

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO BẢNG PHÂN CÔNG CÔNG VIỆC
CSC14119 – NHẬP MÔN KHOA HỌC DỮ LIỆU

Giảng viên hướng dẫn:

Thầy Trần Đại Chí

Thầy Nguyễn Bảo Long

Thầy Lê Nhựt Nam

Thầy Nguyễn Thái Vũ

20120210 – TRẦN THỊ KIM TIỀN

20120258 – LÂM QUỐC CHUNG

20120307 – PHẠM GIA KHIÊM

20120334 – LÝ THÀNH NAM

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO BẢNG PHÂN CÔNG CÔNG VIỆC

|Tên môn học|

CSC14119 – NHẬP MÔN KHOA HỌC DỮ LIỆU

Giảng viên hướng dẫn:

Thầy Trần Đại Chí

Thầy Nguyễn Bảo Long

Thầy Lê Nhựt Nam

Thầy Nguyễn Thái Vũ

MỤC LỤC

MỤC LỤC.....1

SƠ LƯỢC VỀ NHÓM.....2

 1. Giáo viên hướng dẫn:.....2

 2. Sơ lược về cách thành viên trong nhóm:2

PHÂN CHIA CÔNG VIỆC.....3

TỔNG HỢP KẾT QUẢ.....5

SƠ LƯỢC VỀ NHÓM

1. Giáo viên hướng dẫn:

Thầy **Trần Đại Chí**
Thầy **Nguyễn Bảo Long**
Thầy **Lê Nhựt Nam**
Thầy **Nguyễn Thái Vũ**

2. Sơ lược về các thành viên trong nhóm:

- Thông tin nhóm: **NHÓM 06**
- Thông tin thành viên trong nhóm:

STT	Họ và tên	Mã số sinh viên	Email
1	Trần Thị Kim Tiến	20120210	20120210@student.hcmus.edu.vn
2	Lâm Quốc Chung	20120258	20120258@student.hcmus.edu.vn
3	Phạm Gia Khiêm	20120307	20120307@student.hcmus.edu.vn
4	Lý Thành Nam	20120334	20120334@student.hcmus.edu.vn

- Mức độ đóng góp của các thành viên trong nhóm:

STT	Họ và tên	Mức độ đóng góp (%)
1	Trần Thị Kim Tiến	25
2	Lâm Quốc Chung	25
3	Phạm Gia Khiêm	25
4	Lý Thành Nam	25

PHÂN CHIA CÔNG VIỆC

STT	Yêu cầu chính	Phụ		Người đảm nhiệm	Deadline
1	Thu thập dữ liệu			Nam, Chung	26.11
2	Tiền xử lý dữ liệu			Khiêm, Tiến	30.11
3	Trả lời câu hỏi	1	- Đây là top 5 thành phố, quốc gia có mật độ ô nhiễm cao nhất, thấp nhất trong 5 năm 2017 - 2021? Với mỗi thành phố, quốc gia cao nhất, thấp nhất trong năm 2022 mật độ ô nhiễm biến đổi như thế nào theo các tháng? Thử giải thích.	Nam	4.12
		2	- Việt Nam nằm trong top bao nhiêu, với mật độ ô nhiễm là bao nhiêu qua từng năm (2017 - 2021). Có sự thay đổi như nào ở Việt Nam trong giai đoạn này và sự thay đổi này là tốt hay xấu ? Thử giải thích.	Chung	
		3	- Tháng nào trong năm 2022 sẽ có mức độ ô nhiễm cao nhất cũng như thành phố nào ô nhiễm nhất trong 12 tháng qua, và sự thay đổi của không khí Việt Nam từ 2017 đến 2022.	Khiêm	
		4	- Thành phố thuộc quốc gia nào có sự thay đổi nhiều nhất (tăng ô nhiễm nhiều nhất, giảm ô nhiễm nhiều nhất) ? Giai đoạn nào là nhiều nhất? Tại sao?	Tiến	
		5	- Địa điểm, quốc gia nào có nhiều biến động về khí hậu lớn nhất trong 12 tháng của năm 2022, có thể chỉ ra tháng nào hay không ?	Khiêm	7.12
4	Lấy dữ liệu để phân tích			Chung, Nam	9.12

	và xây dựng mô hình			
5	Phân tích đến xây dựng mô hình		Khiêm, Tiến	11.12
6	Đánh giá và phân tích kết quả		Nam, Chung	13.12
7	So sánh với các phương pháp khác		Nam, Chung	14.12
8	Báo cáo phân chia công việc		Tiến	14.12
9	Slide		Khiêm, Tiến	14.12

TỔNG HỢP KẾT QUẢ

Nếu có nhiều thời gian hơn thì nhóm chúng em sẽ có thể lựa chọn bộ dữ liệu từ trang web mới tốt hơn, chính xác hơn. Nghiên cứu thêm về các mô hình có thể áp dụng vào bộ dữ liệu, tìm mô hình học giám sát và học không giám sát với các bộ dữ liệu hiện có. Hoàn thành chỉnh chu hơn về nội dung của bài làm, nâng cao chất lượng của câu hỏi, bộ câu hỏi và cách thức trình bày được chính xác, rõ ràng, rành mạch và có ý nghĩa hơn.

Những khó khăn và những bài học trong quá trình thực hiện đồ án:

STT	Họ và tên	Khó khăn	Điều học được
1	Phạm Gia Khiêm	Tìm ra dữ liệu phù hợp cho nhóm. Lựa chọn các tham số dữ liệu để đưa vào mô hình. Thu thập dữ liệu cho mô hình. Tìm ra mô hình phù hợp cho dữ liệu và đánh giá mô hình.	Biết thêm được các trang web có kiến trúc phức tạp. Và tùy vào kiến trúc của trang web mà ta có thể crawl dữ liệu được hay không. Ví dụ như trong bài thì web được code bằng react nên không thể lấy các url để thu thập dữ liệu cho mô hình được.
2	Trần Thị Kim Tiến	Khó khăn trong việc tìm mô hình học máy phù hợp. Lựa chọn câu hỏi phù hợp với dữ liệu. Khó khăn khi giải quyết câu hỏi vì còn chưa nắm kỹ các cách trực quan hóa, cũng như các thư viện của pandas.	Củng cố kiến thức về các mô hình học máy, trực quan hóa, tiền xử lý dữ liệu. Có kinh nghiệm hơn trong việc tìm dữ liệu.
3	Lý Thành Nam	Khó khăn trong việc tìm hiểu và thao tác github để không bị conflict Việc tìm kiếm, lọc ra các web url phù hợp để có thể crawl dữ liệu về còn khó khăn do cấu trúc trang web khác biệt và có khi có lỗi	Học được cách xác định chính xác từng thành phần web để có thể lấy đúng dữ liệu về, đồng thời biết thêm nhiều loại web khác nhau Có thêm kỹ năng sử dụng lệnh git, xác định conflict và thay đổi code

			để có thể commit phù hợp và push github
4	Lâm Quốc Chung	<p>Tìm cách crawl dữ liệu phục vụ cho mô hình.</p> <p>Lựa chọn câu hỏi phù hợp với dữ liệu thu thập được.</p> <p>Tìm ra hướng giải quyết cho câu hỏi đó.</p>	<p>Biết được thêm các cách crawl dữ liệu mới vì web nhóm tìm có cấu trúc khá phức tạp. Hiểu rõ hơn về mô hình hóa dữ liệu, so sánh và lựa chọn mô hình phù hợp với bài toán cần thực hiện.</p>