

Configuration Manual

MSc Research Project
Msc in Data Analytics

Venkateshwar Rao Telu
Student ID: x23318392

School of Computing
National College of Ireland

Supervisor: Jaswinder Singh

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: Venkateshwar Rao Telu
Student ID: x23318392
Programme: MSC in Data Analytics **Year:** 2024 - 2025
Module: Research Practicum part - 2
Lecturer: Jaswinder Singh
Submission Due Date: September 15, 2025
Project Title: Integrating SHAP and Attention Mechanisms for Transparent Deep Autoencoder Anomaly Detection
Word Count: 481 **Page Count:** 4

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: Venkateshwar Rao Telu

Date: September 15, 2025

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

Venkateshwar Rao Telu
Student ID: x23318392

1. Introduction

The configuration manual has provided the complete details of the software and hardware specification for executing the project along with the running process of the project. The step by step process of the project development has been presented, which has helped for setting up the working environment, including the installation of the required application runtime configuration and the proper execution of the developed model.

2. System Specification

The project has been developed by including the local environment and the runtime of the Google collab for taking the advantage of the hardware accelerators

Local Environment Specification

- Operating system: Windows 10/ Windows 11 (64 bit)
- Storage: 256 GB SSD (Minimum)
- RAM: 8 GB (Minimum)
- Processor: Intel core i5 or above

Google collab Runtime

- Runtime: Python 3
- Hardware specifications: GPU (NVIDIA)
- Minimum of 78 gb disk space available during Runtime

3. Dataset Specification

- The dataset contains the labeled network traffic data for different attack scenario
- There are two different classes, such as Normal and Anomalous
- The CSV file contains the numerical and the categorical features both
- Link <https://www.kaggle.com/deepthisudharsan/network-anomaly-dataset?resource=download>

4. Project development

Google collab platform has been used in this analysis project for the efficient development of the project. After completing the data cleaning and data preprocessing the machine learning model has been developed for validating the data and the proper prediction.

4.1 Importing Libraries

The basic libraries and the machine learning model development libraries have been imported for the efficiency model development and evaluation. Pandas and Numpy have been used for efficient data preparation. Therefore, the Keras and the Scikit learn framework has been used for the efficiency in the machine learning model development and evaluation.

4.2 Importing Data

After importing the data into the program, the dataset has been imported into the program by mounting the google drive into google colab. First the dataset has been uploaded into the drive and the drive has been imported into google colab for accessing the dataset.

4.3 Preprocessing and Validation of the data

- Highly correlated features removed
- Categorical Features has been encoded
- Minmax scaling for the numerical features
- Data splitting into train and test

4.4 Modelling

CBAM Autoencoder with MultiHeadAttention

- Encoder are the MultiHeadAttention, Conv1D layers, CBAM modules
- Loss function is the Mean squared error
- Reverse Conv1D structure has helped in reconstruction

Isolation Forest

- Contamination rate for this model is 0.1
- Random state for this model is 42

Explainability – SHAP

- SHAP has been applied in the trained autoencoder
- Feature importance has been generated by both local and global interpretation

5. Running the project

- Open Google collab
- Load data on the google drive
- Mount google drive into colab
- Load the dataset by defining the path of the dataset from google drive
- Clean data by removing null values and duplicates
- Preprocess the data by Label encoding and Minmax scaler
- Train model
- Model performance evaluation
- SHAP analysis for increasing explainability