SCHOOL OF COMPUTING
Faculty of Engineering

Project Proposal Form MCST1043
Sem: 2 Session: 2024/25

## SECTION A: Project Information.

| | |
|---|---|
| Program Name: | **Masters of Science (Data Science)** |
| Subject Name: | **Project 1 (MCST1043)** |
| Student Name: | Zhao Zhihan |
| Metric Number: | MCS241041 |
| Student Email & Phone: | zzh85642256@gmail.com |
| Project Title: | Research on Spatial Analysis of Soil Environment and Health Risks Based on Machine Learning |
| Supervisor 1: | |
| Supervisor 2 / Industry Advisor(if any): | |

## SECTION B: Project Proposal

### Introduction:

Use machine learning methods to conduct spatial analysis of soil environmental data and health risk data, revealing the internal relationship between soil pollution and regional health problems. The project will integrate multi - source data such as pollutant concentrations in soil, weather conditions, agricultural practices, distribution of nearby industries, disease reports, and demographic statistics. By adopting technologies such as data fusion, spatial statistics, and machine learning modeling, a regional environmental risk assessment and health early - warning model will be constructed. The ultimate goal is to identify high - risk areas within a specific region, provide a scientific basis for the government and relevant departments, and offer decision - making support for prevention, control measures, and emergency responses.

### Problem Background:

Use machine learning methods to conduct spatial analysis of soil environmental data and health risk data, revealing the internal relationship between soil pollution and regional health problems. The project will integrate multi - source data such as pollutant concentrations in soil, weather conditions, agricultural practices, distribution of nearby industries, disease reports, and demographic statistics. By adopting technologies such as data fusion, spatial statistics, and machine learning modeling, a regional environmental risk assessment and health early - warning model will be constructed. The ultimate goal is to identify high - risk areas within a specific region, provide a scientific basis for the

government and relevant departments, and offer decision - making support for prevention, control measures, and emergency responses.

**Problem Statement**:
The main research issues include:

1. With multi - source data inconsistent in format, collection cycle and spatial resolution, how to effectively integrate and pre - process data to meet this major project challenge?
2. Given complex environmental factors and health reports, how to distinguish the true causal link between soil pollution and health risks, avoid confounding factors, and accurately gauge the impact degree?
3. As soil environment and health risks vary regionally, how to use GIS and spatial clustering to reveal internal structures and differences and solve this project problem?
4. The machine - learning model needs good generalization. How to apply it to a real - time early - warning system for dynamic monitoring and risk assessment?

**Aim of the Project**:

In the soil pollution and public health risk research project, the first step is data integration and processing. A comprehensive dataset covering soil pollutants, weather, agriculture, industry, health reports, and demographics is to be created. Meanwhile, issues like data gaps, noise, and heterogeneity must be resolved to underpin further analysis.Post - data processing, spatial statistics and machine - learning methods (e.g., clustering, regression, decision trees, deep neural networks) are employed for spatial data modeling and analysis. This builds a model of the soil - health risk relationship to expose regional risk distributions.Using the model's analysis, a regional environmental and health risk early - warning system is developed. It assesses soil pollution's impact on public health, forecasts risk trends, and enables early warning.

Finally, the project findings are used to offer data - driven advice to government and relevant agencies for environmental governance and health prevention, assisting in informed decision - making.

**Objectives of the Project**:
The research objectives are:

Multi - source Data Integration and Pre - processing
Spatial Modeling and Analysis
Construction of Risk Assessment and Early - warning Model
Decision - making Support and Policy Recommendations

**Scopes of the Project**:
The data of this project includes:
- Environmental data: soil pollutant concentrations, soil types, weather conditions.
- Industrial and agricultural data: types of industries and agriculture, industry distribution.
- Health data: disease types, severity, symptoms and health reports.
- Demographics: gender, age, etc. of the affected population.

**Expected Contribution of the Project**:

Construct a spatial analysis framework for soil environment and health risks based on multi - source data fusion, enriching data analysis methods in the field of environmental health. Utilize advanced machine learning and spatial statistical techniques to effectively handle the integration of multi - dimensional data and the issue of spatial heterogeneity. Build a risk assessment and early - warning system model with dynamic update and real - time early - warning functions. By quantifying the impact of soil pollution on regional public health, provide decision - making support for the government and environmental protection departments to formulate scientific and reasonable intervention measures. Offer specific solutions for agricultural environmental governance and public health risk management, helping to reduce health risks caused by soil pollution.

**Project Requirements**:

Software: Python ,R, GIS,

Hardware: High-performance computers,

Technology/Technique/ Methodology/Algorithm: Data Collection and Integration,Machine Learning Modeling Model Evaluation and Validation

**Type of Project (Focusing on Data Science)**:

[ √ ] Data Preparation and Modeling

[ √ ] Data Analysis and Visualization

[ ] Business Intelligence and Analytics

[ √ ] Machine Learning and Prediction

[ ] Data Science Application in Business Domain

**Status of Project**:

[ √ ] New

[ ] Continued

If continued, what is the previous title?

## SECTION C: Declaration

**I declare that this project is proposed by**:

[ ] Myself

[ ] Supervisor/Industry Advisor ( )

Student Name: Zhao zhihan

**Signature** **Date**

## SECTION D: Supervisor Acknowledgement

The Supervisor(s) shall complete this section.

**I/We agree to become the supervisor(s) for this student under aforesaid proposed title.**

Name of Supervisor 1: .....

**Signature** **Date**

Name of Supervisor 2 (if any): .....

**Signature** **Date**

## SECTION E: Evaluation Panel Approval

The Evaluator(s) shall complete this section.

**Result:**

[ ] FULL APPROVAL [ ] CONDITIONAL APPROVAL (Major)*

[ ] CONDITIONAL APPROVAL (Minor) [ ] FAIL*

* Student has to submit new proposal form considering the evaluators' comments.

**Comments:**

Name of Evaluator 1: .....

**Signature** **Date**

Name of Evaluator 2: .....

**Signature** **Date**