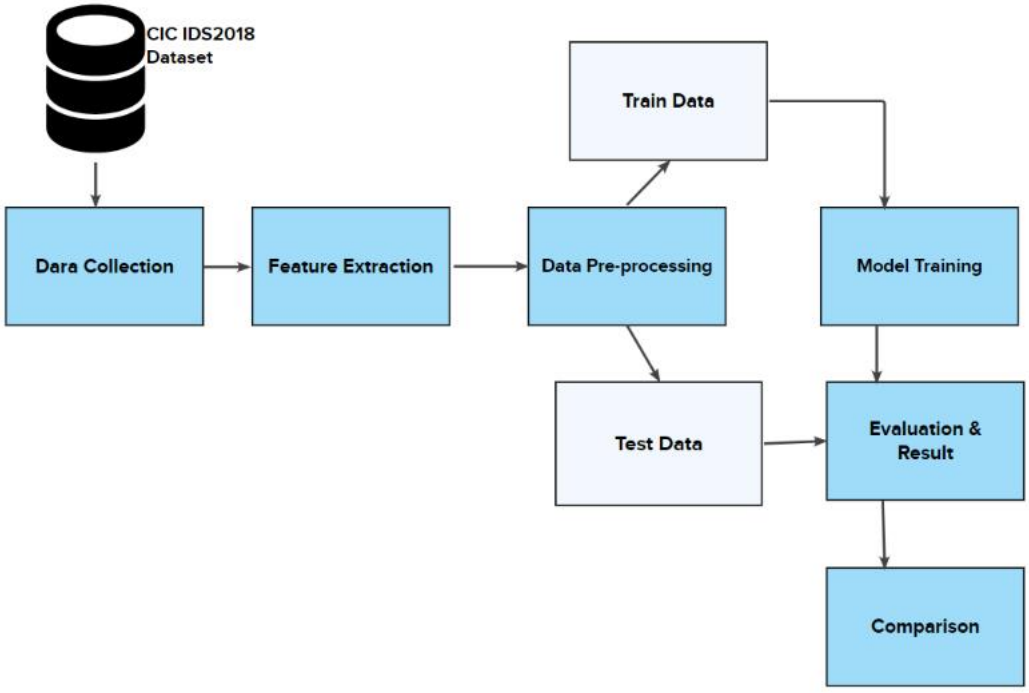

Fig.1. Design Specification

2 Research Lab

My personal computer (HP ProBook), Windows 11 Pro (v21H2) with 512GB storage size and 16GB installed RAM was used for the research development and documentation.

3 Applications

Below are the applications installed and used to carry out the analysis.

3.1 CICFlowMeter

This is a network traffic flow generator implemented in Java and Python and used to generate real-time traffic. It is equally used as an analyser to generate bi-directional flows (UNB, 2023). The application can calculate network features like length of packets, number of bytes, duration etc. The app can be downloaded from the University of New Brunswick website (https://www.unb.ca/cic/research/applications.html)

3.2 AWS CLI

The Amazon Web Services (AWS) Command Line Interface (CLI) is a sophisticated tool that allows developers, administrators, and other users to connect with numerous AWS services using the command-line interface. You can manage AWS resources, automate tasks, and control your AWS services directly from the terminal or command line using the AWS CLI.
The tool was installed and used with the windows command shell to download the dataset.
After the downloading and installation run the following commands to download the complete CSE-CIC-IDS2018 dataset.

```
Microsoft Windows [Version 10.0.22621.2134]
(c) Microsoft Corporation. All rights reserved.

C:\Users\ADMIN>aws s3 sync --no-sign-request --region eu-west-1 "s3://cse-cic-ids2018/" C:\Users\ADMIN\Desktop
```

Fig.2. AWS command to download the dataset.

To download only a particular sub data of the dataset, first use this command to browse directories.

```
C:\Windows\System32>aws s3 ls --no-sign-request "s3://cse-cic-ids2018" --recursive --human-readable --summarize
2018-10-10 12:52:09    0 Bytes Original Network Traffic and Log data/
2018-10-10 12:52:23    0 Bytes Original Network Traffic and Log data/Friday-02-03-2018/
2018-10-10 13:00:39  225.8 MiB Original Network Traffic and Log data/Friday-02-03-2018/logs.zip
2018-10-10 13:00:51   41.7 GiB Original Network Traffic and Log data/Friday-02-03-2018/pcap.zip
2018-10-10 12:52:34    0 Bytes Original Network Traffic and Log data/Friday-16-02-2018/
2018-10-10 13:45:49  148.1 MiB Original Network Traffic and Log data/Friday-16-02-2018/logs.zip
2018-10-10 13:46:01   35.9 GiB Original Network Traffic and Log data/Friday-16-02-2018/pcap.zip
2018-10-10 12:52:41    0 Bytes Original Network Traffic and Log data/Friday-23-02-2018/
2018-10-10 13:46:10  199.8 MiB Original Network Traffic and Log data/Friday-23-02-2018/logs.zip
2018-10-10 13:46:31   55.0 GiB Original Network Traffic and Log data/Friday-23-02-2018/pcap.zip
2018-10-10 12:52:47    0 Bytes Original Network Traffic and Log data/Thursday-01-03-2018/
2018-10-10 14:41:13  217.1 MiB Original Network Traffic and Log data/Thursday-01-03-2018/logs.zip
2018-10-10 14:41:45   48.8 GiB Original Network Traffic and Log data/Thursday-01-03-2018/pcap.zip
2018-10-10 12:52:54    0 Bytes Original Network Traffic and Log data/Thursday-15-02-2018/
2018-10-10 14:41:28  142.6 MiB Original Network Traffic and Log data/Thursday-15-02-2018/logs.zip
2018-10-10 14:41:55   38.4 GiB Original Network Traffic and Log data/Thursday-15-02-2018/pcap.zip
2018-10-10 12:53:01    0 Bytes Original Network Traffic and Log data/Thursday-22-02-2018/
2018-10-10 14:41:42  195.3 MiB Original Network Traffic and Log data/Thursday-22-02-2018/logs.zip
2018-10-10 14:42:27   46.8 GiB Original Network Traffic and Log data/Thursday-22-02-2018/pcap.zip
2018-10-10 12:53:07    0 Bytes Original Network Traffic and Log data/Tuesday-20-02-2018/
2018-10-10 15:39:45  178.9 MiB Original Network Traffic and Log data/Tuesday-20-02-2018/logs.zip
2018-10-10 15:40:40   41.3 GiB Original Network Traffic and Log data/Tuesday-20-02-2018/pcap.rar
2018-10-10 12:53:14    0 Bytes Original Network Traffic and Log data/Wednesday-14-02-2018/
2018-10-10 17:44:20  133.7 MiB Original Network Traffic and Log data/Wednesday-14-02-2018/logs.zip
2018-10-11 13:22:03   37.2 GiB Original Network Traffic and Log data/Wednesday-14-02-2018/pcap.zip
2018-10-10 12:53:21    0 Bytes Original Network Traffic and Log data/Wednesday-21-02-2018/
2018-10-10 17:44:34  185.6 MiB Original Network Traffic and Log data/Wednesday-21-02-2018/logs.zip
2018-10-11 14:35:15   49.8 GiB Original Network Traffic and Log data/Wednesday-21-02-2018/pcap.zip
2018-10-10 12:53:28    0 Bytes Original Network Traffic and Log data/Wednesday-28-02-2018/
2018-10-10 17:44:47  216.1 MiB Original Network Traffic and Log data/Wednesday-28-02-2018/logs.zip
2018-10-11 15:21:03   49.6 GiB Original Network Traffic and Log data/Wednesday-28-02-2018/pcap.zip
2018-10-11 17:02:25    0 Bytes Processed Traffic Data for ML Algorithms/
2018-10-11 17:02:49  336.0 MiB Processed Traffic Data for ML Algorithms/Friday-02-03-2018_TrafficForML_CICFlowMeter.csv
2018-10-11 17:03:10  318.3 MiB Processed Traffic Data for ML Algorithms/Friday-16-02-2018_TrafficForML_CICFlowMeter.csv
2018-10-11 17:03:33  365.1 MiB Processed Traffic Data for ML Algorithms/Friday-23-02-2018_TrafficForML_CICFlowMeter.csv
2018-10-11 17:03:59    3.8 GiB Processed Traffic Data for ML Algorithms/Thuesday-20-02-2018_TrafficForML_CICFlowMeter.csv
2018-10-11 17:08:38  102.8 MiB Processed Traffic Data for ML Algorithms/Thursday-01-03-2018_TrafficForML_CICFlowMeter.csv
2018-10-11 17:08:48  358.5 MiB Processed Traffic Data for ML Algorithms/Thursday-15-02-2018_TrafficForML_CICFlowMeter.csv
```

Fig.3. The complete CSE-CIC-IDS2018 dataset directories

Then use the command below to download the subject file (Wednesday-28-02-2018_TrafficForML_CICFlowMeter.csv)

```
Microsoft Windows [Version 10.0.22621.2134]
(c) Microsoft Corporation. All rights reserved.

C:\Users\ADMIN>aws s3 cp --no-sign-request "s3://cse-cic-ids2018/Processed Traffic Data for ML Algorithms/Wednesday-14-0
2-2018_TrafficForML_CICFlowMeter.csv" C:\Users\ADMIN\Desktop
```

Fig.4. command to download the subject sub dataset.

3.3 ANACONDA

Anaconda is used majorly for data scientist for writing and executing Python programming and R programming languages (Anaconda, 2023).
This application can be downloaded straight from the Anaconda website (Anaconda, 2023).
Install the application following the instruction steps on the page.

For Windows, once successfully installed, to start the application, search for anaconda Navigator on the windows start menu and click to load. The page loads and displays the home page which shows all the environment packages and applications that can be accessed from the application.

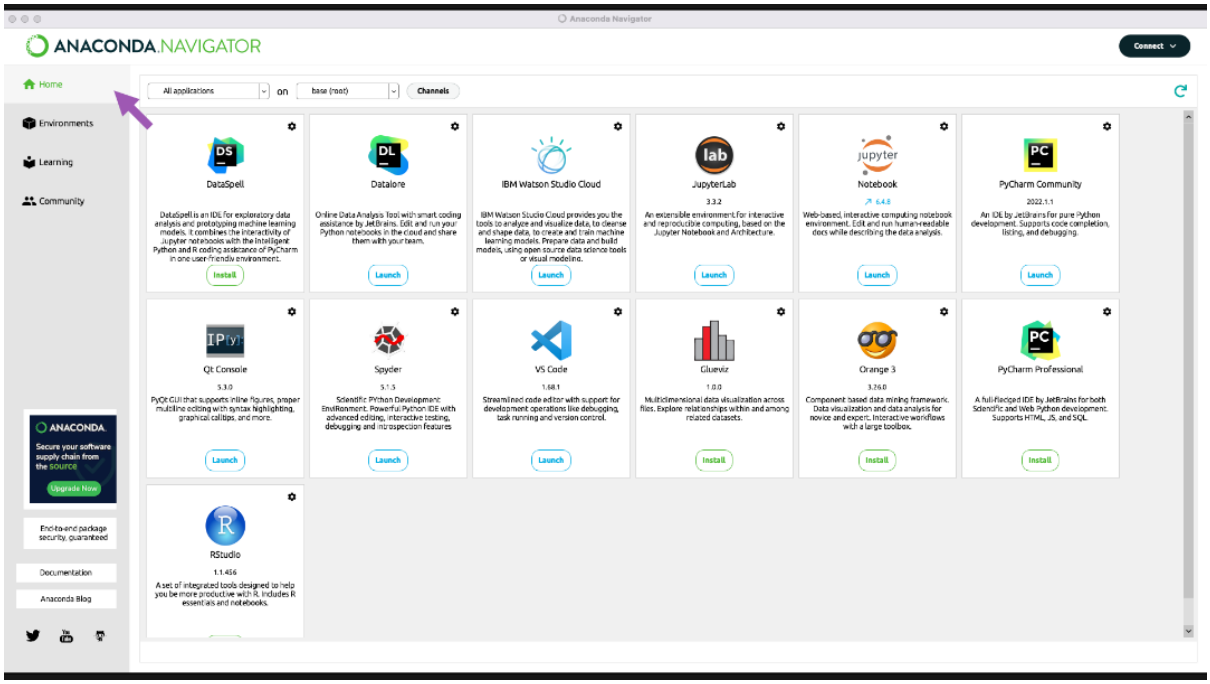

Fig.5. Anaconda Home Page (Anaconda, 2023).

3.4 Jupyter Notebook

Jupyter Notebook is one of the pre-installed Applications in Anaconda. Jupyter notebook is a web-based environment used for scientific computations and data manipulations/analysis (Jupyter, 2023).
To launch the application, click “Launch” on the Jupyer notebook icon displayed on the Anaconda home page. The application opens on the Edge browser as a localhost.

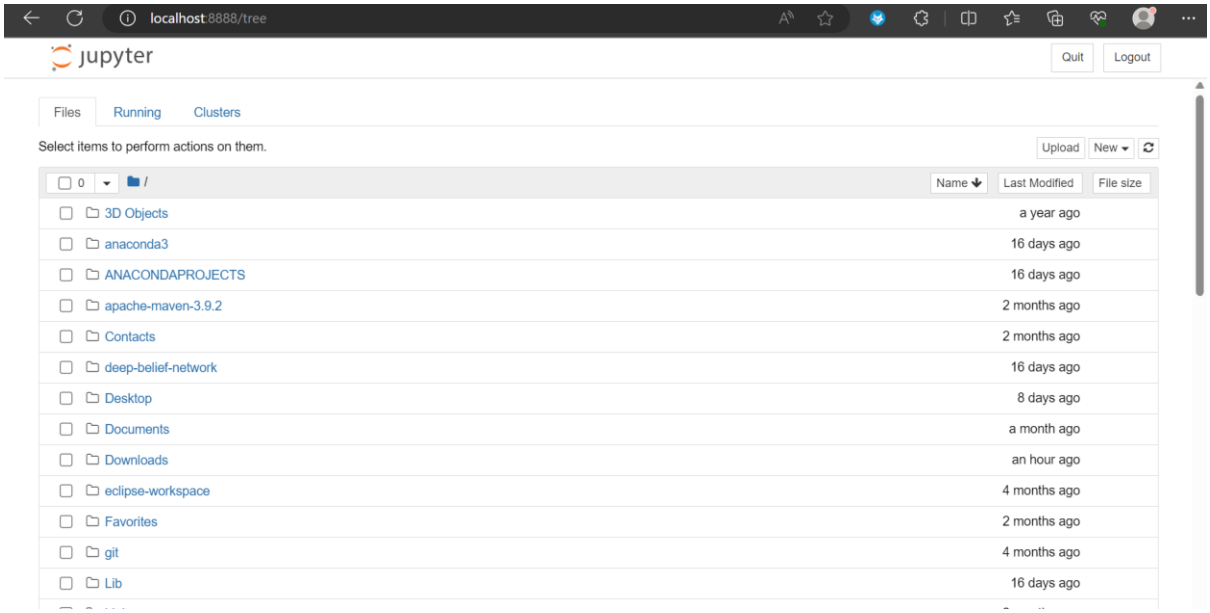

Fig.6. Jupyter Notebook Page

.

3.5 Python

Python is the programming language used for the study analysis. The Python version (3.10.8) can be downloaded and installed from the official web page (Python, 2023). All the python libraries used for the study analysis are already pre-installed in the Jupyter Notebook.

4 Execution of Codes

All codes were written in Python programming language and executed on the Jupyter Notebook which was already pre-installed in Anaconda with all required python libraries.

There are total of 4 codes written for the complete analysis.

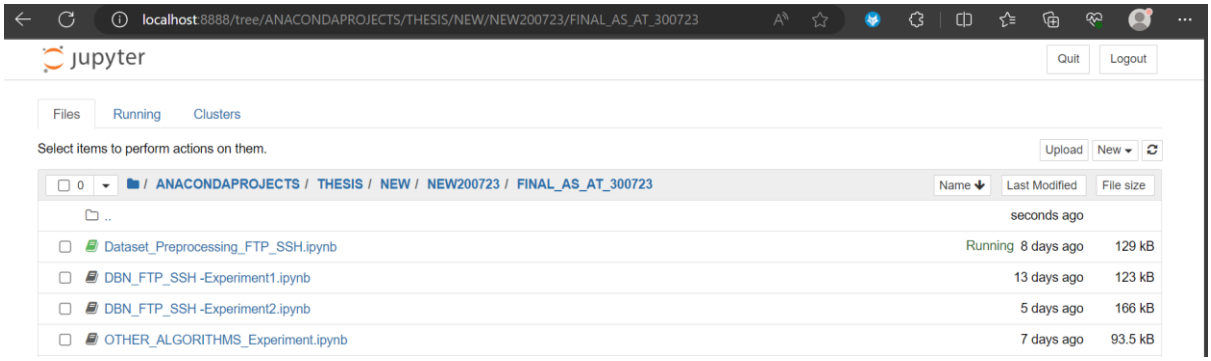

Fig.7. Code files

Save the 4 files in the user's home page (same directory with the installed anaconda application).
Below are the codes and steps to execute them arranged in order.

4.1 Dataset_Preprocessing_FTP_SSH.ipynb

This is the first code written to load the original dataset (Wednesday-28-02-2018_TrafficForML_CICFlowMeter.csv) downloaded for the analysis which consist of Benign, FTP-Bruteforce and SSH network features. The code also handled feature extraction and pre-processing the dataset.

To execute the code, start the anaconda navigator from the windows start page, launch the Jupyter notebook. Click on the code and execute using the Kernel button to run all.

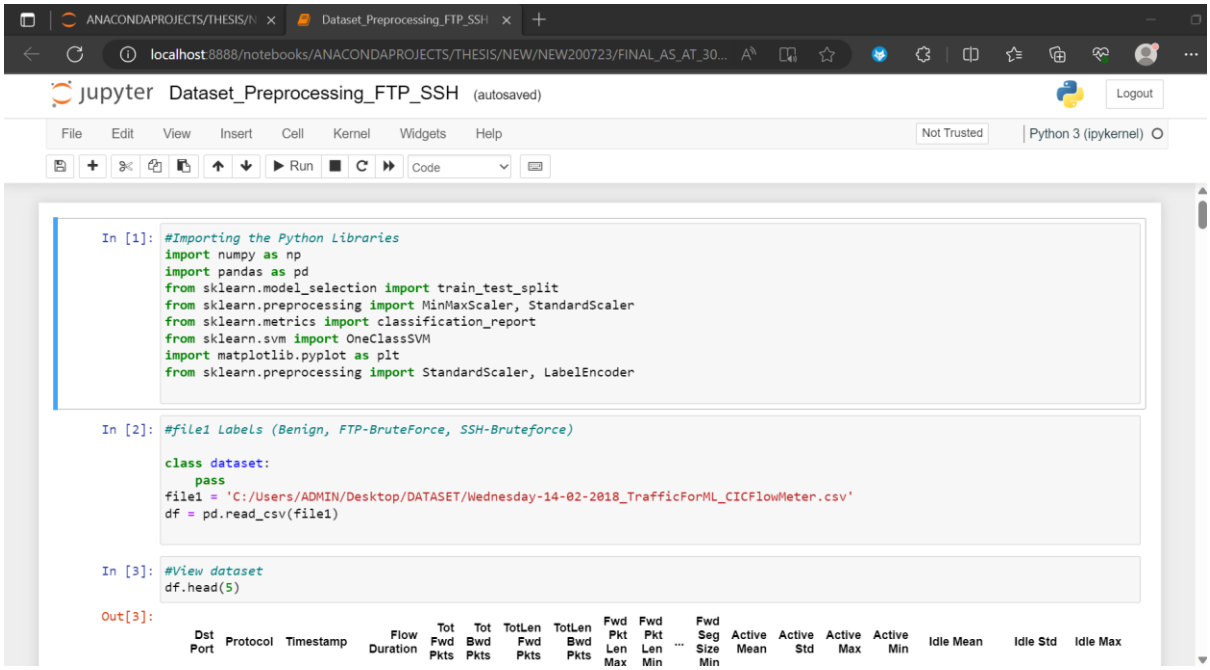

Fig.8. Executing Dataset_Preprocessing_FTP_SSH.ipynb code.

The last part of the code saved the pre-processed dataset as "preprocessed_dataset_FTP_SSH.csv"

Below is the saved file in the directory after successful code execution.

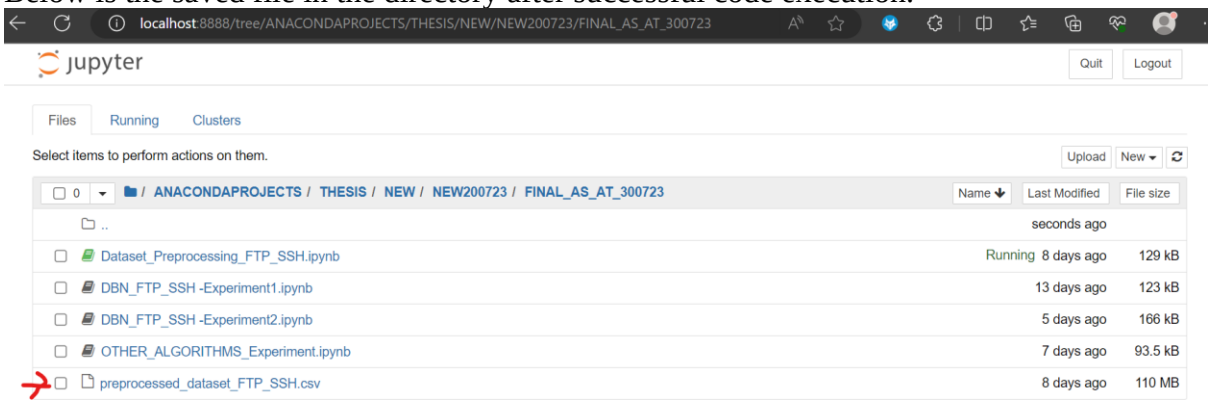

Fig.9. Saved pre-processed dataset.

4.2 DBN_FTP_SSH -Experiment1.ipynb

After the dataset has been pre-processed, we then carry on with the DBN model training. This is the first experiment. The pre-processed dataset saved is used for this experiment.

To execute the code, start the anaconda navigator from the windows start page, launch the Jupyter notebook. Click on the code and execute using the Kernel button to run all.

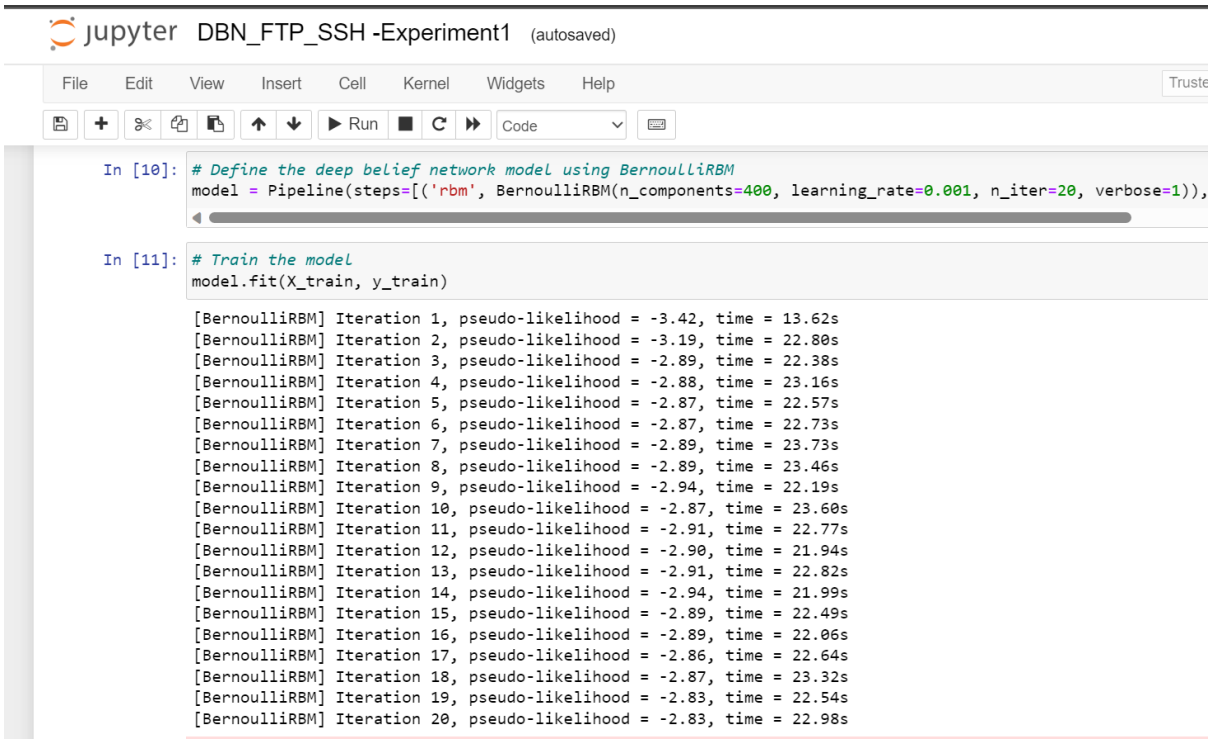

Fig.10. DBN Model training experiment 1.

4.3 DBN_FTP_SSH -Experiment2.ipynb

This is the second and final experiment done to train the DBN model. The pre-processed dataset saved was also used for this experiment with different parameters.

To execute the code, start the anaconda navigator from the windows start page, launch the Jupyter notebook. Click on the code and execute using the Kernel button to run all.

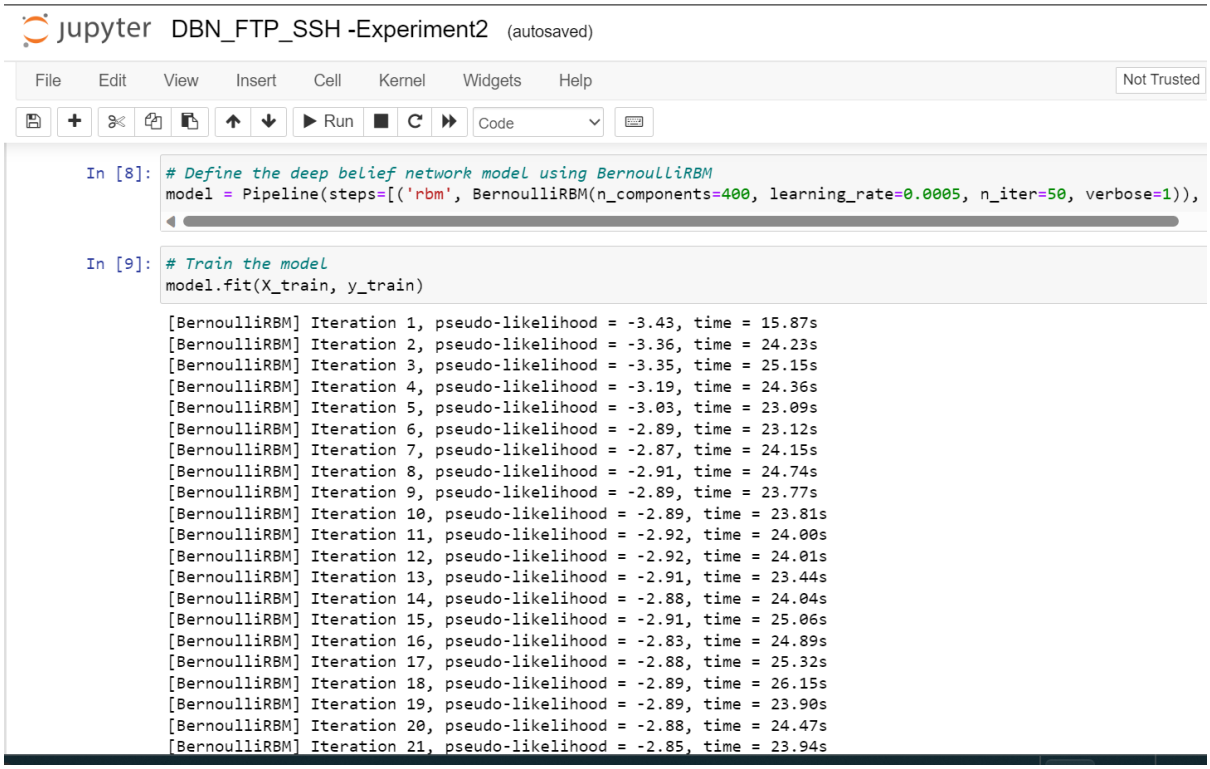

Fig.11. DBN Model training experiment 2.

4.4 OTHER_ALGORITHMS_Experiment.ipynb

In this experiment, the algorithms; Random Forest, Logistic Regression and Decision Tree were trained. The pre-processed dataset saved was also used for this experiment. The result for each of the algorithms are shown.

To execute the code, start the anaconda navigator from the windows start page, launch the Jupyter notebook. Click on the code and execute using the Kernel button to run all.

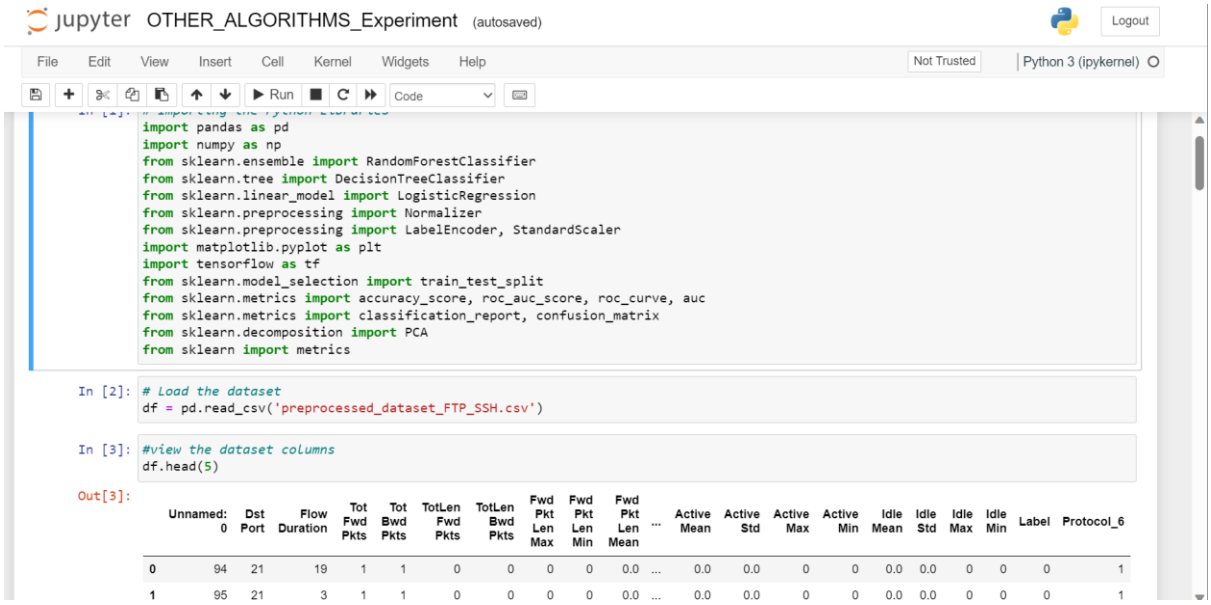

Fig.12. Training Random Forest, Decision Tree, and Logistic Regression Algorithms

References

Anaconda, 2023. *Anaconda.* [Online]
Available at: https://www.anaconda.com/
[Accessed 04 08 2023].
Anaconda, 2023. *Overview.* [Online]
Available at: https://docs.anaconda.com/free/navigator/overview/
[Accessed 04 08 2023].
Jupyter, 2023. *Jupyter.* [Online]
Available at: https://jupyter.org/
[Accessed 04 08 2023].
Python, 2023. *Python Releases for Windows.* [Online]
Available at: https://www.python.org/downloads/windows/
[Accessed 04 08 2023].
UNB, 2023. *CICFlowMeter (formerly ISCXFlowMeter).* [Online]
Available at: https://www.unb.ca/cic/research/applications.html
[Accessed 04 08 2023].