

LỜI CAM ĐOAN

Tôi cam đoan luận văn thạc sĩ công nghệ thông tin “**Ứng dụng khai phá dữ liệu xây dựng hệ thống phân tích hoạt động tiếp thị ngân hàng**” là do tôi nghiên cứu, tổng hợp và thực hiện.

Toàn bộ nội dung luận văn, những điều được trình bày là của chính cá nhân tôi hoặc được tham khảo, tổng hợp từ nhiều nguồn tài liệu khác nhau. Tất cả các tài liệu tham khảo, tổng hợp đều được trích xuất với nguồn gốc rõ ràng. Các số liệu, kết quả nêu trong luận văn là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

TP.HCM, ngày tháng năm 2022

Học viên thực hiện luận văn

Mai Thị Hồng Hạnh

LỜI CẢM ƠN

Trong quá trình nghiên cứu và thực hiện luận văn “**Ứng dụng khai phá dữ liệu để đưa ra những dự đoán trong hoạt động tiếp thị trong ngân hàng**” tôi xin chân thành cảm ơn sâu sắc đến cô PGS.TS. Võ Thị Lưu Phương - Khoa CNTT - Trường Đại học Quốc tế - ĐHQG TP.HCM đã tận tâm dìu dắt, hướng dẫn, góp ý, chỉnh sửa để tôi hoàn thành luận văn này.

Tôi trân trọng gửi lời cảm ơn chân thành đến các Thầy cô Khoa Công nghệ Thông tin - Học viện Bưu chính Viễn thông đã truyền đạt những kiến thức vô cùng quý giá và đã tạo mọi điều kiện thuận lợi cho tôi trong suốt thời gian học tập và nghiên cứu tại trường.

Đồng thời tôi cũng xin cảm ơn cơ quan, gia đình, bạn bè cùng lớp M20CQIS01-N đã luôn ủng hộ, động viên, tạo điều kiện thuận lợi cho tôi hoàn thành luận văn này.

Chân thành cảm ơn./.

TP.HCM, ngày tháng năm 2022

Học viên thực hiện luận văn

Mai Thị Hồng Hạnh

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
MỤC LỤC.....	iii
DANH MỤC CÁC KÝ HIỆU CHỮ VIẾT TẮT	v
DANH MỤC BẢNG BIỂU	vi
DANH MỤC HÌNH ẢNH	vii
MỞ ĐẦU	Trang 1
1. Lý do chọn đề tài	Trang 1
2. Tổng quan về vấn đề nghiên cứu	Trang 3
3. Mục đích nghiên cứu.....	Trang 3
4. Đối tượng và phạm vi nghiên cứu.....	Trang 3
5. Giả thuyết nghiên cứu	Trang 4
6. Câu hỏi nghiên cứu	Trang 4
7. Phương pháp nghiên cứu.....	Trang 4
Chương 1: CƠ SỞ LÝ LUẬN	Trang 5
1.1. Tổng quan về phát hiện tri thức và khai phá dữ liệu.....	Trang 5
1.2. Quá trình phát hiện tri thức và khai phá dữ liệu.....	Trang 5
1.3. Các phương pháp khai phá dữ liệu.....	Trang 7
1.4. Mô hình khai phá dữ liệu	Trang 7
1.5. Kết luận	Trang 9
Chương 2: PHƯƠNG PHÁP TIẾP CẬN THEO HƯỚNG DỮ LIỆU ĐỀ DỰ ĐOÁN SỰ THÀNH CÔNG CỦA TIẾP THỊ QUA ĐIỆN THOẠI NGÂN HÀNG.....	Trang 10
2.1. Tổng quan cơ sở dữ liệu Tiếp thị ngân hàng	Trang 10
2.2. Phân tích yêu cầu chức năng tập dữ liệu	Trang 10

2.3. Thiết kế hệ thống	Trang 11
2.4. Xây dựng cơ sở dữ liệu tiếp thị ngân hàng.....	Trang 12
2.4.1. Cơ sở bộ dữ liệu	Trang 13
2.4.2. Giá trị dữ liệu xây dựng trong các cột của tập dữ liệu	Trang 15
2.5. Mục tiêu của mô hình	Trang 21
2.6. Đánh giá mô hình	Trang 21
Chương 3: KHAI PHÁ DỮ LIỆU BẰNG MÔ HÌNH HỒI QUY	
LOGISTIC, CÂY QUYẾT ĐỊNH	Trang 22
3.1. Mô hình Logistic	Trang 22
3.1.1. Các bước thực hiện mô hình	Trang 22
3.1.2. Kết quả mô hình dự đoán	Trang 43
3.2. Mô hình Cây Quyết định (DT).....	Trang 44
3.2.1. Các bước thực hiện mô hình	Trang 44
3.2.2. Kết quả mô hình dự đoán	Trang 60
3.3. So sánh mô hình Hồi quy Logistic và Cây quyết định (DT).....	Trang 65
KẾT LUẬN	Trang 67
DANH MỤC TÀI LIỆU THAM KHẢO	Trang 69

DANH MỤC CÁC KÝ HIỆU CHỮ VIẾT TẮT

Viết tắt	Tiếng Anh	Tiếng Việt
CNTT	Information technology	Công nghệ thông tin
CSDL	Database	Cơ sở dữ liệu
DM	Data mining	Khai phá dữ liệu
KDD	Knowledge Discovery in Databases	Khám phá kiến thức CSDL
KPDL	Data mining	Khai phá dữ liệu
LR	Logistic regression	Hồi quy logistic
DT	Decision tree	Cây quyết định
DSS	Decision support system	Hệ thống hỗ trợ quyết định

DANH MỤC BẢNG BIỂU

Bảng 2.1: Dữ liệu tiếp thị ngân hàng (Tập dữ liệu bank-full.csv).....	14
Bảng 2.2: Số lượng cột công việc.....	15
Bảng 2.3: Số lượng cột hôn nhân.....	17
Bảng 3.1: Mô tả dữ liệu ngân hàng.....	23
Bảng 3.2: Danh sách khách hàng tiềm năng.....	44
Bảng 3.3: So sánh 2 mô hình.....	69

DANH MỤC HÌNH ẢNH

Hình 1.1: Các giai đoạn trong quá trình khai phá dữ liệu	6
Hình 1.2: Minh họa cây quyết định.....	9
Hình 2.1: Khách hàng có đăng ký tiền gửi hay không đăng ký tiền gửi.....	10
Hình 2.2 – Sơ đồ hệ thống cơ sở dữ liệu.....	11
Hình 2.3: Tỷ lệ khách hàng đăng ký tiền gửi hay không đăng ký theo nghề nghiệp.....	16
Hình 2.4: Tỷ lệ người trong hôn nhân có đăng ký tiền gửi hay không đăng ký	17
Hình 2.5: Tỷ lệ người trong vỡ nợ có đăng ký tiền gửi hay không đăng ký....	18
Hình 2.6: Tỷ lệ người có nhà có đăng ký tiền gửi hay không đăng ký	19
Hình 2.7: Tỷ lệ người có khoản vay có đăng ký tiền gửi hay không đăng ký	19
Hình 2.8: Tỷ lệ người đợt trước đó có đăng ký tiền gửi hay không đăng ký...	20
Hình 3.1: Biểu đồ tương quan với cột mục tiêu.....	27
Hình 3.2: Biểu đồ công việc.....	28
Hình 3.3: Biểu đồ tình trạng hôn nhân.....	28
Hình 3.4: Biểu đồ đợt tiếp thị trước đó	29
Hình 3.5: Tỷ lệ khách hàng đăng ký tiền gửi và không đăng ký	30
Hình 3.6: Biểu đồ tròn tỷ lệ đăng ký tiền gửi.....	31
Hình 3.7: Tỷ lệ đăng ký thường xuyên	32
Hình 3.8: Tỷ lệ tình trạng hôn nhân	33
Hình 3.9: Tầng suất tình trạng hôn nhân so với đăng ký	34
Hình 3.10: Tỷ lệ khách hàng có tín dụng vỡ nợ.....	35
Hình 3.11: Tỷ lệ khách hàng có nhà ở	36
Hình 3.12 – Mô tả tập dữ liệu không cân bằng.....	44
Hình 3.13: Tỷ lệ khách hàng có vay	45
Hình 3.14: Tỷ lệ khách hàng của tiếp thị trước đó.....	46
Hình 3.15: Tỷ lệ ngày.....	46

Hình 3.16: Tỷ lệ khách hàng có hoặc không đăng ký theo tháng	47
Hình 3.17: Tỷ lệ khách hàng có hoặc không đăng ký ở tuổi 38-39	48
Hình 3.18: Biểu đồ tỷ trọng khách hàng có hoặc không đăng ký ở tuổi 38-39.....	48
Hình 3.19: Biểu đồ tỷ trọng thuộc tính số dư của khách hàng.....	49
Hình 3.20: Biểu đồ tỷ trọng thuộc tính thời lượng của khách hàng.....	50
Hình 3.21: Biểu đồ tỷ trọng số lượng liên hệ của đợt	50
Hình 3.22: Biểu đồ tỷ trọng số lượng liên hệ của đợt	51
Hình 3.23: Số ngày trôi qua sau khi khách hàng được liên hệ lần cuối	51
Hình 3.24: Số lượng địa chỉ liên hệ được thực hiện trước đợt này	51
Hình 3.25: Mối tương quan đến biến mục tiêu.....	53
Hình 3.26: Hình dạng cây quyết định.....	58
Hình 3.27: Độ chính xác so với alpha cho bộ đào tạo và thử nghiệm.....	59
Hình 3.28: Kết quả mô hình cây quyết định.....	60
Hình 3.29: Kết quả mô hình độ chính xác cây quyết định 83%	61
Hình 3.30: Kết quả mô hình cây quyết định sau khi cắt tỉa	63
Hình 3.31: Kết quả mô hình độ chính xác cây quyết định 82%	64