Dong Liang 1399853

I am from MGI and I already finished “Geoscripting” and “Geo-information Tools” courses. So, I have experience in Python, R and GIS. But I have not applied my programming skills to data science projects except for those in the GIS/RS domain. So, I chose this course to enrich my experience in programming combined with environmental science and social science. And I am also interested in how to handle different data sources and combine these data in one project using Python. After completing this course, I hope I will be able to combine complex data sources and analyze the data in the environmental science domains.

**Personal learning goals**

Data:

“To be able to acquire and process text data from news websites using Python.” Currently, I have only used Python in geographical domain. At the end of this course, I would like to be able to acquire text data from news websites using web scrapers and analyze these data in Python. Therefore, I will produce an executable Python Jupyter notebook which shows how a web scraper acquires text data from news websites and analyze the text data using word cloud.

Information:

“To be able to do regression analysis between traffic data and emission data using machine learning.” Currently, I do not have any experience in regression analysis. Also, I have never used machine learning in Python. At the end of this course, I want to be able to do multiple regression analysis using Python. Therefore, I will produce a Jupyter notebook to demonstrate the steps of regression analysis between traffic and emission.

Knowledge:

“Understand the pitfalls of using the machine learning to do regression for limited data.” Currently, I only know the basic concept of machine learning but at the end of this course, I will be able to explain the pitfalls of doing regression with machine learning and how it will affect the result of regression. I will demonstrate this in a small essay in which I will provide a literature review about the limitation of machine learning and an analysis of our project procedure to show how machine learning is used and the pros and cons of using it in our project.