

BÁO CÁO GIỮA KỲ

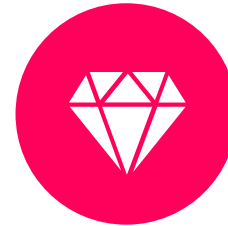
# DỰ ĐOÁN GIÁ NHÀ Ở TP. HCM

Võ Nhật Vinh - 1612815  
Nguyễn Ngọc Khải - 1612909

# DỰ ĐOÁN GIÁ NHÀ Ở TP.HCM



Vấn đề đặt ra



Thu thập dữ liệu



Xử lí dữ liệu

# VẤN ĐỀ ĐẶT RA



- Nhiều người muốn mua nhà/bán nhà nhưng không biết giá cả thị trường thế nào ?
  - Cần có mô hình dự đoán giá nhà phù hợp.
  - Tránh bị lừa đảo khi không hiểu biết về giá nhà.



# THU THẬP DỮ LIỆU

- Thu thập dữ liệu nhà ở trên trang chợ tốt.
- Dữ liệu thu thập hợp pháp.
- Dữ liệu gồm 20 cột, và 17829 dòng.

```
import urllib.robotparser
rp = urllib.robotparser.RobotFileParser()
rp.set_url('https://nha.chotot.com/tp-ho-chi-minh/mua-ban-bat-dong-san')
rp.read()
rp.can_fetch('*', 'https://nha.chotot.com/tp-ho-chi-minh/mua-ban-nha-dat')
# Kết quả sẽ là True hoặc False
```

True

	A	B	
1	link	https://nha.chotot.com//tp-ho-chi-minh/huyen-hoc-mon/mua-ban-nha-dat/64444558.htm	h
2	title	Cần Tiền Bán Gấp Trong Tháng	N
3	price (đồng)	1350000000	
4	type	Môi giới	N
5	time	11/21/2019 9:45	
6	description	Nhà đẹp cần bán gấp trả nợ, ngang 3,6m dài 12m.. gồm 3pn,1 phòng thờ,2wc, hẻm xe hơi tới	N
7	address	41/3h, Đường Trung Đông 1, Xã Thới Tam Thôn, Huyện Hóc Môn, Tp Hồ Chí Minh	3
8	area (m2)	42	
9	bedroom_num	3	
10	house_type	Nhà	N
11	toilet_num	None	N
12	direction	None	N
13	legcal_doc	None	E
14	block_name	None	N
15	total_floor	None	N
16	housing_feature	None	N
17	ward	Xã Thới Tam Thôn	P
18	district	Huyện Hóc Môn	Q
19	city	Tp Hồ Chí Minh	T

# XỬ LÝ DỮ LIỆU



- Lọc trùng.
- Chuẩn hóa.
- Xử lý dữ liệu thiếu.
- Xử lý dữ liệu nhiều.
- Tách cột.
- Label encoding.

