Tên: Đoàn Phúc

Lớp: 11CNPM1

MSSV: 1150080031

BÀI THỰC HÀNH SỐ 7

Tên đề tài: Ứng dụng thuật toán tối ưu ung thư gan trong lựa chọn đặc trưng để xây dựng mô hình chẩn đoán bệnh ung thư gan

Câu 1: Bản tôn chỉ gồm những nội dung nào? Nội dung nào là quan trọng nhất ?

1. Nội dung chính của một bản tôn chỉ dự án:

- Dựa theo mẫu “Employee Recognition Framework”, bản tôn chỉ dự án thường bao gồm 11 mục chính:

1. Project Identification – Thông tin nhận diện dự án → Gồm tên, mô tả, nhà tài trợ, quản lý dự án, nhóm thực hiện.

2. Business Reasons for Project – Lý do kinh doanh, mục tiêu tổ chức thực hiện dự án.

3. Project Objectives (Purpose) – Mục tiêu cụ thể, kết quả mong đợi.

4. Project Scope – Phạm vi công việc (bao gồm và loại trừ).

5. Key Project Deliverables – Các sản phẩm/deliverable đầu ra chính.

6. Milestone Dates – Các mốc thời gian quan trọng.

7. Key Issues – Các vấn đề quan trọng có thể ảnh hưởng đến tiến độ.

8. Risks – Các rủi ro tiềm ẩn và yếu tố ảnh hưởng.

9. Project’s Criteria for Success – Tiêu chí đánh giá thành công của dự án.

10. Critical Success Factors – Những yếu tố then chốt giúp dự án thành công.

11. Signoff – Phần xác nhận và phê duyệt (người bảo trợ, ngày ký).

2. Nội dung quan trọng nhất

- Mục tiêu dự án (Project Objectives) là quan trọng nhất, vì:

+ Nó xác định dự án nhằm đạt điều gì — là kim chỉ nam cho toàn bộ kế hoạch.

+ Nếu mục tiêu không rõ, các phần khác (phạm vi, tiến độ, sản phẩm, thành công) đều sẽ thiếu định hướng.

+ Mục tiêu cũng là căn cứ để đánh giá thành công cuối cùng.

Câu 2: Áp dụng mẫu trên vào đồ án cá nhân

Câu 3: Xây dựng bản tôn chỉ cho dự án cá nhân trên file word (Lab7\_BanTonChi.docx)

# BẢN TÔN CHỈ DỰ ÁN

|  |  |  |
| --- | --- | --- |
| 1.0 nhận diện dự án | | |
| **Tên** | ỨNG DỤNG THUẬT TOÁN TỐI ƯU UNG THƯ GAN TRONG LỰA CHỌN ĐẶC TRƯNG ĐỂ XÂY DỰNG MÔ HÌNH CHẨN ĐOÁN BỆNH UNG THƯ GAN | |
| **Mô tả** | Dự án tập trung nghiên cứu và ứng dụng các thuật toán tối ưu nhằm chọn ra các đặc trưng quan trọng nhất trong tập dữ liệu y học, từ đó xây dựng mô hình máy học giúp chẩn đoán sớm bệnh ung thư gan với độ chính xác cao. | |
| **Người bảo trợ** |  | |
| **Quản lý dự án** | Đoàn Phúc – Hồ Hữu Thịnh | |
| **Tài nguyên nhóm dự án** | Quản lý dự án  Thu thập và nghiên cứu dữ liệu  Lập trình viên  Phân tích mô hình  Kiểm thử và báo cáo | Công nghệ:Python  Datasets y tế công khai |

|  |
| --- |
| 2.0 Lý do thực hiện dự án |
| * Nâng cao kiến thức và kỹ năng ứng dụng thuật toán tối ưu trong lĩnh vực y học, đặc biệt trong chẩn đoán bệnh ung thư gan. * Giúp sinh viên thực hành toàn bộ quy trình của một dự án phần mềm, từ tìm hiểu bài toán, thu thập dữ liệu, xử lý, lựa chọn đặc trưng, đến xây dựng và đánh giá mô hình. * Ứng dụng công nghệ AI vào lĩnh vực chăm sóc sức khỏe, góp phần vào xu hướng phát triển y học thông minh * Tăng cường kỹ năng phân tích, xử lý dữ liệu và đánh giá mô hình học máy, phục vụ cho các nghiên cứu hoặc dự án sau này |

|  |
| --- |
| **3.0 mục tiêu dự án** |
| * Tổng thể là xây dựng một mô hình chẩn đoán bệnh ung thư gan có độ chính xác cao, dựa trên việc ứng dụng các thuật toán tối ưu để lựa chọn tập đặc trưngphù hợp nhất trong dữ liệu y học * Tìm hiểu và phân tích các thuật toán tối ưu phổ biến và đánh giá khả năng ứng dụng của chúng trong lĩnh vực y học. * Nghiên cứu bài toán lựa chọn đặc trưnghọc máy, nhằm giảm số lượng biến đầu vào nhưng vẫn giữ được độ chính xác cao của mô hình. * Ứng dụng thuật toán tối ưu được chọn để tìm ra tập đặc trưng quan trọng nhất từ dữ liệu chẩn đoán bệnh gan * Xây dựng và huấn luyện mô hình chẩn đoán ung thư gan, sử dụng các thuật toán học máy * Đánh giá, so sánh kết quả mô hình giữa các phương pháp có và không áp dụng lựa chọn đặc trưng tối ưu, thông qua các chỉ số như Accuracy, Precision, Recall, F1-score, ROC-AUC |

| **4.0 Phạm vi dự án** |
| --- |
| * Nghiên cứu và tìm hiểu các thuật toán tối ưu có khả năng ứng dụng trong bài toán lựa chọn đặc trưng dữ liệu y học. * Tìm hiểu bài toán lựa chọn đặc trưng trong học máy, áp dụng cho dữ liệu bệnh nhân ung thư gan. * Tiền xử lý và chuẩn hóa dữ liệu y học * Ứng dụng thuật toán tối ưu để lựa chọn tập đặc trưng tối ưu phục vụ cho mô hình chẩn đoán. * Xây dựng và huấn luyện mô hình chẩn đoán bệnh ung thư gan, sử dụng các thuật toán học máy * Đánh giá, so sánh và phân tích kết quả mô hình trước và sau khi áp dụng lựa chọn đặc trưng tối ưu. * Viết báo cáo, tài liệu kỹ thuật và trình bày kết quả nghiên cứu * Không thực hiện thu thập dữ liệu y học thực tế tại bệnh viện (sử dụng bộ dữ liệu công khai sẵn có). * Không nghiên cứu chi tiết các thuật toán học sâu (Deep Learning) do giới hạn thời gian và phạm vi đồ án môn học. * Dự án được thực hiện trong 14 tuần bao gồm các giai đoạn từ nghiên cứu, cài đặt, huấn luyện, đến báo cáo kết quả. |

|  |  |
| --- | --- |
| **5.0 các sản phẩm chính của dự án** | |
| **Tên** | **Mô tả** |
| Khung nghiên cứu | Kế hoạch tổng thể mô tả các bước thực hiện là thu thập dữ liệu, tiền xử lý, lựa chọn đặc trưng bằng thuật toán tối ưu, xây dựng và đánh giá mô hình chẩn đoán. |
| Bản tôn chỉ dự án | Tài liệu mô tả phạm vi, mục tiêu, thời gian, rủi ro và các thành phần của dự án |
| Mô hình logic dự án | Trình bày mối quan hệ giữa đầu vào (dữ liệu, thuật toán), hoạt động (training, test), đầu ra (mô hình, đặc trưng tối ưu) và kết quả đạt được |
| Tài liệu cơ sở | Tài liệu bổ sung cho bản tôn chỉ, bao gồm mô tả chi tiết về bài toán chẩn đoán ung thư gan, dữ liệu sử dụng, các biến y học và lý do chọn thuật toán tối ưu. |
| Tổng quan nghiên cứu | |  | | --- | |  |  |  |  |  | | --- | --- | --- | | |  | | --- | |  |  |  | | --- | | Tài liệu tổng hợp các nghiên cứu, thuật toán tối ưu, phương pháp lựa chọn đặc trưng và các  công trình liên quan trong y học và AI | | |
| Hướng dẫn | Hướng dẫn chi tiết quy trình: tiền xử lý dữ liệu, áp dụng thuật toán tối ưu, xây dựng mô hình và đánh giá độ chính xác. |
| Công cụ và tài nguyên | Bao gồm mã nguồn Python, notebook Jupyter, thư viện sử dụng (Scikit-learn, NumPy, Pandas...), cùng với hướng dẫn chạy mô hình. |
| Kế hoạch báo cáo và triển khai | Lộ trình báo cáo tiến độ, chuẩn bị slide trình bày, hướng dẫn cách chạy thử mô hình khi báo cáo đồ án. |
| Bảng kết quả trực quan | Biểu đồ, ma trận nhầm lẫn, đường cong ROC, và các kết quả đánh giá mô hình được trực quan hóa để phục vụ báo cáo. |

| **6.0 Thời gian cột mốc** | | |
| --- | --- | --- |
| **STT** | **Sự kiện/Giai đoạn chính** | **Thời gian dự kiến** |
| 1. | Xác định mục tiêu, phạm vi, phương pháp và công cụ nghiên cứu | Tuần 1 |
| 2. | Tìm hiểu các thuật toán tối ưu và bài toán lựa chọn đặc trưng. | Tuần 2, 3 |
| 3. | Làm sạch, mã hóa và chuẩn hóa dữ liệu bệnh gan. | Tuần 4,5 |
| 4. | Áp dụng thuật toán tối ưu để lựa chọn đặc trưng quan trọng nhất | Tuần 6,7 |
| 5. | Xây dựng và huấn luyện mô hình chẩn đoán | Tuần 8,9 |
| 6. | Đánh giá và so sánh kết quả mô hình | Tuần 10,11 |
| 7. | Hoàn thiện tài liệu báo cáo tiến trình dự án | Tuần 12,13 |
| 8. | Trình bày và nộp báo cáo dự án | Tuần 14 |

|  |  |
| --- | --- |
| **7.0 Các vấn đề chính** | |
| Mức độ | **Mô tả** |
| Cao | Thiếu nguồn dữ liệu y học chất lượng cao: Bộ dữ liệu công khai có thể không đủ đa dạng hoặc chứa nhiều giá trị thiếu, gây khó khăn cho việc huấn luyện mô hình. |
| Cao | Độ phức tạp của thuật toán tối ưu: Một số thuật toán (như GA, PSO, ACO) có thể tiêu tốn nhiều thời gian tính toán và yêu cầu tinh chỉnh tham số phức tạp. |
| Trung bình | Khó xác định tập đặc trưng tối ưu thực sự: Kết quả lựa chọn đặc trưng phụ thuộc vào chất lượng dữ liệu và cách đánh giá độ phù hợp của mô hình. |
| Trung bình | Giới hạn tài nguyên phần cứng và thời gian xử lý: Việc chạy mô hình nhiều lần để thử nghiệm có thể tốn thời gian nếu không có GPU hoặc máy tính mạnh |
| Thấp | Thiếu kinh nghiệm trong lĩnh vực y sinh học: Việc diễn giải các đặc trưng y học đòi hỏi hiểu biết chuyên ngành, trong khi nhóm dự án là sinh viên công nghệ. |
| Thấp | Khó khăn trong việc đánh giá và so sánh các mô hình: Sự khác biệt nhỏ trong kết quả (accuracy, F1-score, ROC) có thể khó xác định mô hình tối ưu nhất. |
| Thấp | Giới hạn về thời gian học kỳ: Dự án phải hoàn thành trong thời gian ngắn ảnh hưởng đến việc thử nghiệm nhiều phương pháp khác nhau. |

|  |  |
| --- | --- |
| **8.0 RỦI RO** | |
| **Mức độ** | **Mô tả** |
|  | **Rủi ro về dữ liệu** – Bộ dữ liệu y học có thể không đủ lớn hoặc mất cân bằng giữa các lớp (bệnh / không bệnh), dẫn đến mô hình học máy bị lệch và độ chính xác thấp. |
|  | **Rủi ro về hiệu quả của thuật toán tối ưu** – Thuật toán có thể không hội tụ hoặc không tìm được tập đặc trưng thật sự tối ưu, ảnh hưởng đến kết quả chẩn đoán |
|  | **Rủi ro về giới hạn tính toán** – Thời gian xử lý lâu hoặc thiếu tài nguyên phần cứng (CPU/GPU) có thể làm chậm tiến độ huấn luyện và đánh giá mô hình |
|  | **Rủi ro về lựa chọn mô hình học máy** – Việc chọn mô hình (SVM, RF, ANN…) không phù hợp với dữ liệu có thể dẫn đến kết quả không ổn định |
|  | **Rủi ro về kỹ năng chuyên môn** – Nhóm sinh viên có thể thiếu kinh nghiệm trong lĩnh vực y học, gây khó khăn khi phân tích ý nghĩa các đặc trưng bệnh lý. |
|  | **Rủi ro về tiến độ thực hiện** – Dự án có nhiều giai đoạn nghiên cứu, thử nghiệm, đánh giá nên có thể bị trễ tiến độ nếu không quản lý thời gian tốt. |

| 1. **Tiêu chí thành công của dự án** |
| --- |
| * Mô hình đạt độ chính xác ≥ 90% khi đánh giá bằng dữ liệu kiểm thử. * Số lượng đặc trưng được giảm ít nhất 30% so với tập ban đầu, nhưng độ chính xác mô hình không giảm quá 2%. * Thuật toán tối ưu hội tụ trong số vòng lặp cho phép (≤ 100 vòng) và thời gian huấn luyện không vượt quá giới hạn quy định. * Không còn giá trị thiếu (missing values) hoặc sai lệch vượt quá 5% tổng dữ liệu * Mã nguồn và mô hình có thể chạy lại được trên cùng bộ dữ liệu, cho kết quả tương đương trong ±2% sai số. * Báo cáo, hướng dẫn, biểu đồ và slide trình bày được nộp đầy đủ, đạt yêu cầu học phần. * Nhóm trình bày rõ ràng quy trình nghiên cứu, hiểu đúng ý nghĩa các đặc trưng y học và chỉ số đánh giá mô hình. * Dự án đạt điểm đánh giá ≥ 8/10 dựa trên tiêu chí: nội dung, kết quả, tính ứng dụng và báo cáo |

| **10.0 Yếu tố quan trọng quyết định thành công** |
| --- |
| * Giảng viên cung cấp định hướng, phản hồi kịp thời; các thành viên phối hợp chặt chẽ trong từng giai đoạn. * Các thành viên thường xuyên họp nhóm, chia sẻ tiến độ, phân công công việc rõ ràng, cập nhật thay đổi kịp thời. * Các thành viên chủ động hỗ trợ nhau trong quá trình nghiên cứu, xử lý dữ liệu và triển khai thuật toán. * Nhóm nắm vững mục tiêu, giới hạn phạm vi nghiên cứu và yêu cầu cụ thể của đề tài, tránh đi lệch hướng. * Dữ liệu phải chính xác, đầy đủ và phù hợp để đảm bảo kết quả huấn luyện mô hình đáng tin cậy. * Kết quả đánh giá cần dựa trên các chỉ số chuẩn và được kiểm chứng lại * Dự án không chỉ đạt kết quả kỹ thuật tốt mà còn thể hiện khả năng phân tích, trình bày, và ứng dụng kiến thức vào thực tế. |

| **11.0 Phê duyệt** |
| --- |
| Người thực hiện: Đoàn Phúc – Hồ Hữu Thịnh  Tên dự án: Ứng dụng thuật toán tối ưu ung thư gan trong lựa chọn đặc trưng để xây dựng mô hình chẩn đoán bệnh ung thư gan  Giảng viên hướng dẫn: Ths.Phạm Trọng Huynh  Ngày phê duyệt: 24/11/2025 |