

ĐẠI HỌC BÁCH KHOA HÀ NỘI  
TRƯỜNG CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

PHIẾU GIAO NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP HỆ CỬ NHÂN  
KỲ 2023.1

Thông tin về sinh viên

Họ và tên sinh viên: Trương Văn Hiến MSSV: 20194276  
Điện thoại liên lạc: 0904 371 156 Lớp: Kỹ thuật máy tính 03 - K64  
Email: hien.tv194276@sis.hust.edu.vn Mã lớp: 733525

Thông tin giáo viên hướng dẫn

Họ và tên GVHD: Nguyễn Hồng Quang  
Đồ án được thực hiện tại: Trường Công Nghệ thông tin và Truyền thông  
Thời gian làm ĐATN: Từ ngày 26/09/2023 đến ngày 15/01/2024

1. Tên đề tài:

Dự đoán bệnh di truyền của thai nhi dựa vào dữ liệu bản ghi y tế

2. Lĩnh vực đề tài:

- Lựa chọn 1: Công nghệ y tế (MedTech)
- Lựa chọn 2:
- Lựa chọn 3:
- Nếu lĩnh vực không nằm trong danh sách có sẵn, giáo viên hướng dẫn có thể đề xuất:

3. Mục tiêu của ĐATN:

3.1. Kiến thức sinh viên thu thập được:

- Tìm hiểu làm quen với bài toán trong lĩnh vực Tin sinh học
- Phương pháp tổng hợp, thu thập và thống kê dữ liệu
- Các phương pháp xử lý dữ liệu dạng bảng
- Kiến trúc, hoạt động, phương pháp đánh giá của các mô hình học máy tiên tiến trong lĩnh vực Medtech

3.2. Công nghệ sinh viên thu thập được:

- Ngôn ngữ lập trình Python
- Thư viện, công cụ hỗ trợ lập trình cho mô hình AI và ML

3.3. Kỹ năng sinh viên phát triển được:

- Kỹ năng quản lý, phân bổ thời gian hợp lý
- Kỹ năng giao tiếp, trao đổi vấn đề, đề nghị
- Kỹ năng nghiên cứu, tìm kiếm, giải quyết vấn đề

3.4. Sản phẩm kỳ vọng:

- Bộ dữ liệu chuẩn để dự đoán bệnh di truyền
- Mô hình được triển khai dự đoán bệnh di truyền dựa vào dữ liệu bản ghi y tế với độ chính xác tối thiểu 80%
- Báo cáo triển khai công việc

3.5. Vấn đề thực tiễn đồ án giải quyết:

- Ứng dụng công nghệ hiện của trí tuệ nhân tạo để thực hiện chăm sóc y tế, cung cấp giải pháp thiết thực cho cộng đồng về lĩnh vực di truyền học
- Triển khai các mô hình học máy, học sâu mạnh mẽ để có thể dự đoán chính xác các bệnh di truyền trong khoảng thời gian thai kì sớm nhất
- Có thể được tích hợp hệ thống thông tin y tế để thực hiện chẩn đoán và điều trị hiệu quả

4. Các nội dung sẽ thực hiện và kế hoạch triển khai:

Lưu ý: khối lượng yêu cầu đối với đồ án tốt nghiệp hệ cử nhân là 6(0-0-12-12), i.e. 12 tiết làm việc/tuần trong 17 tuần.

Nội dung 1: Tìm hiểu tổng quan về bài toán

từ Tuần 1 đến Tuần 2

Chi tiết:

- Nghiên cứu bước đầu về lĩnh vực Medtech và Tin sinh học
- Tìm hiểu xây dựng đề tài: Dự đoán bệnh di truyền của thai nhi dựa vào dữ liệu bản ghi y tế
- Tìm hiểu nghiệp vụ y học liên quan đến đề tài bài toán

Nội dung 2: Tìm hiểu tổng quan về công nghệ liên quan,

từ Tuần 3 đến Tuần 5

Chi tiết:

- Tìm hiểu, thử nghiệm các công nghệ dự đoán bệnh di truyền hiện có

- Thử nghiệm các thư viện hỗ trợ lập trình mô hình AI (Pytorch, Tensorflow...)

### Nội dung 3: Thu thập và xử lý dữ liệu

từ Tuần

6

đến Tuần

9

Chi tiết:

- Thu thập và phân tích dữ liệu
- Tìm hiểu nghiệp vụ y tế liên quan đến bộ dữ liệu
- Gán nhãn dữ liệu
- Tiền xử lý dữ liệu

### Nội dung 4: Xây dựng chương trình

từ Tuần

10

đến Tuần

14

Chi tiết:

- Tìm hiểu các mô hình dự đoán phổ biến hiện nay
- Đánh giá kết quả mô hình đã được triển khai: Random Forest, XGBoost, LightGBM, Catboost, ANNs
- Đề xuất mô hình phù hợp (FT-Transformer...)
- Thử nghiệm trên bộ dữ liệu và tối ưu mô hình đề xuất

### Nội dung 5: Thử nghiệm và đánh giá,

từ Tuần

15

đến Tuần

17

Chi tiết:

- Đánh giá mô hình dựa trên bộ dữ liệu thu thập
- Nhận ý kiến đánh giá, phản hồi từ giảng viên và bên nghiệp vụ y học
- Hoàn thiện báo cáo

### 5. Lời cam đoan của sinh viên đã nhận được nhiệm vụ

*Em xin cam kết sẽ hoàn thành các nhiệm vụ theo đúng kế hoạch.*

Hà Nội, ngày 27 tháng 10 năm 2023  
Sinh viên

Trương Văn Hiến

### 6. Xác nhận của giáo viên hướng dẫn về việc giao nhiệm vụ cho sinh viên

Hà Nội, ngày 27 tháng 10 năm 2023  
Giảng viên hướng dẫn

Nguyễn Hồng Quang