

MOUSAMI SONI

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

PROFESSIONAL SUMMARY

Airflow Developer and MLAI Developer with 5+ years of experience in Python, SQL, and data workflows, including 3+ years of hands-on Apache Airflow. Skilled in designing and managing ETL pipelines, automating ML workflows, and integrating data systems with cloud services (AWS S3, DynamoDB, PostgreSQL). Experienced in lightweight **single-node Airflow deployments**, with strong knowledge of DAG authoring, operators, sensors, retries, and monitoring. Background includes training professionals in Python, SQL, and Airflow fundamentals, as well as delivering real-world projects for event data automation (Ceamless) and audio data pipelines with ML retraining workflows (Xounds). Adept at building reliable, reproducible pipelines that scale from simple teaching projects to production-ready workflows.

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 (EXIST-Funded Startup) – Berlin, Germany |

March 2025 – Present

Airflow Developer & MLAI Developer (Junior Technical Officer) |

Dec 2024 – Feb 2025

- Used Apache Airflow running on single-node setup (no cluster) to schedule and manage small audio processing pipelines.
- Created simple DAGs that picked up audio files from an S3 bucket, cleaned them, and stored the result back.
- Wrote Python tasks to do preprocessing like trimming, converting audio formats, and removing noise.
- Kept metadata (e.g. file name, status) in AWS DynamoDB using small Python scripts wrapped as Airflow tasks.
- Configured basic monitoring: retries for failed tasks and email/Slack alerts, 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 Startup) – Berlin, Germany

Airflow/Data Developer (Volunteer) |

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.
- Strengthened **teamwork and collaboration skills** in a startup environment.

- Delivered training in **Python, data workflow fundamentals Data Science, and ML** to students and professionals.
- Designed and delivered introductory Airflow Sessions, covering
 - DAG creation.
 - Using PythonOperator and BashOperator for small tasks.
 - Introduction to Sensors (file arrival triggers) and XComs.
- Guided real-world projects in **data analysis, ML pipelines, and deployment**.
- Designed interactive assignments, dashboards, and project frameworks.

- Conducted training on Python, **SQL, database management, and IT skills** for 200+ students.
- Enhanced analytical and technical skills of learners.

KEY PROJECTS

- **Audio Abuse Detection & Speaker Diarization (Xounds) –**
 - ❖ Designed **Airflow workflows** to ingest user audio files from **AWS S3**, preprocess them (trimming, noise removal, format conversion), and store cleaned files back into S3.
 - ❖ Used **PythonOperator** for audio processing tasks and **custom operators** for integration with **DynamoDB** metadata storage.
 - ❖ Deployed everything on a **single-node Airflow instance**, optimized for startup-scale workloads.
 - ❖ Built ML pipelines for real-time audio intelligence; deployed with AWS + FastAPI.
- **AI Event Management System (Ceamless Events) –**
 - ❖ Built Airflow DAGs to ingest attendee CSV/Excel files from S3 into PostgreSQL.
 - ❖ 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) –**
 - ❖ Created sample Airflow DAGs as part of Python/SQL training programs.
 - ❖ Projects included:
 - Simple ETL: Load CSV → clean with Python → insert into MySQL/Postgres.
 - File-based workflows: trigger DAGs when files appeared in local/S3 folders.
 - Daily reporting jobs: generate summaries and send via EmailOperator.
 - ❖ 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**

INTERSHIPS & 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.