

AQI -PERDITION SYSTEM REPORT



**MADE BY
FIZZAH ABDULLAH
INTERN**

🧠 Tech Stack

Frameworks, Libraries & Tools

Web / API / UI

- **FastAPI** – Core API framework serving prediction and health endpoints.
- **Uvicorn** – ASGI server running the FastAPI app.
- **Streamlit** – Interactive dashboard for visualization and user interface.

Machine Learning

- **XGBoost** – Primary AQI prediction model (regression/classification).
- **scikit-learn** – Preprocessing, scaling, and evaluation utilities.
- **joblib** – Model serialization (save/load).
- **SHAP, LIME** – Explainability tools for model interpretation.
- **statsmodels** – Statistical analysis and metrics.
- **pandas, NumPy** – Core data manipulation libraries.

Data & Feature Store

- **Hopsworks / HSML** – Feature store and model registry integration.
- **OpenWeather API** – External weather data source.
- **PyArrow** – Efficient data serialization (Parquet/Arrow).
- **Confluent Kafka** – Listed for streaming data ingestion.
- **requests / httpx** – HTTP clients for API calls.

Visualization

- **Plotly, matplotlib, seaborn** – Data visualization and charting.

Server Utilities / Validation

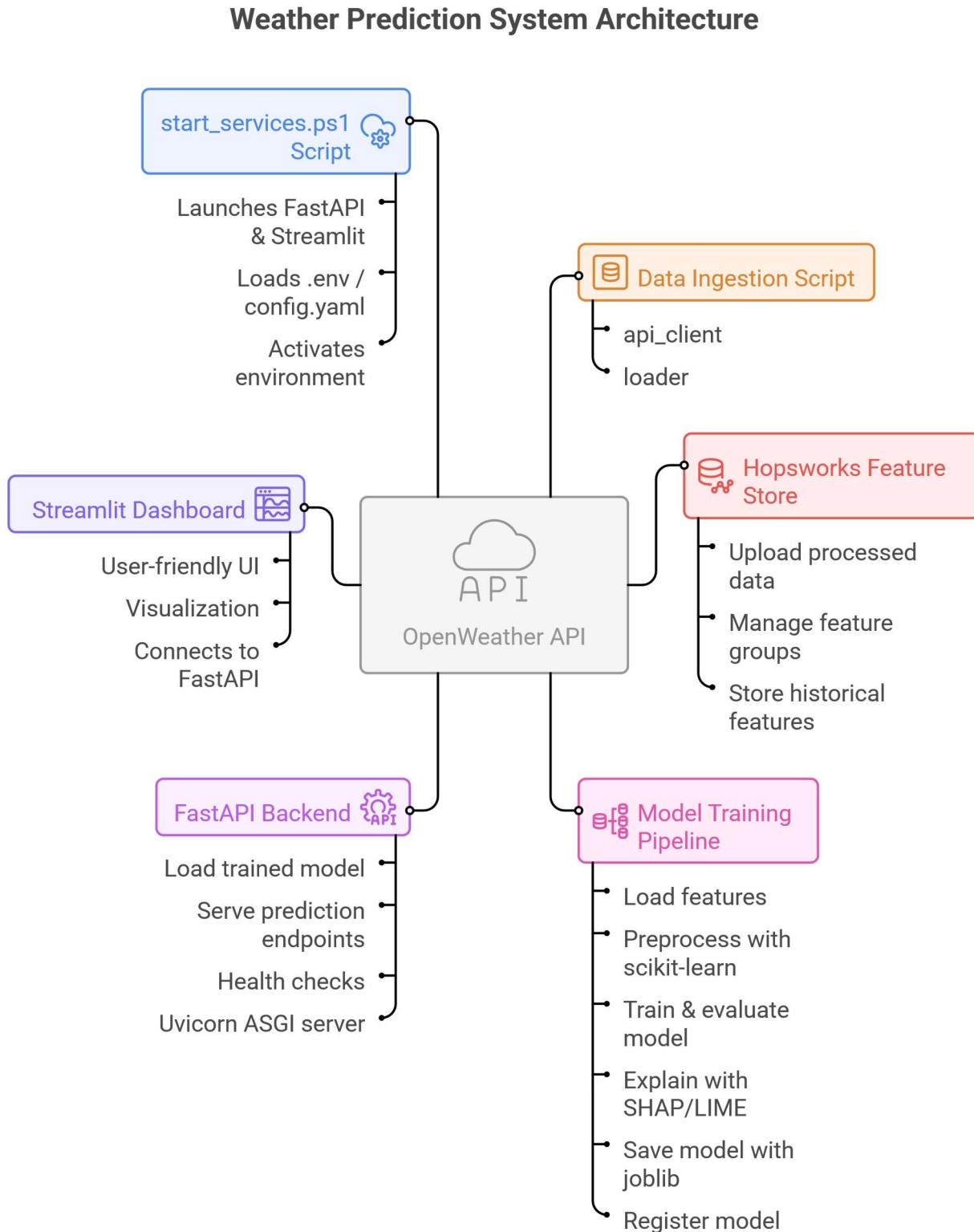
- **pydantic** – Data validation for FastAPI models.
- **aiofiles, python-multipart** – Asynchronous file and multipart handling.
- **python-dotenv / PyYAML** – Environment and configuration management.

Development / Testing

- **pytest, pytest-cov** – Automated testing and coverage.
- **tqdm** – Progress visualization during training.

Orchestration / Deployment

- Launch script for backend (Uvicorn) and dashboard (Streamlit).
- Training orchestration and Hopsworks model registration.



PEARLS-AQI PREDICTION-SYSTEM

HOPS WORKS REGISTRATION

FIZZAH ABDULLAH

2 feature groups

feature logging sort by first created C

weather_features

Feature group for weather_features

Version Last updated
2 3 days ago

pollutant_features

Feature group for pollutant_features

Version Last updated
2 4 days ago

3 models

name	latest version	author	deployed versions	description
aqi_random_forest	2	fu		Random Forest regressor for pollutant AQI
aqi_ridge_regression	2	fu		Ridge Regression (alpha=0.1) for pollutant AQI
aqi_pollutant_aqi_regression	2	fu		XGBoost regression model for pollutant_aqi prediction

CI/CD PIPELINE

79 workflow runs

Event Status Branch Actor

- Feature Ingestion Pipeline Feature Ingestion Pipeline #45: Scheduled main 17 minutes ago 1m 30s ...
- Hourly feature pipeline Hourly feature pipeline #41: Scheduled main 19 minutes ago 1m 6s ...
- Feature Ingestion Pipeline Feature Ingestion Pipeline #44: Scheduled main Today at 3:16 PM 1m 35s ...

