**TRƯỜNG ĐẠI HỌC MỎ - ĐỊA CHẤT**

**KHOA CÔNG NGHỆ THÔNG TIN**  
  
  
  
  
  
  
  
  
  
  


**ĐỀ CƯƠNG ĐỒ ÁN KHOA HỌC MÁY TÍNH**

**Tên đề tài: XÂY DỰNG MÔ HÌNH HỌC MÁY ĐỂ DỰ ĐOÁN GIÁ NHÀ**

**Sinh viên thực hiện : Nguyễn Thị Trà My**

**Lớp : Khoa học máy tính ứng dụng K66\_07D**

**Giảng viên hướng dẫn :**

**ThS.Nguyễn Thùy Dương**

**Mã sinh viên : 2121051509**

**Hà Nội – 2024**

**MỤC LỤC**

[TRƯỜNG ĐẠI HỌC MỎ - ĐỊA CHẤT 6](#_Toc190337029)

[LỜI CẢM ƠN 6](#_Toc190337030)

[CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI 6](#_Toc190337031)

[**1.1. Lý do lựa chọn đề tài** 6](#_Toc190337032)

[**1.2.** **Nội dung nghiên cứu** 6](#_Toc190337033)

[**1.3.** **Cơ sở khoa học và tính thực tiễn của đề tài** 6](#_Toc190337034)

[**1.4.** **Cấu trúc báo cáo** 6](#_Toc190337035)

[CHƯƠNG 2: CƠ SỞ LÝ THUYẾT 6](#_Toc190337036)

[**2.1. Tổng quan về dự đoán giá nhà** 6](#_Toc190337037)

[*2.1.1. Tình hình thị trường bất động sản* 6](#_Toc190337038)

[*2.1.2. Các yếu tố ảnh hưởng đến giá nhà* 6](#_Toc190337039)

[**2.2. Lý thuyết phân tích dữ liệu** 6](#_Toc190337040)

[*2.2.1. Phân loại các thuật toán học máy* 6](#_Toc190337041)

[*2.2.2. Quy trình phân tích dữ liệu* 6](#_Toc190337042)

[*2.2.3. Các kỹ thuật tiền xử lý dữ liệu* 6](#_Toc190337043)

[*2.2.4. Khám phá dữ liệu, các vấn đề cân bằng dữ liệu và đánh giá chất lượng dữ liệu.* 6](#_Toc190337044)

[**2.3. Các phương pháp và thuật toán học máy liên quan** 7](#_Toc190337045)

[*2.3.1. Kỹ thuật Linear Regression* 7](#_Toc190337046)

[*2.3.2. Kỹ thuật Decision Tree* 7](#_Toc190337047)

[*2.3.3. Kỹ thuật Random Forest* 7](#_Toc190337048)

[CHƯƠNG 3: PHƯƠNG PHÁP NGHIÊN CỨU 7](#_Toc190337049)

[**3.1. Thu thập và xử lý dữ liệu** 7](#_Toc190337050)

[*3.1.1. Nguồn dữ liệu (Zillow, Kaggle, …)* 7](#_Toc190337051)

[*3.1.2. Tiền xử lý dữ liệu.* 7](#_Toc190337052)

[**3.2. Phân tích dữ liệu ban đầu** 7](#_Toc190337053)

[*3.2.1. Phân tích thống kê mô tả* 7](#_Toc190337054)

[*3.2.2. Xác định mối quan hệ giữa các biến.* 7](#_Toc190337055)

[**3.3. Xây dựng và lựa chọn mô hình dự đoán** 7](#_Toc190337056)

[*3.3.1. Mô hình hồi quy tuyến tính* 7](#_Toc190337057)

[*3.3.2. Mô hình phi tuyến tính (Tree-based methods)* 7](#_Toc190337058)

[*3.3.3. Các bước tối ưu hóa mô hình* 7](#_Toc190337059)

[**3.4. Đánh giá hiệu quả mô hình** 8](#_Toc190337060)

[*3.4.1. Các chỉ số đánh giá (MAE, RMSE, R²)* 8](#_Toc190337061)

[*3.4.2. So sánh kết quả giữa các mô hình.* 8](#_Toc190337062)

[CHƯƠNG 4: KẾT QUẢ THỰC NGHIỆM VÀ PHÂN TÍCH 8](#_Toc190337063)

[**4.1. Mô hình dự đoán hiệu quả nhất** 8](#_Toc190337064)

[*4.1.1. Thông số tối ưu của mô hình* 8](#_Toc190337065)

[*4.1.2. Đánh giá kết quả trên tập kiểm tra* 8](#_Toc190337066)

[**4.2. Phân tích xu hướng giá nhà từ dữ liệu** 8](#_Toc190337067)

[*4.2.1. Các yếu tố ảnh hưởng mạnh nhất đến giá nhà* 8](#_Toc190337068)

[*4.2.2. Biểu đồ trực quan hóa kết quả* 8](#_Toc190337069)

[**4.3. Ứng dụng thực tiễn của mô hình** 8](#_Toc190337070)

[CHƯƠNG 5: KẾT LUẬN 8](#_Toc190337071)

[**5.1. Tóm tắt nội dung nghiên cứu** 8](#_Toc190337072)

[**5.2. Đóng góp và ý nghĩa thực tiễn** 8](#_Toc190337073)

[**5.3. Tiềm năng và hạn chế của nghiên cứu** 8](#_Toc190337074)

[PHỤ LỤC 9](#_Toc190337075)

[**6.1. Mã nguồn chương trình** 9](#_Toc190337076)

[**6.2. Các biểu đồ, bảng dữ liệu, hình ảnh chi tiết** 9](#_Toc190337077)

[**6.3. Bảng từ viết tắt** 9](#_Toc190337078)

[TÀI LIỆU THAM KHẢO 9](#_Toc190337079)

[TRƯỜNG ĐẠI HỌC MỎ - ĐỊA CHẤT 6](#_Toc190337029)

[LỜI CẢM ƠN 6](#_Toc190337030)

[CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI 6](#_Toc190337031)

[**1.1. Lý do lựa chọn đề tài** 6](#_Toc190337032)

[**1.2.** **Nội dung nghiên cứu** 6](#_Toc190337033)

[**1.3.** **Cơ sở khoa học và tính thực tiễn của đề tài** 6](#_Toc190337034)

[**1.4.** **Cấu trúc báo cáo** 6](#_Toc190337035)

[CHƯƠNG 2: CƠ SỞ LÝ THUYẾT 6](#_Toc190337036)

[**2.1. Tổng quan về dự đoán giá nhà** 6](#_Toc190337037)

[*2.1.1. Tình hình thị trường bất động sản* 6](#_Toc190337038)

[*2.1.2. Các yếu tố ảnh hưởng đến giá nhà* 6](#_Toc190337039)

[**2.2. Lý thuyết phân tích dữ liệu** 6](#_Toc190337040)

[*2.2.1. Phân loại các thuật toán học máy* 6](#_Toc190337041)

[*2.2.2. Quy trình phân tích dữ liệu* 6](#_Toc190337042)

[*2.2.3. Các kỹ thuật tiền xử lý dữ liệu* 6](#_Toc190337043)

[*2.2.4. Khám phá dữ liệu, các vấn đề cân bằng dữ liệu và đánh giá chất lượng dữ liệu.* 6](#_Toc190337044)

[**2.3. Các phương pháp và thuật toán học máy liên quan** 7](#_Toc190337045)

[*2.3.1. Kỹ thuật Linear Regression* 7](#_Toc190337046)

[*2.3.2. Kỹ thuật Decision Tree* 7](#_Toc190337047)

[*2.3.3. Kỹ thuật Random Forest* 7](#_Toc190337048)

[CHƯƠNG 3: PHƯƠNG PHÁP NGHIÊN CỨU 7](#_Toc190337049)

[**3.1. Thu thập và xử lý dữ liệu** 7](#_Toc190337050)

[*3.1.1. Nguồn dữ liệu (Zillow, Kaggle, …)* 7](#_Toc190337051)

[*3.1.2. Tiền xử lý dữ liệu.* 7](#_Toc190337052)

[**3.2. Phân tích dữ liệu ban đầu** 7](#_Toc190337053)

[*3.2.1. Phân tích thống kê mô tả* 7](#_Toc190337054)

[*3.2.2. Xác định mối quan hệ giữa các biến.* 7](#_Toc190337055)

[**3.3. Xây dựng và lựa chọn mô hình dự đoán** 7](#_Toc190337056)

[*3.3.1. Mô hình hồi quy tuyến tính* 7](#_Toc190337057)

[*3.3.2. Mô hình phi tuyến tính (Tree-based methods)* 7](#_Toc190337058)

[*3.3.3. Các bước tối ưu hóa mô hình* 7](#_Toc190337059)

[**3.4. Đánh giá hiệu quả mô hình** 8](#_Toc190337060)

[*3.4.1. Các chỉ số đánh giá (MAE, RMSE, R²)* 8](#_Toc190337061)

[*3.4.2. So sánh kết quả giữa các mô hình.* 8](#_Toc190337062)

[CHƯƠNG 4: KẾT QUẢ THỰC NGHIỆM VÀ PHÂN TÍCH 8](#_Toc190337063)

[**4.1. Mô hình dự đoán hiệu quả nhất** 8](#_Toc190337064)

[*4.1.1. Thông số tối ưu của mô hình* 8](#_Toc190337065)

[*4.1.2. Đánh giá kết quả trên tập kiểm tra* 8](#_Toc190337066)

[**4.2. Phân tích xu hướng giá nhà từ dữ liệu** 8](#_Toc190337067)

[*4.2.1. Các yếu tố ảnh hưởng mạnh nhất đến giá nhà* 8](#_Toc190337068)

[*4.2.2. Biểu đồ trực quan hóa kết quả* 8](#_Toc190337069)

[**4.3. Ứng dụng thực tiễn của mô hình** 8](#_Toc190337070)

[CHƯƠNG 5: KẾT LUẬN 8](#_Toc190337071)

[**5.1. Tóm tắt nội dung nghiên cứu** 8](#_Toc190337072)

[**5.2. Đóng góp và ý nghĩa thực tiễn** 8](#_Toc190337073)

[**5.3. Tiềm năng và hạn chế của nghiên cứu** 8](#_Toc190337074)

[PHỤ LỤC 9](#_Toc190337075)

[**6.1. Mã nguồn chương trình** 9](#_Toc190337076)

[**6.2. Các biểu đồ, bảng dữ liệu, hình ảnh chi tiết** 9](#_Toc190337077)

[**6.3. Bảng từ viết tắt** 9](#_Toc190337078)

[TÀI LIỆU THAM KHẢO 9](#_Toc190337079)