Configuration Manual

Academic internship
MSc in Cybersecurity

Varun Gowda Doddakarade Nagendra
Student ID: x22171541

School of Computing
National College of Ireland

Supervisor: Prof. Vikas Sahni

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Varun Gowda Doddakarade Nagendra		
Student ID:	X22171541		
Programme:	MSc in Cyber Security	**Year:**	2023-2024
Module:	Academic internship		
Supervisor:	Prof. Vikas Sahni		
Submission Due Date:	31/01/2024		
Project Title:	A new approach to Cloud security posture management and anomaly detection in cloud traffic using gradient boosting classifier algorithm		
Word Count:	732	**Page Count**	3

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: Varun Gowda Doddakarade Nagendra

Date: 31/01/2024

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):	

3

Configuration Manual

Varun Gowda Doddakarade Nagendra
Student ID: x22171541

The configuration manual outlines a step-by-step guide to prepare and evaluate a machine-learning model using Python, Jupyter Notebook, and essential libraries. This comprehensive guide is designed to assist users in setting up their local environment and seamlessly executing each stage of the model development process.

It offers users a systematic approach, guiding users through the critical steps required for successful model creation. From the installation of the requisite software, such as Python and Jupyter Notebook, to the installation of essential libraries for data manipulation and visualization, the manual ensures a well-prepared development environment.

Hardware configurations used:

Hardware configuration	Specification
Processor	Intel(R) Core (TM) i5-7200U CPU @ 2.50GHz 2.70 GHz
RAM	8 GB
Storage	1TB HDD
Operating system	Windows 10 Home v22H2

1 Section 1: Prerequisites to prepare the model

Step 1: Python Installation: Download the python v3.11.1 installation file from https://www.python.org/downloads/ and install it on the local machine.

Step 2: Jupyter Notebook Installation: Open the command prompt and install Jupyter Notebook using the following command:

- **pip install notebook**

Step 3: Required Libraries Installation: Install the necessary Python libraries by running the following commands:

- **pip install pandas**
- **pip install scikit-learn**
- **pip install matplotlib**
- **pip install seaborn**
- **pip install numpy**
- **pip install TensorFlow**

Tabel 1: Software version summary

Software Name	Version	Download URL
Python	3.11.0	https://www.python.org/ftp/python/3.11.0/python-3.11.0-amd64.exe
Jupyter	1.0.0	https://jupyter.org/install

Libraries version summary table:

Library Name	Version	Download URL
numpy	1.26.2	https://files.pythonhosted.org/packages/dd/2b/205ddff2314d4eea852e31d53b8e55eb3f32b292efc3dd86bd827ab9019d/numpy-1.26.2.tar.gz
pandas	2.1.3	https://files.pythonhosted.org/packages/86/ff/662dde2193fc93b8547b073db20472b9676f944d907247a46c9c5bc45bfc/pandas-2.1.3.tar.gz
matplotlib	3.8.2	https://files.pythonhosted.org/packages/fb/ab/38a0e94cb01dacb50f06957c2bed1c83b8f9dac6618988a37b2487862944/matplotlib-3.8.2.tar.gz
seaborn	0.13.0	https://files.pythonhosted.org/packages/06/6f/caf0741c5787358b0efba3b4db7f8235e3a48e719ad2444bbd51485f966c/seaborn-0.13.0.tar.gz
scikit-learn	1.3.2	https://files.pythonhosted.org/packages/88/00/835e3d280fdd7784e76bdef91dd9487582d7951a7254f59fc8004fc8b213/scikit-learn-1.3.2.tar.gz
TensorFlow	2.15.0	https://files.pythonhosted.org/packages/93/21/9b035a4f823d6aee2917c75415be9a95861ff3d73a0a65e48edbf210cec1/tensorflow-2.15.0-cp311-cp311-win_amd64.whl

2 Section 2: Steps to Prepare the Model

Step 1: Download and save the dataset - CSE-CIC-IDS2018[1] in the same directory where Jupyter Notebook is installed.

Step 2: Jupyter Notebook Setup: Open the command prompt navigate to the directory containing the dataset and run the below command,

- **jupyter notebook**

Step 3: Run the Data Preprocessing Code: Copy and paste the code or write the code for loading, preprocessing, and exploring the dataset.

Step 4: Run the Data Visualization Code: Execute the code for visualizing label distribution and attack types using Matplotlib and Seaborn.

[1] https://aws.amazon.com/marketplace/pp/prodview-qkyroawpr2aw6#resources

Step 5: Run the Label Transformation Code: Implement the code for transforming labels into numerical values.

Step 6: Run the Gradient Boosting Classifier Model Code: Copy and paste the code for loading the dataset, preprocessing, selecting features, converting labels, and splitting the dataset. Ensure the provided features align with the dataset's column names. Run the Gradient Boosting Classifier model code.

3 Section 3: Evaluate the model

Step 1: Execute the code for evaluating the model accuracy and visualize the confusion matrix.

Step 2: Review Results: Analyse the accuracy precision, F1-score, confusion matrix generated by the model to assess its performance.