

QUY ĐỊNH THỰC HIỆN ĐỒ ÁN

MÔN PHÂN TÍCH DỮ LIỆU VÀ DỰ BÁO

I. TÊN ĐỒ ÁN

Sinh viên tự chọn đề án theo quy định. Đề án thực hiện 3SV/nhóm.

II. CÁC YÊU CẦU CHUNG CỦA ĐỒ ÁN:

Sau khi chọn xong tập dữ liệu, các nhóm cần thực hiện các yêu cầu sau và trình bày vào một quyển báo cáo đồ án (File word, khổ A4, trình bày như một luận văn tốt nghiệp). Các nội dung yêu cầu như sau:

- Giới thiệu tập dữ liệu: Giới thiệu chi tiết liên quan đến tập dữ liệu, các thuộc tính, kiểu dữ liệu, số dòng, số cột, ...
- Nêu rõ mục đích, yêu cầu của khách hàng mong muốn nhận được đối với tập dữ liệu khi giao cho chúng ta phân tích. Đặt ra những giả thuyết cần thiết trước khi thực hiện phân tích tiếp theo.
- Thực hiện quản lý dữ liệu: Làm sạch, Tạo biến phụ, Mã hóa lại, ...
- Xét mối quan hệ giữa các biến (các mối quan hệ này có liên quan đến mục đích của khách hàng). Nêu rõ lý do mình cần xét các mối quan hệ này.
- Thực hiện kiểm định các giả thuyết liên quan đến các mối quan hệ bên trên.
- Xây dựng mô hình hồi quy để mô hình hóa mối quan hệ (mô hình hồi quy này cần làm rõ yêu cầu của khách hàng, không cần thực hiện những mô hình không cần thiết). Xác định chất lượng của mô hình hồi quy thông qua các độ đo. Lựa chọn các biến giải thích để chọn mô hình có chất lượng tốt nhất.
- Xây dựng mô hình cây quyết định để thực hiện dự đoán các vấn đề của khách hàng yêu cầu. Xác định chất lượng của mô hình hồi quy thông qua các độ đo. Lựa chọn các biến giải thích và các thông số cần thiết để chọn mô hình có chất lượng tốt nhất.
- Có thể thực hiện thêm mô hình dự đoán khác để so sánh.
- So sánh các mô hình với nhau và rút ra kết luận. Viết báo cáo cho khách hàng.

III. HÌNH THỨC BÁO CÁO

- Thời gian nộp bài: Trước Buổi 15 (Theo dõi trên Classroom)
- Sinh viên nộp bài trên Classroom bao gồm: Source Code (Các notebook) và file word.
- Sinh viên nộp 1 quyển báo cáo bản in (in 1 mặt, đóng bìa, không cần bìa kiếng)
- File word gồm các nội dung:
 - Trang bìa (Có danh sách thành viên)
 - Trang phân công công việc
 - Mục lục
 - Lời mở đầu
 - Chương 1: Tổng quan và giới thiệu tập dữ liệu
 - Chương 2: Yêu cầu của khách hàng và các giả thuyết
 - Chương 3: Quản lý dữ liệu
 - Chương 4: Xem xét quan hệ giữa các biến
 - Chương 5: Mô hình hồi quy
 - Chương 6: Mô hình cây quyết định
 - Chương 7: Các mô hình khác (Nếu có)
 - Chương 8: Báo cáo khách hàng (Trường hợp bạn đã hoàn thành phân tích và xây dựng mô hình dự báo thành công. Hãy viết một chương để báo cáo cho khách hàng, những người không có chuyên ngành CNTT, về các kết quả mình đạt được và hướng dẫn sử dụng mô hình mình đưa ra.
 - Kết luận
- Hình thức trình bày file word: Sử dụng hình thức của một luận văn tốt nghiệp