



ĐẠI HỌC KHOA HỌC TỰ NHIÊN TP HCM
KHOA CÔNG NGHỆ THÔNG TIN

NHẬP MÔN KHOA HỌC DỮ LIỆU

BÁO CÁO LÀM VIỆC NHÓM

ĐỒ ÁN CUỐI KỲ

Mục lục

1. Thông tin thành viên nhóm:.....	3
2. Tiến độ công việc:	3
2.1. Trước khi bắt đầu đề án (trước 28/12):	3
2.2. Tuần 1 (28/12 – 03/01):	4
2.3. Tuần 2 (04/12 – 10/01):	4
2.4. Tuần 3 (04/12 – 10/01):	5

1. Thông tin thành viên nhóm:

STT	MSSV	Họ và tên
1	18120058	Phạm Công Minh
2	18120090	Phạm Nguyên Minh Thy

2. Tiến độ công việc:

2.1. Trước khi bắt đầu đồ án (trước 28/12):

Họ tên	Công việc	Tự đánh giá
Phạm Nguyên Minh Thy + Phạm Công Minh	- Suy nghĩ chủ đề	100%
Phạm Công Minh	- Tìm API hoặc Web để Parse HTML nếu không có API	100%
Phạm Nguyên Minh Thy	- Gửi mail xin key cho API, đọc documents của API chuẩn bị cho quá trình thu thập dữ liệu	100%

Tiến độ:

- Đã chọn được chủ đề: Định giá laptop dựa trên các thuộc tính của laptop
- API đã chọn: <https://noteb.com/?public/api.php>
- Đã xin được key cho API: iH0PvxW6fR6Dng8

2.2. Tuần 1 (28/12 – 03/01):

Họ tên	Công việc	Tự đánh giá
Phạm Nguyên Minh Thy	- Thu thập dữ liệu dựa vào API	100%
Phạm Công Minh	- Tiền xử lý dữ liệu	100%

Tiến độ:

- Đã hoàn thành code thu thập dữ liệu từ API
- Đã thu thập được khoảng 1000 mẫu dữ liệu
- Đã hoàn thành code tiền xử lý dữ liệu

2.3. Tuần 2 (04/12 – 10/01):

Họ tên	Công việc	Tự đánh giá
Phạm Nguyên Minh Thy + Phạm Công Minh	- Mô hình hóa dữ liệu và đánh giá mô hình	70%

Tiến độ:

- Code mô hình hóa đã hoàn thành (sử dụng thuật toán SGDRegressor để mô hình hóa)
- Tuy nhiên độ lỗi còn cao khoảng 40%

2.4. Tuần 3 (04/12 – 10/01):

Họ tên	Công việc	Tự đánh giá
Phạm Nguyên Minh Thy	- Tìm cách cải thiện độ chính xác của mô hình	100%
Phạm Công Minh	- Hoàn thiện file “Teamwork.pdf”	100%
Phạm Nguyên Minh Thy + Phạm Công Minh	- Hoàn thiện các câu hỏi trong file Notebook - Viết file “Slide.pdf”	100%

Tiến độ:

- Thử nghiệm mô hình hóa với các thuật toán và các tham số khác nhau (Linear Regression, MLP Regression,...) nhưng tỷ lệ lỗi vẫn không cải thiện => Thêm vào dữ liệu các thuộc tính khác => cải thiện được độ lỗi còn khoảng 30% => Đã giải quyết được vấn đề độ chính xác của mô hình

- Hoàn thành các câu trả lời trong file notebook

- Đã hoàn thành các file Slide.pdf và Teamwork.pdf