

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN TP HỒ CHÍ MINH**  
**MÔN HỌC: KHOA HỌC DỮ LIỆU**

# **BÁO CÁO ĐỒ ÁN CUỐI KỲ**



**Nhóm 21**

1612272 - Trần Nhật Huy

1512222 - Nguyễn Duy Hưng

## I.   Hợp nhóm và phân công công việc

- Lần 1

Ngày: 24/11/2019

Địa điểm: Facebook

<b>Công việc</b>	<b>Ngày thực hiện</b>	<b>Người làm</b>
Crawl dữ liệu	26/11/2019	Nguyễn Duy Hưng
Viết báo cáo	30/11/2019	Trần Nhật Huy

- Lần 2

Ngày: 02/01/2020

Địa điểm: Facebook

<b>Công việc</b>	<b>Ngày thực hiện</b>	<b>Người làm</b>
Thống nhất kế hoạch	02/01/2020	Nguyễn Duy Hưng
làm		Trần Nhật Huy
Viết kế hoạch và chỉnh sửa	03/01/2020	Nguyễn Duy Hưng
		Trần Nhật Huy
Thí nghiệm 1	05/01/2020	Trần Nhật Huy
Thí nghiệm 2	05/01/2020	Nguyễn Duy Hưng
Thí nghiệm 3	05/01/2020	Nguyễn Duy Hưng
Thí nghiệm 4	05/01/2020	Nguyễn Duy Hưng
Thí nghiệm 5	05/01/2020	Nguyễn Duy Hưng
Thí nghiệm 6	06/01/2020	Trần Nhật Huy

Thí nghiệm 07	06/01/2020	Trần Nhật Huy
Viết báo cáo	06/01/2020	Nguyễn Duy Hưng
Slide	06/01/2020	Trần Nhật Huy

## II. Bài toán đặt ra và hướng giải quyết

### 1. Bài toán

- Nhóm muốn dự đoán độ ẩm của ngày hôm sau dựa vào dữ liệu của ngày hôm trước
- Input: Dữ liệu thời tiết của ngày hôm trước (temperature, humidity, ...)
- Output: Humidity của ngày hôm sau
- Lợi ích: Chúng ta có thể dự đoán trước được humidity của một ngày nào đó. Phục vụ cho nông nghiệp hoặc dự báo thời tiết
- Nguồn gốc: Nhóm tự đặt ra câu hỏi này

### 2. Thu thập dữ liệu

- Dữ liệu của nhóm được lấy từ API Dark Sky từ 01-01-2010 đến 31-12-2011 hành phố Hồ Chí Minh, Việt Nam

### 3. Tiền xử lý dữ liệu

- Dữ liệu có nhiễu, tiến hành gom các nhóm gần nhau và lấy trung bình
- Điền giá trị thiếu, nhóm sử dụng SimpleImputer với các strategy lần lượt là mean, median, most\_frequent

- Lựa chọn đặc trưng nhóm sử dụng PCA, SelectKBest, Pearson + với tự tạo thêm đặc trưng ở một số thí nghiệm
- Chuẩn hóa dữ liệu nhóm sử dụng StandardScalar

#### 4. Các thí nghiệm đã thực hiện

- Thí nghiệm 1

Nhóm tách thuộc tính time thành day và month. Bỏ thuộc tính visibility, apparentTemperature. Dùng PCA với n\_components từ [1:số lượng cột].

Độ lỗi MAE trên tập validation là 0.054. Không tốt vì std của cột Humidity là 0.08.

- Thí nghiệm 2

Nhóm giữ nguyên các thuộc tính ban đầu, không thêm hay xóa bất kỳ cột nào

Độ lỗi MAE trên tập validation là 0.055

- Thí nghiệm 3

Nhóm lựa chọn đặc trưng bằng Pearson's Correlation

Độ lỗi MAE trên tập validation là 0.056

- Thí nghiệm 4

Nhóm lựa chọn đặc trưng bằng SelectKBest

Độ lỗi MAE trên tập validation là 0.056

- Thí nghiệm 5

Nhóm lựa chọn đặc trưng bằng Feature Importance

Độ lỗi MAE trên tập validation là 0.057

- Thí nghiệm 6

Nhóm sử dụng mô hình Neural Network

Độ lỗi MAE trên tập validation là 0.07

Độ lỗi của mô hình này còn lớn hơn trước. Có thể bị overfitting

- Thí nghiệm 7

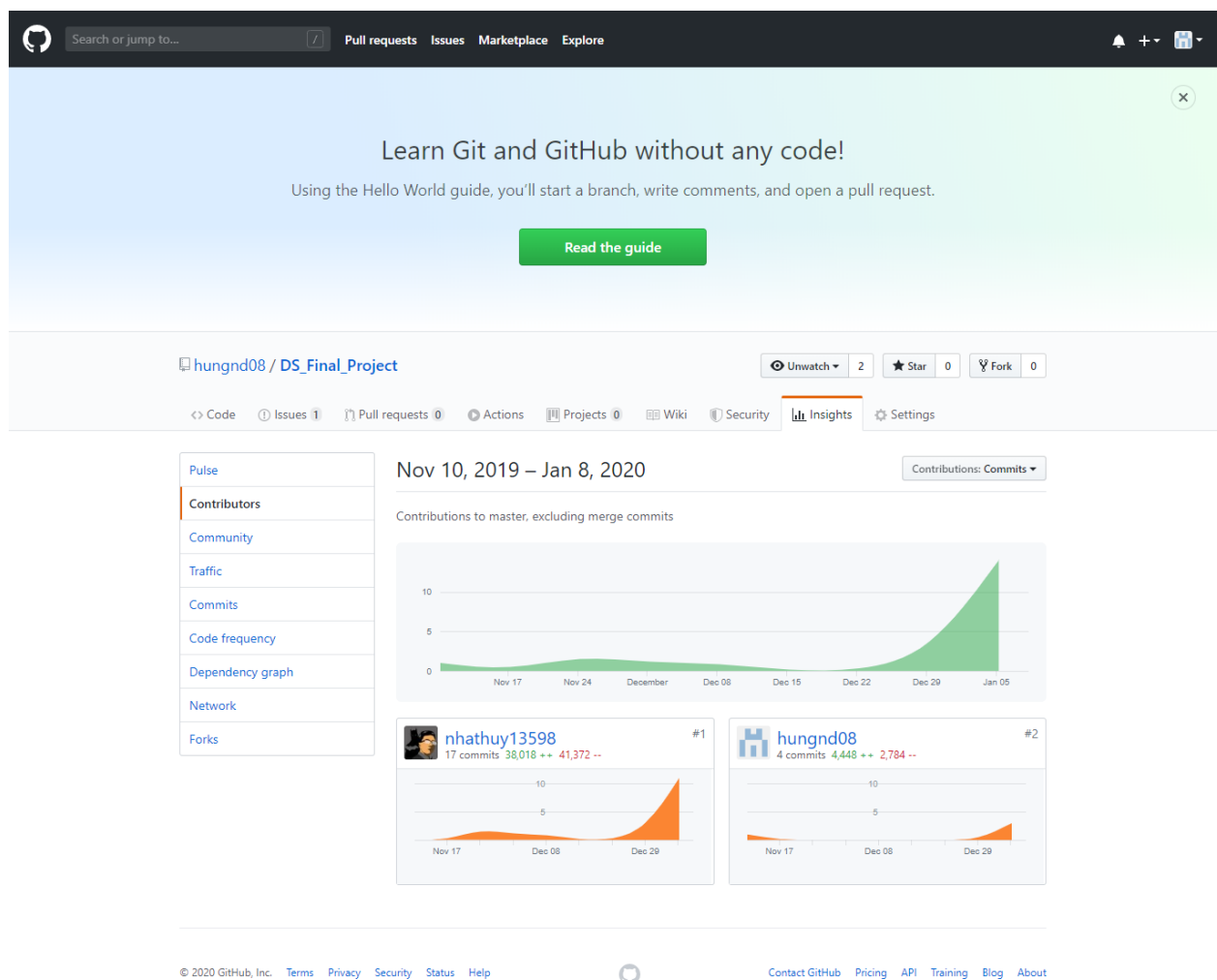
Nhóm dự đoán có thể sử dụng dữ liệu của một ngày chưa đủ nên thử thêm độ ẩm của k ngày trước đó. Nhóm chỉ lấy cột humidity còn các cột khác nhóm sẽ bỏ. Nhóm thử nghiệm với  $k = 14$  và chọn được số ngày before tốt nhất là 7

Độ lỗi MAE trên tập validation là 0.045

### III. Tổng kết và đánh giá đóng góp của các thành viên trong nhóm

- Các thí nghiệm mà nhóm thực hiện có kết quả không được tốt
- Nhóm nghĩ có thể nguyên nhân là:
  - ✦ Dữ liệu có nhiễu
  - ✦ Dữ liệu humidity bị lệch, không cân bằng
  - ✦ Mô hình không phù hợp, bị underfitting
- Link github: [https://github.com/hungnd08/DS\\_Final\\_Project](https://github.com/hungnd08/DS_Final_Project)
- Đánh giá đóng góp của các thành viên trong nhóm

Ở mục contributors account “hungnd08” bị thiếu mất 6 commit vì em kết nối với project bằng SSH key và trong file .gitconfig trên máy cấu hình name em để là “hungnd” và email của em khác với email đăng ký account trên github nên khi commit github nó không tính vào mục contributors, phía dưới em có chụp thêm phần lịch sử commit để thầy xem ạ.



Search or jump to...

Pull requestsIssuesMarketplaceExplore

Learn Git and GitHub without any code!

Using the Hello World guide, you'll start a branch, write comments, and open a pull request.

Read the guide

hungnd08 / DS\_Final\_Project

Unwatch2Star0Fork0

CodeIssuesPull requestsActionsProjectsWikiSecurityInsightsSettings

Branch: master

Commits on Jan 7, 2020

Fix methods

hungnd08 committed yesterday

8795f61

add gitignore

shelley13398 committed 2 days ago

289e42b

- Update folder

shelley13398 committed 2 days ago

a088f62

Commits on Jan 6, 2020

Fix report file

hungnd08 committed 2 days ago

Verifiedc176875

Add report file

hungnd08 committed 2 days ago

Verifieda5d80ad

- Fix Neural Network

shelley13398 committed 2 days ago

d987a9f

Merge branch 'huy'

shelley13398 committed 2 days ago

79a8731

- Delete file

shelley13398 committed 2 days ago

a913d8f

- Finish slide v1

shelley13398 committed 2 days ago

3a53946

- Update slide

shelley13398 committed 2 days ago

387c17b

- Merge huy

shelley13398 committed 2 days ago

cfa2bc9

- Add slide

shelley13398 committed 2 days ago

6a683d3

Fix selectKBest method and add comments

hungnd committed 2 days ago

4e23999

Try another methods

hungnd committed 2 days ago

7959dc7

Try another methods

hungnd committed 2 days ago

ca89a01

- Add method 3

shelley13398 committed 2 days ago

6239638

- Finish method 1

shelley13398 committed 3 days ago

a8663d2

Commits on Jan 5, 2020

- Finish first version

shelley13398 committed 3 days ago

8d58358

- Add preprocessing transformer

shelley13398 committed 3 days ago

e9d5546

Commits on Jan 4, 2020

Handling data

hungnd committed 4 days ago

a57889e

Commits on Jan 3, 2020

- Update

shelley13398 committed 6 days ago

068a093

- Update plan

shelley13398 committed 6 days ago

532142d

Commits on Dec 8, 2019

- Sua lai cau hoi ma nhom dat ra

shelley13398 committed on Dec 8, 2019

2b44368

Commits on Dec 2, 2019

- Sua lai file bao cao va lam ro cau hoi

shelley13398 committed on Dec 2, 2019

8faef3d

Commits on Nov 30, 2019

- Sua lai dia diem thu thap du lieu trong file bao cao

shelley13398 committed on Nov 30, 2019

579831c

+ Them file bao cao lan 1

shelley13398 committed on Nov 30, 2019

6ac756c

Commits on Nov 26, 2019

Merge branch "master" of github.com:hungnd08/DS\_Final\_Project

hungnd committed on Nov 26, 2019

6a3ef1c

Collect weather data

hungnd committed on Nov 26, 2019

8ae9227

Commits on Nov 14, 2019

Initial commit

hungnd08 committed on Nov 14, 2019

Verified875f178

Newer

Older

© 2020 GitHub, Inc.

TermsPrivacySecurityStatusHelp

Contact GitHubPricingAPITrainingBlogAbout