**Danh sách đề tài**

01. Dự đoán doanh số bán hàng

02. Dự đoán bệnh tiểu đường

03. Dự đoán với chi phí bảo hiểm sức khoẻ

04. Dự đoán bệnh ung thư phổi

05. Dự đoán người sống sót trên tàu Titanic

06. Dự đoán trị giá kim cương

07. Dự đoán bệnh tim

08. Phát hiện gian lận thẻ tín dụng

09. Phân lớp thể loại nhạc

10. Dự đoán cho vay tín dụng

11. Dự đoán chất lượng rượu

12. Dự đoán đột quỵ

13. Dự đoán giá xe oto

14. Phân loại tin tức thông qua headline

15. Phân lớp thư rác qua email

16. Phân lớp khách hàng

17. Dự đoán chữ viết tay với mạng noron nhân tạo

18. Dự đoán khả năng vay tiền của khách hàng

19. Dự đoán khách hàng rời bỏ dịch vụ trong thương mại điện tử

20. Phát hiện tin tức giả mạo

**Nội dung triển khai**

1. Các đề tài sử dụng GoogleColab, Kaggle, Anacoda hoặc Jupyter trên máy cục bộ để triển khai theo các bước sau:

* Bước 0: Giới thiệu đề tài
* Bước 1: Dọc dữ liệu từ file csv, txt
* Bước 2: Phân tích dữ liệu
* Bước 3: Làm sạch dữ liệu
* Bước 4: trích chọn đặc trưng
* Bước 5: Xây dựng mô hình
* Bước 6: Huấn luyện mô hình
* Bước 7: Đánh giá thuật toán

1. Các đề tài sẽ chọn

* Cách 1: ít nhất từ 3 mô hình đề triển khai và so sánh đánh giá độ chính xác giữa các mô hình. (scikit-learn)
* Cách 2: cài đặt lại thuật toán sau đó so sánh khi triển khai với thư viện scikit-learn tương ứng với thuật toán đã chọn

1. Viết báo cáo

* Trang bìa
* Phân công đánh giá
* Tóm tắt ( Viết tóm tắt khoảng 1000 – 2000 từ)
* Giới thiệu (Tổng quan của đề tài)
* Xây dựng mô hình (Qui trình xây dựng đề tài)
* Phân tích dữ liệu
  + Giới thiệu về dữ liệu (dữ liệu được trích từ nguồn nào, các thành phần dữ liệu là gì,sử dụng pandas và matplotlib, seaborn để trực quan hoá dữ liệu)
  + Trích chọn đặc trưng (dựa vào độ tương quan dữ liệu)
  + Chuẩn hoá dữ liệu ( nêu dữ liệu dạng category hoặc miền dữ liệu không đồng nhất)
  + Phân chia dữ liệu (chia theo tỷ lệ nào )
* Huấn luyện mô hình ( huấn luyện theo các mô hình nào)
* Đánh giá mô hình ( tuỳ theo dạng sẽ lựa chọn cách đánh giá )
* Kết luận
* Tài liệu tham khảo

1. Thang điểm đánh giá

|  |  |
| --- | --- |
| Nội dung | Điểm |
| Source code (ipynb) – CLO6 | 2đ |
| Báo cáo (docx) – CLO6 | 1đ |
| Phân tích dữ liệu ( xử lý dữ liệu, làm sạch dữ liệu, trích chọn đặc trưng) - CLO1, CLO2 | 3đ |
| Huấn luyện, Kiểm tra, đánh giá mô hình – CLO3,CLO4 | 3đ |
| Trình bày - CLO5 | 1đ |