TEAM - 01

The Spectral Soil Modeler
An Automated ML Workflow

Software Systems Development[CS6.302] - Phase 1 Submission

Akshat Kotadia	2025201005
Jewel Joseph	2025201047
Gaurav Patel	2025201065
Parv Shah	2025201093
Eshwar Pingili	2025204030

INTERNATIONAL INSTITUTE OF
INFORMATION TECHNOLOGY
HYDERABAD

Supervisor : Dr. Abhishek Singh
Submission Date: 30th September, 2025

1. Project Understanding

1.1. Core Problem and Scientific Context

The current workflow for soil scientists at the Laboratory for Spatial Informatics (LSI) presents a significant research bottleneck. To build predictive models, researchers must manually test every combination of **3 spectral preprocessing techniques** and **5 machine learning algorithms** for each soil property. This process is repetitive, time-consuming, and limits the speed at which new hypotheses can be explored.

Our solution, **The Spectral Soil Modeler**, is a purpose-built web application designed to eliminate this inefficiency. It fully automates the pipeline creation, training, and validation process. By simply uploading their dataset and selecting a target property, researchers can systematically evaluate all **15 model combinations** in a single, automated run.

The application's core output is an interactive dashboard that allows researchers to instantly compare all pipelines via diagnose model performance with visualizations like feature importance plots, and export the best-performing model for future use. This tool transforms the workflow from tedious manual labor to efficient, data-driven analysis, directly accelerating the scientific research cycle at LSI.

1.2. Objective and Scope

Our objective is to engineer "The Spectral Soil Modeler," a sophisticated yet user-friendly web application that automates this entire workflow. The tool will serve as an automatic platform tailored specifically for soil spectroscopy.

In Scope:

- A Streamlit-based graphical user interface (GUI) for all operations.
- Automated training and validation of all 15 pipelines (3 preprocessing * 5 ML models).
- Rigorous model evaluation using k-fold cross-validation.
- An interactive dashboard for visualizing, comparing, and diagnosing models.
- Functionality to export the best-performing, fully trained model for future use.

2. Personas Involved

1. *Soil Scientists*
    - **Role:** End-user who conducts soil spectroscopy research and builds predictive models for soil properties.
    - **Primary Interaction:** Uses the GUI to upload datasets, select target properties, run automated ML pipelines, visualize results, and export models.

2. *Field Agronomist*
    - **Role**: A professional working directly with farmers to implement soil management practices based on research findings.
    - **Primary Interaction**: Leverages the tool's exported models to predict soil properties, aiding in on-site decision-making for crop management.

3. *Environmental Policy Analyst*
    - **Role**: A professional working for government agencies or NGOs focused on environmental conservation, using soil data to inform land-use policies or sustainability initiatives.
    - **Primary Interaction**: Uses the tool's outputs to assess soil health trends and advocate for science-based policy decisions.

3. Detailed Solution Description and Workflow

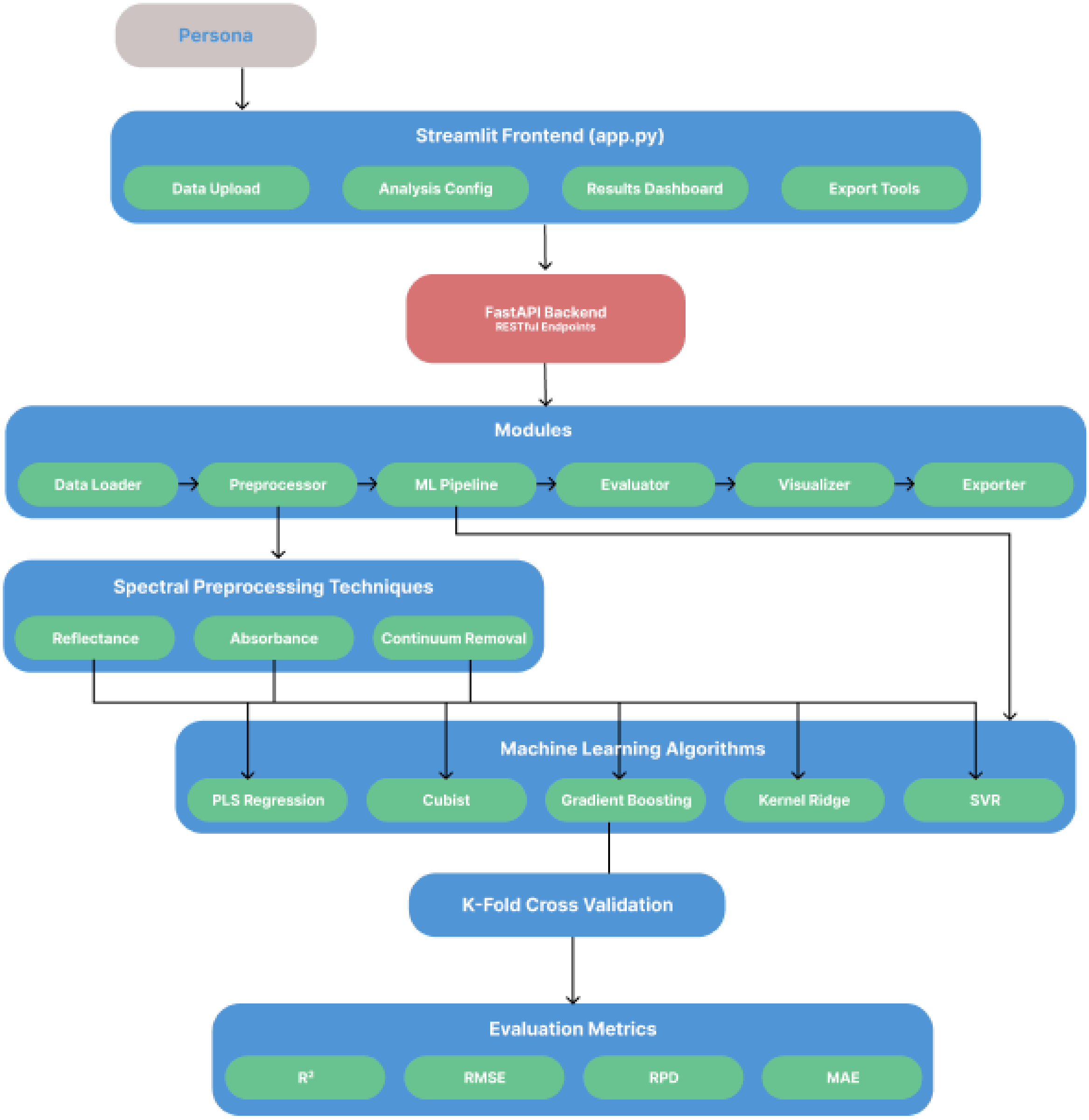
1. *Data Ingestion:* The user begins by uploading a .csv dataset containing spectral data. The system performs automated validation to ensure data integrity, checking for missing values, correct formatting, and compatibility with the expected structure.
2. *Analysis Configuration:* Through an intuitive graphical user interface (GUI), the user selects a target soil property from a dropdown menu populated with available target variables

3. *Automated Training and Validation:* With one click, the backend systemically evaluates all 15 pipelines using k-fold cross-validation, calculating a comprehensive suite of metrics (R2, RMSE, RPD, MAE, Bias, CCC).
4. *Results Dashboard:* Once the evaluation is complete, the user is presented with a dashboard that visualizes the results in a clear and actionable manner.
5. *Export:* Users can download the best-performing model from the dashboard and export it in a portable format.

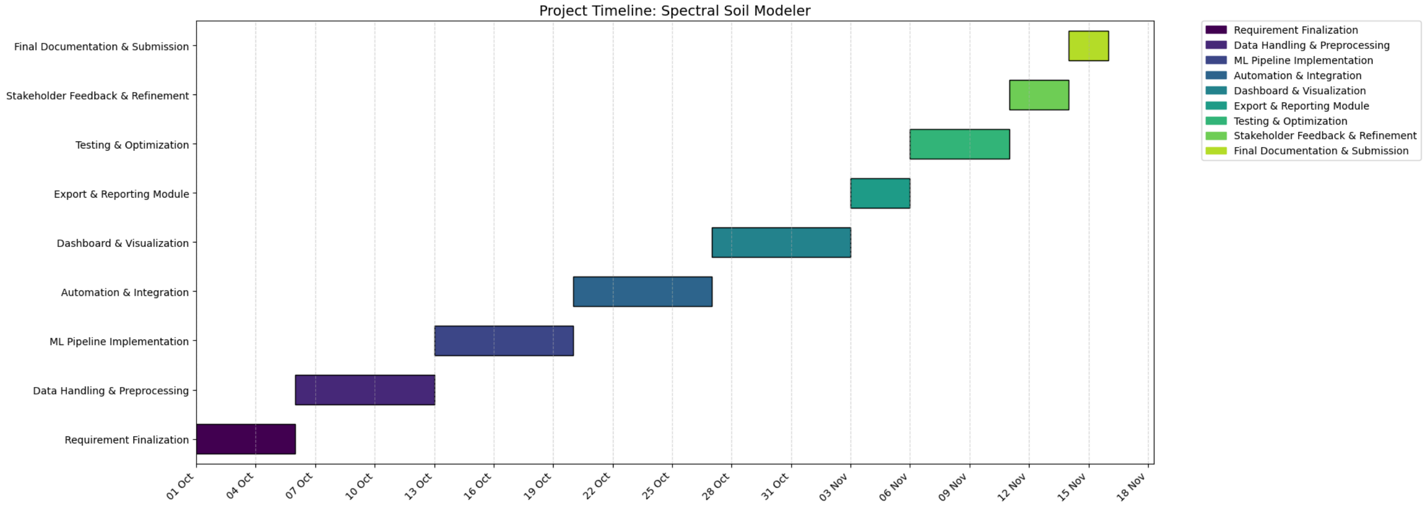
4. Technology Stack and Justification

Python - FastAPI	Backend Framework	High-performance, asynchronous web framework that efficiently handles the computationally intensive task.
Streamlit	Frontend Framework	Enables rapid development of interactive data apps with pure Python. Perfect for research tools.
Machine Learning Libraries	ML & Data Processing	Provides the tools for building, training, and evaluating predictive models. Includes libraries for classical ML (scikit-learn), and data handling (NumPy, Pandas, SciPy).
Git & GitHub	Version Control	Essential for managing code development within our 5-person team and enabling collaborative work.

5. Solution Diagram



6. Project-Timeline


Project Timeline: Spectral Soil Modeler
Final Documentation & Submission
Stakeholder Feedback & Refinement
Testing & Optimization
Export & Reporting Module
Dashboard & Visualization
Automation & Integration
ML Pipeline Implementation
Data Handling & Preprocessing
Requirement Finalization
01 Oct
04 Oct
07 Oct
10 Oct
13 Oct
16 Oct
19 Oct
22 Oct
25 Oct
28 Oct
31 Oct
03 Nov
06 Nov
09 Nov
12 Nov
15 Nov
18 Nov
Requirement Finalization
Data Handling & Preprocessing
ML Pipeline Implementation
Automation & Integration
Dashboard & Visualization
Export & Reporting Module
Testing & Optimization
Stakeholder Feedback & Refinement
Final Documentation & Submission