

DANH SÁCH ĐỀ TÀI ĐỒ ÁN CHUYÊN NGÀNH NGÀNH KHDL (2TC) HỌC KỲ II NĂM HỌC 2025-2026

STT ĐỀ TÀI	TÊN ĐỀ TÀI	NỘI DUNG YÊU CẦU	GIÁO VIÊN HƯỚNG DẪN	EMAIL	BỘ MÔN	Định hướng đề tài (Ứng dụng, Nghiên cứu, Kết hợp, MMT)	GHI CHÚ
1	Hệ thống phát hiện tin giả và có khả năng giải thích	+ Tìm hiểu hiện trạng ứng dụng đồ thị tri thức (Knowledge Graph) và mô hình ngôn ngữ lớn (Large Language Models – LLMs) trong xây dựng các hệ thống phát hiện và kiểm chứng tin giả. Đánh giá khả năng áp dụng các phương pháp này vào bài toán phát hiện tin giả tiếng Việt có khả năng giải thích, từ đó xác định tính khả thi và tiềm năng ứng dụng thực tế. + Nghiên cứu các mô hình và phương pháp tiếp cận hiện đại như Knowledge Graph, Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Explainable AI, ... Tìm hiểu ưu điểm và nhược điểm của từng phương pháp để lựa chọn giải pháp phù hợp cho bài toán phát hiện tin giả tiếng Việt kèm theo lời giải thích và bằng chứng kiểm chứng. + Xây dựng kiến trúc hệ thống phát hiện tin giả tiếng Việt có khả năng giải thích dựa trên mô hình ngôn ngữ lớn kết hợp truy hồi tri thức, bao gồm quy trình thu thập – tiền xử lý dữ liệu, tổ chức và truy hồi tri thức, suy luận và sinh lời giải thích, đánh giá và triển khai hệ thống. Đề xuất giải pháp và mô hình nhằm nâng cao độ chính xác, tính minh bạch và khả năng giải thích của kết quả phát hiện tin giả. + Thu thập và chuẩn hóa dữ liệu tin tức tiếng Việt từ nguồn chính thống và các bộ dữ liệu công khai về tin giả. Triển khai cơ chế truy hồi tri thức (sử dụng cơ sở dữ liệu văn bản hoặc đồ thị tri thức) để hỗ trợ mô hình ngôn ngữ lớn trong quá trình kiểm chứng thông tin. + Thử nghiệm giao diện người dùng thân thiện, cho phép nhập nội dung tin cần kiểm tra, hiển thị kết quả phát hiện tin giả kèm theo lời giải thích và nguồn đối chiếu + Tổng hợp kết quả nghiên cứu, đánh giá hiệu suất của hệ thống phát hiện tin giả và chất lượng lời giải thích, so sánh với các phương pháp truyền thống; viết báo cáo chi tiết và hoàn chỉnh, trình bày toàn bộ quá trình thực hiện và kết quả đạt được của đề tài.	Huỳnh Thị Cẩm Dung	dunghtc@huit.edu.vn	KTPM	Nghiên cứu	
2	Phát hiện lỗi cú pháp và lỗi logic cơ bản trong mã code bằng mô hình LLM	Mục tiêu là ứng dụng LLM để tự động phát hiện lỗi trong đoạn code ngắn (<30 dòng lệnh) Yêu cầu cụ thể: - SV tự chọn ngôn ngữ lập trình cho mã code cần thực nghiệm. Chọn 1 trong các NNLT sau: python, C/C++, C#, Java... - Tìm hiểu các mô hình LLM hỗ trợ phát hiện lỗi trong code - Khảo sát các nghiên cứu liên quan bài toán - Thu thập tập dữ liệu code có lỗi và code đúng (có bộ dataset chuẩn, các nguồn github,...) - Tiền xử lý mã nguồn (tokenization, normalization) - Sử dụng LLM để phân loại: code lỗi / không lỗi - Phạm vi giới hạn lỗi code: + Lỗi cú pháp (Syntax Errors) • Thiếu dấu đóng/mở (, , { , } , ; , ...) • Sai cấu trúc câu lệnh • Lỗi khai báo biến, hàm + Một số lỗi logic cơ bản (Simple Logical Errors) – chọn tối đa 1–2 loại: • Sai điều kiện trong câu lệnh rẽ nhánh (if, else) • Điều kiện dừng vòng lặp không hợp lý (dẫn đến vòng lặp vô hạn) • So sánh sai toán tử (= vs == trong một số ngôn ngữ) - Đánh giá độ chính xác của mô hình dùng các metric: Precision, Recall, F1-score. - Xây dựng baseline để so sánh, nhận xét : • Rule-based syntax checker (compiler/interpreter, regex, static rules) • Hoặc mô hình ML truyền thống: Logistic Regression, SVM, Naive Bayes với đặc trưng token - Demo kết quả trên 1 nền tảng cơ bản (web,...)	Nguyễn Thị Bích Ngân	nganntb@huit.edu.vn	KTPM	Nghiên cứu	

3	Nghiên cứu và ứng dụng các mô hình tạo sinh trong bài toán phát hiện cộng đồng trên mạng xã hội	Mục tiêu: Ứng dụng các mô hình tạo sinh (Generative Models) nhằm nâng cao hiệu quả phát hiện cộng đồng trong mạng xã hội thông qua việc học biểu diễn dữ liệu đồ thị và khai thác cấu trúc ẩn của mạng. Yêu cầu cụ thể: 1. Tìm hiểu bài toán phát hiện cộng đồng (Community Detection) trên dữ liệu mạng xã hội dạng đồ thị. - Khảo sát một số phương pháp cơ bản và học sâu, bao gồm: Louvain, Spectral Clustering, Graph Autoencoder, Variational Autoencoder (VAE) Variational Graph Autoencoder (VGAE) 2. Thu thập và sử dụng 1–2 bộ dữ liệu mạng xã hội công khai (ưu tiên Karate Club và Cora hoặc Facebook). 2.1 Tiền xử lý dữ liệu đồ thị: Chuẩn hóa adjacency matrix, Chuẩn hóa node features, Xử lý graph sparse cơ bản 2.2 Triển khai các mô hình chính: VAE cho học biểu diễn đặc trưng, VGAE cho học biểu diễn cấu trúc đồ thị và kết hợp thuật toán phân cụm (K-Means hoặc Spectral Clustering) để phát hiện cộng đồng. 2.3 So sánh kết quả với các phương pháp baseline: Louvain, Spectral Clustering 2.4 Đánh giá hiệu quả mô hình bằng các chỉ số chính: NMI, ARI, Modularity Score 3. Trình quan hóa kết quả nhận cụm và xây dựng demo đơn giản bằng Streamlit hoặc Flask hoặc tùy	Bùi Công Danh		KTPM	Nghiên cứu	
4	Tối ưu hóa tham số mô hình học máy để phân loại bệnh ung thư dựa trên thuật toán tiến hóa vi phân	Mục tiêu: nghiên cứu và ứng dụng thuật toán tiến hóa vi phân để tối ưu hóa các siêu tham số của mô hình học máy, qua đó nâng cao hiệu quả phân loại bệnh ung thư. • Khảo sát và hiểu rõ nguyên lý hoạt động của DE và vai trò của nó trong tối ưu toàn cục; phân tích đặc trưng, ưu điểm và hạn chế của DE. • Ứng dụng DE để tìm bộ tham số tối ưu cho các mô hình học máy (đặc biệt là SVM, Decision Tree và Random Forest) trên các bộ dữ liệu tiêu biểu: Breast Cancer Wisconsin (Diagnostic) và Gene Expression Cancer RNA-Seq. • Đánh giá và so sánh hiệu quả mô hình tối ưu bằng DE với các phương pháp tối ưu truyền thống như Grid Search, Random Search. • Trình bày và trực quan hóa kết quả thông qua biểu đồ và giao diện minh họa bằng công cụ Web/Streamlit /Flask.	Đinh Thị Tâm		KTPM	Nghiên cứu	
5	Nghiên cứu thuật toán khai thác chuỗi clickstream phổ biến có trọng số	Mục tiêu: - Khảo sát và tìm hiểu bài toán khai thác mẫu clickstream phổ biến có trọng số - Tìm hiểu bài toán khai thác chuỗi tuần tự và các tính chất liên quan - Tìm hiểu thuật toán CM-WSPADE và Compact-SPADE - Cài đặt thực nghiệm và đánh giá - xây dựng ứng dụng demo minh họa kết quả của bài toán trên ứng dụng web - Bộ dữ liệu sử dụng là; Chess, Retail	Vũ Văn Vinh	vinhvv@huit.edu.vn	KTPM	Nghiên cứu	

6	Hệ thống phát hiện cháy và khói từ ảnh camera giám sát có giám cảnh báo giả	+ Tìm hiểu hiện trạng và nhu cầu ứng dụng AI trong phát hiện cháy/khói từ camera giám sát nhằm cảnh báo sớm. Khảo sát các hướng tiếp cận truyền thống (xử lý ảnh, đặc trưng màu/chuyển động) và hướng học sâu (object detection/segmentation) để đánh giá tính khả thi. + Khảo sát các mô hình phát hiện đối tượng hiện đại phù hợp cho bài toán cháy/khói như YOLOv5/YOLOv8, Faster R-CNN, EfficientDet. Phân tích ưu/nhược điểm (độ chính xác, tốc độ suy luận) để lựa chọn mô hình baseline. + Thu thập và chuẩn hóa dữ liệu ảnh (và/hoặc trích frame từ video) từ các bộ dữ liệu công khai về fire/smoke; có thể bổ sung một tập dữ liệu tự thu thập. Thực hiện gán nhãn (bounding box) cho 2 lớp tối thiểu: fire, smoke. + Tiền xử lý và tăng cường dữ liệu: chuẩn hóa kích thước ảnh, cân bằng lớp, augmentation (blur, noise, thay đổi độ sáng/độ tương phản), chia tập train/validation/test, kiểm soát trùng lặp và rò rỉ dữ liệu. + Xây dựng mô hình baseline và đề xuất một cơ chế giám cảnh báo giả: ví dụ kết hợp mô hình detection với một mô hình phân loại hậu kiểm (fire/smoke vs non-fire) trên vùng nghi ngờ, hoặc hiệu chỉnh ngưỡng theo bối cảnh, hoặc hậu xử lý NMS/temporal smoothing (nếu dùng video). + Đánh giá mô hình theo các tiêu chí phù hợp: mAP@0.5 (hoặc mAP@0.5:0.95), Precision, Recall, F1-score; đo thời gian suy luận/FPS (nếu có). So sánh kết quả giữa baseline và phương pháp đề xuất, phân tích lỗi và nguyên nhân. + Xây dựng giao diện demo website (Flask/Streamlit) cho phép tải ảnh (hoặc video ngắn), hiển thị	Nguyễn Tuấn Anh	anhngt@huit.edu.vn	KTPM	Nghiên cứu	
7	Phân loại sản phẩm nông nghiệp dựa trên hình ảnh	Mục tiêu: Đề tài xây dựng ứng dụng demo cho phép phân loại các sản phẩm nông nghiệp (trái cây) dựa trên hình ảnh đầu vào. Thông qua đề tài, sinh viên được tiếp cận bài toán phân loại ảnh (image classification) trong lĩnh vực nông nghiệp, phát triển ứng dụng CNTT phục vụ nông nghiệp thông minh. Nội dung: - Khảo sát bài toán phân loại ảnh trong nông nghiệp (trái cây) và các ứng dụng thực tế. - Nghiên cứu cấu trúc bộ dữ liệu Fruits-360 và lựa chọn tập trái cây phù hợp cho thực nghiệm. - Tiền xử lý dữ liệu: chuẩn hóa kích thước, chia tập huấn luyện và kiểm thử. - Sử dụng mô hình học sâu tiền huấn luyện để trích xuất đặc trưng ảnh. - Xây dựng mô hình phân loại sản phẩm nông nghiệp dựa trên hình ảnh. - Huấn luyện và kiểm thử mô hình với bộ dữ liệu đã chọn. - Đánh giá hiệu quả mô hình phân loại trái cây (Accuracy). - Xây dựng giao diện cho phép người dùng tải ảnh và xem kết quả phân loại trái cây. - Phân tích kết quả thực nghiệm. Yêu cầu: - Ngôn ngữ: Python, Java, C#	Nguyễn Thị Định	dinght@huit.edu.vn	HTTT	Nghiên cứu	

8	Tối ưu hóa mô hình chú thích ảnh cho thiết bị di động dựa trên chiết xuất tri thức	Mục tiêu của đề tài: Đề tài nhằm nghiên cứu và xây dựng mô hình chú thích ảnh hiệu quả cho thiết bị di động dựa trên kỹ thuật chiết xuất tri thức (Knowledge Distillation). Trong đó, một mô hình chú thích ảnh có hiệu suất cao được xây dựng hoặc lựa chọn từ các mô hình đã được huấn luyện sẵn, đóng vai trò là mô hình teacher nhằm học các biểu diễn thị giác-ngôn ngữ giàu thông tin. Trên cơ sở đó, tri thức từ mô hình teacher được truyền sang một mô hình student có kiến trúc gọn nhẹ hơn. Mô hình student được thiết kế và tối ưu hóa để phù hợp với các thiết bị tài nguyên hạn chế như thiết bị di động hoặc thiết bị nhúng, đồng thời vẫn duy trì chất lượng chú thích ảnh ở mức chấp nhận được. Nội dung nghiên cứu chính - Khảo sát các phương pháp chú thích ảnh dựa trên mạng học sâu. - Tổng quan các kỹ thuật chiết xuất tri thức (Knowledge Distillation) và các chiến lược distillation trong học sâu. - Xây dựng và huấn luyện, hoặc khai thác mô hình chú thích ảnh hiệu suất cao đã được huấn luyện sẵn làm mô hình teacher. - Thiết kế và xây dựng mô hình chú thích ảnh gọn nhẹ (student model) phù hợp với thiết bị di động. - Áp dụng kỹ thuật chiết xuất tri thức để huấn luyện mô hình student dưới sự hướng dẫn của mô hình teacher. - Đánh giá mô hình teacher và student trên các tập dữ liệu chuẩn, sử dụng các độ đo đánh giá phổ biến trong bài toán chú thích ảnh. - Phân tích và so sánh kết quả thực nghiệm nhằm đánh giá hiệu quả của kỹ thuật chiết xuất tri thức đối với chất lượng chú thích ảnh và chi phí tính toán. - Tối ưu hóa và triển khai thử nghiệm mô hình student trên thiết bị di động, thiết bị nhúng hoặc môi trường mô phỏng, đánh giá thời gian suy luận và khả năng ứng dụng thực tế. Yêu cầu đối với sinh viên - Có kiến thức cơ bản về học máy, học sâu. - Có khả năng lập trình Python và các thư viện học sâu (PyTorch/TensorFlow).	Đinh Thị Mận Nguyễn Văn Thịnh	mandt@huit.edu.vn	HTTT	Nghiên cứu	
9	Phân loại cây xanh đường phố phường Bến Nghé và xây dựng ứng dụng	"Đề tài nghiên cứu và ứng dụng các mô hình máy học và học sâu trong bài toán nhận dạng, phân loại cây xanh đường phố tại phường Bến Nghé, nhằm hỗ trợ cung cấp thông tin trực quan cho khách du lịch quốc tế. Thông qua việc nhận dạng lá cây từ hình ảnh, ứng dụng giúp người dùng dễ dàng tra cứu tên gọi và đặc điểm của các loại cây xanh đô thị. -Khảo sát và tổng hợp các loài cây xanh đường phố trên địa bàn phường Bến Nghé. -Thu thập và xây dựng bộ dữ liệu hình ảnh lá cây phục vụ huấn luyện và đánh giá mô hình. -Nghiên cứu các kỹ thuật và mô hình phân lớp trong lĩnh vực máy học và học sâu. -Cài đặt và huấn luyện mô hình phân lớp trên bộ dữ liệu đã xây dựng. -Thực hiện đánh giá thực nghiệm nhằm phân tích hiệu quả và độ chính xác của các mô hình (accuracy, loss, F1-score...) -Đề xuất mô hình phù hợp với bộ dữ liệu thu thập được. -Xây dựng giao diện minh họa cho bài toán phân loại cây xanh đường phố. Ngôn ngữ lập trình: python (NumPy, OpenCV, TensorFlow,...) Môi trường: Google Colab/Visual Studio Code,.... Yêu cầu: sinh viên lập trình python tốt, tiếng Anh tốt, có khả năng nghiên cứu và đọc tài liệu tiếng Anh"	Trần Như Ý	ytn@huit.edu.vn	HTTT	Nghiên cứu	

10	Nghiên cứu mô hình Graph Convolutional Network cho bài toán định vị trong nhà dựa trên Wi-Fi fingerprint	Mục tiêu: Nghiên cứu và đánh giá khả năng áp dụng Graph Convolutional Network (GCN) để dự đoán vị trí trong nhà (tọa độ ECoord, NCoord) dựa trên Wi-Fi fingerprint, đồng thời so sánh với các mô hình baseline nhằm xác định hiệu quả và điều kiện áp dụng phù hợp. Nội dung: -Khảo sát bài toán định vị trong nhà dựa trên Wi-Fi fingerprint và các hướng tiếp cận học sâu. -Nghiên cứu các bộ dữ liệu HCXY / SYL / UJLIB và thực hiện tiền xử lý dữ liệu, bao gồm: +Xử lý mất tín hiệu RSSI +Chuẩn hóa đặc trưng +Lọc và lựa chọn Access Point (AP) -Xây dựng mô hình dự đoán vị trí trong nhà dựa trên Graph Convolutional Network (GCN). -Huấn luyện và kiểm thử mô hình với bộ dữ liệu đã lựa chọn. -Đánh giá hiệu quả mô hình dự đoán vị trí thông qua các độ đo: MSE, MAE, Mean Error, %<2m, %<5m. -Xây dựng giao diện minh họa cho phép người dùng nhập dữ liệu Wi-Fi fingerprint và hiển thị kết quả dự đoán vị trí. -Phân tích và thảo luận kết quả thực nghiệm. Yêu cầu: - Ngôn ngữ: Python, Java hoặc C#	Trần Thị Vân Anh	anhstv@huit.edu.vn	HTTT	Nghiên cứu	
11	Xây dựng hệ thống dự đoán nguy cơ bệnh tiểu đường từ dữ liệu lâm sàng bằng học máy	Đề tài tập trung xây dựng một hệ thống hỗ trợ dự đoán nguy cơ bệnh tiểu đường dựa trên dữ liệu lâm sàng của bệnh nhân như Đa niệu (Polyuria), Chứng khát nhiều (Polydipsia), Ăn nhiều (Polyphagia), Sụt cân đột ngột, Suy nhược, Mờ mắt, Ngứa, Dễ kích động, Vết thương lâu lành, Liệt nhẹ, Cứng cứng cơ, Rụng tóc và Béo phì,... Hệ thống ứng dụng các mô hình học máy phổ biến để ước lượng nguy cơ mắc bệnh, đồng thời so sánh hiệu quả giữa các mô hình nhằm lựa chọn giải pháp phù hợp cho triển khai thực tế. Kết quả được tích hợp vào ứng dụng web demo phục vụ mục đích hỗ trợ sàng lọc ban đầu. Mục tiêu ứng dụng - Xây dựng pipeline xử lý dữ liệu y tế hoàn chỉnh - Phát triển mô hình dự đoán nguy cơ bệnh tiểu đường - Triển khai hệ thống dự đoán trực quan, dễ sử dụng - Hỗ trợ người dùng (bác sĩ/sinh viên/y tá) trong việc đánh giá nguy cơ ban đầu Hướng triển khai kỹ thuật - Dữ liệu: Kaggle Diabetes Dataset: https://www.kaggle.com/code/scifosamahosney/diabetes - Mô hình sử dụng: Logistic Regression (baseline), Random Forest, XGBoost, MLP - Thực nghiệm: + Xử lý dữ liệu thiếu, chuẩn hóa đặc trưng + Thử nghiệm các tập đặc trưng khác nhau + So sánh các cấu hình tham số cơ bản của mô hình + Huấn luyện và đánh giá trên cùng tập dữ liệu - Đánh giá: Accuracy, Recall (ưu tiên do bài toán y tế), F1-score, AUC Sản phẩm ứng dụng - Web demo dự đoán nguy cơ bệnh tiểu đường - Báo cáo so sánh hiệu quả các mô hình	Lê Thị Thùy Lan	lanlttt@huit.edu.vn	HTTT	Nghiên cứu	

12	Phát hiện gian lận giao dịch tài chính trên dữ liệu mất cân bằng	1. Yêu cầu về Xác định bài toán - Phát hiện giao dịch gian lận (fraud) - Ý nghĩa thực tiễn trong ngân hàng – fintech 2. Yêu cầu về Khảo sát nghiên cứu liên quan - Fraud detection methods - Imbalanced learning (SMOTE, ADASYN) - Các mô hình phổ biến 3–5 năm gần đây 3. Yêu cầu về Phân tích dữ liệu - Dataset Credit Card Fraud - Phân tích mức độ mất cân bằng - Visualization phân bố lớp 4. Yêu cầu về Tiền xử lý và cân bằng dữ liệu - Chuẩn hóa - Thực nghiệm: + Không cân bằng + SMOTE + ADASYN + Undersampling 5. Yêu cầu về Xây dựng mô hình - Random Forest - XGBoost - LightGBM 6. Yêu cầu về Đánh giá mô hình - Precision, Recall, F1, AUC - So sánh trước và sau cân bằng dữ liệu 7. Yêu cầu về Phân tích kết quả	Trần Văn Thọ	thotv@huit.edu.vn	KHDL		
13	Xây dựng hệ thống dự đoán tiêu thụ điện năng dựa trên mô hình học sâu chuỗi thời gian	1. Yêu cầu về bài toán và mục tiêu - Xác định rõ bài toán dự đoán tiêu thụ điện năng theo thời gian (Regression) - Nêu rõ: + Đối tượng dự đoán (hộ gia đình / khu vực / tòa nhà) + Khoảng thời gian dự đoán (ngắn hạn hoặc trung hạn) - Trình bày ý nghĩa thực tiễn: + Hỗ trợ quản lý năng lượng + Tối ưu vận hành hệ thống điện -> Trọng tâm: bài toán thực tế và khả năng triển khai 2. Yêu cầu về khảo sát công trình liên quan - Khảo sát các nghiên cứu và ứng dụng: + Dự đoán chuỗi thời gian điện năng + Ứng dụng LSTM/GRU trong năng lượng - Ưu tiên tài liệu 3–5 năm gần đây - Không yêu cầu phân tích sâu lý thuyết, nhưng phải: + Nêu phương pháp đã dùng + Chỉ ra mô hình phổ biến -> Mục đích: làm cơ sở lựa chọn mô hình 3. Yêu cầu về dữ liệu - Sử dụng dữ liệu thực tế hoặc dữ liệu công khai + Household Power Consumption + Energy Load Dataset - Mô tả: + Nguồn dữ liệu + Số lượng mẫu + Tần suất (giờ/ngày/tháng)	Trần Văn Thọ	thotv@huit.edu.vn	KHDL		

14	Ứng dụng Clear Sky Optimization trong tối ưu chọn đặc trưng cho mô hình Naive Bayes trong bài toán phân loại email spam	Mục tiêu - Hiểu bài toán phân loại email spam trong khai thác dữ liệu văn bản - Phân tích vai trò của chọn đặc trưng đối với mô hình Naive Bayes - Áp dụng Clear Sky Optimization để lựa chọn tập đặc trưng tối ưu Đánh giá hiệu quả mô hình trước và sau khi tối ưu Nội dung thực hiện - Tiền xử lý dữ liệu email (chuyển văn bản sang vector đặc trưng) - Xây dựng mô hình Naive Bayes với tập đặc trưng ban đầu - Áp dụng Clear Sky Optimization để chọn tập đặc trưng tối ưu - So sánh hiệu năng mô hình dựa trên các chỉ số đánh giá Kết quả mong đợi - Mô hình phân loại spam có độ chính xác cao hơn - Giảm số lượng đặc trưng nhưng vẫn giữ hiệu năng tốt - Bảng so sánh kết quả trước và sau tối ưu - Báo cáo giải thích rõ vai trò của chọn đặc trưng Dataset đề xuất: UCI – Spambase Data Set	Đinh Nguyễn Trọng Nghĩa	nghiadnt@huit.edu.vn	KHDL		
15	Ứng dụng Cyclone Optimization Algorithm trong tối ưu siêu tham số cho mô hình Random Forest dự đoán khách hàng rời bỏ dịch vụ viễn thông	Mục tiêu - Hiểu bài toán dự đoán khách hàng rời bỏ trong phân tích dữ liệu kinh doanh - Hiểu được các siêu tham số quan trọng của mô hình Random Forest - Áp dụng Cyclone Optimization để tối ưu các siêu tham số - Đánh giá mức cải thiện hiệu năng mô hình Nội dung thực hiện - Phân tích tập dữ liệu khách hàng viễn thông - Xây dựng mô hình Random Forest với cấu hình mặc định - Tối ưu các siêu tham số bằng Cyclone Optimization - So sánh kết quả giữa mô hình ban đầu và mô hình tối ưu Kết quả mong đợi - Mô hình dự đoán churn có độ chính xác và F1-score cao hơn - Bộ siêu tham số tối ưu cho Random Forest - Biểu đồ và bảng so sánh hiệu năng - Nhận xét về hiệu quả của tối ưu metaheuristic Dataset đề xuất: Kaggle – Telco Customer Churn	Đinh Nguyễn Trọng Nghĩa	nghiadnt@huit.edu.vn	KHDL		

16	Ứng dụng Drizzle Optimization Algorithm trong tối ưu trọng số đặc trưng cho mô hình SVM trong bài toán chẩn đoán bệnh tim	Mục tiêu - Hiểu bài toán chẩn đoán bệnh tim dựa trên dữ liệu lâm sàng - Phân tích vai trò của từng đặc trưng trong mô hình SVM - Áp dụng Drizzle Optimization để tối ưu trọng số đặc trưng - Cải thiện độ chính xác của mô hình chẩn đoán Nội dung thực hiện - Phân tích tập dữ liệu bệnh tim (tuổi, huyết áp, cholesterol, ECG, ...) - Xây dựng mô hình SVM với tập đặc trưng ban đầu - Sử dụng Drizzle Optimization để gán trọng số cho các đặc trưng - Đánh giá mô hình dựa trên Accuracy, Precision, Recall Kết quả mong đợi - Mô hình SVM chẩn đoán bệnh tim có hiệu năng cao hơn - Danh sách trọng số đặc trưng sau tối ưu - Phân tích đặc trưng quan trọng trong chẩn đoán bệnh - Báo cáo thực nghiệm có ý nghĩa thực tiễn Dataset đề xuất: UCI – Heart Disease Data Set	Đinh Nguyễn Trọng Nghĩa	nghiadnt@huit.edu.vn	KHDL		
17	Xây dựng hệ thống gợi ý sản phẩm dựa trên hành vi người dùng	Mục tiêu: - Xây dựng hệ thống gợi ý sản phẩm tự động dựa trên lịch sử và hành vi tương tác của người dùng. - Nâng cao trải nghiệm cá nhân hóa, giúp người dùng nhanh chóng tiếp cận sản phẩm phù hợp. - Hỗ trợ doanh nghiệp tăng tỷ lệ chuyển đổi, doanh thu và mức độ hài lòng của khách hàng. Nội dung: - Xác định bài toán hệ gợi ý (Recommendation System) trong thương mại điện tử. - Mô tả tập dữ liệu sử dụng: lịch sử mua hàng, lượt xem, đánh giá, thời gian tương tác của người dùng; nguồn gốc, kích thước và các vấn đề của dữ liệu (thừa, thiếu, nhiễu, cold-start...). - Khảo sát các phương pháp tiêu biểu: + Content-based Filtering + Collaborative Filtering (User-based, Item-based) - Phân tích ưu điểm, hạn chế và khả năng áp dụng của từng phương pháp. - So sánh các cách tiền xử lý dữ liệu, biểu diễn đặc trưng người dùng – sản phẩm và xử lý dữ liệu thừa. Thực hiện: - Thu thập hoặc sử dụng bộ dữ liệu công khai (MovieLens, Amazon Reviews, Retail Dataset...). - Tiền xử lý dữ liệu: làm sạch, chuẩn hóa, xây dựng ma trận người dùng – sản phẩm, xử lý giá trị thiếu và dữ liệu thừa. - Cài đặt và áp dụng các mô hình gợi ý đã khảo sát. - Thực nghiệm với các phương pháp khác nhau và tính chính tham số mô hình. - Đánh giá hệ thống bằng các chỉ số phù hợp: Precision@K, Recall@K, F1@K, MAP, NDCG. - So sánh kết quả giữa các mô hình và phân tích hiệu quả. - Xây dựng giao diện demo (Streamlit/Flask) cho phép nhập người dùng và hiển thị danh sách sản phẩm được gợi ý. Kết quả mong đợi - Một hệ thống gợi ý hoạt động ổn định, đưa ra danh sách sản phẩm phù hợp với từng người dùng. - Bảng so sánh và biểu đồ thể hiện hiệu quả của các phương pháp gợi ý khác nhau.	Phan Thị Ngọc Mai	maiphantn@huit.edu.vn	KHDL		

18	Xây dựng mô hình phân nhóm khách hàng phục vụ Marketing cá nhân hóa	Mục tiêu: - Xây dựng mô hình phân cụm nhằm chia nhóm khách hàng dựa trên hành vi mua sắm và đặc điểm tiêu dùng. - Hỗ trợ doanh nghiệp hiểu rõ cấu trúc tập khách hàng, từ đó đề xuất các chiến lược marketing cá nhân hóa, nâng cao hiệu quả kinh doanh và chăm sóc khách hàng. Nội dung: - Xác định bài toán phân cụm (Clustering) trong bối cảnh phân tích khách hàng. - Mô tả tập dữ liệu sử dụng: dữ liệu lịch sử mua hàng (tần suất mua, giá trị đơn hàng, loại sản phẩm, thời gian mua, v.v.), nguồn gốc, kích thước và các vấn đề tồn tại của dữ liệu (thiếu giá trị, nhiễu, chênh lệch thang đo...). - Khảo sát các phương pháp phân cụm: K-means, DBSCAN, Hierarchical Clustering. - Phân tích ưu điểm, hạn chế và khả năng áp dụng của từng phương pháp trong bài toán phân nhóm khách hàng. - So sánh các cách tiền xử lý, chuẩn hóa dữ liệu và lựa chọn đặc trưng phù hợp cho bài toán. Thực hiện: - Thu thập hoặc sử dụng bộ dữ liệu thực tế/công khai về hành vi mua sắm của khách hàng. - Tiến hành tiền xử lý dữ liệu: làm sạch dữ liệu, xử lý giá trị thiếu, chuẩn hóa và biến đổi đặc trưng. - Áp dụng và cài đặt các thuật toán phân cụm đã khảo sát. - Thực nghiệm với nhiều cấu hình tham số và phương pháp tiền xử lý khác nhau để đánh giá ảnh hưởng đến kết quả phân cụm. - Đánh giá chất lượng phân cụm bằng các chỉ số phù hợp, tiêu biểu là Silhouette Score. - Trực quan hóa kết quả phân cụm thông qua biểu đồ và giao diện hỗ trợ phân tích (sử dụng Python với Streamlit/Flask hoặc công cụ tương đương). Kết quả mong đợi: - Xác định được các nhóm khách hàng có đặc trưng hành vi rõ ràng và có ý nghĩa thực tiễn. - Xây dựng các biểu đồ trực quan thể hiện cấu trúc và đặc điểm của từng cụm khách hàng.	Phan Thị Ngọc Mai	maiphantn@huit.edu.vn	KHDL		
19	Dự báo và phân tích xu hướng giá Bitcoin	Mục tiêu - Xây dựng hệ thống dự báo giá Bitcoin và phân tích xu hướng biến động thị trường để hỗ trợ quyết định đầu tư. Đánh giá hiệu năng giữa các mô hình học sâu chuỗi thời gian và thuật toán lọc nhiễu. Nội dung thực hiện Tiền xử lý dữ liệu Bitcoin lịch sử và phân tích các đặc trưng xu hướng giá. Huấn luyện mô hình RNN và LSTM để học sự phụ thuộc dài hạn của chuỗi dữ liệu. Áp dụng bộ lọc Kalman (Kalman Filter) để giảm nhiễu hoặc làm mịn mô hình tham chiếu. Thực nghiệm tinh chỉnh và đánh giá độ chính xác dự báo qua RMSE và MAE. Kết quả mong đợi Mô hình nắm bắt được xu hướng biến động chính của thị trường tiền mã hóa. Đồ thị so sánh trực quan để chỉ ra phương pháp nào tốt nhất dựa trên hai chỉ số RMSE và MAE.	Trần Việt Hùng	hungtv@huit.edu.vn	KHDL		

20	Dự báo biến động giá chứng khoán dùng LSTM	Mục tiêu Xây dựng mô hình dự báo xu hướng giá chứng khoán dựa trên dữ liệu lịch sử để hỗ trợ quyết định đầu tư. Đánh giá hiệu quả giữa các phương pháp học sâu chuỗi thời gian và kỹ thuật ước lượng trạng thái. Nội dung thực hiện Thu thập và tiền xử lý dữ liệu giá cổ phiếu (chuẩn hóa, sliding window) phục vụ huấn luyện. Cài đặt các mô hình mạng nơ-ron hồi quy RNN và LSTM để nắm bắt đặc trưng dài hạn. Áp dụng bộ lọc Kalman (Kalman Filter) để khử nhiễu dữ liệu hoặc làm mô hình đối sánh. Thực nghiệm tinh chỉnh tham số và đánh giá sai số dự báo qua chỉ số RMSE và MAE. Kết quả mong đợi Xác định được mô hình tối ưu cho bài toán dự báo giá chứng khoán. Đồ thị so sánh trực quan để chỉ ra phương pháp nào tốt nhất dựa trên hai chỉ số RMSE và MAE.	Trần Việt Hùng	hungtv@huit.edu.vn	KHDL		
21	Dự báo giá Ethereum kết hợp phân tích cảm xúc thị trường	Mục tiêu Xây dựng hệ thống dự báo giá Ethereum kết hợp dữ liệu giao dịch và Sentiment analysis để hỗ trợ đầu tư. Đánh giá hiệu năng giữa các mô hình học sâu và thuật toán ước lượng trạng thái. Nội dung thực hiện Thu thập dữ liệu lịch sử Ethereum và xử lý đặc trưng cảm xúc từ tin tức/mạng xã hội. Huấn luyện mô hình RNN và LSTM với đầu vào đa dạng (giá, sentiment) để học xu hướng. Áp dụng bộ lọc Kalman (Kalman Filter) để giảm nhiễu biến động giá hoặc làm mô hình cơ sở. Thực nghiệm tinh chỉnh và đánh giá độ chính xác qua chỉ số RMSE và MAE. Kết quả mong đợi Xác định được mức độ cải thiện dự báo khi kết hợp chỉ số cảm xúc. Đồ thị so sánh trực quan để chỉ ra phương pháp nào tốt nhất dựa trên hai chỉ số RMSE và MAE.	Trần Việt Hùng	hungtv@huit.edu.vn	KHDL		
22	Dự báo chất lượng không khí (AQI) và nồng độ bụi mịn (PM2.5) sử dụng mô hình Boosting hiện đại	Mục tiêu: - Nghiên cứu bài toán dự báo chất lượng không khí, tập trung vào chỉ số AQI và nồng độ bụi mịn PM2.5 - Khảo sát và ứng dụng các mô hình Boosting hiện đại trong bài toán dự báo dữ liệu môi trường. - Phân tích ảnh hưởng của tiền xử lý dữ liệu, lựa chọn đặc trưng và cấu hình tham số đến hiệu quả dự báo. - Đánh giá khả năng áp dụng mô hình Boosting cho dữ liệu môi trường thực tế và nêu các hạn chế, hướng mở rộng. Nội dung thực hiện: - Xác định bài toán dự báo AQI và PM2.5 từ dữ liệu môi trường - Khảo sát các nghiên cứu liên quan và mô hình Boosting phù hợp - Thu thập, phân tích và tiền xử lý dữ liệu chất lượng không khí - Xây dựng và huấn luyện mô hình Boosting - Thực nghiệm điều chỉnh tham số và lựa chọn đặc trưng - Đánh giá kết quả dự báo bằng các chỉ số phù hợp - Phân tích, trực quan hóa kết quả và rút ra nhận xét Kết quả đạt được: - Mô hình hóa được bài toán dự báo AQI và PM2.5 từ dữ liệu môi trường - Xây dựng được mô hình Boosting cho bài toán dự báo - Thu được kết quả dự báo và đánh giá bằng các chỉ số phù hợp - Phân tích và trực quan hóa được kết quả thực nghiệm	Nguyễn Thị Huyền Trang	trangthuyen@huit.edu.vn	KHDL		

23	Nghiên cứu ứng dụng Transfer Learning trong bài toán phân loại bệnh võng mạc tiểu đường từ ảnh đáy mắt	Mục tiêu: - Nghiên cứu bài toán phân loại bệnh võng mạc tiểu đường từ ảnh đáy mắt. - Khảo sát và ứng dụng Transfer Learning trong phân loại ảnh y sinh. - Phân tích ảnh hưởng của tiền xử lý dữ liệu và cấu hình mô hình đến kết quả phân loại. - Đánh giá khả năng áp dụng mô hình trên dữ liệu ảnh thực tế. Nội dung thực hiện: - Xác định bài toán, phạm vi nghiên cứu và dữ liệu ảnh đáy mắt - Khảo sát các nghiên cứu liên quan và mô hình Transfer Learning phù hợp - Phân tích đặc điểm dữ liệu và thực hiện tiền xử lý ảnh - Xây dựng và huấn luyện mô hình Transfer Learning - Thực nghiệm điều chỉnh tham số và cải thiện kết quả phân loại - Đánh giá, phân tích và trực quan hóa kết quả Kết quả: - Mô hình hóa được bài toán phân loại bệnh võng mạc tiểu đường từ ảnh đáy mắt. - Xây dựng được mô hình phân loại ảnh dựa trên Transfer Learning. - Thu được kết quả đánh giá mô hình bằng các chỉ số phù hợp. - Phân tích và trực quan hóa được kết quả thực nghiệm. - Nhận xét được hạn chế và đề xuất hướng phát triển tiếp theo.	Nguyễn Thị Huyền Trang	trangthuyen@huit.edu.vn	KHDL		
24	Phân tích và tiền xử lý dữ liệu RSSI để cải thiện độ chính xác định vị trong hệ thống IPS	1. Mục tiêu đề tài: - Nghiên cứu đặc điểm dữ liệu RSSI trong môi trường trong nhà - Đánh giá ảnh hưởng của nhiễu, đa đường, vật cản - Đề xuất phương pháp xử lý dữ liệu RSSI trước khi tiến hành định vị 2. Nội dung thực hiện: - Tổng quan về IPS và vai trò của RSSI - Tìm hiểu một vài bộ dữ liệu sẵn có - Phân tích thống kê RSSI + Trung bình, phương sai, phân bố + RSSI theo thời gian và khoảng cách - So sánh các kỹ thuật tiền xử lý: + Loại bỏ outlier + Median filter + Kalman filter + Moving average - Đánh giá ảnh hưởng tiền xử lý tín hiệu đến sai số định vị 3. Sản phẩm: - Bộ dữ liệu RSSI sau khi đã được làm sạch - Các biểu đồ, bảng biểu so sánh trước và sau khi xử lý - Nhận xét kỹ thuật nào giúp RSSI ổn định hơn.	Huỳnh Thị Châu Lan	lanhtc@huit.edu.vn	KHDL		

25	So sánh các mô hình định vị trong nhà dựa trên RSSI bằng phương pháp học máy	1. Mục tiêu đề tài: - Giúp sinh viên tiếp cận bài toán định vị trong nhà IPS theo hướng phân tích dữ liệu và học máy - Cung cấp cái nhìn so sánh rõ ràng giữa các mô hình định vị dựa trên RSSI - Có khả năng ứng dụng trong các môi trường như: trường học, bệnh viện, trung tâm thương mại. 2. Nội dung thực hiện: - Nghiên cứu cơ sở lý thuyết + Tìm hiểu bài toán định vị trong nhà IPS dựa trên RSSI + Phân tích các đặc điểm của dữ liệu RSSI trong môi trường trong nhà như nhiễu, suy hao tín hiệu, hiện tượng đa đường (multipath) + Nghiên cứu phương pháp định vị fingerprinting và vai trò của học máy trong IPS - Chuẩn bị dữ liệu RSSI + Tổ chức dữ liệu fingerprint RSSI theo từng vị trí cụ thể + Thực hiện tiền xử lý dữ liệu (chuẩn hóa, lọc nhiễu, xử lý dữ liệu thiếu) nhằm nâng cao chất lượng dữ liệu đầu vào cho mô hình học máy - Xây dựng các mô hình định vị dựa trên học máy + Triển khai các thuật toán học máy phổ biến như K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest + Thiết kế mô hình dự đoán vị trí dựa trên vector RSSI đầu vào - Đánh giá và so sánh hiệu quả các mô hình + Đánh giá độ chính xác định vị thông qua sai số khoảng cách giữa vị trí dự đoán và vị trí thực tế + So sánh các mô hình theo các tiêu chí: độ chính xác, độ ổn định, thời gian huấn luyện và khả năng mở rộng + Phân tích ảnh hưởng của số lượng điểm phát (Access Point/Beacon) và mật độ fingerprint đến kết quả định vị - Đề xuất giải pháp và hướng phát triển + Xác định mô hình học máy phù hợp nhất cho hệ thống định vị trong nhà dựa trên RSSI	Huỳnh Thị Châu Lan	lanhtc@huit.edu.vn	KHDL		
26	Mô hình học sâu 3D kết hợp Attention cho bài toán dự báo mưa từ dữ liệu vệ tinh	Yêu cầu nội dung: - Khảo sát và nghiên cứu các phương pháp học sâu cho dữ liệu không gian–thời gian trong bài toán dự báo mưa (CNN 2D/3D, RNN, ConvLSTM, Attention). - Nghiên cứu đặc điểm dữ liệu mưa vệ tinh (TRMM/GPM) và các thách thức trong dự báo mưa (tính phi tuyến, không liên tục, mưa lớn hiếm). - Biểu diễn dữ liệu mưa vệ tinh dưới dạng tensor 3D không gian–thời gian ($\text{time} \times \text{latitude} \times \text{longitude}$). - Tiền xử lý dữ liệu: đồng bộ thời gian, chuẩn hóa, cắt vùng quan tâm và xây dựng tập huấn luyện/kiểm tra. - Xây dựng mô hình nền dựa trên ConvLSTM để học đồng thời đặc trưng không gian và động lực thời gian của trường mưa. - Tích hợp cơ chế Attention nhằm giúp mô hình tập trung vào các vùng mưa quan trọng và cải thiện khả năng dự báo. - Đề xuất và áp dụng hàm loss có trọng số để nâng cao hiệu quả học trên toàn dải cường độ mưa, đặc biệt là mưa lớn. - Thực nghiệm, đánh giá kết quả dự báo bằng các độ đo phù hợp (RMSE, MAE, CSI, POD, FAR, ...).	Ngô Dương Hà	hand@huit.edu.vn	KHDL		SV đã đăng ký: Trương Ngô Phát - 2045230075 Trần Hoàng Khánh - 2045230050
27	Hệ thống phát hiện trạng thái buồn ngủ của người lái xe dựa trên thị giác máy tính và học sâu	Yêu cầu nội dung: - Khảo sát và nghiên cứu các phương pháp CNN, YOLO, LSTM trong bài toán nhận diện hành vi - Thu thập và xử lý dữ liệu - Phát hiện đặc trưng khuôn mặt thông qua các kỹ thuật Yolo hoặc CNN - Xây dựng mô hình huấn luyện để dự đoán trạng thái buồn ngủ với kỹ thuật LSTM/GRU - Thực nghiệm đánh giá kết quả với các độ đo - Xây dựng ứng dụng phù hợp cho bài toán phát hiện trạng thái buồn ngủ của người lái xe	Ngô Dương Hà	hand@huit.edu.vn	KHDL		

28	Cải tiến thuật toán gom cụm mờ Fuzzy C-Means bằng Particle Swarm Optimization trong phân loại bệnh	Yêu cầu nội dung Đề tài cần đảm bảo các nội dung sau: - Trình bày tổng quan về bài toán phân loại bệnh trên dữ liệu y tế; nêu rõ ý nghĩa thực tiễn và khoa học của bài toán. - Khảo sát tình hình nghiên cứu trong và ngoài nước liên quan đến các phương pháp phân cụm mờ và các thuật toán tối ưu bầy đàn. - Trình bày cơ sở lý thuyết của thuật toán Fuzzy C-Means (FCM) và Particle Swarm Optimization (PSO). - Xây dựng và mô tả chi tiết mô hình kết hợp PSO-FCM trong phân loại bệnh. - Cài đặt thuật toán PSO-FCM và các thuật toán so sánh (FCM, GA-FCM, WOA-FCM, SSA-FCM,...). - Thực nghiệm trên các bộ dữ liệu y tế thực (hoặc dữ liệu giả lập). - Đánh giá hiệu quả mô hình thông qua các chỉ số: độ chính xác, thời gian hội tụ, hàm mục tiêu, các chỉ số phân cụm (XB, DB, Silhouette,...). - So sánh kết quả với các phương pháp khác và phân tích ưu – nhược điểm của mô hình đề xuất. - Xây dựng chương trình và giao diện minh họa cho hệ thống phân loại bệnh dựa trên PSO-FCM.	Nguyễn Thị Thủy Trang	trangntthuy@huit.edu.vn	KHDL		
29	Ứng dụng mô hình học sâu trong phân loại bệnh ngoài da	Mục tiêu: Áp dụng mô hình học sâu để phân loại bệnh ngoài da dựa trên hình ảnh, góp phần hỗ trợ nhận dạng và đánh giá tình trạng da liễu một cách tự động. Đối tượng và phạm vi nghiên cứu: - Đối tượng nghiên cứu: Các mô hình học sâu và dữ liệu hình ảnh bệnh ngoài da. - Phạm vi nghiên cứu: Tập trung vào bài toán phân loại một số loại bệnh da phổ biến dựa trên ảnh chụp vùng da bị tổn thương. Nội dung nghiên cứu: - Tổng quan về bệnh ngoài da và các phương pháp phân loại dựa trên học sâu. - Nghiên cứu, lựa chọn và thử nghiệm một số mô hình học sâu cho bài toán phân loại. - Huấn luyện, đánh giá và so sánh hiệu quả của các mô hình trên tập dữ liệu ảnh. Kết quả dự kiến: Lựa chọn được mô hình học sâu có khả năng phân loại với độ chính xác phù hợp, đồng thời trình bày và phân tích các kết quả đánh giá, hiệu năng của mô hình.	Phùng Thế Bảo	baopt@huit.edu.vn	KHDL		SV thực hiện: 1. Tô Anh Pháp (2045230074) 2. Nguyễn Thị Mỹ Duyên (2045230020) 3. Đoàn Mỹ Hằng (2045230027)
30	Ứng dụng nhận diện cảm xúc tài xế trong thời gian thực sử dụng mô hình học sâu CNN và LSTM	Xây dựng mô hình học sâu kết hợp CNN và LSTM nhận diện cảm xúc khuôn mặt. + Xử lý và phân tích dữ liệu hình ảnh/video + Nhận diện cảm xúc tài xế (tức giận,...) trong thời gian thực với độ chính xác cao. + Cảnh báo khi phát hiện trạng thái cảm xúc nguy hiểm bằng âm thanh hoặc tín hiệu.'	Phùng Thế Bảo	baopt@huit.edu.vn	KHDL		
31	Ứng dụng mô hình học sâu trong nhận dạng chữ số viết tay từ hình ảnh	- Thu thập dữ liệu bộ ký tự chữ số viết tay, tiền xử lý dữ liệu và làm giàu dữ liệu tăng độ chính xác cho mô hình - Lựa chọn mô hình học sâu phù hợp cho bài toán - Huấn luyện mô hình, đánh giá thực nghiệm, so sánh với kết quả các mô hình CNN đã áp dụng - Xây dựng dashboard nhận diện được chữ số từ dữ liệu phiếu điểm của GV	Phùng Thế Bảo	baopt@huit.edu.vn	KHDL		Sinh viên thực hiện:' - Nguyễn Phạm Duy 2045230018 -Nguyễn Đình Tùng 2045230089

32	Ứng dụng thuật toán COATI để cải tiến gom cụm mờ trong phân loại khách hàng	- Trình bày tổng quan về bài toán phân loại khách hàng và vai trò của các phương pháp gom cụm trong phân tích dữ liệu khách hàng. - Khảo sát tình hình nghiên cứu trong và ngoài nước liên quan đến gom cụm mờ và các thuật toán tối ưu hóa bầy đàn trong phân cụm dữ liệu. - Trình bày cơ sở lý thuyết của thuật toán Fuzzy C-Means (FCM), bao gồm nguyên lý hoạt động, hàm mục tiêu, ưu điểm và hạn chế. - Giới thiệu và phân tích thuật toán COATI: nguồn cảm hứng sinh học, cơ chế hoạt động, đặc điểm nổi bật và khả năng áp dụng trong bài toán tối ưu hóa phân cụm. - Xây dựng và mô tả chi tiết mô hình kết hợp COATI-FCM trong phân loại khách hàng. - Cài đặt thuật toán COATI-FCM và các thuật toán so sánh (FCM, PSO-FCM, GA-FCM, WOA-FCM). - Thực nghiệm mô hình trên các tập dữ liệu khách hàng thực tế và/hoặc dữ liệu giả lập. - Đánh giá hiệu quả mô hình thông qua các chỉ số: hàm mục tiêu FCM, độ chính xác phân loại, các chỉ số phân cụm (Silhouette, Davies-Bouldin, Xie-Beni), thời gian hội tụ. - So sánh kết quả giữa COATI-FCM và các phương pháp khác; phân tích ưu điểm, hạn chế và khả năng ứng dụng thực tiễn của mô hình đề xuất. - Xây dựng chương trình và giao diện minh họa hỗ trợ phân loại khách hàng dựa trên mô hình COATI-FCM.	Nguyễn Thị Thủy Trang	trangntthuy@huit.edu.vn	KHDL		
33	Tìm hiểu thuật toán Phân đoạn và ứng dụng nhận diện cảnh báo hư hỏng mặt đường (ổ gà, vết nứt) từ camera hành trình.	(1) Tìm hiểu thuật toán Mask RCNN và YOLO_Seg; (2) Tìm hiểu và thu thập bộ dữ liệu (mặt đường nứt, ổ gà) cho bài toán; (3) Xây dựng thực nghiệm tạo mô hình và đánh giá kết quả đạt được; (4) Xây dựng ứng dụng web hoặc nền tảng khác dựa trên mô hình đã đạt được, kết hợp với các thuật toán khác để hoàn thiện ứng dụng. - Sinh viên: Biết sử dụng code python và các ngôn ngữ phù hợp khác. - Sinh viên có tư duy lập trình, làm việc nghiêm túc, có kế hoạch thực hiện công việc, có khả năng đọc tài liệu chuyên ngành, có kỹ năng làm việc nhóm.	Trần Đình Toàn	toantd@huit.edu.vn	KHDL		Tối đa 3 sinh viên
34	Tìm hiểu Instance Segmentation và ứng dụng đếm tế bào máu (hồng cầu, bạch cầu) nhằm hỗ trợ chẩn đoán bệnh.	(1) Tìm hiểu Instance Segmentation; (2) Tìm hiểu và thu thập bộ dữ liệu ảnh tế bào máu cho bài toán; (3) Xây dựng thực nghiệm tạo mô hình và đánh giá kết quả đạt được; (4) Xây dựng ứng dụng web hoặc nền tảng khác dựa trên mô hình đã đạt được, kết hợp với các thuật toán khác để hoàn thiện ứng dụng. - Sinh viên: Biết sử dụng code python và các ngôn ngữ phù hợp khác. - Sinh viên có tư duy lập trình, làm việc nghiêm túc, có kế hoạch thực hiện công việc, có khả năng đọc tài liệu chuyên ngành, có kỹ năng làm việc nhóm.	Trần Đình Toàn	toantd@huit.edu.vn	KHDL		Tối đa 3 sinh viên
35	Đánh giá hiệu năng của kiến trúc FaceNet và FaceNet512 trong framework DeepFace trên bộ dữ liệu LFW	- Triển khai và đánh giá hai biến thể FaceNet (128-d) và FaceNet512 (512-d), xác định sự ảnh hưởng của kích thước vector đặc trưng đến độ chính xác và tốc độ nhận diện sử dụng framework DeepFace trên bộ dữ liệu chuẩn LFW. Đánh giá các mô hình dựa trên các chỉ số: độ chính xác (Accuracy), tỷ lệ nhận diện sai (FAR/FRR) và thời gian xử lý (Inference Time). - Xây dựng ứng dụng cho phép đăng ký mới và nhận diện trực tiếp qua Webcam, đánh giá tính khả thi / hiệu quả của mô hình khi áp dụng thực tế.	Vũ Phú Lộc	locvp@huit.edu.vn	CNS		

36	So sánh hiệu quả phương pháp tinh chỉnh ResNet và EfficientNet trên tập dữ liệu CIFAR-10	Đề tài thực hiện phân loại ảnh CIFAR-10 gồm 10 lớp đối tượng, với mục tiêu so sánh hiệu quả của hai kiến trúc mạng nơ-ron tích chập phổ biến là ResNet và EfficientNet khi áp dụng fine-tuning. Dữ liệu sử dụng: bộ CIFAR-10 gồm 60.000 ảnh màu kích thước 32×32, được tiền xử lý bằng chuẩn hóa và data augmentation cơ bản nhằm cải thiện khả năng tổng quát hóa. Phương pháp thực hiện: fine-tuning hai mô hình ResNet-18 và EfficientNet-B0. Hiệu năng mô hình được đánh giá thông qua accuracy và confusion matrix, đồng thời quá trình huấn luyện và kết quả được trực quan hóa bằng các biểu đồ loss, accuracy và confusion matrix. Từ kết quả thực nghiệm, đề tài phân tích mức độ phù hợp của từng mô hình đối với CIFAR-10 và mối quan hệ giữa độ chính xác đạt được và chi phí huấn luyện.	Trần Khải Thiện	thientk@huit.edu.vn	CNS		
37	Khảo sát và triển khai thử nghiệm mô hình đa phương thức sử dụng vLLM-Omni	Đề tài tập trung tìm hiểu và triển khai thử nghiệm mô hình đa phương thức dựa trên vLLM-Omni, với mục tiêu phân tích kiến trúc, quy trình inference và khả năng tích hợp nhiều loại dữ liệu. Dữ liệu sử dụng gồm tập văn bản ngắn (prompt) tự xây dựng và tập ảnh công khai quy mô nhỏ (ví dụ COCO hoặc Flickr8k) phục vụ cho các kịch bản inference đa phương thức. Phương pháp thực hiện: cài đặt và cấu hình vLLM-Omni để chạy inference cho các mô hình đa phương thức có sẵn, không thực hiện huấn luyện mô hình. Thực nghiệm được thiết kế theo hướng so sánh các cấu hình inference khác nhau (batch size, độ dài prompt) và so sánh với cách inference truyền thống, nhằm phân tích ảnh hưởng của cấu hình hệ thống đến chất lượng và tính ổn định của đầu ra. Chất lượng kết quả được đánh giá thông qua đánh giá định tính bởi con người (human evaluation) và chỉ số BLEU ở mức tham khảo trong các kịch bản có dữ liệu tham chiếu, kết hợp với đánh giá hiệu năng cơ bản về thời gian phản hồi. Các kết quả được trực quan hóa bằng biểu đồ và giao diện demo đơn giản.	Trần Khải Thiện	thientk@huit.edu.vn	CNS		
38	Hệ thống nhận diện trang bị bảo hộ lao động (PPE) và nhận thương hiệu nhà thầu dựa trên kiến trúc Hybrid YOLO11 và Vision-Language Model (Ollama)	Mục tiêu đề tài: Đề xuất hướng giải quyết bài toán "Cold-start" khi triển khai tại các công trình mới, nơi logo thương hiệu nhà thầu chưa có dữ liệu huấn luyện. Phương pháp: - Bước 1: Sử dụng YOLO11n (bản Nano) để phát hiện các vùng đối tượng (Region of Interest - ROI) như mũ bảo hiểm, áo bảo hộ, và vị trí đặt logo trên thiết bị. - Bước 2: Cắt các vùng ROI này và gửi tới Ollama Server chạy mô hình Vision nhỏ (như Moondream2 hoặc Qwen2.5-VL) để nhận diện văn bản hoặc ký hiệu thương hiệu bằng kỹ thuật Zero-shot/Few-shot prompting. Đề xuất hệ thống: Sử dụng Camera IP giá rẻ hoặc Webcam truyền luồng MJPEG về máy chủ Edge (Laptop/PC tại công trình) chạy Ollama. Kết quả đánh giá: Đánh giá độ chính xác (mAP, F1-score) giữa việc huấn luyện lại YOLO11 (truyền thống) và việc dùng Prompting trên VLM cho các thương hiệu mới xuất hiện lần đầu (Cold-start).	Võ Hoàng Hải	haivh@huit.edu.vn	CNS		
39	Xây dựng quy trình thích nghi nhanh (Rapid Site Adaptation) cho nhận diện thiết bị bảo hộ có gắn thương hiệu công ty (PPE+BRANDNAME) sử dụng Few-shot Learning và Ollama	Mục tiêu: Đề xuất cách xây dựng hệ thống có thể "học" được màu áo bảo hộ và logo của một công trình mới chỉ với 5-10 ảnh mẫu (Few-shot) trong vòng 4-8 giờ đầu tiên triển khai hệ thống giám sát ở công trình Phương pháp: - Sử dụng YOLO11 làm "Base Model" đã huấn luyện trên tập dữ liệu PPE công khai (như Construction-PPE dataset). https://data.mendeley.com/datasets/zkzghjvnpn2/6 - Áp dụng Few-shot Object Detection (FsDet) https://github.com/YanxingLiu/SAE-FSDet hoặc Fine-tuning lớp Head của YOLO11 với dữ liệu mới thu thập tại công trường. - Tích hợp Ollama để tự động dán nhãn (Auto-labeling) cho dữ liệu mới thu thập, giúp giảm thời gian gán nhãn thủ công cho sinh viên. Hệ thống vận hành: Giao diện ứng dụng hiển thị luồng video từ camera giá rẻ, tích hợp nút "Thêm đối tượng mới" để hệ thống tự cập nhật qua Ollama. So sánh: Đánh giá sự chênh lệch về \$mAP\$ (mean Average Precision) và \$Recall\$ giữa việc chỉ dùng dữ liệu Few-shot thực tế (bị nhiễu, thiếu góc cạnh) so với việc dùng dữ liệu tổng hợp từ khung AIGC-HWD (chuẩn hóa nhãn, đa dạng tư thế). https://www.mdpi.com/2075-5309/15/22/4080	Võ Hoàng Hải	haivh@huit.edu.vn	CNS		

40	Hệ thống tối ưu hóa giám sát an toàn công trình bằng Edge Computing: So sánh hiệu năng giữa Tiny-YOLO11 và Local VLM (Ollama) trên thiết bị cấu hình thấp	Mục tiêu: Tìm hiểu khả năng thiết lập hệ thống nhận diện PPE thời gian thực và phân tích khả năng áp dụng của các mô hình ngôn ngữ thị giác (VLM) tại máy tính biên cấu hình thấp. Phương pháp: - Triển khai YOLO11n đã được lượng tử hóa (INT8 Quantization) bằng OpenVINO để chạy trên CPU/GPU giá rẻ. https://github.com/ultralytics/ultralytics - Thiết lập một Ollama Server nội bộ chạy mô hình LLaVA-v1.6-7B hoặc Gemma3-4B-Vision để thực hiện các truy vấn phức tạp như: "Công nhân có đang đội nón bảo hiểm có LOGO X không ?" (bài toán Few-shot reasoning). Hệ thống vận hành: Dùng ESP32-CAM làm thiết bị thu hình, đẩy dữ liệu qua Wi-Fi về máy chủ xử lý tại biên. So sánh: So sánh tốc độ (FPS) và độ chính xác trong việc phát hiện vi phạm PPE giữa mô hình Object Detection thuần túy (YOLO) và mô hình suy luận ngữ cảnh (Ollama VLM).	Võ Hoàng Hải	haivh@huit.edu.vn	CNS		
41	Xây dựng hệ thống dự đoán nguy cơ bệnh tim mạch từ dữ liệu lâm sàng bằng học máy	Mô tả ngắn: Đề tài tập trung xây dựng một hệ thống hỗ trợ dự đoán nguy cơ bệnh tim mạch dựa trên dữ liệu lâm sàng của bệnh nhân như tuổi, giới tính, huyết áp, cholesterol, kết quả ECG,... Hệ thống ứng dụng các mô hình học máy phổ biến để ước lượng nguy cơ mắc bệnh, đồng thời so sánh hiệu quả giữa các mô hình nhằm lựa chọn giải pháp phù hợp cho triển khai thực tế. Kết quả được tích hợp vào ứng dụng web demo phục vụ mục đích hỗ trợ sàng lọc ban đầu. Mục tiêu ứng dụng - Xây dựng pipeline xử lý dữ liệu y tế hoàn chỉnh - Phát triển mô hình dự đoán nguy cơ bệnh tim mạch - Triển khai hệ thống dự đoán trực quan, dễ sử dụng - Hỗ trợ người dùng (bác sĩ/sinh viên/y tá) trong việc đánh giá nguy cơ ban đầu Hướng triển khai kỹ thuật - Dữ liệu: + UCI Heart Disease: https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data + Kaggle Heart Dataset: https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset - Mô hình sử dụng: Logistic Regression (baseline), Random Forest, XGBoost, MLP - Thực nghiệm: + Xử lý dữ liệu thiếu, chuẩn hóa đặc trưng + Thử nghiệm các tập đặc trưng khác nhau + So sánh các cấu hình tham số cơ bản của mô hình + Huấn luyện và đánh giá trên cùng tập dữ liệu - Đánh giá: Accuracy, Recall (ưu tiên do bài toán y tế), F1-score, AUC Sản phẩm ứng dụng - Web demo dự đoán nguy cơ bệnh tim mạch - Báo cáo so sánh hiệu quả các mô hình	Lữ Thị Cẩm Tú	tultc@huit.edu.vn	CNS		

42	Hệ thống phân tích và gợi ý sản phẩm dựa trên hành vi người dùng	Mô tả ngắn Xây dựng hệ thống phân tích hành vi mua sắm và gợi ý sản phẩm cho người dùng dựa trên dữ liệu thương mại điện tử. Bài toán - Loại bài toán: Recommendation + Clustering - Dữ liệu: Online Retail / E-commerce Dataset (Kaggle) - Ý nghĩa thực tiễn: Cá nhân hóa trải nghiệm mua sắm Phương pháp – mô hình - K-Means (phân cụm khách hàng) - Collaborative Filtering (User-based / Item-based) - So sánh với phương pháp gợi ý phổ biến Dữ liệu: (SV lựa chọn bộ dữ liệu tham khảo) - Online Retail Dataset: https://www.kaggle.com/datasets/ulrikthgepedersen/online-retail-dataset - E-commerce Customer Behavior Dataset: https://www.kaggle.com/datasets/uom190346a/e-commerce-customer-behavior-dataset Thực nghiệm - Chuẩn hóa dữ liệu hành vi - Thử nghiệm số cụm khác nhau - Đánh giá bằng Precision@K, Recall@K Sản phẩm ứng dụng - Dashboard phân tích nhóm khách hàng - Demo gợi ý sản phẩm theo người dùng	Lữ Thị Cẩm Tú	tueltc@huit.edu.vn	CNS		
43	So sánh hiệu quả phương pháp fine-tuning BERT cho bài toán phát hiện clickbait	Đề tài thực hiện phân loại clickbait trong tiêu đề tin tức, với mục tiêu so sánh hiệu quả của các mô hình ngôn ngữ dựa trên BERT khi áp dụng fine-tuning. Dữ liệu sử dụng là tập dữ liệu clickbait gồm các tiêu đề văn bản được gán nhãn, được tiền xử lý bằng chuẩn hóa văn bản và tokenization theo chuẩn BERT. Phương pháp thực hiện: fine-tuning các mô hình BERT-base và DistilBERT cho bài toán phân loại nhị phân. Hiệu năng mô hình được đánh giá thông qua accuracy, precision, recall, F1-score và confusion matrix; đồng thời quá trình huấn luyện và kết quả được trực quan hóa bằng các biểu đồ loss, accuracy và confusion matrix. Từ kết quả thực nghiệm, đề tài phân tích mức độ phù hợp của từng mô hình đối với bài toán phát hiện clickbait và mối quan hệ giữa độ chính xác đạt được và chi phí huấn luyện.	Trần Khải Thiện	thientk@huit.edu.vn	CNS		
44	So sánh hiệu quả phương pháp fine-tuning GPT-OSS cho bài toán phát hiện clickbait	Đề tài thực hiện phân loại clickbait trong tiêu đề tin tức, với mục tiêu đánh giá hiệu quả của mô hình GPT-OSS khi áp dụng fine-tuning cho bài toán phân loại văn bản. Dữ liệu sử dụng là tập dữ liệu clickbait gồm các tiêu đề được gán nhãn, được tiền xử lý và chuẩn hóa phù hợp với đầu vào của mô hình GPT. Phương pháp thực hiện: fine-tuning GPT-OSS cho tác vụ phân loại nhị phân. Hiệu năng mô hình được đánh giá thông qua accuracy, precision, recall, F1-score và confusion matrix; đồng thời quá trình huấn luyện và kết quả được trực quan hóa bằng các biểu đồ loss, accuracy và confusion matrix. Từ kết quả thực nghiệm, đề tài phân tích khả năng áp dụng GPT-OSS cho phát hiện clickbait và mối quan hệ giữa chất lượng dự đoán và chi phí huấn luyện.	Trần Khải Thiện	thientk@huit.edu.vn	CNS		
45	Xây dựng hệ thống Gợi ý sản phẩm cho sàn TMDT thời trang	1. Dữ liệu: Thu thập/Sử dụng bộ dữ liệu >10k mẫu (User-Item rating). Xử lý dữ liệu thưa (sparsity). 2. Mô hình: Cài đặt và so sánh 2 phương pháp: Collaborative Filtering (KNN/Matrix Factorization) và Content-based. 3. Thực nghiệm: Tính chính tham số (K, metrics) để tối ưu RMSE/MAE. 4. Ứng dụng: Xây dựng Web App (Streamlit/Flask) cho phép chọn User và hiển thị list sản phẩm gợi ý.	Nguyễn Huy Liêm	liemnh@huit.edu.vn	KHDL	Ứng dụng	
46	Phân tích và Dự báo khả năng nợ xấu ngân hàng (Credit Scoring)	1. Dữ liệu: Xử lý bộ dữ liệu tín dụng thực tế. Tập trung giải quyết bài toán Mất cân bằng dữ liệu (Imbalanced Data) bằng SMOTE hoặc class weights. 2. Mô hình: So sánh các thuật toán phân loại: Logistic Regression, Random Forest, XGBoost. 3. Đánh giá: Sử dụng ROC-AUC, F1-Score thay vì Accuracy. 4. Ứng dụng: Web App nhập thông tin khách hàng -> Trả về xác suất vỡ nợ và giải thích lý do (Feature Importance).	Nguyễn Huy Liêm	liemnh@huit.edu.vn	KHDL	Ứng dụng	

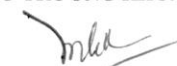
47	Phân đoạn khách hàng (Customer Segmentation) cho chiến dịch Marketing	1. Dữ liệu: Lịch sử mua hàng (RFM: Recency, Frequency, Monetary). Chuẩn hóa dữ liệu. 2. Mô hình: Áp dụng thuật toán Gom cụm (Clustering): K-Means, DBSCAN, Hierarchical Clustering. 3. Thực nghiệm: Sử dụng Elbow Method và Silhouette Score để tìm số cụm (K) tối ưu. 4. Ứng dụng: Dashboard hiển thị biểu đồ phân bố các nhóm khách hàng và đặc tính của từng nhóm.	Nguyễn Huy Liêm	liemnh@huit.edu.vn	KHDL	Ứng dụng	
48	Phát hiện gian lận giao dịch thẻ tín dụng (Credit Card Fraud Detection)	1. Dữ liệu: Dataset giao dịch (thường bị ẩn danh PCA). Xử lý dữ liệu nhiễu và cực kỳ mất cân bằng. 2. Mô hình: Sử dụng các thuật toán phát hiện bất thường (Anomaly Detection) như Isolation Forest, Autoencoder. 3. Thực nghiệm: So sánh hiệu quả giữa phương pháp học có giám sát và không giám sát. 4. Ứng dụng: Hệ thống cảnh báo real-time giả lập trên giao diện web	Nguyễn Huy Liêm	liemnh@huit.edu.vn	KHDL	Ứng dụng	

NGƯỜI LẬP



Lương Thị Quỳnh Mai

KT. TRƯỞNG KHOA
PHÓ TRƯỞNG KHOA



Nguyễn Thị Bích Ngân