National College of Ireland

Configuration Manual

MSc Research Project
Msc in Artificial Intelligence

MURUGAPPAN KRISHNAN
Student ID: 23187905

School of Computing
National College of Ireland

Supervisor: ABDUL SHAHID

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	MURUGAPPAN KRISHNAN
Student ID:	23187905
Programme:	Msc in Artificial Intelligence
Year:	2024
Module:	MSc Research Project
Supervisor:	ABDUL SHAHID
Submission Due Date:	12/12/2024
Project Title:	Configuration Manual
Word Count:	589
Page Count:	7

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:	MURUGAPPAN KRISHNAN
Date:	12.12.2024

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

MURUGAPPAN KRISHNAN
23187905

Introduction

This configuration manual is going to walk the user through setting up and using the Federated Learning framework to predict which telecom customers will churn. The set-up, creation of SSH to AWS, retrieval of session keys, and running scripts to train and test FL models constitute some of the work to be done. Ensuring that the data private, can handle large data amounts, and works fine with Non-IID is thus what shall be made certain.

System Requirements

Hardware Requirements

- **Laptop/PC:** Any system with Windows 11 or Ubuntu OS.
- **RAM:** Minimum 16 GB (recommended for smooth training).
- **Processor:** Quad-core or higher.
- **Storage:** Minimum 256 GB (to store datasets and configurations).
- **Internet:** Stable high-speed connection (required for AWS).

Cloud Resources

- **AWS EC2:** Hosting Federated Learning processes across regional and central servers.
- **AWS S3:** Storing updates and aggregated models.
- **AWS IAM:** Managing access roles and permissions.
- **AWS Key Pair:** Securely access EC2 instances via SSH.

Software and Tools

Local Setup

- **IDE:** Visual Studio Code (VS Code) with AWS Toolkit extension.

- **Python Version:** Python 3.9+.
- **Python Libraries:**
  - `boto3`, `paramiko`, `numpy`, `pandas`
  - `torch`, `scikit-learn`, `matplotlib`

Configuration Setup

AWS EC2 Instances

- **Regions:** Mumbai (Central Server), Mumbai, Ireland, and Ohio (Regional Nodes).
- **Instance Configuration:**
  - Instance Type: `t2.large`
  - Storage: 50 GB (EBS root volume attached as `/dev/xvda`).
  - Operating System: Amazon Linux 2.
  - Public IP Enabled: Yes.
  - IAM Role: `s3accessRole` (full access to S3 buckets).

S3 Buckets

- **Central Bucket:** `fl-central-bucket` (aggregated model storage).
- **Regional Buckets:**
  - `flmumbai-bucket`
  - `flireland-bucket`
  - `flohio-bucket`

AWS Resource	Purpose
`fl-central-bucket`	Central bucket for storing updates and aggregated global models.
`flmumbai-bucket`	Bucket for Mumbai, initially storing raw data and later receiving aggregated updates.
`flireland-bucket`	Bucket for Ireland, initially storing raw data and later receiving aggregated updates.
`flohio-bucket`	Bucket for Ohio, initially storing raw data and later receiving aggregated updates.
EC2 Instances	Execute local training and retraining processes for each region. Connected to respective S3 buckets.

Table 1: AWS Resources and Their Purpose

1 AWS S3 Bucket Configuration

1.1 Dataset Availability

The datasets are distributed across different AWS S3 buckets corresponding to regional nodes. Below is the mapping of datasets to their respective locations:

- **Location1.csv** - Mumbai (flmumbai-bucket)
- **Location2.csv** - Ireland (flireland-bucket)
- **Location3.csv** - Ohio (flohio-bucket)

For testing purposes, a consolidated dataset is stored in the central bucket:

- **Location2.csv** - Central Bucket (fl-central-bucket)

1.2 Mumbai Bucket

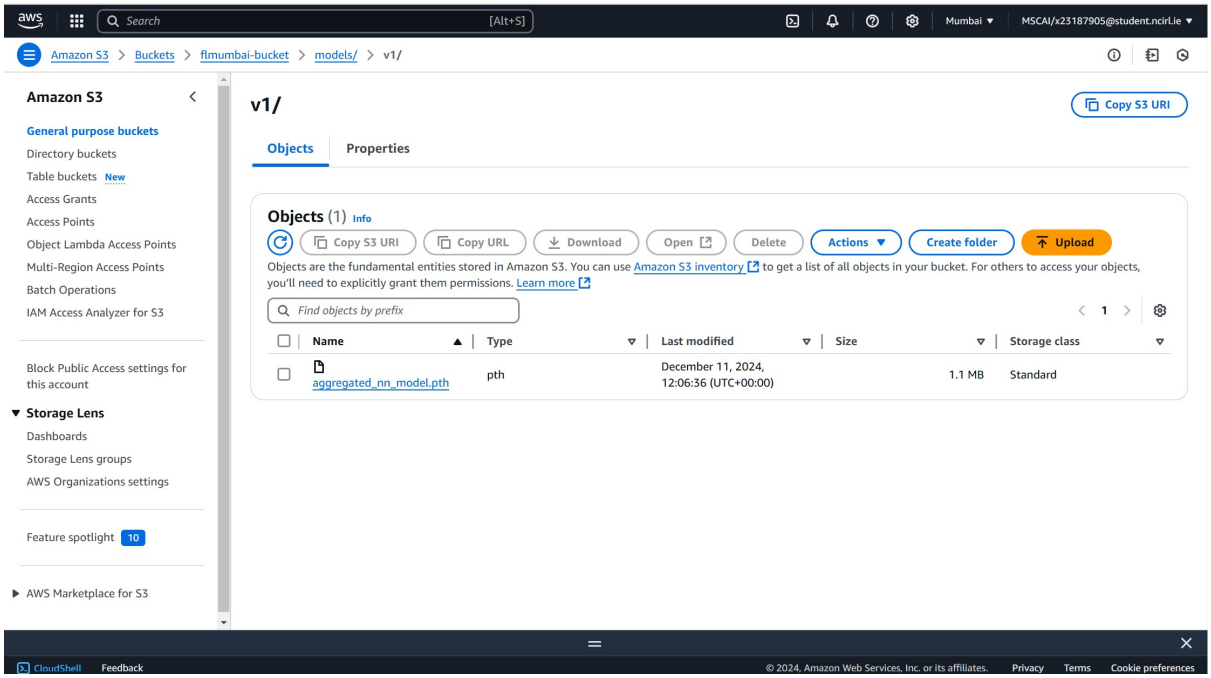

Figure 1: Mumbai Bucket in AWS S3.

1.3 Central Bucket

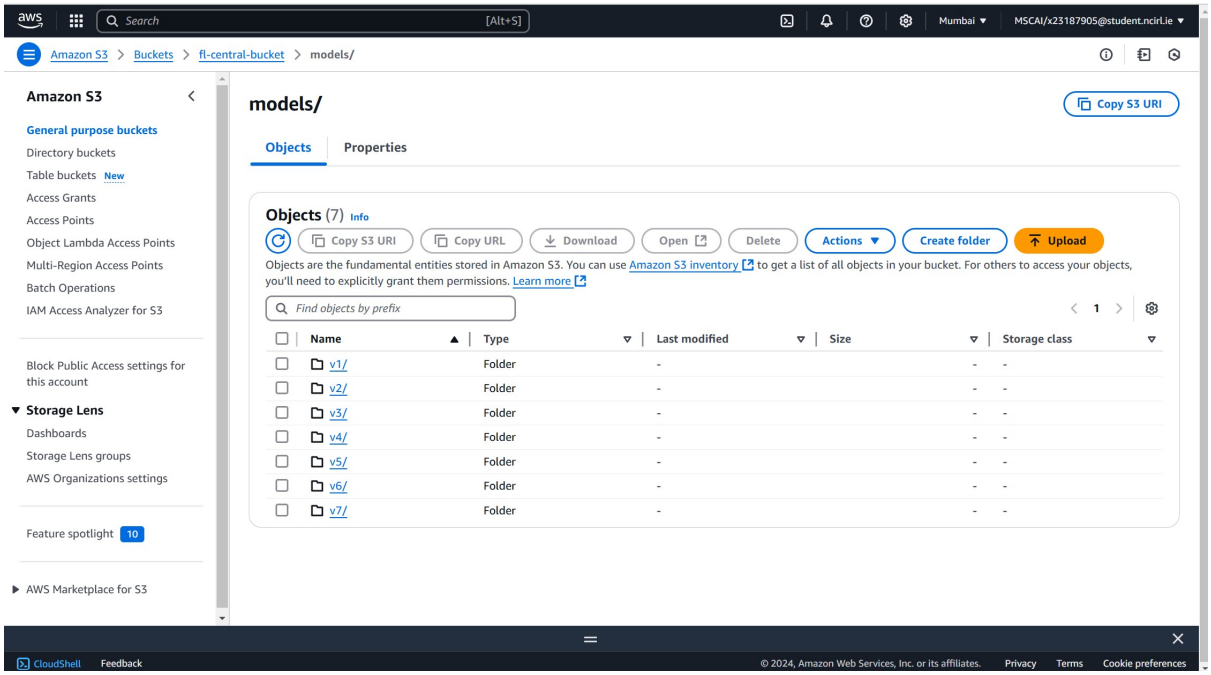

Figure 2: Central Bucket in AWS S3.

1.4 Central Bucket Updates and Aggregates

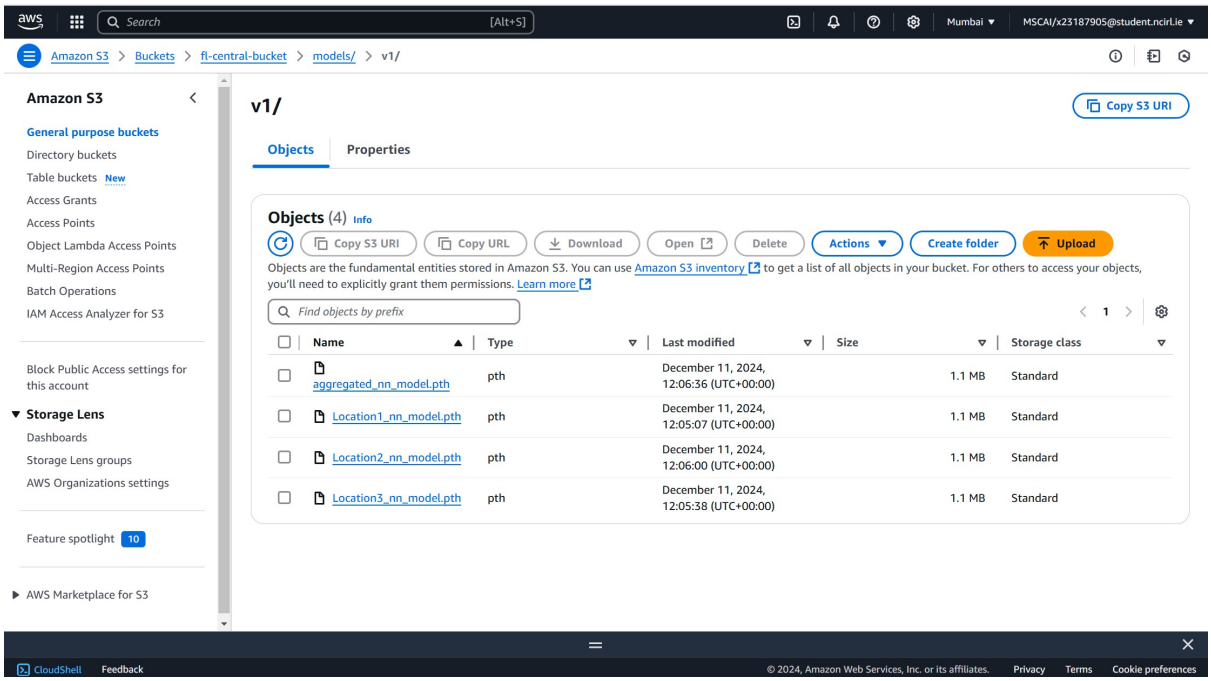

Figure 3: Central Bucket-updates and aggregated models.

2 Example Script Execution

Below is the step-by-step process for running the Federated Learning workflow.

2.1 Step-by-Step Workflow for the Mumbai Node

1. **Connect to the Mumbai Server:**

   ```
   ssh -i "C:\Users\murug\Downloads\mumbaiserver.pem" ec2-user@43.205.230.72
   ```

2. Export session tokens:

   - `export AWS_ACCESS_KEY_ID="your_key"`
   - `export AWS_SECRET_ACCESS_KEY="your_secret_key"`
   - `export AWS_SESSION_TOKEN="your_token"`

3. **Run the Initial Training Script:**

   ```
   python mumbaitraining.py
   ```

4. **Connect to the Central Server and Perform Aggregation:**

   ```
   ssh -i "C:\Users\murug\Downloads\centralserver.pem" ec2-user@34.123.45.67
   python federatedavg.py
   ```

5. **Retrain the Model on Mumbai Node:**

   ```
   ssh -i "C:\Users\murug\Downloads\mumbaiserver.pem" ec2-user@43.205.230.72
   python mumbai_retrain.py
   ```

6. **Run Continuous Training Rounds on Mumbai Node:**

   ```
   python mumbai_rounds.py --rounds 5
   ```

7. **Run Final Aggregation on Central Server:**

   ```
   ssh -i "C:\Users\murug\Downloads\centralserver.pem" ec2-user@34.123.45.67
   python aggregate_iterations.py
   ```

8. **Evaluate the Model:**

   ```
   python calculate.py
   ```

2.2 Repeat for Other Locations

Repeat the above steps for Ireland and Ohio by connecting to their respective servers and running the appropriate scripts.

2.3 Notes

- Ensure AWS session tokens are refreshed periodically (valid for 12 hours).
- Always connect to the appropriate node/server before running scripts.
- Follow the sequence strictly to ensure consistent model updates and aggregation.

Federated Learning Process

Script Workflow

- **Training Phase:**
  - `mumbaitraining.py`
  - `irelandtraining.py`
  - `ohiotraining.py`
  - `federatedavg.py`
- **Retraining Phase:**
  - `mumbairetrain.py`
  - `irelandretrain.py`
  - `ohioretrain.py`
  - `updatedavg.py`
- **Iterative Rounds:**
  - `mumbairounds.py`
  - `irelandrounds.py`
  - `ohiorounds.py`
  - `aggregatesiteration.py`
- **Metrics Evaluation:**
  - `calculate.py`

Conclusion

The purpose of this manual is to provide a walkthrough of how the Federated Learning framework should be set up and used for telecom churn prediction. Users can thereby get flexible solutions that assure privacy, being in accordance with GDPR and other similar standards.

References

[1] D. Azzam, M. Hamed, N. Kasiem, Y. Eid, and W. Medhat. Customer churn prediction using apriori algorithm and ensemble learning. In *2023 5th Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, pages 377–381, 2023.

[2] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konecný, S. Mazzocchi, H. B. McMahan, T. V. Overveldt, D. Petrou, D. Ramage, and J. Roselander. Towards federated learning at scale: System design, 2019.

[3] A. Gaur and R. Dubey. Predicting customer churn in telecom sector using various machine learning techniques. In *2018 International Conference on Advanced Computation and Telecommunication (ICACAT)*, pages 1–5, 2018.

[4] X. Hu, Y. Yang, L. Chen, and S. Zhu. Research on a customer churn combination prediction model based on decision tree and neural network. In *2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, pages 129–132, 2020.

[5] J. Huh and W. Lee. Privacy-preserving consumer churn prediction in telecommunication through federated machine learning. In *2024 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 355–356, 2024.

[6] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, et al. Advances and open problems in federated learning, 2021.