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

Nguyễn Như Thế

NGHIÊN CỨU CÁC PHƯƠNG PHÁP PHÂN LỚP DỮ LIỆU VÀ ỨNG DỤNG TRONG BÀI TOÁN DỰ BÁO THUÊ BAO RỜI MẠNG VIỄN THÔNG

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên - 2016

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

Nguyễn Như Thế

NGHIÊN CỨU CÁC PHƯƠNG PHÁP PHÂN LỚP DỮ LIỆU VÀ ỨNG DỤNG TRONG BÀI TOÁN DỰ BÁO THUÊ BAO RỜI MẠNG VIỄN THÔNG

Chuyên ngành: Khoa học máy tính
Mã số: 60 48 0101

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS NGUYỄN LONG GIANG

Thái Nguyên - 2016

LỜI CAM ĐOAN

Tên tôi là: **Nguyễn Như Thế**

Sinh ngày: 12/12/1989

Học viên lớp cao học: CHK13E - Trường Đại học Công nghệ thông tin và Truyền thông – Đại học Thái Nguyên.

Hiện đang công tác tại: Sở Thông tin và Truyền thông tỉnh Phú Thọ

Xin cam đoan: Đề tài ***"Nghiên cứu các phương pháp phân lớp dữ liệu và ứng dụng trong bài toán dự báo thuê bao rời mạng viễn thông"*** do Thầy giáo **TS. Nguyễn Long Giang** hướng dẫn là công trình nghiên cứu của riêng tôi. Tất cả tài liệu tham khảo đều có nguồn gốc, xuất xứ rõ ràng.

Tác giả xin cam đoan tất cả những nội dung trong luận văn đúng như nội dung trong đề cương và yêu cầu của thầy giáo hướng dẫn. Nếu sai tôi hoàn toàn chịu trách nhiệm trước hội đồng khoa học và trước pháp luật.

Thái Nguyên, ngày 28 tháng 6 năm 2016

HỌC VIÊN

Nguyễn Như Thế

LỜI CẢM ƠN

Sau một thời gian nghiên cứu và làm việc nghiêm túc, được sự động viên, giúp đỡ và hướng dẫn tận tình của Thầy giáo hướng dẫn **TS. Nguyễn Long Giang**, luận văn với đề tài ***"Nghiên cứu các phương pháp phân lớp dữ liệu và ứng dụng trong bài toán dự báo thuê bao rời mạng viễn thông"***đã hoàn thành.

Tôi xin bày tỏ lòng biết ơn sâu sắc đến:

Thầy giáo hướng dẫn **TS. Nguyễn Long Giang** đã tận tình chỉ dẫn, giúp đỡ tôi hoàn thành luận văn này.

Tôi xin bày tỏ lòng biết ơn đến các thầy cô trong Trường Đại học Công nghệ thông tin và Truyền thông – Đại học Thái Nguyên đã giúp đỡ tôi trong quá trình học tập cũng như thực hiện luận văn.

Tôi xin cảm ơn Chi nhánh Mobifone Phú Thọ đã nhiệt tình giúp đỡ, cung cấp thông tin trong quá trình nghiên cứu, thực nghiệm chương trình luận văn.

Tôi xin chân thành cảm ơn bạn bè, đồng nghiệp và gia đình đã động viên, khích lệ, tạo điều kiện giúp đỡ tôi trong suốt quá trình học tập, thực hiện và hoàn thành luận văn này.

Thái Nguyên, ngày 28 tháng 6 năm 2016

HỌC VIÊN

Nguyễn Như Thế

MỤC LỤC

DANH MỤC CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT

TT	Thuật ngữ	Định nghĩa
1.	KPDL	Khai phá dữ liệu
2.	KDD	Knowledge Discovery and Data Mining
3.	NB	Naïve Bayes
4.	SVM	Support vector machine
5.	NN	Neural Networks

DANH MỤC HÌNH ẢNH

DANH MỤC BẢNG BIỂU

MỞ ĐẦU

Sự bùng nổ và phát triển của ngành công nghệ thông tin đã làm lượng dữ liệu được thu thập và lưu trữ ở các hệ thống thông tin tăng lên một cách nhanh chóng. Trước tình hình đó, việc khai thác và chọn lọc những dữ liệu có ích, tiền ẩn từ lượng dữ liệu khổng lồ này là rất cần thiết. Các tri thức trích lọc từ dữ liệu sẽ giúp các cơ quan, tổ chức đưa ra những dự báo và điều hành hiệu quả.

Khai phá dữ liệu và khám phá tri thức (Data mining and Knowledge discovery) là một lĩnh vực quan trọng của ngành Công nghệ thông tin với mục tiêu là tìm kiếm các tri thức có ích, cần thiết, tiềm ẩn và chưa được biết trước trong cơ sở dữ liệu lớn. Đây là lĩnh vực đã và đang thu hút đông đảo các nhà khoa học trên thế giới và trong nước tham gia nghiên cứu. Phân lớp (classification) là một trong những bài toán cơ bản trong khai phá dữ liệu với mục tiêu là phân loại các đối tượng vào các lớp cho trước. Theo tiếp cận học máy, phân lớp là phương pháp học có giám sát với hai giai đoạn: Giai đoạn 1 là xây dựng mô hình phân lớp dựa trên tập dữ liệu huấn luyện có đầu vào và đầu ra mong muốn (gọi là nhãn lớp); Giai đoạn 2 là sử dụng mô hình phân lớp để phân loại các tập dữ liệu chưa có nhãn lớp vào các lớp đã cho và có ứng dụng trong nhiều bài toán dự báo trong thực tế. Phân lớp được sử dụng rộng rãi trong các bài toán thực tiễn như trong y tế, ngân hàng, viễn thông, kinh tế, tài chính...

Ngày nay, cùng với sự phát triển mạnh mẽ của thị trường viễn thông là sự ra đời của nhiều nhà cung cấp và kinh doanh dịch vụ mạng viễn thông. Thị trường viễn thông đang đi vào giai đoạn bão hòa, khách hàng có nhiều sự lựa chọn, dẫn đến họ có thể thay đổi sử dụng dịch vụ bất cứ khi nào, kết quả là số