ĐẠI HỌC QUỐC GIA TP HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC BÁCH KHOA

Khoa: Khoa học và Kỹ thuật Máy tính

**Biên bản các cuộc họp cho Bài tập lớn 1**

Môn học: Học máy (Machine Learning)

Mã môn học: CO3117

Học kỳ: 251, năm học 2025 – 2026

# Cuộc họp 1

**Thời gian:** 10/9/2025 – 21h00 – Google Meet

**Thành viên tham dự:** Phạm Khánh Duy, Tăng Hồng Ái, Nguyễn Trần Yến Nhi

**Mục tiêu:** Làm quen với đề bài, xác định yêu cầu của Bài tập lớn 1 và lựa chọn bộ dữ liệu phù hợp.

## Nội dung cuộc họp:

– Trao đổi chi tiết về yêu cầu bài tập, cấu trúc báo cáo, và cách chấm điểm.

– Tìm hiểu các bộ dữ liệu trên Kaggle để chọn dữ liệu có độ tin cậy và phù hợp.

– Phân công trách nhiệm bước đầu: một thành viên tìm dữ liệu, một thành viên đọc kỹ yêu cầu, một thành viên lập dàn ý sơ bộ.

## Kết quả:

– Nhóm thống nhất chọn **bộ dữ liệu Heart Disease** (nguồn Kaggle).

A screenshot of a computer

AI-generated content may be incorrect.

*Hình 1. Buổi họp thứ nhất*

# Cuộc họp 2

**Thời gian:** 12/9/2025 – 21h30 – Google Meet

**Thành viên tham dự:** Phạm Khánh Duy, Tăng Hồng Ái, Nguyễn Trần Yến Nhi

**Mục tiêu:** Xác định cách xử lý dữ liệu và xây dựng pipeline cơ bản.

## Nội dung cuộc họp:

– Thảo luận hướng xử lý dữ liệu: kiểm tra giá trị thiếu, xác định kiểu dữ liệu, chuẩn hóa các thuộc tính liên tục, mã hóa thuộc tính phân loại.

– Xây dựng pipeline sơ bộ: khám phá dữ liệu, tiền xử lý, chia dữ liệu, chọn mô hình thử nghiệm ban đầu.

## Kết quả:

– Hoàn thành pipeline cơ bản, chuẩn bị cho việc thử nghiệm huấn luyện và đánh giá ở bước tiếp theo

A screenshot of a video chat

AI-generated content may be incorrect.

*Hình 2. Buổi họp thứ hai*

# Cuộc họp 3

**Thời gian:** 15/9/2025 – 20h00 – Google Meet

**Thành viên tham dự:** Phạm Khánh Duy, Tăng Hồng Ái, Nguyễn Trần Yến Nhi

**Mục tiêu:** Hoàn thiện pipeline cho từng mô hình, tối ưu tham số và viết báo cáo sơ bộ.

## Nội dung cuộc họp:

– Hoàn thiện pipeline cho từng mô hình, tìm các tham số tối ưu hóa

– Bắt đầu viết báo cáo sơ bộ: mô tả dữ liệu, tiền xử lý, phương pháp tìm tham số tối ưu

## Kết quả:

– Có kết quả chạy tốt nhất cho Logistic Regression, Random Forest, SVM, KNN, Naive Bayes, Decision Tree. Báo cáo được viết các phần đầu.

A screenshot of a computer

AI-generated content may be incorrect.

*Hình 3. Buổi họp thứ ba*

# Cuộc họp 4

**Thời gian:** 18/9/2025 – 21h30 – Google Meet

**Thành viên tham dự:** Phạm Khánh Duy, Tăng Hồng Ái, Nguyễn Trần Yến Nhi

**Mục tiêu:** Rà soát toàn bộ kết quả, hoàn thiện báo cáo và tổ chức repo GitHub.

## Nội dung cuộc họp:

– Rà soát lại toàn bộ kết quả chạy mô hình và thống nhất chọn ra kết quả cuối cùng

– Hoàn thiện báo cáo: bổ sung kết quả tìm tham số, đánh giá, so sánh giữa các mô hình, thảo luận kết quả và kết luận

– Bổ sung mô tả chi tiết cho từng bước xử lý và huấn luyện mô hình trong Colab notebook

– Thảo luận cách trình bày repo GitHub

## Kết quả:

– Báo cáo hoàn chỉnh với đầy đủ nội dung, repo GitHub được tổ chức, tạo GitHub Page để trình bày trực quan

A screenshot of a video chat

AI-generated content may be incorrect.

*Hình 4. Buổi họp thứ tư*