

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



**KẾ HOẠCH LÀM VIỆC NHÓM**  
**ĐỒ ÁN CUỐI KỲ**  
**NHẬP MÔN KHOA HỌC DỮ LIỆU**

Tháng 1 - 2021

# Mục lục

<b>I.</b>	<b>Thông tin nhóm .....</b>	<b>3</b>
<b>II.</b>	<b>Kế hoạch làm việc .....</b>	<b>3</b>
1.	Giai đoạn chọn đề tài .....	3
2.	Giai đoạn thu thập dữ liệu .....	3
3.	Giai đoạn khám phá, tiền xử lý và mô hình hóa dữ liệu .....	3
4.	Giai đoạn hoàn thiện đồ án .....	4

## I. Thông tin nhóm

	STT	Họ và tên	MSSV	Github
Nhóm 11	1	Nguyễn Văn Hậu	18120359	<a href="#">HauNguyen</a>
	2	Nguyễn Tấn Thìn	18120085	<a href="#">TanThinNguyen</a>

## II. Kế hoạch làm việc

### 1. Giai đoạn chọn đề tài

Thời gian	Người thực hiện	Công việc	Kết quả
25/12/2020- 26/12/2020	Cả nhóm	Suy nghĩ, tìm ý tưởng chủ đề	Chọn được ý tưởng: Dự đoán độ ẩm tương đối từ các thông số thời tiết khác như nhiệt độ, điểm sương, tốc độ gió,...

### 2. Giai đoạn thu thập dữ liệu

Thời gian	Người thực hiện	Công việc	Kết quả
27/12/2020- 29/12/2020	Thìn	Tìm API cung cấp dữ liệu thời tiết trong quá khứ	<p>Tìm được một số API như Open Weather API, Visual Crossing API, Underground Weather API</p> <p>Kết quả cuối cùng chọn Visual Crossing API vì:</p> <ul style="list-style-type: none"><li>• Không tốn phí</li><li>• Dữ liệu cung cấp khá chi tiết</li></ul> <p>Lí do không chọn 2 API kia vì:</p> <ul style="list-style-type: none"><li>• Open Weather API tốn phí để có thể lấy được dữ liệu thời tiết quá khứ</li><li>• Underground Weather API đã bị đóng cửa không thể sử dụng</li></ul>
30/12/2020- 1/1/2021	Thìn	Tiến hành dùng API để thu thập dữ liệu	Thu thập được dữ liệu thời tiết với hơn 4000 dòng

### 3. Giai đoạn khám phá, tiền xử lý và mô hình hóa dữ liệu

Thời gian	Người thực hiện	Công việc	Kết quả
02/01/2021- 03/01/2021	Hậu	Tìm hiểu ý nghĩa và quan hệ của các	Nhận ra dữ liệu thời tiết có tính liên tục từ quá khứ đến hiện tại.

		dòng và các cột dữ liệu	Điều chỉnh câu hỏi thành: Dự đoán độ ẩm tương đối của ngày kế tiếp từ thông số nhiệt độ, độ ẩm tương đối, điểm sương, ... của ngày hiện tại và những ngày trước
04/01/2021-07/01/2021	Hậu	Tách các tập dữ liệu Khám phá dữ liệu Tiền xử lý dữ liệu	Cột dữ liệu thời tiết của những ngày trước được thêm vào từng dòng của tập dữ liệu  Dữ liệu được chia thành 3 tập huấn luyện, tập validation, tập kiểm tra  Xây dựng được pipeline tiền xử lý dữ liệu <ul style="list-style-type: none"> <li>Loại bỏ các cột không cần thiết và giữ lại các cột quan trọng</li> <li>Xử lý các giá trị thiếu</li> <li>Chuẩn hóa dữ liệu</li> </ul>
08/01/2021-11/01/2021	Thìn	Xây dựng, thử nghiệm, đánh giá nhiều mô hình	Xây dựng được full_pipeline kết hợp quá trình tiền xử lý và mô hình hóa dữ liệu  Tìm được bộ siêu tham số tốt nhất cho mô hình  Đánh giá được độ lỗi của mô hình trên tập kiểm tra

#### 4. Giai đoạn hoàn thiện đồ án

Thời gian	Người thực hiện	Công việc	Kết quả
12/01/2021-13/01/2021	Hậu	Tổng hợp file Teamwork	Hoàn thiện file Teamwork từ những ghi chép phân công trong suốt quá trình làm đồ án
	Thìn	Viết Slide tóm tắt file notebook	Hoàn thiện file Slide
14/01/2021-15/01/2021	Cả nhóm	Ghi nhận ý kiến đóng góp	Hoàn thành các file notebook, Slide, Teamwork
	Hậu	Tinh chỉnh, hoàn thiện file notebook, Teamwork	Hoàn thành đồ án
	Thìn	Tinh chỉnh, hoàn thiện file Slide	