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

ĐÀO MỸ HẠNH

CỤM DỮ LIỆU VÀ ỨNG DỤNG TRONG PHÂN TÍCH LƯƠNG CỦA CÁN BỘ TRƯỜNG CAO ĐẲNG NGHỀ HÀ NAM

Chuyên ngành: KHOA HỌC MÁY TÍNH

Mã số chuyên ngành: 60 48 0101

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên - 2015

LỜI CẢM ƠN

Tôi xin chân thành cảm ơn tập thể các thầy cô trong khoa đào tạo sau đại học trường Đại học Công nghệ Thông tin và Truyền thông Thái Nguyên đã trang bị cho tôi những kiến thức cơ bản trong những năm học tập tại trường để tôi có thể hoàn thành tốt bản luận văn tốt nghiệp này.

Tôi xin cảm ơn các đồng nghiệp và người thân đã động viên, giúp đỡ tôi trong quá trình nghiên cứu và thực hiện luận văn.

Đặc biệt, tôi xin cảm ơn **GS.TS Vũ Đức Thi**, người đã trực tiếp, tận tâm hướng dẫn, giúp đỡ, cung cấp tài liệu và tạo mọi điều kiện thuận lợi cho tôi nghiên cứu thành công luận văn tốt nghiệp của mình.

Thái Nguyên, ngày … tháng … năm 2015

Tác giả luận văn

Đào Mỹ Hạnh

LỜI CAM ĐOAN

Tôi xin cam đoan toàn bộ nội dung bản luận văn này là do tôi tự sưu tầm, tra cứu và sắp xếp cho phù hợp với nội dung yêu cầu của đề tài.

Nội dung luận văn này chưa từng được công bố hay xuất bản dưới bất kỳ hình thức nào và cũng không được sao chép từ bất kỳ một công trình nghiên cứu nào.

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. Tôi cũng xin cam đoan rằng mọi sự giúp đỡ cho việc thực hiện luận văn này đã được cảm ơn và các thông tin trích dẫn trong luận văn đã được chỉ rõ nguồn gốc.

Nếu sai tôi xin hoàn toàn chịu trách nhiệm.

Thái Nguyên, ngày … tháng … năm 2015

Người cam đoan

Đào Mỹ Hạnh

DANH MỤC TỪ VIẾT TẮT

CSDL: Cơ sở dữ liệu

KPDL: Khai phá dữ liệu

PCDL: Phân cụm dữ liệu

DANH MỤC CÁC BẢNG

DANH MỤC HÌNH VẼ

MỤC LỤC

MỞ ĐẦU

Khám phá tri thức - Khai phá dữ liệu (Knowledge discovery - Data mining) là một lĩnh vực quan trọng của ngành Công nghệ thông tin, đã và đang thu hút sự quan tâm đông đảo các nhà khoa học trên thế giới và trong nước tham\gia nghiên cứu. Khai phá dữ liệu ra đời vào những năm cuối thập kỷ 80 của thế kỷ XX, nó là lĩnh vực được nghiên cứu nhằm tự động khai thác thông tin, tri thức mới hữu ích, tiềm ẩn từ các CSDL lớn, kho dữ liệu,... Những vấn đề được quan tâm trong khai phá dữ liệu là phân lớp nhận dạng mẫu, luật kết hợp, phân cụm dữ liệu, ... Trong đó, phân cụm dữ liệu (Data Clustering) là một trong những kỹ thuật khai thác dữ liệu có hiệu quả. Phân cụm dữ liệu là quá trình tìm kiếm và phát hiện ra các cụm hoặc các mẫu dữ liệu tự nhiên trong cơ sở dữ liệu lớn. Phân cụm dữ liệu đã được ứng dụng trong nhiều lĩnh vực khác nhau như giáo dục, y tế, kinh tế, bảo hiểm, phân đoạn ảnh, ...

Việc áp dụng phân cụm dữ liệu để phân tích trong ngành kế toán hiện nay là rất cần thiết, bởi lượng dữ liệu lưu trữ lương khá lớn, việc phân tích đánh giá lương để đưa ra các chiến lược cân đối nguồn chi phí của đơn vị, dự báo quỹ lương và có kế hoạch cân đối tài chính cho phù hợp cũng gặp nhiều khó khăn. Ngoài ra việc phân tích lương còn phục vụ công tác quản lý nhân sự, giúp nắm được tình hình sử dụng con người của đơn vị từ đó đưa ra các chính sách tuyển dụng phù hợp, có các giải pháp tạo động lực cho người lao động bằng các chính sách tài chính.

Việc phân cụm dữ liệu để phân tích lương cho kết quả thu được sẽ phân loại theo giá trị lương của mỗi cán bộ, phân loại ra các mức thu nhập cao thấp khác nhau từ đó đưa ra các chính sách cân đối thu chi để có những chính sách ưu đãi phù hợp mà vẫn đảm bảo tài chính của đơn vị.

Với các lý do như vậy tôi chọn đề tài: **"Một số phương pháp phân cụm dữ liệu và ứng dụng trong phân tích lương của cán bộ trường Cao đẳng Nghề Hà Nam"** làm đề tài luận văn tốt nghiệp. Bố cục luận văn gồm có 3 chương:

Chương I: Tổng quan về khai phá dữ liệu và phân cụm dữ liệu.

Chương II: Một số thuật toán phân cụm dữ liệu điển hình

Chương III: Ứng dụng phương pháp phân nhóm dữ liệu vào phân tích lương của cán bộ trường Cao đẳng Nghề Hà Nam.