Big Data Course

Capstone Project   
Action Plan

For students (instructor’s review required)

ⓒ2023 SAMSUNG. All rights reserved.

Samsung Electronics Corporate Citizenship Office holds the copyright of this document.

This document is a literary property protected by copyright law so reprint and reproduction without permission are prohibited.

To use this document other than the curriculum of Samsung Innovation Campus, you must receive written consent from copyright holder.

| Course | Big Data Course |
| --- | --- |
| Team Name | BD3 - Nhóm 6 |
| Team Leader/  Members | Võ Minh Thông,  Ngô Phước Hậu,  Lương Minh Thông,  Trần Ngọc Phương Uyên,  Lê Nhật Vy,  Lê Dư Như Ý. |
| Project Title | Dự đoán giá nhà thành phố Hồ Chí Minh |
| Goal |  |
| Giá nhà tại các quận ở thành phố Hồ Chí Minh | |
| Abstract |  |
| Vì thị trường bất động sản hiện nay đang có sự biến động lớn về giá cả ở các thành phố lớn, đặc biệt là ở thành phố Hồ Chí Minh. Vì vậy nhóm đã quyết định xây dựng một mô hình sử dụng trí tuệ nhân tạo để dự đoán giá nhà tại các quận, huyện ở thành phố Hồ Chí Minh. Với mô hình này nhóm kỳ vọng sẽ dự đoán chính xác khoảng 70% đến 85% so với giá thực tế. Ngoài ra mô hình này có tiềm năng phát triển ở các nước khác, không chỉ ở Việt Nam. | |
| Method |  |
| * Sử dụng Selenium để thu thập dữ liệu từ website batdongsan.com.vn * Xử lý - phân tích dữ liệu. * Trực quan hoá dữ liệu. * Xây dựng mô hình AI. * Kiểm tra, chạy thử mô hình. * Kết luận. | |

| Data |  |
| --- | --- |
| Sử dụng dữ liệu chính thu thập từ website: batdongsan.com.vn | |
| Expected  Outcome |  |
| Xây dựng được mô hình dự đoán giá nhà tại thành phố Hồ Chí Minh nói riêng, và phát triển ứng dụng cho các thành phố khác trong và ngoài nước nói chung. Mô hình được kỳ vọng sẽ có độ chính xác trên 70% | |
| Role by  Member |  |
| Võ Minh Thông (trưởng nhóm): thu thập, xử lý, phân tích dữ liệu; xây dựng mô hình AI; thực hiện báo cáo.  Ngô Phước Hậu: thu thập dữ liệu, tìm hiểu các thuật toán phù hợp cho mô hình, hỗ trợ xử lý dữ liệu.  Lương Minh Thông: Tìm hiểu thuật toán KNN, nghiên cứu giao diện.  Trần Ngọc Phương Uyên: Tìm hiểu thuật toán Linear Regression,nghiên cứu giao diện.  Lê Nhật Vy: hỗ trợ thu thập dữ liệu, làm slide, tìm hiểu thuật toán Random Forest.  Lê Dư Như Ý: hỗ trợ thu thập dữ liệu, tìm hiểu thuật toán KNN, làm slide. | |

| Schedule  Summary |  |
| --- | --- |
| **Tháng 6 - Tháng 9:** Thu thập dữ liệu (có thể sử dụng script hoặc tự thu thập).   * *(1/6/2024 - 1/7/2024)*: Thu thập bộ dữ liệu hoàn chỉnh. * *(1/7/2024 - 1/8/2024):* Tiếp tục thu thập dữ liệu để bổ sung thêm vào bộ dữ liệu. * **Yêu cầu:** thu thập được bộ dữ liệu đủ lớn để máy có thể học được. * *(1/7/2024) - (7/7/2024):* Ôn tập kiến thức về python và làm việc với dữ liệu * *(8/7/2024) - (15/7/2024):* Tìm hiểu về các thuật toán phù hợp. * *(16/7/2024) - (31/7/2024):* Bắt đầu xây dựng mô hình dạy cho máy học. * **Yêu cầu:** Máy có thể nhận dữ liệu và học được. * *(1/8/2024) - (7/8/2024):* Đánh giá sơ bộ mô hình, tuỳ chỉnh lại để có thể đưa vào dự đoán và chạy thử mô hình * **Yêu cầu:** có thể dự đoán số liệu với độ chính xác từ 70% trở lên. * *(8/8/2024) - (14/8/2024):* Viết báo cáo + Slide | |
| Comment &  Assessment |  |
| <Comment and assessment **by the instructor.**> | |