Capstone Project\_Week 1

Requirement:

1. A description of the problem and a discussion of the background. (**15 marks**)
2. A description of the data and how it will be used to solve the problem. (**15 marks)**

# Introduction

## 1.1 Background

In recent years, due to natural causes and protection problems, a series of environmental problems, such as soil erosion, heavy rain, drought and land desertification, have emerged resulting in the rapid reduction of forests and the destruction of animal and plant resources. Therefore, the protection of the rainforest is an important part of the world's environmental protection.

Rainforests are located mainly in the tropics near the equator, where abundant rainfall makes the forests flourish. And, because of its abundant rainfall, almost all the year round hot and humid, most of the time is flooded. Rainforests are also home to more than half of the world’s flora and fauna, including millions of undiscovered species in addition to the many rare ones known to man. According to statistics, more than 25 percent of modern medicines are extracted from tropical rain forest plants, so the tropical rain forest is known as the ”world's largest pharmacy", is a precious resource given to human beings on earth. In addition, rainforests can regulate the climate, prevent soil erosion, purify the air, and ensure the material circulation of the earth's biosphere.

Especially, the Amazon rainforest plays a vital role in the health of all living things on earth. The rainforest climate has great impact on the world, plants absorb carbon dioxide and produce oxygen for photosynthesis, carbon dioxide influence on earth's climate, and photosynthesis to produce oxygen, and all humans and animals substances necessary to sustain life, therefore, the amazon rainforest and its reputation as the "lungs of the earth. The Amazon rainforest contains the most abundant and diverse biological resources in the world, with millions of species of insects, plants, birds and other creatures. It is also known as the "animal and plant kingdom of the world".

## 1.2 Problem

In this project, I will investigate the data set of different vegetation types covering the rainforest surface. I will use the data to structure a classification model and determine whether the vegetation is suitable for the region in order to get better earning.

# 2. Data

## 2.1 Data Description

I got the data set from Kaggle website, it is an open source one.

The data contains 581012 rows and 55 columns.

Following are the columns in the data set. I will use different features to structure the classification model. Details will be updated in the following report.