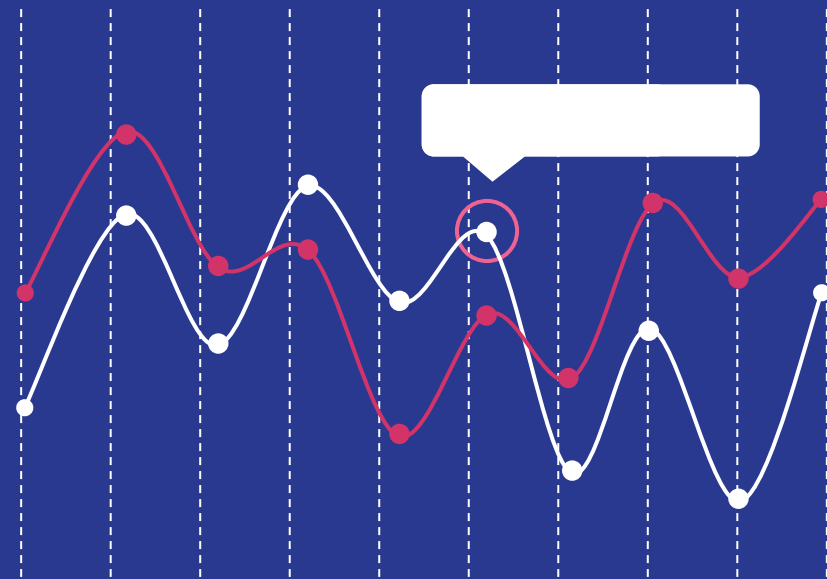


# BÁO CÁO TIẾN ĐỘ ĐỒ ÁN

KHOA HỌC DỮ LIỆU & ỨNG DỤNG



1712202 - Nguyễn Trọng Văn  
1712449 - Lê Đức Hòa

# ***NỘI DUNG TRÌNH BÀY***



# 1. THÔNG TIN ĐỀ TÀI

**TÊN:** *Phân loại nhà dựa trên đặc điểm mua bán.*

## Vấn Đề

- Dự đoán giá nhà là bài toán quen thuộc.
- Thông tin giá nhà thông qua nhiều thuộc tính: *diện tích, phòng ngủ, loại nhà,...*
- Ngược lại, người mua **có thể biết** được **loại nhà** nào có thể mua **trong khả năng**? (*số tiền có, mong muốn vị trí, chiều dài, diện tích,...*)

## Câu Hỏi

*Có thể phân loại nhà ở (mặt tiền, ngõ hẻm, biệt thự, ...) theo các đặc điểm mua bán (diện tích, vị trí, giá thành, ...) hay không?*

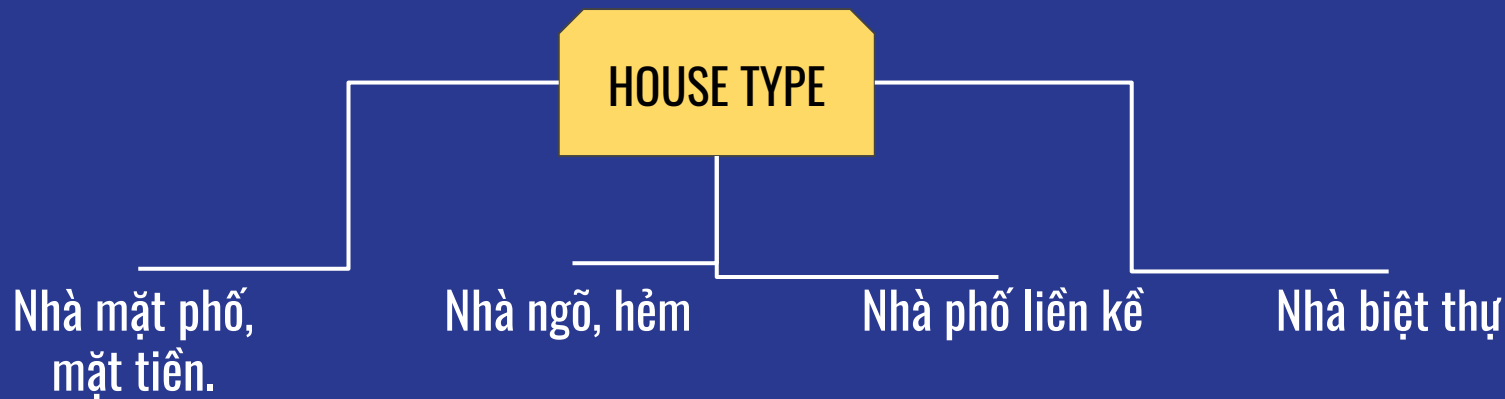
## Ý Nghĩa

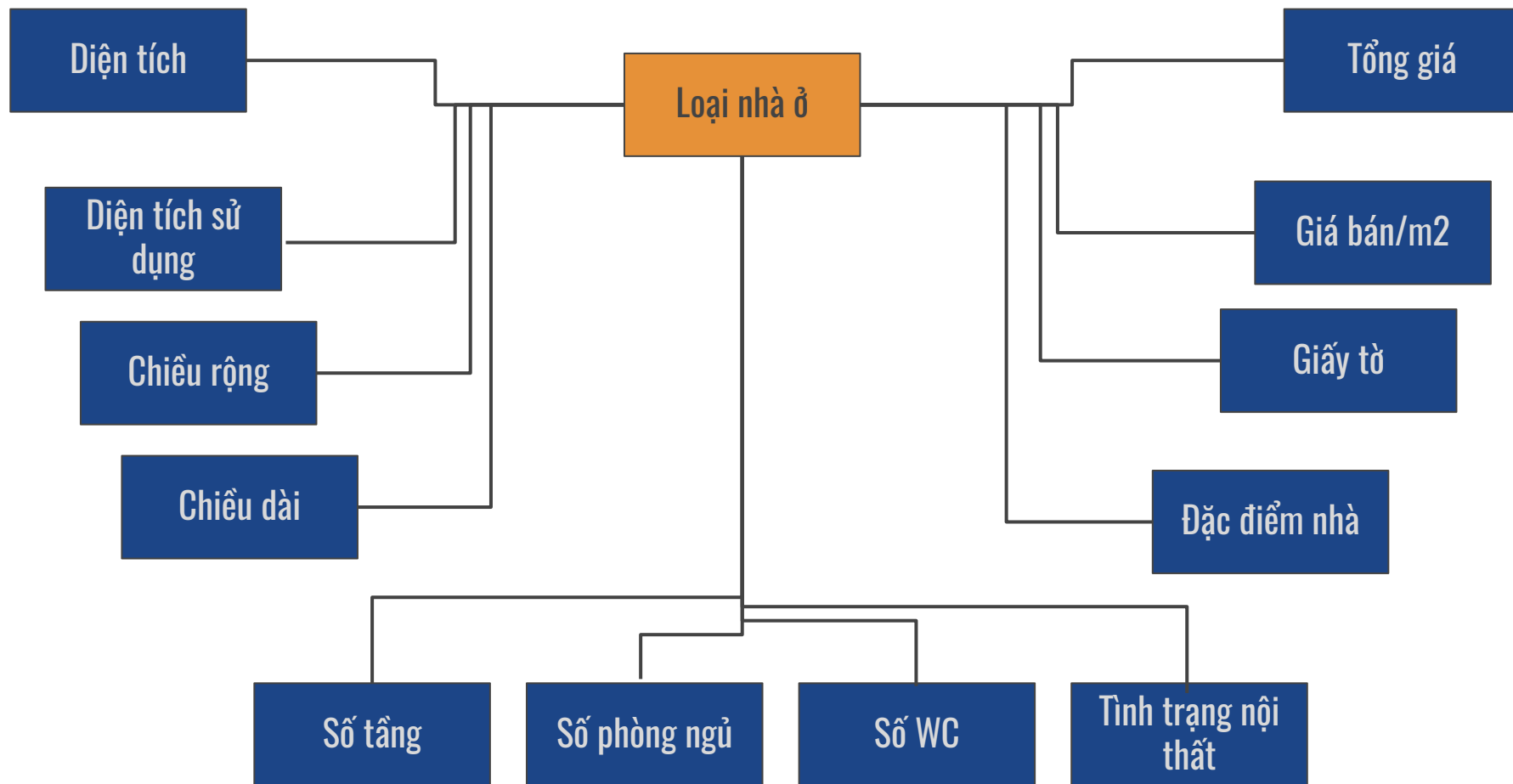
- Người mua có thể biết được loại nhà ở mình có thể giao dịch với khả năng và mong muốn hiện có.
- Hỗ trợ tư vấn mua bán giao dịch nhà đất.
- Tự động phân loại nhà ở theo đặc điểm trên các website, hệ thống mua bán BĐS.

## 2. PHƯƠNG PHÁP THỰC HIỆN

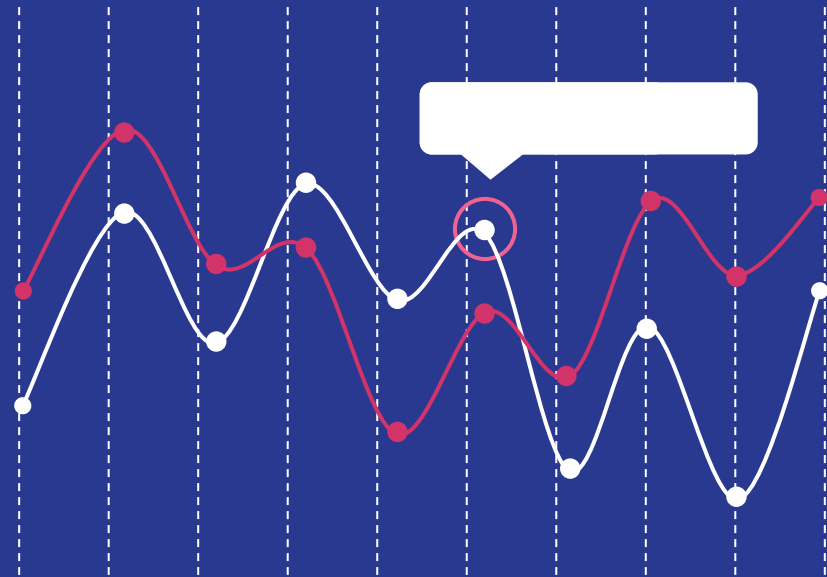
1. Dựa trên câu hỏi, tìm nguồn dữ liệu (các Website nhà đất, BĐS)
2. Thu thập dữ liệu (parse HTML).
3. Tìm hiểu thông tin dữ liệu (ý nghĩa từng trường, các giá trị trên từng thuộc tính).
4. Tiền xử lý dữ liệu (xóa trùng, chuẩn hóa dữ liệu số - phi số, lọc thuộc tính không mang ý nghĩa, điền dữ liệu thiếu, rời rạc hóa,...).
5. Trực quan hóa (xem phân bố dữ liệu, đặc điểm).
6. Lựa chọn mô hình học - Kiểm thử đánh giá mô hình.
7. Đánh giá kết quả đạt được và trả lời \_\_\_\_\_ cho câu hỏi đặt ra.

### 3. THUỘC TÍNH DỰ ĐOÁN





## 4. MÔ HÌNH DỰ ĐOÁN





**CÂY QUYẾT  
ĐỊNH**

**DECESION  
TREE.**

**01**

**03**

**SUPPORT  
VECTOR  
MACHINE**

**02**

**RANDOM  
FOREST**

*\_Tinh chỉnh tham số.  
\_Chọn mô hình cho  
kết quả tốt nhất trên  
tham số tốt nhất của  
từng mô hình.*