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: WANG TONG

Metric Number: MCS241052

Student Email & Phone: tong20@graduate.utm.my +601114182936

Project Title: Prediction and Analysis of Typhoon Landfall Points Based on Random Forest

Supervisor 1:

Supervisor 2 / Industry Advisor(if any):

SECTION B: Project Proposal

Introduction:

Typhoon, being a type of tropical cyclone, typically accompanies strong winds and heavy rain, exerting a considerable influence on the housing of residents, people's properties, urban construction, road traffic and other economic aspects in coastal regions. In Southeast Asia, particularly in the surrounding sea areas of Malaysia, the generation and trajectory of typhoons are characterized by complexity and uncertainty, which renders predicting the landing point of typhoons an extremely challenging task. Predicting the landing point of typhoons is of paramount importance for disaster early warning, disaster emergency response and the corresponding resource allocation.

In recent years, machine learning techniques have exhibited strong potential in the domain of meteorological forecasting. Random forest, as a classical machine learning algorithm, has been extensively utilized in various types of prediction issues. It is capable of effectively capturing the complex relationship directly between typhoon paths and meteorological variables.

This research endeavors to employ the random forest model, in conjunction with typhoon data and meteorological characteristics, to establish a prediction model for typhoon landfall points in the surrounding waters of Malaysia, with the aim of facilitating better typhoon prediction and disaster emergency responses by relevant departments.

Problem Background:

Typhoons constitute one of the most severe natural disasters globally, particularly in Southeast Asia where they recurrently assail, presenting a substantial threat to coastal nations. Malaysia is located in the tropical region of Southeast Asia, bordering the Strait of Malacca to the west and the South China Sea to the east. Although Malaysia is not a country frequently hit by typhoons, its eastern coast (such as Kelantan, Terengganu, and Pahang) may still be affected by typhoons formed in the South China Sea or the remnants of typhoons moving northward from the East China Sea, causing strong winds, heavy rain, and floods.

For example:

• In 2017, Typhoon Hato, although it did not make a direct landfall, its peripheral circulation caused floods in many areas along the eastern coast of Malaysia.

• In 2021, Tropical Depression Haitang brought continuous heavy rain to Johor, forcing the evacuation of tens of thousands of people.

Accurate predictions of typhoon landfall points are crucial for Malaysia's disaster warning systems, coastal infrastructure protection, and evacuation plans.

Problem Statement:

Traditional typhoon trajectory prediction primarily depends on conventional numerical models (e.g., WRF, ECMWF), yet their computational complexity is considerable and they are highly sensitive to initial conditions, thereby failing to satisfy the timeliness demands of disaster emergency responses. Moreover, studies on typhoons in the surrounding waters of Malaysia are scarce. The existing models mostly concentrate on the Northwest Pacific or the North Atlantic, leading to insufficient analysis of the crucial factors in the waters around Malaysia and inadequate consideration of the specific geographical and climatic conditions of the Malaysian sea area.

Aim of the Project:

1. How can a typhoon landing point prediction model (with longitude and latitude coordinates as the output) be constructed based on the random forest algorithm by utilizing the relevant data in the IBTrACS dataset?

2. Which key factors drive the prediction outcomes of typhoon landfall points? How do the particular geographical conditions of the sea areas surrounding Malaysia influence the distribution of typhoon landfall points?

Objectives of the Project:

1.Construct a random forest model: Based on the typhoon trajectories and environmental variables in the IBTrACS dataset, develop a prediction model for typhoon landing points in the surrounding sea areas of Malaysia (with the output of longitude and latitude coordinates).

2. Analysis of Key Influencing Factors: Through feature importance analysis and nonlinear modeling, quantify the dynamic influence of sea surface temperature, topography and atmospheric conditions on the landing locations of typhoons.

Scopes of the Project:

1. In terms of the geographical range, the typhoon data from the surrounding sea areas of Malaysia (latitude and longitude range: 0° - 10° N, 95° - 120° E) are employed, with the typhoon data from other regions excluded.

2. Regarding the definition of landfall, a typhoon is regarded as having landed if the distance between its center and the coastline of Malaysia is no more than 50 kilometers. The potential landfall points on the east coast (the side facing the South China Sea) are of key concern.

3. Concerning the prediction duration, with the typhoon entering the study area (latitude 0° – 20° N, longitude 100° – 120° E) as the starting point, the landing point within the next 72 hours is predicted.

4. Employing the relevant data in the IBTrACS dataset, a typhoon landing point prediction model is established.

5. The impacts of key factors, such as climate and terrain, on the landing points of typhoons will be analyzed.

Expected Contribution of the Project:

1.Enhancing disaster early warning capabilities: The model can assist the Meteorological Department of Malaysia in optimizing the prediction of typhoon landing points, locking high-risk areas 48 hours in advance, and reducing economic losses in coastal communities due to heavy rainfall and typhoon landings.

2. Supporting emergency decision-making: Through analysis and clarification of key driving factors, it provides a scientific basis for the allocation of disaster prevention resources (such as the planning of evacuation routes).

Project Requirements:

Software: Python、 Anaconda

Hardware:

Technology/Technique/ Methodology/Algorithm: Machine Learning, Random Forest

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: WANG TONG

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**