Software Requirements Specification (SRS)

AI-based Downscaling of Satellite Air Quality Data

1. Introduction

1.1 Purpose

The purpose of this project is to develop a software system that enhances low-resolution satellite-based air quality data (e.g., NO_2 concentration) into higher spatial resolution maps using Machine Learning techniques. This system is developed as part of an academic software project inspired by the SIH 2024 problem statement.

1.2 Scope

The system will:

- Take coarse-resolution satellite air quality data as input (from the dataset from ISRO)
- Apply ML-based downscaling techniques
- Produce fine-resolution air quality maps for localized analysis

1.3 Definitions

- **Downscaling**: Increasing spatial resolution of data using computational methods
- **NO_2**: Nitrogen Dioxide, a common air pollutant
- **ML**: Machine Learning

2. Overall Description

2.1 Product Perspective

The system is a standalone analytical tool that processes geospatial datasets and outputs enhanced-resolution air quality maps.

2.2 User Characteristics

- User is expected to have basic knowledge of Python and data science

- No advanced GIS expertise required

2.3 Operating Environment

- Python 3.x
- Jupyter Notebook / VS Code
- OS: Windows / Linux
- Libraries: NumPy, Pandas, Scikit-learn, Matplotlib

3. Functional Requirements

FR1: Data Ingestion

The system shall accept satellite-based air quality data in CSV or NetCDF format.

FR2: Data Preprocessing

The system shall clean missing values and normalize input features.

FR3: Feature Integration

The system shall combine satellite data with auxiliary datasets such as meteorological and land-use data.

FR4: Model Training

The system shall train an ML model to learn the mapping between low-resolution and high-resolution data.

FR5: Prediction

The system shall generate high-resolution air quality maps as output.

FR6: Visualization

The system shall visualize results using plots or heatmaps.

4. Non-Functional Requirements

NFR1: Performance

The system should process datasets within reasonable time on a personal computer.

NFR2: Usability

The system should be easy to run using clearly documented scripts or notebooks.

NFR3: Scalability

The system may support additional pollutants in the future.

5. Constraints

- Availability of high-resolution ground truth data is limited
- Computational resources are limited to a student environment

6. Assumptions

- Input datasets are correctly formatted
- Auxiliary data is spatially aligned with satellite data