

MOUSAMI SONI

[Berlin, Germany](#) | [+4915158852196](#) | soni.mousami7@gmail.com
<https://www.linkedin.com/in/mousami7/> | <https://github.com/Mousami7>

PROFESSIONAL SUMMARY

Passionate **Airflow/MLAI Developer** with 5+ years of experience in Python, SQL, and data workflows, including 3 years of hands-on Apache Airflow. Skilled in **Big Data integration, ETL pipeline design, and workflow orchestration**. Adept at working with **single-node Airflow deployments** for lightweight, production-ready pipelines and integrating cloud services such as AWS S3, DynamoDB, and PostgreSQL. Strong background in **Data Science and ML modules**, supporting model training and automation alongside data engineering tasks. Experienced trainer who has taught Python, SQL, and Airflow fundamentals, and delivered projects ranging from **event data automation (Ceamless)** to **audio data pipelines with ML workflows (Xounds)**.

SKILLS

Programming Languages: Python, SQL.

Workflow Orchestration: Apache Airflow (DAG Authoring, scheduling. Operators, sensors, XComs).

Databases: PostgreSQL, DynamoDB, MySQL, SQLite.

Machine Learning & AI Libraries: Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn, SpeechBrain, Whisper, TensorFlow, PyTorch.

Backend Frameworks: FastAPI, Flask, Streamlit.

Cloud & DevOps: Amazon Web Services (AWS – Lambda, S3, EC2), Git, Docker, RESTful APIs, CI/CD.

Specialized Expertise: Audio AI (Speaker Diarization, Abuse Detection), Natural Language Processing (NLP), Predictive Analytics, ETL Pipelines.

WORK EXPERIENCE

Xounds Berlin, Germany

March 2025 – Present

Airflow & MLAI Developer

Dec 2024 – Feb 2025

- Used Apache Airflow running on single-node setup to schedule and manage small audio processing pipelines.
- Developed and scheduled workflows for handling audio data.
- Created pipelines to collect data from storage and prepare it for further use.
- Wrote Python-based tasks to organize and manage incoming data.
- Managed basic metadata storage and tracking for files.
- Contributed to backend services and cloud integration for the product.
- Created simple DAGs that picked up audio files from an S3 bucket, cleaned them, and stored the result back.
- Kept metadata (e.g. file name, status) in AWS DynamoDB using small Python scripts wrapped as Airflow tasks.
- Configured basic monitoring: retries for failed, so we knew if something broke.
- Leading ML feature development for an **AI-powered mobile audio app**.
- Developing ML pipelines for **abuse detection, speaker diarization, and speaker verification**.
- Built scalable backend APIs using **FastAPI** and **AWS (Lambda, S3, EC2)**.

Ceamless Events, Berlin, Germany

Airflow/Data Developer

June 2024 – November 2024

- Set up a lightweight Airflow instance (single node, LocalExecutor) to automate event data workflows.
- Built DAGs to ingest attendee data (CSV/Excel files) from an S3 bucket and load it into PostgreSQL.
- Used FileSensor/S3KeySensor to trigger pipelines when new files arrived.
- Added Python tasks for data cleaning (removing duplicates, formatting phone numbers, normalizing names).

- Scheduled event-based jobs to generate summary reports (attendees stats, ticket sales) and email them to the team.
- Worked closely with the team to understand business data needs.
- Strengthened **teamwork and collaboration skills** in a startup environment.

Yess Infotech (Training Institute) – Pune, India
Data Science & Python Trainer |

Aug 2022 – Aug 2023

- Performed in-house training sessions on **Python, SQL, Data Workflow fundamentals and Data Science** topics.
- Designed and delivered introductory Airflow Sessions, covering
 - DAG creation.
 - Using PythonOperator and BashOperator for small tasks.
 - Introduction to Sensors (file arrival triggers) and XComs.
- Trained learners on best practices for **performance optimization** in data tasks.
- Introduced students to **Apache Airflow** with hands-on examples.
- Guided participants through **POC projects in Airflow** to demonstrate automation.
- Designed project frameworks for ETL, reporting, and basic data workflows.
- Supported learners in building end-to-end mini-projects.
- Created training material, exercises, and evaluations for classroom delivery.

Vocational Education Project – MP Govt. – India
IT Trainer |

Sep 2016 – Sep 2019

- Conducted training on Python, **SQL, database management, and IT skills** for 200+ students.
- Enhanced analytical and technical skills of learners.
- Supported the **District Magistrate office's IT department** in **election database management**.
- Assisted in maintaining departmental datasets and ensuring data accuracy.

KEY PROJECTS

-
- **Audio Abuse Detection & Speaker Diarization (Xounds) –**
 - ❖ Developed workflows in **Apache Airflow** to collect audio files from AWS S3 and process them for further use.
 - ❖ Used **PythonOperator** for audio processing tasks and **custom operators** for integration with **DynamoDB** metadata storage.
 - ❖ Deployed everything on a **single-node Airflow instance**, for small workloads.
 - ❖ Organized and tracked basic metadata using AWS DynamoDB.
 - ❖ Set up simple monitoring with retries and alerts to ensure reliability.
 - ❖ Built ML pipelines for real-time audio intelligence; deployed with AWS + FastAPI.
 - **AI Event Management System (Ceamless Events) –**
 - ❖ Built Airflow pipelines to process attendee files (CSV/Excel) and load cleaned data into PostgreSQL.
 - ❖ Deployed workflows on a **lightweight single-node Airflow setup**.
 - ❖ Developed Python tasks for data cleaning (removing duplicates, formatting, normalization).
 - ❖ Ran pipelines on a single node Airflow Setup (LocalExecutor) to keep infrastructure lightweight.
 - **Introductory Airflow Projects (Yess Infotech) –**
 - ❖ Loading and cleaning CSV data into SQL databases.
 - ❖ Guided learners through **POC projects** in Airflow to demonstrate workflow automation.
 - ❖ Created sample Airflow DAGs as part of Python/SQL training programs.
 - ❖ Taught students the basics of DAGs, Operators, Sensors, XComs, and Scheduling, providing early exposure to workflow orchestration platform.
 - ❖ Classification models to predict student conversions; Flask/Streamlit deployment.

EDUCATION

- **Masters in Computer Science**
International University of Applied Sciences Germany **2022 - 2024**
- **Master of Technology in Digital communication**
Sagar Institute of Research & Technology Indore **2015-2017**
- **Bachelor of Engineering in Electronics and Communication**
Jawaharlal Institute of Technology Borawan **2010-2013**

INTERNSHIPS & Certifications

- Data Analytics and Machine Learning | Puneri Pattern Pvt. Ltd. Pune India | **Jan 2022 - June 2022**
- Data Science with Python | SimpliLearn | **Apr 2021 - Sep 2021**

LANGUAGES

English (proficient) IELTS C1 certified.

German (basic conversational) A2.