

No.	Project name	Description	Student ID of your group 's leader	1st Student name, ID, email	2nd Student name, ID, email	3rd Student name, ID, email	4th Student name, ID, email	5th Student name, ID, email
1	Đào Hoài Nam	Predicate rating of house in mystay.vn website Tool : Scrapy Algorithm: Random Forest, Adaboost... Use Apache Hadoop, Apache Hbase and Apache Spark open source to storage and perform machine learning algorithm	20173271	Đào Hoài Nam 20173271 hoainam25699@gmail.com	Hoàng Thanh Lâm 20173217 hoangthanhlam@gmail.com	Lê Tuấn Linh 20142567 linhhd96@gmail.com	Đặng Minh Đức 20173037 dangminhduc99@gmail.com	Đỗ Thị Ngọc Huyền 20173188 huyendtn28@gmail.com
2	Dự đoán giá nhà đất	Project của nhóm em là thu thập dữ liệu về giá nhà đất và đặc điểm khác liên quan, từ các trang web rao bán nhà. Sau đó xử lý dữ liệu thu thập được và sử dụng để huấn luyện mô hình dự đoán giá nhà đất. Đầu vào của mô hình là đặc điểm của nhà cần bán, đầu ra mô hình là giá cả nhà mô hình dự đoán được.	20172953	Nguyen Tien Anh - 20172953 - nguyentienanh2303@gmail.com	Nguyễn Văn Tiến - 20173402 - tien.nv173402@sis.hust.edu.vn	Nguyễn Văn Hùng - 20173152 - hung.nv173152@sis.hust.edu.vn	Nguyễn Phú Vương - 20173473 - vuong.np173473@sis.hust.edu.vn	
3	Đề tài: Phân loại bài báo	BÀI toán cần thực hiện gán nhãn cho bài báo vào các chủ đề khác nhau như thời sự, thể thao, giải trí, công nghệ...	20173279	Nguyễn Tuấn Nam - 20173269 - nguyentuannam0007@gmail.com	Vũ Duy Anh - 20172951 - vuduyanh153@gmail.com	Vũ Nam Sơn - 20173348 - vunamson231@gmail.com	Nguyễn Văn Nga - 20173279 - nga.nv173279@sis.hust.edu.vn	Hoàng Công Tuệ - 20173448 - tue.hc173448@sis.hust.edu.vn
4	Classifying newspapers	Classifying articles in Vietnamese to determine their category with data which be crawled from the internet.	20173429	Tran Dinh Truong - 20173429 - truong.td173429@sis.hust.edu.vn	Phan Ba Hoang - 20173141 - hoang.pb173141@sis.hust.edu.vn	Nguyen Tien Dat - 20173012 - dat.nt173012@sis.hust.edu.vn	Nguyen Tran Hieu - hieu.nt173118@sis.hust.edu.vn - 20173118	
5	Dự đoán giá bất động sản.	Dự đoán bất động sản tại Hà Nội. Input: Mỗi bất động sản gồm:tên bất động sản, địa chỉ, giá, diện tích, tiện ích, mô tả, ngày đăng bán, số phòng ngủ, số nhà vệ sinh,... Output: Giá bất động sản. Data type: Text. Future application: sử dụng để dự đoán giá bất động sản cho các chủ bất động sản. Thuật toán sử dụng: Hồi quy. Tập dữ liệu: được crawl từ các trang web bất động sản như : chotot, batdongsan.com, batdongsan.com.vn, alonhadat,...	20173280	Dương Kiều Nga, 20173280, duongkieunga9928@gmail.com	Hoàng Thị Hào, 20173100, hao.ht173100@sis.hust.edu.vn	Nguyễn Quang Phúc , 20173304 , phuc.nq173304@sis.hust.edu.vn	T là Hoàng Thế Anh , 20172945 , anh.ht172945@sis.hust.edu.vn	
6	Phân tích cảm xúc bình luận	-Phân tích thái độ và cảm xúc của người dùng qua các bình luận(tích cực hoặc tiêu cực) -Input: Câu comment. -Output: Dự đoán bình luận tích cực hay tiêu cực. -Dataset:Nhóm sẽ crawl dữ liệu từ các bình luận trên các trang bán hàng.	20173453	Ma Việt Tùng- 20173453-tung.mv173453@sis.hust.edu.vn	Vũ Tiến Quang- 20163335- quang.vt163335@sis.hust.edu.vn	Nguyễn Văn Cao-20183485- CAO.NV183485@sis.hust.edu.vn	Mai Xuân Thắng- 20173369- thang.mx173369@sis.hust.edu.vn	

7	Trích xuất thông tin nhà đất	<ul style="list-style-type: none">+ Mục đích: trích xuất thông tin nhà đất (giá cả, khu vực, diện tích, mặt tiền, phong thủy, ...) từ đó đưa ra thông báo cho người dùng, nhận định về thị trường bất động sản tại Việt Nam.+ Đầu vào: một bài đăng mua/bán/thuê nhà+ Đầu ra: các trường thông tin cơ bản về ngôi nhà. Từ đó đưa ra thông báo cho người mua/bán dựa trên tiêu chí của họ+ Dạng dữ liệu: text+ Ứng dụng tương lai: đưa ra thông báo, gợi ý cho người dùng, tích hợp cho các trang web bất động sản ở VN+ Thuật toán: Diet-Classifier, CRF, MLP (có thể thử nghiệm thêm trong quá trình làm)+ Công cụ sử dụng: Scrapy để crawl, RASA+ Tập dữ liệu sử dụng: crawl từ facebook (nguồn chính) các bài đăng, có thể mở rộng thêm ở trang web khác	20173226	Dang Trung Duc Anh - 20172939 - anh.dtd172939@sis.hust.edu.vn	Le Duc Cuong - 20172990 - cuong.ld172990@sis.hust.edu.vn	Tran Tung Lam - 20173226 - lam.tt173226@sis.hust.edu.vn	Nguyen Van Tien - 20173401 - tien.nv173401@sis.hust.edu.vn		
8	CRAWLING DATA FROM IMDB AND PREDICTION OF MOVIE'S RATING	<ul style="list-style-type: none">* PROBLEM:<ul style="list-style-type: none">- Data is the new oil.- With internet use growing, there is a massive amount of data on different websites.<ul style="list-style-type: none">- Web scraping is a way to grab data from websites.- Scrapy makes it easy for us to quickly prototype and develop web scrapers.- Study to build a system that can make accurate prediction about the average rating for a movie, using some descriptions about the movie.<ul style="list-style-type: none">* INPUT:<ul style="list-style-type: none">- Some descriptions about the movie.* OUTPUT:<ul style="list-style-type: none">- Average rating from users for a given movie.	20173290	Lê Công Nguyên, 20173290, nguyen.lc173290@sis.hust.edu.vn	Mai Văn Hòa, 20173122, hoa.mv173122@sis.hust.edu.vn	Phạm Viết Bằng, 20172965, bang.pv172965@sis.hust.edu.vn	Kiều Minh Hiếu, 20173111, hieu.km173111@sis.hust.edu.vn	Lê Trung Hoàng Long, 20173242, long.lth173242@sis.hust.edu.vn	
9	CRAWLING DATA FROM IMDB AND PREDICTION OF MOVIE'S RATING	<ul style="list-style-type: none">* PROBLEM:<ul style="list-style-type: none">- Data is the new oil.- With internet use growing, there is a massive amount of data on different websites.<ul style="list-style-type: none">- Web scraping is a way to grab data from websites.- Scrapy makes it easy for us to quickly prototype and develop web scrapers.- Study to build a system that can make accurate prediction about the average rating for a movie, using some descriptions about the movie.<ul style="list-style-type: none">* INPUT:<ul style="list-style-type: none">- Some descriptions about the movie.* OUTPUT:<ul style="list-style-type: none">- Average rating from users for a given movie.	20173290	Lê Công Nguyên, 20173290, nguyen.lc173290@sis.hust.edu.vn	Mai Văn Hòa, 20173122, hoa.mv173122@sis.hust.edu.vn	Phạm Viết Bằng, 20172965, bang.pv172965@sis.hust.edu.vn	Kiều Minh Hiếu, 20173111, hieu.km173111@sis.hust.edu.vn	Lê Trung Hoàng Long, 20173242, long.lth173242@sis.hust.edu.vn	

10	Dự đoán giá phòng trọ	<p>1. Mô tả: Từ một số thông tin của phòng trọ (ví dụ: diện tích, vị trí, giá điện nước,...) đưa ra giá phòng sát với giá thị trường nhất dựa trên danh sách các phòng trọ đang được cho thuê.</p> <p>2. Ứng dụng thực tiễn: Việc xây dựng được mô hình dự đoán tốt sẽ giúp người cho thuê trọ có cơ sở đưa ra giá phòng phù hợp. Bên cạnh đó, người thuê cũng có góc nhìn tốt hơn về giá phòng, ước lượng tốt hơn về các tiện ích tương ứng với giá phòng cần tìm.</p> <p>3. Input, output: Input: Tập các thuộc tính của phòng trọ, vd: diện tích, vị trí, tiện ích, ... Output: Giá dự đoán của phòng trọ tương ứng.</p> <p>4. Dữ liệu: - Thu thập dữ liệu từ website: https://www.ohanaliving.vn/#/. - Tiến hành crawl thuộc tính, giá của các phòng trọ và đưa về dưới dạng file csv. Mỗi hàng trong file csv thể hiện 1 phòng trọ, mỗi cột sẽ biểu diễn các thuộc tính hoặc giá của phòng trọ đó. - Dữ liệu có thể chứa cả kiểu số (giá, ...) và chữ, kí tự (địa chỉ, ...).</p> <p>5. Phương pháp đề xuất: - Hồi quy tuyến tính</p>	20170058	Thái Văn Chiến - 20170042	Trần Công Đạo - 20173000	Nguyễn Thanh Đức - 20170058 - ducnt9907@gmail.com	Nguyễn Nhật Hoàng - 20173138	
11	Bài toán dự đoán giá ô tô	<p>Bài toán dự đoán giá ô tô</p> <p>Input Đầu vào là 1 tập các thuộc tính về ô tô như: -hãng -giá tiền -năm sản xuất -nhiên liệu -số km đã đi -số chỗ -ghế -xuất xứ -tình trạng -số cửa -hộp số -màu xe -màu nội thất -kiểu dáng -dòng xe -Tên xe -địa chỉ</p> <p>Output Giá dự đoán của ô tô phù hợp với giá thị trường.</p>	20173275	Cao Hải Nam - 20173275 - nam.ch173275@sis.hust.edu.vn	Nguyễn Quỳnh Lộc - 20173238 - loc.nq173238@sis.hust.edu.vn	Trần Hoài Nhật - 20173293 - nhath173293@sis.hust.edu.vn	Đặng Văn Nam - 20173268 - nam.ch173268@sis.hust.edu.vn	Trịnh Hoàng Hiệp - 20173105 - hiepth173105@sis.hust.edu.vn
12	Brand logo detection and recognition	We build a system which takes image input and returns the brand name and brand logo's position in the image. Our system combines two module: detection and recognition. Module detection finds the bounding boxes of the logo instances and then module recognition finds the brand that owns this logos. Data is crawled and labeled by us from websites.	20173042	Nguyễn Sỹ Đức - 20173042 - duc.ns173042@sis.hust.edu.vn	Vũ Hồng Phúc - 20173305 - phuc.vh173305@sis.hust.edu.vn	Nguyễn Việt Hoài - 20173124 - hoai.nv173124@sis.hust.edu.vn	Nguyễn Hồng Sơn - 20173349 - son.nh173349@sis.hust.edu.vn	
13	Phân cụm báo chí từ đó xây dựng hệ tìm kiếm báo chí.	Data bao gồm các bài báo + tên bài báo tương ứng. Phân cụm các bài báo = việc phân cụm tên bài báo (KNN, HAT...). Học 1 encoder từ nội dung bài báo sang vector tên bài báo tương ứng. Với bộ encoder đã học bên trên này thì encode search string của user thành vector tương ứng. Tìm min khoảng cách của vector đầu với các cụm. Cụm nào gần nhất thì khả năng cao search string nói về các bài báo trong cụm đấy.	20173274	Nguyễn Hoàng Nam - 20173274 - nam.nh173274@sis.hust.edu.vn	Nguyễn Trí Quân, - 20173312 - quan.nt173312@sis.hust.edu.vn	Vũ Tùng Lâm - 20173221 - lam.vt173221@sis.hust.edu.vn	Trần Văn Kiên - 20173208 - kien.tv173208@sis.hust.edu.vn	

14	Phân tích dữ liệu thô từ các trang thương mại điện tử để gợi ý sản phẩm cho khách hàng	<ul style="list-style-type: none"> - Input: Dữ liệu thô được crawl từ các trang thương mại điện tử gồm có thông tin về các sản phẩm (giá, loại sản phẩm, nhà cung cấp,...), và đánh giá của người đã mua sản phẩm đó. - Output: gợi ý cho người dùng những sản phẩm mà họ có thể thích. - Ứng dụng của project này có thể giúp các trang thương mại điện tử gợi ý sản phẩm cho khách hàng tốt hơn và tăng doanh thu. - Về công nghệ sử dụng trong đề tài này, nhóm em dự kiến gồm có : + Spark để xử lý dữ liệu + Các phương pháp gợi ý sản phẩm dựa trên tương tác user-item. - Về dữ liệu sử dụng : nhóm em sử dụng dữ liệu thô tự crawl từ trang thương mại điện tử (ví dụ như "tiki.vn") . Dữ liệu thô bao gồm thông tin về các sản phẩm và rating của người dùng về sản phẩm 	20173393	Nguyen Hoang Thuan - 20173393 - thuan.nh173393@sis.hust.edu.vn	Nguyen The Duc - 20170057 - duc.nt170057@sis.hust.edu.vn	Nguyen Minh Hieu - 20173115 - hieu.nm173115@sis.hust.edu.vn	Vu Quang Huy - 20173178 - huy.vq173178@sis.hust.edu.vn	Nguyen Ky Tung - 20173455 - tung.nk173455@sis.hust.edu.vn
15	Dự đoán điểm đánh giá của 1 app trên google store	<ul style="list-style-type: none"> -Xây dựng mô hình dự đoán điểm đánh giá của một app trên google store dựa các thuộc tính thu thập được trên google store . -Input : Một số thuộc tính của app -Output: Điểm đánh giá app -Datasets : Tập dữ liệu thu thập từ google store -Algorithms : Sử dụng thuật toán hồi quy 	20173405	Đoàn Tiến Huy Hoàng - 20173130 - hoang.dth173130@sis.hust.edu.vn	Đoàn Văn Lợi - 20173241 - loi.dv173241@sis.hust.edu.vn	Phạm Thế Tài - 20173351 - tai.pt173351@sis.hust.edu.vn	Nguyễn Văn Toàn - 20173405 - toan.nv173405@sis.hust.edu.vn	
16	Dự đoán giá chứng khoán	Dự đoán giá chứng khoán	20173120	Nguyễn Văn Hiếu, 20173120, hieu.nv173120@sis.hust.edu.vn	Vũ Khắc Chinh, 20172979, chinh.vk172979@sis.hust.edu.vn	Đặng Thái Hùng, 20173156, hung.dt173156@sis.hust.edu.vn	Phạm Trí Ninh, 20173295, ninh.pt173295@sis.hust.edu.vn	
17	Hanoi's real estate	<p>In this project, we want to build a price suggestion application. Where the user can enter information about the apartment they want in order to get the suggested price. Information may include: location, area, balcony direction, number of bedrooms, number of bathrooms, investor, time of purchase, etc. Then they will receive a real number, which is the predicted price for that house. This application will provide useful reference information for buyers, sellers and investors when participating in real estate transactions.</p> <p>About the datasets: We do not use the existing datasets, instead, we collect data from real estate websites, as well as real estate forums on social networks. The real estate objects here are apartments in the inner city of Hanoi.</p> <p>Technologies that our team has been and intend to use in our work include: collecting and extract information with request library, scrapy, beautifulsoup, selenium, ... Data analyzing with pandas, seaborn, tableau, ... and using spatial regression models in price prediction suggested by Mr. Tran Viet Trung.</p>	20173333	Vũ Trí An - 20172933 - an.vt172933@sis.hust.edu.vn	Trần Thị Dinh - 20173015 - dinh.tt173015@sis.hust.edu.vn	Nguyễn Văn Đức - 20173033 - duc.nv173033@sis.hust.edu.vn	Trần Thị Quyên - 20173333 - quyên.tt173333@sis.hust.edu.vn	

18	Hanoi's real estate	<p>In this project, we want to build a price suggestion application. Where the user can enter information about the apartment they want in order to get the suggested price. Information may include: location, area, balcony direction, number of bedrooms, number of bathrooms, investor, time of purchase, etc. Then they will receive a real number, which is the predicted price for that house. This application will provide useful reference information for buyers, sellers and investors when participating in real estate transactions.</p> <p>About the datasets: We do not use the existing datasets, instead, we collect data from real estate websites, as well as real estate forums on social networks. The real estate objects here are apartments in the inner city of Hanoi.</p> <p>Technologies that our team has been and intend to use in our work include: collecting and extract information with request library, scrapy, beautifulsoup, selenium, ... Data analyzing with pandas, seaborn, tableau, ... and using spatial regression models in price prediction suggested by Mr. Tran Viet Trung.</p>	20173333	Vu Tri An - 20172933 - an.vt172933@sis.hust.edu.vn	Tran Thi Dinh - 20173015 - dinh.tt173015@sis.hust.edu.vn	Nguyen Van Duc - 20173033 - duc.nv173033@sis.hust.edu.vn	Tran Thi Quyen - 20173333 - quyen.tt173333@sis.hust.edu.vn	
19	Building a movie recommendation system	<p>https://drive.google.com/file/d/1N6VfwAf_UJnH0iclp8-aUUd99Yk2Dhgm/view?usp=sharing</p>	20172998	Dang Nguyen Minh - 20172998 - dang.nm172998@sis.hust.edu.vn	Quan Vo Duc - 20173320 - quan.vd173320@sis.hust.edu.vn	Luong Nguyen Van - 20173249 - luong.nv173249@sis.hust.edu.vn	Dung Bui Viet - 20173045 - dung.bv173045@sis.hust.edu.vn	Anh Nguyen Duc - 20172937 - anh.nd172937@sis.hust.edu.vn
20	Gợi ý tin tức	<p>- Tóm tắt: Gợi ý tin tức cho người đọc trên trang báo online. Đầu vào: tập người dùng, tập bài báo, các thông tin của bài báo, lịch sử đọc báo của người dùng. Đầu ra: K bài báo có thể phù hợp với người dùng. Kịch bản: Trên các trang báo có một vùng sidebar, dùng để hiển thị các dòng tin tức được gợi, tạm gọi là vùng gợi ý. Khi người dùng xem một bài báo, hệ thống sẽ chọn K bài báo và hiển thị ở vùng gợi ý để người dùng click vào xem nếu họ muốn. Mục đích: Đưa ra các gợi ý phù hợp nhất với người dùng để họ chọn tìm được thông tin phù hợp với sở thích và mong muốn của mình. - Đề xuất chi tiết: https://husteduvn-my.sharepoint.com/:w:/g/personal/hoang_nv173142_sis_hust_edu_vn/ETgPeymzVmhPoBN6tG_dWpoBiHUYxQ9tGL8sRbHvcnOJgQ?e=Gc9eii</p>	20173142	Ngô Việt Hoàng - 20173142 - hoang.nv173142@sis.hust.edu.vn	Nguyễn Thị Nhung - 20173294 - nhung.nt173294@sis.hust.edu.vn	Đàm Trọng Tuyền - 20173463 - tuyen.dt173463@sis.hust.edu.vn		
21	Rate hotel	A app for rate hotels	20173011	Nguyen Dinh Dat - 20173011 - dat.nd173011@sis.hust.edu.vn	mung.vt173261@sis.hust.edu.vn	hai.db173095@sis.hust.edu.vn	thuy.tt173394@sis.hust.edu.vn	

22	Xây dựng mô hình học máy Dự đoán giá chứng khoán	https://drive.google.com/file/d/1OuUq2jXc5DDbcY7p38Ktjt6cvZhgdI-8/view?usp=sharing	20173062	Nguyễn Bình Dương - 20173062 - duong.nb173062@sis.hust.edu.vn	Vũ Ngọc Hiền - 20173103 - hien.vn173103@sis.hust.edu.vn	Lê Văn Linh - 20173235 - linh.lv173235@sis.hust.edu.vn	Nghiêm Văn Nghĩa - 20173283 - nghia.nv173283@sis.hust.edu.vn	Lê Thế Tài - 20174180 - tai.lt174180@sis.hust.edu.vn
23	Sử dụng mô hình học sâu và biểu đồ nền Nhật để dự đoán sự dịch chuyển của cổ phiếu	Thực hiện dự đoán sự tăng giảm cổ phiếu của Apple và Facebook dựa trên dữ liệu chứng khoán của 100 công ty top đầu sàn giao dịch chứng khoán Nasdaq. Project thực hiện thu thập dữ liệu, phân tích dữ liệu và áp dụng các mô hình học sâu và đánh giá thực nghiệm mô hình.	20173084	Van Giang Nguyen - 20173084 - giang.nv173084@sis.hust.edu.vn	Ngoc Minh Phung - 20173257 - minh.pn173257@sis.hust.edu.vn	Tung Lam Le - 20173218 - lam.lt173218@sis.hust.edu.vn		
24	Phân loại văn bản theo chủ đề.	Thu thập và phân cụm các văn bản theo các chủ đề được định sẵn. Dữ liệu đầu vào: văn bản được thu thập từ các trang báo điện tử. Dữ liệu đầu ra: văn bản đã được phân loại vào các chủ đề. Kiểu dữ liệu: văn bản, dạng text. Thuật toán dự kiến: Naive Bayes, Support Vector Machine. Công cụ dự kiến: Python, sklearn, nltk, hadoop.	20170115	Lê Anh Thành - 20170115 - thanh.la170115@sis.hust.edu.vn	Lê Hải Nam - 20173264 - nam.lh173264@sis.hust.edu.vn	Trần Tất Đắc - 20172992 - dac.tt172992@sis.hust.edu.vn		
25	IMDb rating prediction	Dự đoán rating những bộ phim trên IMDb dựa vào những tiêu chí như lượt vote, giải thưởng, rated, thể loại, đạo diễn, diễn viên chính, biên kịch,... Dữ liệu được crawl từ imdb.com (qua API hỗ trợ bởi omdbapi.com).	20172948	Hà Hải Phong - 20173299 - phong.hh173299@sis.hust.edu.vn	Trần Tiến Anh - 20172948 - anh.tt172948@sis.hust.edu.vn	Vũ Quang Đại - 20172993 - dai.vq172993@sis.hust.edu.vn	Nguyễn Văn Long - 20173244 - long.nv173244@sis.hust.edu.vn	
26	Prediction of hotel's rating	dự đoán số sao của khách sạn dựa trên bộ dữ liệu	20173302	Nguyễn Quý Phúc, 20173302, phuc.nq173302@sis.hust.edu.vn	Đào Minh Tuấn, 20173437, tuan.dm173437@sis.hust.edu.vn	Đào Minh Đức, 20173029, duc.dm173029@sis.hust.edu.vn	Đồng Văn Hiệp, 20173104, hiệp.dv173104@sis.hust.edu.vn	Nguyễn Thế Tùng Dương, 20173060, duong.ntt173063@sis.hust.edu.vn
27	Movie Recommendation System	https://docs.google.com/document/d/1HAsS_MuQ6Gk0hKsqU59RPIh-JJGZdWz4YXDQ9Y_ckA8/edit?fbclid=IwAR2KM9795eyMYKlZtCtHbPYvYHcbi86pRIKcquDz2TYcrYIGehMch7vB0D0	20172983	Phạm Văn Chung - 20172983 - chung.pv172983@sis.hust.edu.vn	Nguyễn Đình Mạnh - 20173255 - manh.nd173255@sis.hust.edu.vn	Hoàng Mạnh Hùng - 20173155 - hung.hm173155@sis.hust.edu.vn		
28	Dự đoán giá nhà đất	do em không quen ai nên không tìm được nhóm, em sẵn sàng làm nhóm 1 hoặc open cho các bạn sau với 1 điểm số thấp hơn	20173325	Đỗ Duy Quang- 20173325- quang.doduy99@gmail.com	Đỗ Duy Quang- 20173325- quang.doduy99@gmail.com	Đỗ Duy Quang- 20173325- quang.doduy99@gmail.com		