National College of Ireland

Configuration Manual

MSc Research Project
MSc Cloud Computing

Aishwarya Pokharkar
Student ID: 23141808

School of Computing
National College of Ireland

Supervisor: Prof. Vikas Sahni

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Aishwarya Pokharkar		
Student ID:	23141808		
Programme:	Cloud Computing	**Year:**	2024
Module:	MSc Research Project		
Lecturer:	Prof. Vikas Sahni		
Submission Due Date:	12/8/2024		
Project Title:	Configuration Manual		
Word Count:	**536**	**Page Count: 5**	

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: Aishwarya Pokharkar

Date: 12/8/2024

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):	

Configuration Manual

Aishwarya Pokharkar
Student ID: 23141808

1 System Requirements

The configuration manual provides detailed descriptions of the basic software and hardware prerequisites necessary for the project's successful operation. The hardware requirements and software versions required to execute the code were thoroughly outlined.

1.1 Software

- TensorFlow: TensorFlow 2.4.1
- NumPy: NumPy 1.19.5
- Pandas: Pandas 1.1.5
- Matplotlib: Matplotlib 3.3.3
- Seaborn: Seaborn 0.11.1

1.2 Hardware

- OS Name: Microsoft Windows 11 Home Single Language
- Processor: 12th Gen Intel(R) Core (TM) i5-1235U, 1300 Mhz, 10 Core(s), 12 Logical Processor(s)
- RAM: 16G
- Storage (SSD): 512GB

2 ML Packages used

This section discussed the ML packages used in the project:

1. Keras : Keras, a high-level neural networks API running on TensorFlow. It is used in the code to define and train the neural network model. It involves building a Sequential model, adding layers (LSTM, Dense, Dropout), and defining a custom attention layer. Keras also includes utilities for label conversion and model optimization with Adam.

2. Imbalanced-learn (imblearn): Imbalanced-learn is a library that provides tools for handling imbalanced datasets in machine learning. The code utilized SMOTE (Synthetic Minority Over-sampling Technique) from this library to oversample the minority class, creating synthetic samples to balance class distribution and improve classifier performance.

3. TensorFlow (tf): TensorFlow, an open-source machine learning library, used as the backend for Keras in this code. It provides optimizers like Adam for model compilation and includes specific layers such as tf.keras.layers.Bidirectional for adding bidirectional LSTM layers.

4. Scikit-learn (sklearn): Scikit-learn is a machine learning library which provides efficient tools for data mining and analysis. In the code, modules like LabelEncoder, MinMaxScaler, and train_test_split are used for data preprocessing and splitting, while metrics such as classification_report and confusion_matrix are used for model evaluation.

3 Dataset

This dataset was used in the Third International Knowledge Discovery and Data Mining Tools Competition NSL KDD-99[1]. This dataset has been reorganized, enhanced, and updated primarily for malware analysis. It incorporates the latest information on various malware attacks, including DoS, R2L, U2R, and probing which makes it an excellent resource for developing an intrusion detection system.

4 Software Tools

- To start with the Project a working Azure account is needed. You can run the python Notebooks on AWS sage maker or google collab as well.
- Now, Create a separate Workspace for the Project.

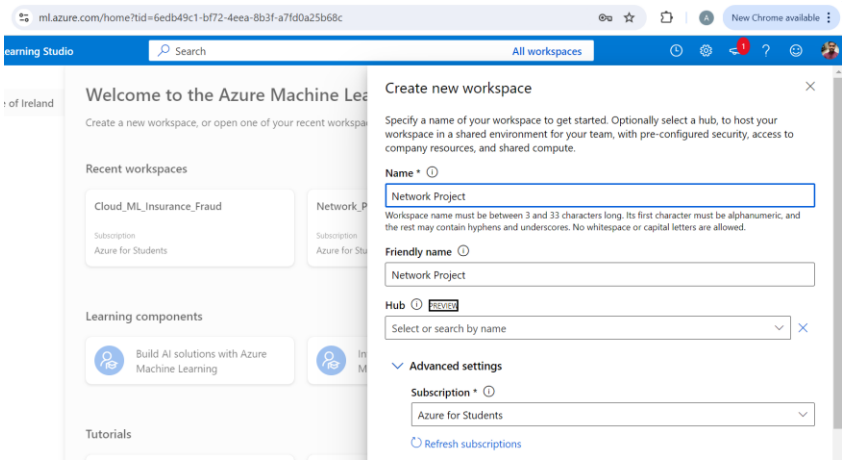

Fig.1: Create a workspace for project

- After that create a folder in Notebook and upload all the CSV files and notebook files which are necessary.

[1] https://www.kaggle.com/datasets/galaxyh/kdd-cup-1999-data

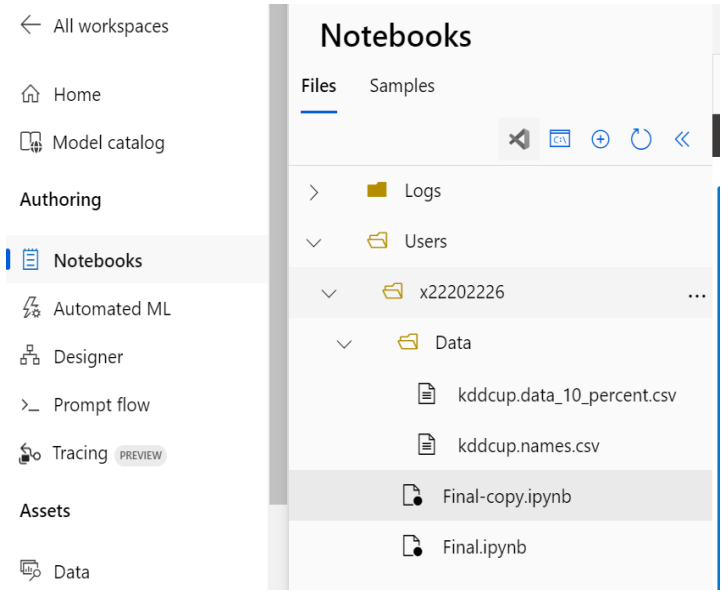

Fig. 2: Uploading Files to Notebook

- Create Compute Instance for code to execute in Notebook

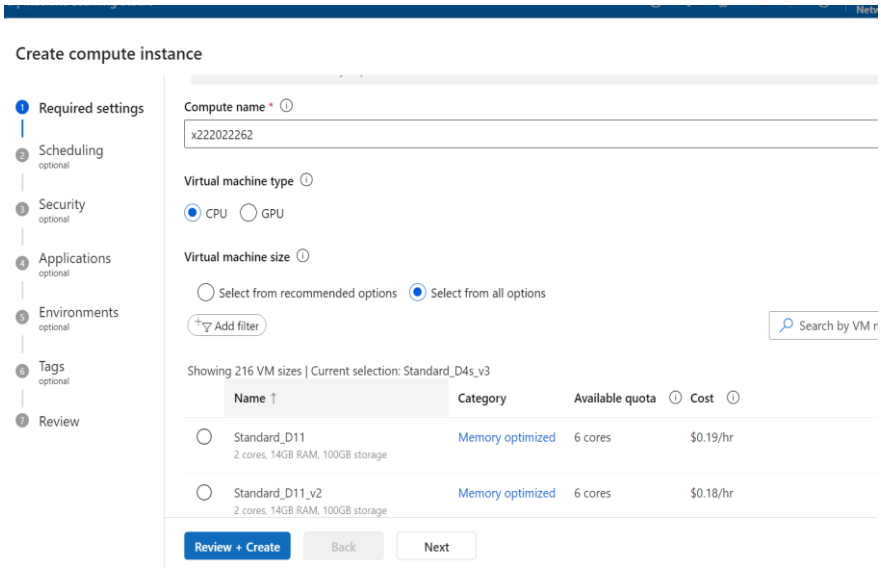

Fig. 3: Create Compute Instance

- Open the Final-copy.ipynb file and click on cell operation and this will run the whole code which is there in the file.
- Once execution of all cells gets over, the results will be executed in the form of graphs, plots and tables. This Files executes LSTM and Bi-LSTM Models and Bi-LSTM along with Attention Mechnisam.

5 Results

Classification Report :

	precision	recall	f1-score	support
0	0.92	0.98	0.95	78292
1	0.93	0.88	0.90	78291
2	0.99	0.91	0.95	78292
3	0.74	0.78	0.76	78291
4	0.77	0.78	0.77	78292
accuracy			0.86	391458
macro avg	0.87	0.86	0.87	391458
weighted avg	0.87	0.86	0.87	391458

Fig. 4: Classification Report for LSTM Model

Classification Report :

	precision	recall	f1-score	support
0	0.99	0.99	0.99	78292
1	0.91	0.89	0.90	78291
2	0.98	0.94	0.96	78292
3	0.80	0.91	0.85	78291
4	0.85	0.78	0.81	78292
accuracy			0.90	391458
macro avg	0.91	0.90	0.90	391458
weighted avg	0.91	0.90	0.90	391458

Fig. 5: Classification Report for Bi-LSTM Model

Classification Report :

	precision	recall	f1-score	support
0	1.00	1.00	1.00	78292
1	1.00	0.99	1.00	78291
2	1.00	1.00	1.00	78292
3	1.00	0.98	0.99	78291
4	0.98	1.00	0.99	78292
accuracy			1.00	391458
macro avg	1.00	1.00	1.00	391458
weighted avg	1.00	1.00	1.00	391458

Fig.6: Classification Report for Attention Mechanism Model

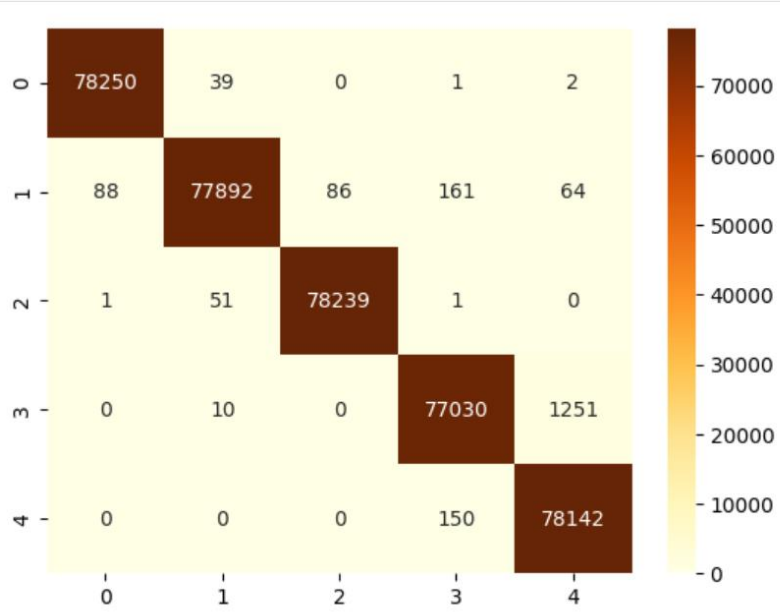

Fig. 7: Confusion Matrix for Bi-LSTM with attention mechanism

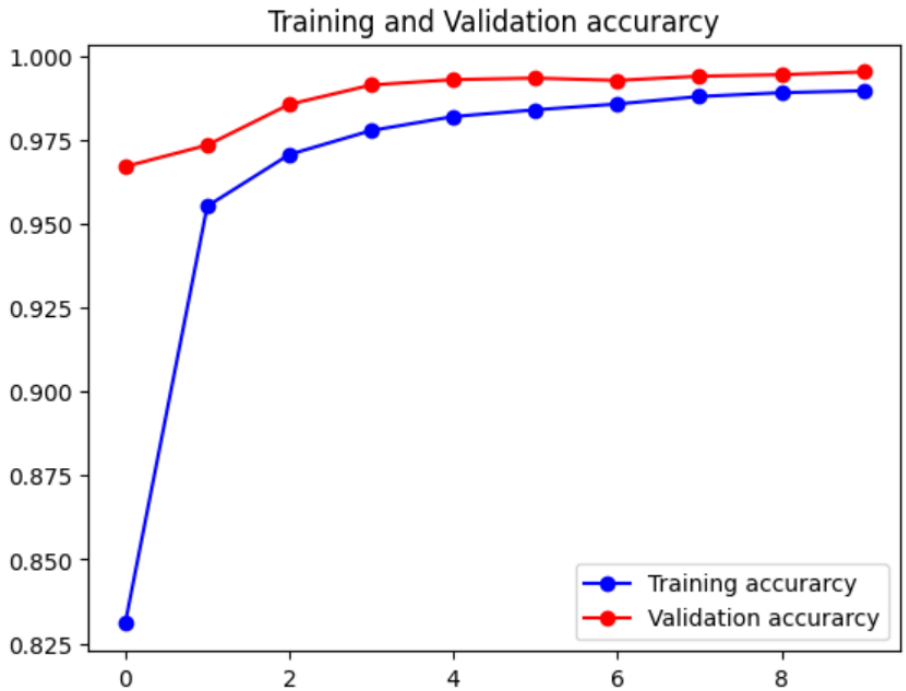

Fig. 8: Training and validation accuracy for attention mechnisam