

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

----- o0o -----

VŨ THỊ LAN ANH

NGHIÊN CỨU PHƯƠNG PHÁP PHÂN TÍCH HỒI QUY
ỨNG DỤNG TRONG PHÂN TÍCH DỮ LIỆU KÊ KHAI NỘP
THUẾ PHỤC VỤ THANH TRA

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN - 2016

Số hóa bởi Trung tâm Học liệu – ĐHTN <http://www.lrc.tnu.edu.vn>

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

----- o0o -----

VŨ THỊ LAN ANH

NGHIÊN CỨU PHƯƠNG PHÁP PHÂN TÍCH HỒI QUY
ỨNG DỤNG TRONG PHÂN TÍCH DỮ LIỆU KÊ KHAI NỘP
THUẾ PHỤC VỤ THANH TRA

Chuyên ngành: Khoa học máy tính

Mã số: 60480101

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Người hướng dẫn khoa học

GS.TS.Vũ Đức Thi

Số hóa bởi Trung tâm Học liệu – ĐHTN <http://www.lrc.tnu.edu.vn>

THÁI NGUYÊN - 2016

Số hóa bởi Trung tâm Học liệu – ĐHTN <http://www.lrc.tnu.edu.vn>

LỜI CẢM ƠN

Trong thời gian hai năm của chương trình đào tạo thạc sỹ, trong đó gần một nửa thời gian dành cho các môn học, thời gian còn lại dành cho việc lựa chọn đề tài, giáo viên hướng dẫn, tập trung vào nghiên cứu, viết, chỉnh sửa và hoàn thiện đề tài. Với quỹ thời gian như vậy và với vị trí công việc đang phải đảm nhận, không riêng bản thân em mà hầu hết các sinh viên cao học muốn hoàn thành tốt luận văn của mình trước hết đều phải có sự sắp xếp thời gian hợp lý, có sự tập trung học tập và nghiên cứu với tinh thần nghiêm túc, nỗ lực hết mình; tiếp đến cần có sự ủng hộ về tinh thần, sự giúp đỡ về chuyên môn một trong những điều kiện không thể thiếu quyết định đến việc thành công của đề tài.

Để hoàn thành được đề tài này trước tiên em xin gửi lời cảm ơn đến thầy giáo hướng dẫn GS.TS Vũ Đức Thi, thầy đã có những định hướng cho em về nội dung và hướng phát triển, đã có những đóng góp quý báu cho em về những vấn đề chuyên môn của đề tài, giúp em tháo gỡ kịp thời những vướng mắc trong quá trình làm luận văn.

Em xin gửi lời cảm ơn tới Tập thể cán bộ nghiên cứu nhóm thực hiện đề tài do TS Nguyễn Long Giang Viện Công nghệ Thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam làm chủ nhiệm đã cung cấp đầy đủ các số liệu thu được từ Chi cục thuế Hà Nội hỗ trợ và giúp đỡ Em tiến hành các thử nghiệm thành công.

Em cũng xin cảm ơn các thầy cô giáo Trường Đại học Công nghệ thông tin và Truyền thông Thái Nguyên, cũng như bạn bè cùng lớp đã có những ý kiến đóng góp bổ sung cho đề tài luận văn của em. Xin cảm ơn gia đình, người thân cũng như đồng nghiệp luôn quan tâm, ủng hộ hỗ trợ về mặt tinh thần trong suốt thời gian từ khi nhận đề tài đến khi hoàn thiện đề tài này.

Trong nội dung của luận văn chắc chắn còn nhiều thiếu sót. Em rất mong các Thầy cô cùng bạn bè đóng góp để bản luận văn của Em được hoàn thiện hơn.

Em xin trân trọng cảm ơn.

Thái Nguyên, ngày 20 tháng 08 năm 2016

Sinh viên

Vũ Thị Lan Anh

LỜI CAM ĐOAN

Em xin cam đoan đây là công trình nghiên cứu của riêng em.

Các số liệu, trong luận văn được sử dụng từ nguồn số liệu được cung cấp bởi chi cục thuế Hà Nội thực hiện năm 2014. Kết quả nghiên cứu này chưa được sử dụng trong bất kể nghiên cứu nào khác.

Thái Nguyên, ngày 20 tháng 8 năm 2016

TÁC GIẢ

Vũ Thị Lan Anh

MỤC LỤC

MỞ ĐẦU	1
Chương 1. Tổng quan hồi quy, bài toán tính điểm rủi ro thuế	3
1.1. Kiến thức cơ bản về phân tích hồi quy	3
1.1.1 Khái niệm cơ bản về hàm nội suy	3
1.1.2 Bài toán hồi quy	8
1.2. Phát biểu bài toán phân tích điểm rủi ro	13
1.2.1 Mục tiêu của bài toán.....	13
1.2.2 Yêu cầu của bài toán phân tích tính điểm rủi ro.....	13
Chương 2. Mô hình hồi quy tuyến tính.....	16
2.1. Mô hình hồi quy tuyến tính đơn	16
2.1.1 Vấn đề mô hình hồi quy.....	16
2.1.2 Ước lượng hệ số hồi quy	18
2.1.3 Tính chất của ước lượng của các hệ số hồi quy.....	19
2.1.4 Kiểm định giả thuyết	20
2.1.5 Khoảng tin cậy	23
2.1.6 Tính phù hợp của mô hình.....	27
2.1.7 Tuyến tính hóa một số mô hình	31
2.2. Mô hình hồi quy tuyến tính bội	32
2.2.1 Phương trình hồi quy	32
2.2.2 Ước lượng hệ số hồi quy và tính chất của ước lượng	34
2.2.3 Kiểm định giả thuyết	35
2.2.4 Ước lượng và dự đoán	37
2.2.5 Phân tích phần dư	38
2.2.6 Sử dụng phần mềm.....	38
2.2.7 Lựa chọn biến và xây dựng mô hình	41
2.3. Phân tích hồi quy logistic.....	49
2.3.1 Mô hình hồi quy logistic	49
2.3.2 Phân tích hồi quy logistic đa thức	51
Chương 3. Bài toán tính điểm rủi ro của doanh nghiệp phục vụ thanh tra, kiểm tra thuế	53
3.1. Tính cấp thiết xây dựng hệ thống tính điểm rủi ro của doanh nghiệp phục vụ thanh tra, kiểm tra thuế	53
3.2. Bài toán tính điểm rủi ro của doanh nghiệp.....	54
3.3. Các bước xây dựng mô hình giải quyết bài toán	56

3.3.1	<i>Xây dựng mô hình hồi quy tuyến tính bội</i>	57
3.3.2	<i>Mô hình tổng thể bài toán tính điểm rủi ro cho doanh nghiệp</i>	59
3.3.3	<i>Thực thi mô hình để tính điểm rủi ro cho doanh nghiệp</i>	60
3.3.4	<i>Tiền xử lý dữ liệu</i>	61
3.4.	<i>Thử nghiệm và đánh giá kết quả</i>	63
3.4.1	<i>Mục tiêu thử nghiệm</i>	63
3.4.2	<i>Công cụ thử nghiệm và môi trường thử nghiệm</i>	63
3.4.3	<i>Quy trình thử nghiệm</i>	64
DANH MỤC TÀI LIỆU THAM KHẢO		70
PHỤ LỤC		71

Danh mục hình vẽ

Hình 2.1. Đồ thị rải điểm, đường hồi quy cho số liệu độ sạch của oxy	17
Hình 2.2. Độ lệch và các đường hồi quy lý thuyết, thực nghiệm	18
Hình 2.3. Khoảng tin cậy (2 đường Hyperbol giữa) và khoảng dự đoán (2 đường hyperbol ngoài) cho mức tiêu thụ nhiên liệu	26
Hình 2.4. Độ lệch và các đường hồi quy lý thuyết, thực nghiệm	27
Hình 2.5. Dáng điệu phần dư	28
Hình 2.6. Đồ thị phần dư chuẩn hóa cho số liệu độ kéo	31
Hình 2.7. Miền biến thiên của các biến hồi quy	38
Hình 2.8. Đồ thị xác suất chuẩn và phần dư chuẩn hóa của số liệu lực kéo	41
Hình 2.9. Phần dư chuẩn hóa theo quan sát của số liệu độ tan	48
Hình 2.10. Mối liên hệ giữa $\text{logit}(p)$ và p , cho $0 < p < 1$	49

Danh mục bảng

Bảng 1.1. Độ sạch của oxy ứng với tỷ lệ phần trăm hydrocarbon.....	16
Bảng 2.1. Phân tích phương sai để kiểm định tính hiệu quả của hồi quy.....	23
Bảng 2.2. Phân tích hệ số và phân tích phương sai cho Ví dụ 2.2.....	30
Bảng 2.4. Số liệu cho mô hình hồi quy bội.....	32
Bảng 2.5. Kết quả xử lý với số liệu lực kéo dây dẫn.....	39
Bảng 2.6. Tóm tắt, phân tích phương sai và phân tích hệ số cho Ví dụ 2.3	40
Bảng 2.7. Số liệu độ tan	45
Bảng 2.8. Tóm tắt, phân tích phương sai và phân tích các hệ số cho mô hình đầy đủ của số liệu độ tan.....	47
Bảng 2.9. Tóm tắt, phân tích phương sai, phân tích hệ số của mô hình cuối cùng theo phương pháp cân nhắc từng bước của số liệu độ tan	48

MỞ ĐẦU

Dữ liệu thống kê rất đa dạng và phong phú, khai thác được dữ liệu này góp phần đưa ra được những ý kiến tốt cho kết luận đối với hoạt động của các công ty, doanh nghiệp. Dựa trên các phương pháp phân tích dữ liệu hồi quy trên cơ sở đề xuất phương pháp hiệu quả.

Vấn đề gian lận trong thuế có ảnh hưởng lớn trong nguồn thu ngân sách của các quốc gia. Dựa trên phân tích dữ liệu hồi quy phân tích rủi ro nhằm phát hiện các gian lận của người nộp thuế và các doanh nghiệp tham gia đóng thuế để đưa ra quyết định thanh tra trường hợp có biểu hiện gian lận thuế. Kết quả phân tích này nhằm phát hiện trường hợp gian lận đem lại lợi ích tăng nguồn thu ngân sách, thúc đẩy sự phát triển kinh tế, xã hội của đất nước. Phân tích rủi ro là một phương pháp quản lý khoa học và hiệu quả để giải quyết bài toán quản lý thuế trong điều kiện số lượng người nộp thuế ngày càng tăng lên nhanh chóng, mức độ quản lý ngày càng phức tạp do trình độ của người nộp thuế ngày càng cao, các hoạt động kinh tế phát sinh ngày càng đa dạng và phong phú dẫn đến khối lượng công việc của công tác quản lý thuế tăng lên vượt quá khả năng tăng nguồn lực bị hạn chế nhất định của các cơ quan thuế.

Từ các nội dung được nêu trên tôi lựa chọn đề tài “*Nghiên cứu phương pháp phân tích hồi quy ứng dụng trong phân tích dữ liệu kê khai nộp thuế phục vụ thanh tra*” làm luận văn tốt nghiệp thạc sỹ của mình.

Nội dung chính của luận văn gồm 3 chương bao gồm

Chương 1: Luận văn trình bày cơ sở toán học trong việc xác định công thức hàm nội suy và hàm hồi quy cùng các thuật toán tương ứng, đây là một lĩnh vực quan trọng của toán học đối với lớp các bài toán thực nghiệm nhằm xây dựng các công thức gần đúng miêu tả mối ràng buộc giữa các số liệu xuất hiện trong các thí nghiệm tại các phòng thí nghiệm. các kiến thức này là rất cần thiết làm cơ sở để nghiên cứu các nội dung trong luận văn.