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:	**Li Xinya**
Metric Number:	**MCS241029**
Student Email & Phone:	lixinya@graduate.utm.my & 0175703279
Project Title:	Prediction and Tracking of Plastic Pollution in Rivers of Malaysia Based on Machine Learning
Supervisor 1:	
Supervisor 2 / Industry Advisor(if any):	

SECTION B: Project Proposal

Introduction:

Rivers, as a major component of terrestrial water bodies, are crucial for ecological balance and the survival of living organisms. However, with the increase in human activities and the widespread use of plastic products, rivers have become dumping grounds for plastic waste. Plastics are continuously discharged into rivers, which then spread to the oceans, releasing harmful substances that pose a threat to marine ecosystems and the global environment. Traditional methods for monitoring plastic pollution mainly rely on manual sampling and conventional water quality monitoring techniques, which are inadequate in addressing the spatiotemporal variability of river plastic pollution. This paper explores the feasibility and effectiveness of using machine learning and other technological approaches to monitor and predict river plastic pollution, aiming to achieve dynamic tracking of plastic pollution in rivers and improve the timeliness and accuracy of monitoring, thereby providing scientific data support for pollution management and mitigation efforts.

Problem Background:

As the issue of plastic pollution becomes increasingly severe, global environmental awareness is gradually rising. For Malaysia, a tropical nation with abundant river resources, this unique advantage is now facing significant challenges, with plastic pollution in its rivers and surrounding water bodies becoming more pronounced. Influenced by rapid industrialization, urban expansion, and the booming tourism industry, Malaysia is experiencing plastic pollution from various sources. This pollution not only affects water quality and ecosystems but also enters the human food chain, posing a significant

threat to ecological security and public health.

In the face of the plastic pollution challenge, traditional monitoring methods still dominate, such as manual sampling and on-site testing. While these methods can provide data on the current state of river plastic pollution to some extent, they have significant limitations in terms of timeliness, spatial coverage, and accuracy. Additionally, the amount of plastic waste discharge varies with seasons, rainfall, and human activities, and the pollution status differs across various sections of the river. Traditional monitoring methods struggle to capture the changes in plastic pollution in real-time and with precision. Therefore, developing innovative and efficient monitoring methods is crucial for effective plastic pollution management.

Machine learning techniques can uncover underlying patterns and trends in large datasets. By constructing predictive models, they can proactively identify pollution risks, forecast the spread of pollution, and effectively track the location and propagation path of pollution sources. This data-driven approach significantly enhances the accuracy and efficiency of river plastic pollution monitoring, providing scientific support for pollution management and mitigation efforts.

Problem Statement:

Malaysia's river network is widely distributed across multiple geographic regions, and the plastic pollution status is deeply influenced by a combination of factors, such as industrial growth, urbanization, and the booming tourism industry. These factors interact with each other, resulting in significant instability and spatiotemporal variability in the sources of plastic pollution. In light of this, accurately predicting plastic pollution trends and effectively tracking pollution sources, in turn, developing more optimized management strategies, has become a critical issue that needs to be addressed.

Traditional pollution monitoring methods often rely on manual sampling and field detection. However, these methods are limited by factors such as timeliness, spatial coverage, and operational costs, making it difficult to capture the subtle changes in pollution dynamics in real-time. This is particularly challenging in the context of complex river systems, where the migration and transformation processes of pollutants are hard to monitor comprehensively. Moreover, existing monitoring technologies often provide data from a single dimension, making it difficult to fully reveal the overall trends of pollution sources and the spread of pollutants.

To address these issues, this study proposes the use of machine learning technologies to integrate and analyze multi-source data, constructing a data-driven model for river plastic pollution prediction and tracking. By conducting real-time analysis of river basin data, pollution sources, and pollution diffusion paths, the study aims to predict the changing trends of pollution and track the location and spread of pollution sources. This machine learning-based intelligent monitoring system is expected to effectively compensate for the shortcomings of traditional methods, improving the accuracy, timeliness, and efficiency of pollution monitoring and management, thus providing scientific support for the management of plastic pollution in Malaysia's rivers.

Aim of the Project:

Utilizing machine learning to perform real-time analysis on large-scale data enables efficient and accurate monitoring of plastic pollution in Malaysia's rivers, reducing the limitations of traditional manual sampling and on-site testing methods. It enhances data processing speed and analysis efficiency, providing more timely responses and feedback for pollution management, and helping to forecast and control the spread of plastic pollution.

Objectives of the Project:

Utilizing machine learning techniques, such as Support Vector Machines and Random Forests, in conjunction with remote sensing images, water quality data, and sensor data, an efficient plastic pollution prediction model will be developed. This model will be capable of forecasting the variation trends of plastic pollution concentrations in Malaysian rivers and providing pollution warnings for different regions and time periods.

Scopes of the Project:

1.Data Collection: Reliable data will be collected from remote sensing platforms and water quality monitoring data platforms, including remote sensing images, real-time water quality data, and sensor data (e.g., EOSDIS, Malaysian DOE, and GRoW Data).

2. Data Processing: The collected multi-source data will be classified and archived, addressing issues such as missing values, outliers, and noisy data to ensure data quality and reliability. Data from different sources and units will be standardized or normalized to ensure comparability and suitability for machine learning models.

3.Model Construction: Based on multi-source data, machine learning algorithms (e.g., Random Forests, Deep Learning, etc.) will be used to construct plastic pollution prediction models, which will forecast the concentration of plastic pollution and its variation trends in different time periods and river sections.

4.Prediction and Analysis: The model will predict the future trends of plastic pollution in the river over different time periods, identifying pollution peaks and their potential impacts.

5.Error Analysis: The prediction errors of the model will be analyzed to identify factors affecting model performance. Algorithm optimization and parameter tuning will be employed to improve the model's prediction accuracy.

Expected Contribution of the Project:

1.The machine learning-based research on plastic pollution prediction and tracking will help Malaysia optimize pollution control strategies, enabling real-time monitoring of pollution sources and improving intervention effectiveness.

2.The study will explore the application of machine learning in data fusion and spatiotemporal modeling, developing data models that address complex river pollution dynamics and advancing the intelligent development of environmental technology.

3.This research combines remote sensing technology and water quality monitoring data to enable real-time monitoring of river water quality changes, promptly detecting plastic pollution, and providing scientific solutions for water quality protection.

Project Requirements:

Software:	Python ; ArcGIS
Hardware:	Personal computer with sufficient processing power and storage capacity
	Machine Learning

Technology/Technique/ Methodology/Algorithm:	Statistical Analysis
	Data Visualization

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:

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**