National College of Ireland

# Configuration Manual

MSc Research Project
MSc in Cloud Computing


## Sandesh Muralidhar
Student ID: X20195737

School of Computing
National College of Ireland


Supervisor: Shreyas Setlur Arun

**National College of Ireland**
**Project Submission Sheet**
**School of Computing**

| | |
|---|---|
| **Student Name:** | Sandesh Muralidhar |
| **Student ID:** | x20195737 |
| **Programme:** | MSc in Cloud Computing |
| **Year:** | 2023 |
| **Module:** | MSc Research Project |
| **Supervisor:** | Shreyas Setlur Arun |
| **Submission Due Date:** | 14/12/2023 |
| **Project Title:** | SSD and HDD Failure Detection using Advance Deep Learning Algorithms |
| **Word Count:** | XXX |
| **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:** | Sandesh Muralidhar |
| **Date:** | 13th December 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

Sandesh Muralidhar
X20195737

# 1 Notebook Instance

## 1.1 Requirements:

**Software:**

- Python: Any Python Version above 3.x (Installing python will also install other Library like Pandas, Numpy,Matplotlib, Seaborn, Calender, Counter)
- Jupyter Notebook
- XGBoost(pip install xgboost)
- Tensorflow(pip install tensorflow) - To Execute Deep Learning model

**Hardware:** Any Instance with 8 GB RAM or more to run the models smoothly

## 1.2 Configuration:

**Local Execution:**

- Install python of version above 3.x on Local Machine
- Install Jupyter Notebook by using pip install jupyter.
- Open Notebook by launching - http://localhost:8888.
- Upload the Dataset from Blackblaze (n.d.)
- Install all the Mentioned Software Requirements

**Cloud Execution:**

- Create S3 Bucket
- Upload the Dataset from Blackblaze (n.d.)
- Create Policy and make the Data in the S3 bucket public AWS (n.d.)
- Create Sagemaker Instance
- Use S3 Boto CLient to access the dataset files in S3 bucket from Notebook

Running Notebook involves training and evaluation of 3 deep learning models which are convolutional neural networks (CNN), gated recurrent units (GRU), and a combination of CNN and GRU. The data is manipulated using Pandas and NumPy, visualized with Plotly, and oversampled using SMOTE for handling imbalanced datasets. XGBoost for gradient boosting and TensorFlow with Keras for building neural network models. The validation of models is through using different metrics such as accuracy, precision, recall, and F1 score.

# 2 Application Configuration

## 2.1 Requirement:

**Software Requirement**

- Python 3.11
- Flask
- TensorFlow

```
python -m pip install flask tensorflow
```

## 2.2 Configuration:

**Local execution:**

- Install all the software requirements
- Make sure to keep the index.html file in the Template folder
- Run the Application
  Navigate to the application directory and run the application using: `python app.py`

**Cloud execution:**

- Create an EC2 Instance using an Amazon Linux machine
- Download the .pem file while creating Keypair and save it in the local Instance
- Create an Inbound rule allowing Port 5000 in the Security Group (This should allow traffic on Port 5000)
- Establish an SSH session with the EC2 instance
- Run the below command to install dependencies on EC2 `sudo yum install python3`
  `sudo yum install python3-pip`
  `pip3 install flask numpy tensorflow`
- Transfer the files into the EC2 Instance from the Local PC `scp -i (path-of-downloaded-key.p`
  `/path-to-index.html /path-to-CONV_GRU.h5 /path-to-app.py ec2-user@ec2-instance-ip`
- Make a Directory called Template in the l̃ocation and move the index.html file to the Template folder

- From the root directory run the command `python3 app.py`
- Access the Web application `http://ec2-instance-ip:5000`

This web application utilizes a Convolutional Gated Recurrent Unit (CONV-GRU) model to predict disk failures. It's designed to showcase the practicality of the research allowing users to input specific parameters and receive predictions of disk failure.

# References

AWS (n.d.). read-access-objects-s3-bucket.
**URL:** *https://www.backblaze.com/cloud-storage/resources/hard-drive-test-data*

Blackblaze (n.d.). Blackblaze dataset.
**URL:** *https://www.backblaze.com/cloud-storage/resources/hard-drive-test-data*