TRƯỜNG ĐẠI HỌC THỦ DẦU MỘT
VIỆN KỸ THUẬT – CÔNG NGHỆ
****

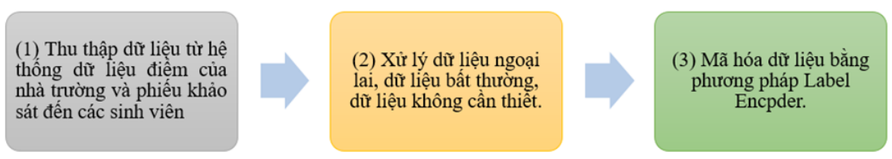
Ứng Dụng Học Máy Dự Đoán Khả Năng Tốt Nghiệp Của Sinh Viên Nhóm Ngành Hệ Thống Thông Tin

NTH: NGUYỄN QUỐC CƯỜNG

TÓM TẮT

"Ứng Dụng Học Máy Dự Đoán Khả Năng Tốt Nghiệp Của Sinh Viên Nhóm Ngành Hệ Thống Thông Tin" được xây dựng bằng các phương pháp học máy như sau: Phân tích, so sánh và trực quan hóa các dữ liệu được thu thập để thấy được sự phân hóa tập trung của các trường yếu tố; phân tích các yếu tố ảnh hưởng đến kết quả tốt nghiệp của sinh viên bằng các độ đo Pearson và Spearman; sau đó sử dụng mô hình học máy Hồi quy tuyến tính (Logistic Regression) để huấn luyện model.

Xử lý dữ liệu



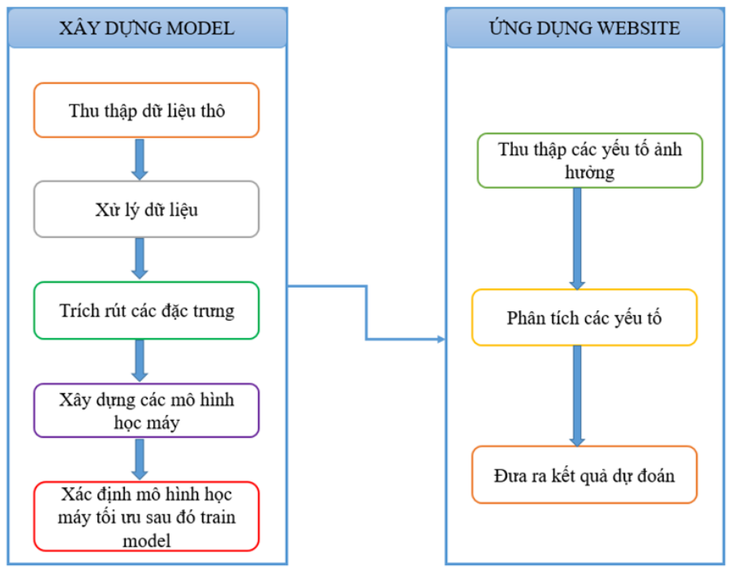
Phân tích đặc trưng

Để phân tích các đặc trưng ảnh hưởng đến mô hình dự đoán khả năng tốt nghiệp của sinh viên nhóm ngành hệ thống thông tin dựa trên đánh giá của 2 độ đo Pearson và Pearman.

+1	Mối tương quan tích cực hoàn toàn
+0.8	Mối tương quan tích cực mạnh mẽ
+0.6	Mối tương quan dương vừa phải
0	Không có mối tương quan
-0.6	Mối tương quan âm vừa phải
-0.8	Mối tương quan tiêu cực mạnh mẽ
-1	Hoàn thành mối tương quan tiêu cực

STT	Yếu tố	Pearson	Pearman
1	Điểm rèn luyện năm 2018-2019	-0.131253	-0.120080
2	Điểm rèn luyện năm 2019-2020	-0.162948	-0.170915
3	Điểm rèn luyện năm 2020-2021	-0.063039	-0.075719
4	Điểm rèn luyện trung bình các năm	-0.214139	-0.224890
5	Điểm trung bình môn năm 2018-2019	0.366668	0.349188
6	Điểm trung bình môn năm 2019-2020	0.545840	0.537225
7	Điểm trung bình môn năm 2020-2021	0.587375	0.566444
8	Số giờ làm thêm	-0.63317	-0.604897
9	Số môn chưa học	-0.690331	-0.831886
10	Số môn còn nợ	-0.700046	-0.869359
11	Điểm trung bình tích lũy các năm	0.550909	0.548107

Bảng 1: Đánh giá hệ số tương quan

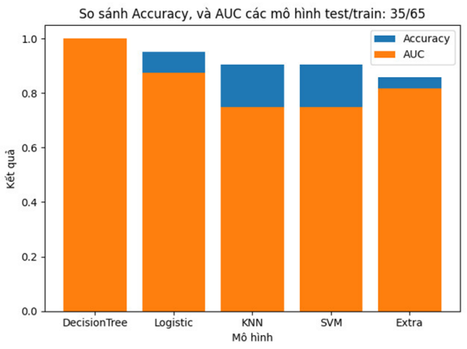


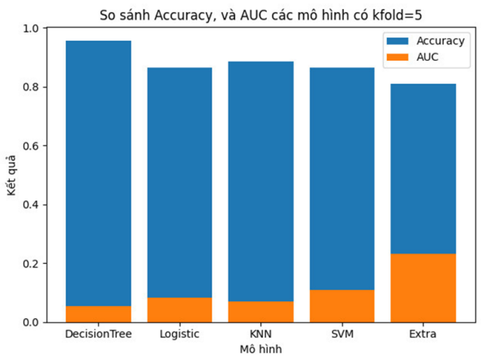
Mô hình tổng quát

Nhiệm vụ của hệ thống dự đoán khả năng tốt nghiệp của sinh viên là xử lý dữ liệu dự trên các yếu tố ảnh hưởng, từ đó sử dụng thuật toán học máy tối ưu để phân tích và đưa ra tỷ lệ phầm trăm tương đối chính xác so với dữ liệu được đưa vào.

Mô hình dự đoán

Để đánh giá mô hình xây dựng tối ưu nhất ta cần các phương pháp đánh giá chuẩn xác. Đối với hệ thống này chúng tôi đã sử dụng 2 phương pháp đánh giá đưa ra độ chính xác cao nhất và chuẩn xác nhất trong các loại phương pháp đó là Accucary và AUC.

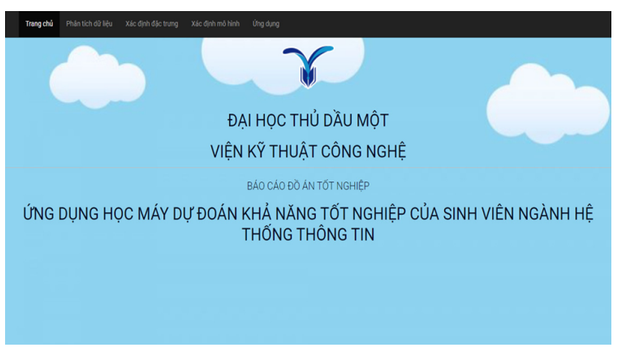




KẾT QUẢ

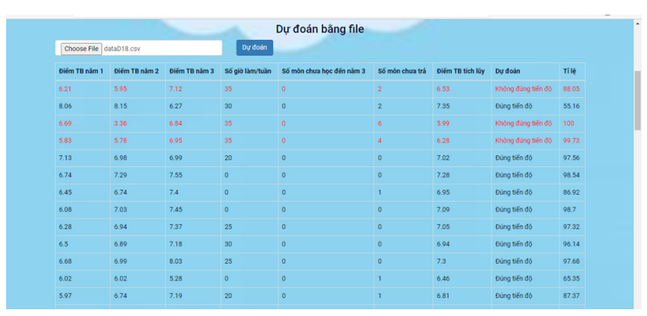
Từ các kết quả thực nghiệm trên dù thay đổi tỉ lệ Test/Train và Kfold = 5. Cho ta thấy mô hình Decision Tree Classifier (Cây quyết định) cho độ chính xác Accucary và AUC với giá trị cao nhất. Tuy nhiên xét về tổng thể mô hình Decision Tree Classifier bị hạn chế thì mức độ chính xác không được ổn định như mô hình Logistic Regression. Vì vậy, ta quyết định chọn mô hình Logistic Regression là mô hình tối ưu nhất để xây dựng ứng dụng dự đoán khả năng tố nghiệp của sinh viên ngành Hệ thống thông tin.

Website



Hình 5: Giao diện trang chủ

Hình 6: Giao diện dự đoán



KẾT LUẬN

xây dựng thành công hệ thống dự đoán khả năng tốt nghiệp của sinh viên bằng các phương pháp học máy. Tôi mong muốn sẽ phát triển hệ thống dự đoán khả năng tốt nghiệp của sinh một cách hoàn thiện hơn, nhiều chức năng bổ ích và giao diện được đẹp mắt hơn. Website không chỉ phục vụ cho nhóm ngành Hệ thống thông tin mà chúng tôi còn mong muốn nó có thể phát triển cho tất cả các bạn sinh viên trong và ngoài trường. Ngoài ra, nó còn có thể phát triển thành một tính năng trên trang đăng ký môn học của trường.