

NHẬP MÔN KHOA HỌC DỮ LIỆU

Phân công công việc và tổng hợp kết quả

Nguyễn Duy Khang – 20120505

Nguyễn Hữu Anh Khoa – 20120510

Lê Quang Thọ - 20120588

Nguyễn Anh Quốc - 19120633



Bộ môn Khoa học dữ liệu

Khoa Công nghệ thông tin

Đại học Khoa học tự nhiên TP HCM

MỤC LỤC

1	Bảng đánh giá thành viên	3
2	Phân công công việc	4
3	Tổng hợp kết quả:	5

1 Bảng đánh giá thành viên

MSSV	Họ Tên	% đóng góp (tối đa 100%)	Chữ ký
20120505	Nguyễn Duy Khang	100%	Khang
20120510	Nguyễn Hữu Anh Khoa	100%	Khoa
20120588	Lê Quang Thọ	100%	Thọ
19120633	Nguyễn Anh Quốc	100%	Quốc

2 Phân công công việc

Phần công việc	Họ Tên	Hoàn thành
Phần 1A: Thu thập dữ liệu.	Khoa	100%
Phần 1B: Khám phá dữ liệu.	Khang: 3 câu đầu. Thọ, Quốc: Dữ liệu phân bố như thế nào? (Kết hợp vẽ các biểu đồ). Tiền xử lý: Khang, Thọ, Quốc.	100%
Phần 1C: Đặt và trả lời câu hỏi.	Cả nhóm.	100%
Phần 2: Mô hình và đánh giá.	Thọ: LinearRegression. Khang: DecisionTree Regressor. Quốc: RandomForest. Khoa: KneighborsRegressor.	100%
Phần 3: Tổng hợp kết quả.	Cả nhóm.	100%

3 Tổng hợp kết quả:

- Thọ:
 - * Khó khăn:
 - Khó đặt được câu hỏi để đào sâu dữ liệu đã thu thập.
 - Khó thống nhất trong việc tiền xử lí dữ liệu.
 - * Học được:
 - Học được thêm nhiều câu lệnh để xử lí, vẽ biểu đồ trong các thư viện của python.
 - Đặt câu hỏi để tìm hiểu lợi ích của việc thu thập và phân tích dữ liệu từ đó đưa ra những thông tin mà ta có thể suy ra được từ dữ liệu.
- Khoa:
 - * Khó khăn:
 - Trong việc hiểu các tham số truyền vào của các mô hình.
 - * Học được:
 - Cách đặt câu hỏi và hiểu biểu đồ.
 - Cách sử dụng github hiệu quả.
- Khang:
 - * Khó khăn:
 - Khó khăn trong việc đưa ra các câu hỏi.
 - Khó khăn trong việc đánh giá mô hình.
 - Chưa có nhiều kiến thức về xe.
 - * Học được:
 - Học được thêm về python.
 - Biết được cách cơ bản để đưa ra các câu hỏi.
 - Học được các thao tác đơn giản với github.
 - Ôn lại và biết được một số kiến thức khoa học dữ liệu.

- Biết được cách cơ bản xây dựng và đánh giá một mô hình đơn giản.
- Quốc:
 - * Khó khăn:
 - Chưa có nhiều kinh nghiệm tham gia thị trường xe nên chưa hiểu nhiều dữ liệu thu thập dẫn đến khó khăn trong việc đặt câu hỏi đáp ứng yêu cầu khách hàng.
 - Khó khăn trong việc biểu diễn các biểu đồ theo ý muốn, đẹp, dễ hiểu.
 - * Học được:
 - Sử dụng các câu lệnh cơ bản trong version control (git).
 - Hiểu rõ các quá trình trong ngành khoa học dữ liệu.
 - Biết được cách thức hoạt động của k fold cross validation, và tìm hiểu cơ bản một vài giải thuật hồi quy.
- Nếu có thêm thời gian nhóm sẽ:
 - Đặt thêm các câu hỏi từ dữ liệu và trả lời.
 - Tìm hiểu sâu hơn về các mô hình.