



Luận văn tốt nghiệp

**Khai thác dữ liệu hồ sơ nhân
sự hiện nay**

LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình nghiên cứu của bản thân, được xuất phát từ yêu cầu phát sinh trong công việc để hình thành hướng nghiên cứu. Các số liệu có nguồn gốc rõ ràng tuân thủ đúng nguyên tắc và kết quả trình bày trong luận văn được thu thập được trong quá trình nghiên cứu là trung thực chưa từng được ai công bố trước đây.

Hà Nội, tháng 5 năm 2008

Tác giả luận văn
Phạm Đức Chiến

Lời cảm ơn



Luận văn được thực hiện dưới sự hướng dẫn của PGS. TS. Hà Quang Thụy - Trường Đại học Công nghệ. Em xin bày tỏ lòng biết ơn sâu sắc tới Thầy đã hướng dẫn và có ý kiến chỉ dẫn quý báu trong quá trình em làm luận văn. Em xin chân thành cảm ơn Tiến sĩ Nguyễn Hải Châu và Tiến sĩ Nguyễn Hà Nam Bộ môn Các Hệ thống Thông tin, Khoa Công nghệ Thông tin đã cho nhiều ý kiến đóng góp quý báu để bản luận văn được hoàn thiện hơn. Em xin chân thành cảm ơn các Thầy giáo trong bộ môn Các Hệ thống Thông tin, nhóm seminar “Khai phá dữ liệu và ứng dụng”. Em cũng xin cảm ơn các thầy cô giáo trong Khoa, cán bộ thuộc phòng Khoa học và Đào tạo sau Đại học, Trường Đại học Công nghệ đã tạo điều kiện trong quá trình học tập và nghiên cứu tại Trường. Cuối cùng xin bày tỏ lòng cảm ơn tới những người thân trong gia đình, bạn bè đã động viên và giúp đỡ để tôi hoàn thành bản luận văn này.

Hà Nội, Tháng 5/2008
Học viên thực hiện
Phạm Đức Chiến

MỤC LỤC	Trang
LỜI CAM ĐOAN	1
MỤC LỤC	3
DANH MỤC HÌNH VẼ MINH HỌA.....	5
MỞ ĐẦU	7
CHƯƠNG 1. TỔNG QUAN.....	10
1.1. Ý nghĩa:	10
1.2.Các yêu cầu đặt ra trong công tác quản lý nguồn nhân lực:	11
1.2.1 Yêu cầu chung	11
1.2.2 Yêu cầu cụ thể:	12
1.2.3. Thông tin quản lý :	12
1.2.4. Khảo sát thực trạng yêu cầu vấn đề quản lý nguồn nhân lực của cơ quan Hải quan: .	15
1.2.5. Thực trạng dữ liệu, đề xuất yêu cầu cần đạt được và kiến trúc của hệ thống:	27
Kết luận chương 1.....	31
CHƯƠNG 2. NGHIÊN CỨU CÔNG NGHỆ, KỸ THUẬT VÀ CÔNG CỤ PHỤC VỤ YÊU CẦU.....	32
2.1. Khai phá dữ liệu:	32
2.1.1. Khái niệm	32
2.1.2. Ưu thế khai phá dữ liệu	33
2.2. Các kỹ thuật khai phá dữ liệu:	34
2.3. Cây quyết định:.....	37
2.3.1. Sức mạnh của cây quyết định:	37
2.3.2.Nhược điểm của cây quyết định:	38
2.4. Các phần mềm công cụ khai phá dữ liệu:	38
2.4.1. Phân tích số liệu bằng R:	38
2.4.2. Phân tích số liệu bằng phần mềm weka.....	40
2.4.3 Phân tích số liệu bằng See5/C5.0	43
2.4.4. Phân tích số liệu bằng DTREG1	44
2.4.5.Phân tích số liệu sử dụng công cụ của Microsoft:	45
2.5. Công cụ lựa chọn:	47
Kết luận chương 2.....	52
CHƯƠNG 3.PHÂN LỚP DỮ LIỆU SỬ DỤNG CÂY QUYẾT ĐỊNH	53
3.1. Tổng quan về phân lớp dữ liệu trong khai phá dữ liệu.....	53

3.1.1. Phân lớp dữ liệu.....	53
3.1.2. Các vấn đề liên quan đến phân lớp dữ liệu.....	56
3.1.3 Các phương pháp đánh giá độ chính xác của mô hình phân lớp.....	58
3.2 Cây quyết định ứng dụng trong phân lớp dữ liệu	59
3.2.1. Định nghĩa:	59
3.2.2. Các vấn đề trong khai phá dữ liệu sử dụng cây quyết định.....	60
3.2.3. Xây dựng cây quyết định.....	62
Kết luận chương 3.....	63
CHƯƠNG 4. THỰC NGHIỆM	64
4.1. Giới thiệu về mô hình xây dựng:	64
4.1.1. Sơ đồ luồng dữ liệu thông tin nhân sự.....	64
4.1.2. Giải quyết vấn đề:.....	65
4.1.3. Các mô hình được xây dựng:.....	66
4.2. Minh họa kết quả hỗ trợ thu được từ mô hình xây dựng	68
4.2.1 Minh họa hỗ trợ công tác tuyển lựa và đào tạo-cây lựa chọn cán bộ đào tạo quản lý nhà nước	68
4.2.2. Minh họa công tác kiểm tra thông tin hồ sơ nhập máy-cây phân lớp vị trí công tác (lãnh đạo, nhân viên)	71
4.2.3. Minh họa công tác hoạch định mô hình tổ chức.....	74
4.2.4. Minh họa mô hình giám sát theo dõi công việc phục vụ công tác đánh giá:.....	77
4.2.5. Minh họa hỗ trợ xây dựng các quy trình quản lý nguồn nhân lực:	78
Kết luận chương 4.....	80
KẾT LUẬN	81
TÀI LIỆU THAM KHẢO	83
PHỤ LỤC 1: MÔ TẢ BẢNG DỮ LIỆU SỬ DỤNG	84
PHỤ LỤC 2: THÔNG TIN VỀ QUẢN LÝ QUÁ TRÌNH CÁ NHÂN CỦA CHỨC NĂNG HỒ SƠ HIỆN TẠI	85

DANH MỤC HÌNH VẼ MINH HỌA

Hình 1.1	Thông kê tổng hợp số liệu bằng việc sử dụng phần mềm tin học	14
Hình 1.2	Ví dụ bảng tham chiếu	28
Hình 1.3	Đề xuất kiến trúc tổng thể của hệ thống quản lý nhân sự	30
Hình 2.1	Quá trình phát hiện tri thức trong cơ sở dữ liệu.....	33
Hình 2.2	Hình mô tả chức năng Export dữ liệu từ dữ liệu nhân sự.....	39
Hình 2.3	Minh họa chức năng chọn phân lớp dữ liệu trong R.....	40
Hình 2.4	Minh họa chọn phân lớp trong weka.....	41
Hình 2.5	Liệt kê các phương pháp phân lớp của công cụ.....	42
Hình 2.6	Liệt kê kết quả phân lớp.....	42
Hình 2.7	Dạng kết quả của DTREE.....	44
Hình 2.8	Kết quả cây quyết định.....	45
Hình 2.9	Bảng khuyến cáo lựa chọn thuật toán của Microsoft.....	46
Hình 2.10	Màn hình Analysis Manager	47
Hình 2.11	Chọn bảng dữ liệu đầu vào cho mô hình	48
Hình 2.12	Chọn kỹ thuật “Data Mining”	49
Hình 2.13	Lựa chọn các cột dữ liệu cho mô hình dự đoán.....	49
Hình 2.14	Kết quả cây quyết định.....	50
Hình 2.15	Lược đồ minh họa cho dự đoán thuộc tính liên tục	51
Hình 2.16	Minh họa cho dự đoán thuộc tính liên tục	51
Hình 3.1	Bài toán phân lớp	53
Hình 3.2	Quá trình phân lớp dữ liệu - (a) Bước xây dựng mô hình phân lớp	55
Hình 3.3	Quá trình phân lớp dữ liệu - (b1)Ước lượng độ chính xác của mô hình	56
Hình 3.4	Quá trình phân lớp dữ liệu - (b2) Phân lớp dữ liệu mới	56
Hình 3.5	Ước lượng độ chính xác của mô hình phân lớp với phương pháp holdout.....	59
Hình 3.6	Ví dụ về cây quyết định	60

Hình 4.1 Sơ đồ luồng dữ liệu thông tin nhân sự	64
Hình 4.2 Hình minh họa là các mô hình được xây dựng hỗ trợ cho công tác	66
Hình 4.3 Hình minh họa các thành phần của giao diện hỗ trợ.....	67
Hình 4.4 Minh họa công văn đào tạo	68
Hình 4.5 Minh họa yêu cầu đối tượng đào tạo.....	68
Hình 4.6 Cây phân lớp cán bộ, nhân viên	69
Hình 4.7 Minh họa thông tin một node của cây.....	70
Hình 4.8 mô hình phân lớp vị trí công tác	71
Hình 4.9 Quan hệ giữa các nút.....	72
Hình 4.10 Cây phân lớp chức vụ.....	73
Hình 4.11 Cây hoạch định mô hình tổ chức	75
Hình 4.12 Thông tin nút đơn vị cần bổ sung nhân sự	76
Hình 4.13 Thông tin đơn vị không cần điều chỉnh bổ sung nhân sự	76
Hình 4.14 Cây mô tả công việc hiện tại	77

MỞ ĐẦU

Quản lý cán bộ là mảng công tác quan trọng phối hợp một cách tổng thể các hoạt động hoạch định, tuyển mộ, tuyển chọn, duy trì, phát triển, động viên và tạo điều kiện thuận lợi cho tài nguyên nhân sự trong tổ chức, nhằm đạt được mục tiêu chiến lược và định hướng viễn cảnh của tổ chức. Một số công tác cán bộ điển hình là tổ chức, sắp xếp cán bộ, đánh giá cán bộ, quy hoạch cán bộ, lựa chọn cán bộ, bồi dưỡng quản lý, hoạch định mô hình tổ chức... [4], trong đó, công tác đánh giá hồ sơ cán bộ là công tác đầu tiên quan trọng xuyên suốt trong công tác cán bộ. Chỉ khi có đánh giá đúng cán bộ thì mới có thể sắp xếp đúng và người cán bộ có điều kiện phát huy được hết khả năng. Về phía người quản lý cán bộ thì họ cần đưa ra những quyết định đúng: lựa chọn đúng để đào tạo, để đề bạt, bổ nhiệm... Công tác quản lý hồ sơ cán bộ phổ biến là quản lý theo mô hình thủ công; đánh giá cán bộ dựa vào cảm tính và tự đánh giá của cá nhân để xem xét đã phát sinh nhiều tiêu cực làm suy giảm sức mạnh của bộ máy quản lý. Từ thực trạng đó, lộ trình tin học hóa dữ liệu nhân sự đã được tiến hành theo hướng số hóa hồ sơ nhân sự để xây dựng ứng dụng khai thác dữ liệu nhanh chóng hiệu quả phục vụ cho công tác nghiệp vụ.

Trong quá trình quản lý, cập nhật, bổ sung, thay đổi thông tin trong hồ sơ thì dữ liệu được tích lũy đã tăng trưởng ngày càng nhiều, và có thể chứa nhiều thông tin ẩn dạng những quy luật chưa được khám phá. Dữ liệu nhân sự là một cơ sở dữ liệu có nhiều thông tin cần quản lý, với mỗi trường hợp có nhiều thuộc tính (Biểu mẫu 2C/TCTW-98 hồ sơ cán bộ Bộ nội vụ đã quy định thông tin quản lý gồm 31 thuộc tính) và đặc tính phải phân loại đánh giá một trường hợp dựa trên các thuộc tính. Chính vì vậy, kho dữ liệu nhân sự hình thành đặt ra nhu cầu cần tìm cách trích rút ra các luật trong dữ liệu hay dự đoán những xu hướng mới của dữ liệu tương lai. Yêu cầu phương pháp khai thác kho dữ liệu này một cách khoa học hiệu quả và thuận tiện để có cơ sở thông tin hỗ trợ công tác quản lý nguồn nhân lực, đánh giá một con người cụ thể từ những thông tin đã được lưu trữ.

J. Han và M. Kamber [6] đã trình bày quá trình tiến hóa của lĩnh vực công nghệ cơ sở dữ liệu, mà trong đó, công nghệ khai phá dữ liệu (Data Mining) đã trở thành dạng tiến hóa mới của công nghệ cơ sở dữ liệu. Một quan niệm khác về công nghệ khai phá dữ liệu của Fayyad, Piatetsky-Shapiro là việc nghiên cứu phát triển lĩnh vực khai phá dữ liệu nhằm giải quyết tình trạng “ngập tràn thông tin mà thiếu thốn tri thức”. Khai phá dữ liệu có nhiều ứng dụng là một phương

pháp đơn vị Able Danger của Quân đội Mỹ đã dùng để xác định kẻ đứng đầu cuộc tấn công ngày 11/9, Mohamed Atta, và ba kẻ tấn công ngày 11/9 khác là các thành viên bị nghi ngờ thuộc lực lượng al Qaeda hoạt động ở Mỹ hơn một năm trước cuộc tấn công [7]. Đối với dữ liệu nhân sự, khi cập nhật một hồ sơ nhân sự mới vào cơ sở dữ liệu thì việc phân lớp nhân sự đó một cách tự động thực sự có ý nghĩa hỗ trợ cho việc đánh giá ban đầu. Những nghiên cứu công nghệ thông tin và những sản phẩm phần mềm về quản lý nhân sự, quản lý nguồn nhân lực hiện có mới chỉ đạt được mức độ thu thập hồ sơ lý lịch cán bộ và in ra các biểu mẫu báo cáo phục vụ quản lý, chưa có sản phẩm nào áp dụng kỹ thuật để phát hiện những thông tin tiềm ẩn trong dữ liệu nhân sự. Minh chứng cho việc này có thể kể đến một loạt các sản phẩm quản lý hồ sơ nhân sự như chương trình “Quản lý cán bộ phiên bản 4.0” của công ty sản xuất công nghệ phần mềm CSE, sản phẩm đã được sử dụng Hệ quản trị Oracle phiên bản 9i để cập nhật, quản lý hồ sơ nhân sự của Bộ Nội vụ, Bộ Tài chính và các cơ quan trực thuộc Bộ Tài chính trong đó có Tổng cục Hải quan... Vì vậy, việc nghiên cứu các giải pháp khai thác các thông tin tiềm ẩn trong các kho dữ liệu nhân sự là hết sức cần thiết.

Luận văn nghiên cứu tổng quan về đặc tính công nghệ khai phá dữ liệu, các kỹ thuật khai phá dữ liệu (phân cụm, phân lớp...), các phần mềm thông dụng khai phá dữ liệu và giải pháp phân lớp dựa trên cây quyết định. Luận văn tập trung vào thuật toán tiêu biểu ứng dụng cho phạm vi phân tích dữ liệu là “Microsoft Decision Tree”, sử dụng công cụ phân tích dữ liệu của Microsoft. Đây là công cụ rất thuận tiện trong việc kết nối với cơ sở dữ liệu nhân sự dùng phần mềm Hệ quản trị SQL Server của Microsoft, công cụ có khả năng phân tích trực tuyến qua mạng (có quyền truy cập hợp pháp có thể phân tích từ bất kỳ máy tính nào có trong mạng) và là một công cụ mạnh khai thác nhanh đáp ứng được phân tích theo mô hình tăng trưởng dữ liệu. Ta biết rằng các tập dữ liệu được bổ sung và tăng trưởng theo thời gian, do vậy các tập thường xuyên và các luật kết hợp đã được tính toán không còn giá trị trên tập dữ liệu mới. Ngoài ra, với một dữ liệu ổn định, khi cần tìm các tập thường xuyên với độ hỗ trợ khác, công việc phải tính lại từ đầu.

Luận văn đã chạy thực nghiệm trên bộ dữ liệu nhân sự thử nghiệm tại Tổng cục Hải quan (việc sử dụng dữ liệu này chấp hành đúng quy tắc bảo quản thông tin hồ sơ cán bộ). Dữ liệu đầu vào của bài toán là cơ sở dữ liệu thử nghiệm hồ sơ lý lịch của 6978 nhân sự. Kết quả đầu ra là mô hình phân lớp và

đặc tính hỗ trợ của mô hình trong công tác quản lý nguồn nhân lực. Quá trình chạy thử nghiệm đã thu được các mô hình phân lớp trực quan với kết quả khích lệ. Trên cơ sở đó, luận văn đề xuất những cải tiến để hoàn thiện quan điểm quản lý nguồn nhân lực của ngành Hải quan và cấu trúc tổng thể cho hệ thống ứng dụng quản lý nguồn nhân lực. Phương pháp của luận văn đã nêu ra một hướng đi mới trong phân tích số liệu khác không chỉ phục vụ cho công tác thống kê nhà nước về hải quan mà còn phục vụ cho việc hình thành hệ hỗ trợ ra quyết định trong tương lai.

Bài toán phân lớp dữ liệu nhân sự để hỗ trợ quyết định đánh giá cán bộ nhằm khám phá được những đặc tính ẩn là rất có ý nghĩa. Đây là hướng giải pháp có hiệu quả cho việc phân tích thông tin phục vụ cho công tác đánh giá nhân sự nói riêng và công tác quản lý nguồn nhân lực nói chung.

Phạm vi nội dung nghiên cứu của đề tài:

Sử dụng phân lớp dữ liệu dựa trên cây quyết định để xây dựng các mô hình phân lớp hỗ trợ việc thực hiện các công việc quản lý nguồn nhân lực: giám sát công việc của nhân viên, hoạch định mô hình tổ chức, theo dõi giám sát số liệu của hồ sơ, hỗ trợ việc ra quyết định lựa chọn cán bộ tham gia chương trình đào tạo.

Luận văn gồm có 4 chương chính:

Chương 1: Tổng quan đề cập tới bối cảnh thực tiễn định hình hướng nghiên cứu của luận văn.

Chương 2: Yêu cầu và nghiên cứu các kỹ thuật, công cụ liên quan để chọn kỹ thuật, công cụ sử dụng.

Chương 3: Luận văn đi sâu vào nghiên cứu kỹ thuật phân lớp dựa trên cây quyết định.

Chương 4: Thực nghiệm trên bộ dữ liệu nhân sự và đưa ra kết quả minh họa cho phương pháp.

Kết luận định hướng phát triển kết quả nghiên cứu.

CHƯƠNG 1. TỔNG QUAN

1.1. Ý nghĩa:

Hiện nay, công tác quản lý nguồn nhân lực đang đặt ra nhiều yêu cầu bức thiết. Thực trạng, công tác hoạch định mô hình tổ chức chưa hình thành, công tác quản lý thay đổi, tình hình đánh giá cán bộ, các điều kiện, yếu tố chi phối chất lượng, hiệu quả đánh giá cán bộ có nhiều vấn đề. Nhu cầu có được bộ máy tổ chức có nhân sự có chuyên môn cao giàu năng lực để triển khai thực hiện các chương trình hiện đại hóa đang đặt ra. Nguồn nhân lực là một trong những yếu tố quan trọng quyết định sự thành công hay không thành công trong phát triển kinh tế - xã hội của quốc gia, do vậy tất cả các nước trên thế giới đều quan tâm đến phát triển nguồn nhân lực. Hơn lúc nào hết, khi nước ta đã trở thành thành viên WTO, đòi hỏi nguồn nhân lực có chất lượng cao đáp ứng quá trình hội nhập. Kinh nghiệm nhiều nước cho thấy, như Trung Quốc, sau 5 năm gia nhập WTO, kinh tế phát triển gần gấp đôi, nhưng kèm theo đó là việc thiếu nhân lực trầm trọng nhất là nhân lực có trình độ tay nghề cao. Mức độ trầm trọng hay không, có thể vượt qua hay không tùy thuộc rất nhiều vào sự nhận diện ra nguy cơ và sự khắc phục như thế nào của chúng ta? [6]. Do vậy, giải pháp nhằm thực hiện tốt công tác quản lý nguồn nhân lực xuất phát từ yêu cầu quản lý thông tin cơ bản của hồ sơ cán bộ, sử dụng công nghệ thông tin hỗ trợ khai thác dữ liệu phải được khẳng định tính cấp thiết để có cơ sở xây dựng mô hình tổ chức, luân chuyển cán bộ hợp lý. Quan điểm giải pháp khắc phục những mặt còn hạn chế, lúng túng trong công tác đánh giá luân chuyển cán bộ, giải quyết tốt mối quan hệ giữa luân chuyển với ổn định và xây dựng đội ngũ cán bộ chuyên môn sâu, vừa đáp ứng yêu cầu, nhiệm vụ công tác, vừa coi trọng mục đích bồi dưỡng, rèn luyện cán bộ, chuẩn bị đội ngũ cán bộ kế cận là vấn đề đang được đặt ra.

Tính nghiên cứu trong công tác nghiên cứu trong quản lý thông tin hồ sơ cán bộ để làm cơ sở hỗ trợ công tác quản lý nguồn nhân lực: đánh giá, lựa chọn, đề bạt cán bộ, hoạch định mô hình tổ chức... là một khâu trọng yếu trong nội dung công tác này. Hiểu cán bộ có hệ thống từ thông tin cơ bản trong hồ sơ, đánh giá cán bộ chính xác là cơ sở quan trọng đầu tiên để bố trí, sử dụng, bồi dưỡng cán bộ. Đánh giá người cán bộ trong sự vận động và phát triển không ngừng, trong các mối quan hệ chính trị và xã hội đa dạng, phức tạp, từ rất nhiều hiện tượng, sự việc cụ thể, riêng lẻ tìm ra mặt bản chất của người cán bộ - đó quả không phải là một công việc dễ dàng, đơn giản. Hơn nữa, những công chức

làm công tác cán bộ có kinh nghiệm nhưng không có kiến thức về công nghệ thông tin rất khó đưa ra một kiến trúc hợp lý cho công tác này. Nhu cầu công tác quản lý nguồn nhân lực không ngừng đổi mới, nâng cao chất lượng công tác mang tính nghiên cứu rất cao.

Quản lý thông tin cán bộ, phát hiện những thông tin tiềm ẩn hỗ trợ công tác quản lý nguồn nhân lực không chỉ là vấn đề nóng hiện nay mà còn là vấn đề khó. Luận văn sẽ đề cập chi tiết hơn những yêu cầu đặt ra trong công tác cán bộ quản lý cán bộ được thu tập từ những nghiên cứu thực tế. Từ đó đề xuất mô hình kiến trúc hệ thống công nghệ thông tin hỗ trợ. Trong phạm vi của đề tài sẽ đề cập tới giải pháp công nghệ, kĩ thuật phục vụ cho việc khai thác thông tin mang tính hỗ trợ cho các chức năng quản lý nguồn nhân lực.

1.2. Các yêu cầu đặt ra trong công tác quản lý nguồn nhân lực:

Công tác quản lý cán bộ gần đây được bổ sung một số yêu cầu mới nó được thường gọi với tên gọi khác là công tác quản lý nguồn nhân lực. Công tác quản lý cán bộ đã quy định trong Chương 5 của Pháp lệnh cán bộ công chức – 2003 sau này được cụ thể hóa bằng quản lý việc quản lý hồ sơ cán bộ công chức được quy định tại Quyết định số: 14/2006/QĐ-BNV ngày 06/11/2006 của Bộ trưởng Bộ Nội vụ về việc ban hành quy chế quản lý hồ sơ cán bộ, công chức. Công tác quản lý cán bộ có các yêu cầu được thể hiện bằng các chỉ tiêu thông tin trong biểu mẫu hồ sơ 2C/TCTW-98 của Ban tổ chức cán bộ chính phủ bao gồm: các thông tin cá nhân, quá trình đào tạo, quá trình công tác, quá trình lương, quan hệ gia đình. Công tác quản lý nguồn nhân lực chưa được cụ thể bằng các văn bản pháp quy nhưng được định nghĩa với một số yêu cầu bổ sung như hoạch định mô hình tổ chức (mở rộng hơn quản lý đến đơn vị, tổ chức). Luận văn sẽ đưa ra dưới đây những khảo sát yêu cầu của công tác quản lý nguồn nhân lực để định hướng nghiên cứu.

1.2.1 Yêu cầu chung

Trong công tác tổ chức cán bộ nói cách khác là công tác quản lý nguồn nhân lực. Bài toán hoạch định nguồn nhân lực cho đơn vị nhằm hoàn thiện và nâng cao hiệu quả trong việc quản lý nhân sự, thu hút nguồn nhân lực, khai thác và sử dụng nguồn nhân lực một cách hiệu quả. Cung cấp các báo cáo, thông tin cho ban lãnh đạo đầy đủ và nhanh chóng nhất, một loạt câu hỏi hóc búa cần phải giải quyết như:

- Làm thế nào để cơ quan quản lý nguồn lực hoạch định được **mô hình tổ chức** phù hợp nhất?

Làm thế nào để **phân tích công việc** cho các vị trí trong đơn vị?

Làm thế nào để phân công “**đúng người, đúng việc**”?

Làm thế nào để **thu hút, bồi dưỡng được nguồn lực** tại đơn vị mình?

Làm thế nào để **lập kế hoạch công việc** và **đánh giá hiệu quả làm việc** của nhân viên một cách chính xác và khách quan nhất?

Làm thế nào để quản lý tổng quan và chi tiết tất cả những **thông tin nhân sự** liên quan và cần thiết?

1.2.2 Yêu cầu cụ thể:

Từ các yêu cầu chung hình thành các yêu cầu cụ thể bước đầu:

- Hoàn chỉnh bước đầu, lưu trữ đầy đủ thông tin cơ bản của hồ sơ liên quan đến cán bộ: Thông tin cơ bản, thông tin cá nhân, thông tin công việc;
- Quản lý được nguồn lực: Đáp ứng được việc hỗ trợ xây dựng quy trình nhân sự như: tuyển dụng, quy trình đánh giá nguồn nhân lực, đào tạo, giám sát bổ sung biên chế...;
- Giám sát được quá trình làm việc của nhân viên, đánh giá;
- Giám sát được số liệu cập nhật trong hệ thống (dữ liệu luân chuyển, dữ liệu bổ sung từ các đơn vị cấp dưới..);
- Cơ sở dữ liệu thiết kế có cấu trúc mở, dễ liên lạc với các hệ thống cơ sở dữ liệu khác liên quan.

1.2.3. Thông tin quản lý :

Hồ sơ cán bộ, công chức là tài liệu pháp lý phản ánh các thông tin cơ bản nhất về cán bộ, công chức bao gồm: nguồn gốc xuất thân, quá trình công tác, hoàn cảnh kinh tế, phẩm chất, trình độ, năng lực, các mối quan hệ gia đình và xã hội của cán bộ, công chức [2].

Tại Điều 6 và Điều 7 của Quyết định số 14/2006/QĐ-BNV ngày 06/11/2006 của Bộ trưởng Bộ nội vụ về ban hành quy chế quản lý hồ sơ cán bộ, công chức đã quy định chi tiết thành phần của hồ sơ cán bộ công chức. Thông tin thành phần hồ sơ cán bộ công chức quản lý bao gồm các tài liệu sau[2] :

1. Quyển “**Lý lịch cán bộ, công chức**” là tài liệu chính và bắt buộc có trong thành phần hồ sơ cán bộ, công chức để phản ánh toàn diện về bản thân, các mối quan hệ gia đình, xã hội của cán bộ, công chức. Quyển lý lịch do cán

bộ, công chức tự kê khai và được cơ quan có thẩm quyền quản lý cán bộ, công chức thẩm tra, xác minh, chứng nhận.

2. Bản **“Sơ yếu lý lịch cán bộ, công chức”** là tài liệu quan trọng phản ánh tóm tắt về bản thân cán bộ, công chức và các mối quan hệ gia đình và xã hội của cán bộ, công chức. Sơ yếu lý lịch do cán bộ, công chức quy định tại khoản 1 Điều này và các tài liệu bổ sung khác của cán bộ, công chức được cơ quan có thẩm quyền quản lý cán bộ, công chức xác minh, chứng nhận.

3. Bản **“Bổ sung lý lịch cán bộ, công chức”** là tài liệu do cán bộ, công chức khai bổ sung theo định kỳ hoặc theo yêu cầu của cơ quan có thẩm quyền quản lý cán bộ, công chức. Bản bổ sung lý lịch được cơ quan có thẩm quyền quản lý cán bộ, công chức thẩm tra, xác minh, chứng nhận.

4. Bản **“Tiểu sử tóm tắt”** là tài liệu do cơ quan, tổ chức có thẩm quyền quản lý cán bộ, công chức tóm tắt từ Quyển lý lịch của cán bộ, công chức quy định tại khoản 1 Điều này để phục vụ cho bầu cử, bổ nhiệm khi có yêu cầu.

5. Bản sao giấy khai sinh; giấy chứng nhận sức khỏe do đơn vị y tế từ cấp huyện trở lên cấp và các văn bản có liên quan đến nhân thân của cán bộ, công chức; các loại giấy tờ có liên quan đến trình độ đào tạo của cán bộ, công chức như: bảng điểm, văn bằng, chứng chỉ về trình độ đào tạo chuyên môn, lý luận chính trị, ngoại ngữ, tin học, bồi dưỡng nghiệp vụ do cơ quan có thẩm quyền chứng nhận,... Trường hợp, văn bằng chứng chỉ được cấp bằng tiếng nước ngoài phải được dịch sang tiếng Việt Nam theo quy định của pháp luật.

6. Các quyết định về việc tuyển dụng, bổ nhiệm, điều động, biệt phái, luân chuyển, nâng ngạch, nâng lương, khen thưởng, kỷ luật,... của cán bộ, công chức.

7. Các bản tự kiểm điểm, nhận xét đánh giá cán bộ, công chức theo định kỳ hoặc theo yêu cầu của cơ quan, tổ chức, đơn vị có thẩm quyền.

8. Các bản nhận xét đánh giá của cơ quan, tổ chức, đơn vị có thẩm quyền đối với cán bộ, công chức (hàng năm, khi hết nhiệm kỳ, bầu cử hoặc bổ nhiệm, giới thiệu ứng cử, thuyên chuyển, khen thưởng, kỷ luật hoặc sau các đợt công tác, tổng kết học tập...).

9. Bản kê khai tài sản theo quy định của pháp luật.

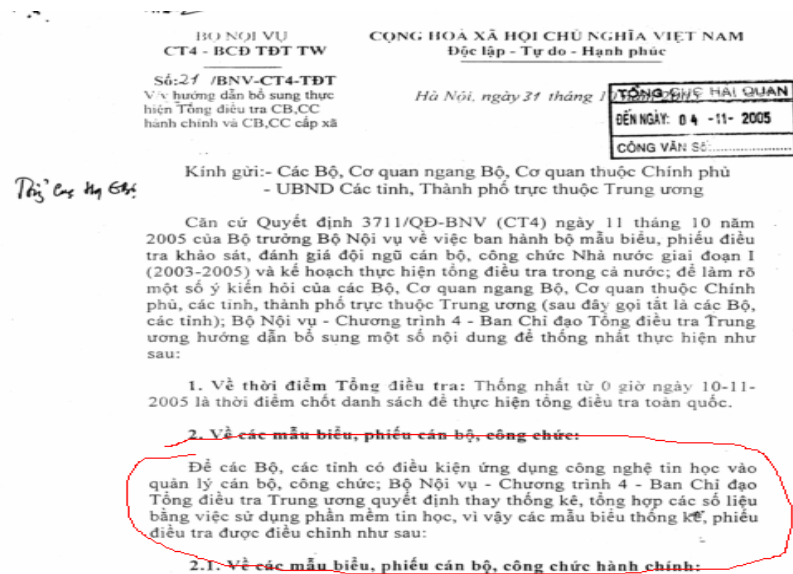
10. Đơn, thư kèm theo các văn bản thẩm tra, xác minh, biên bản, kết luận của cơ quan, tổ chức, đơn vị có thẩm quyền về những vấn đề liên quan đến cán

bộ, công chức và gia đình cán bộ, công chức được phản ánh trong đơn thư. Không lưu trong thành phần hồ sơ những đơn, thư nặc danh; đơn, thư chưa được xem xét, kế luận của cơ quan có thẩm quyền.

11. Những văn bản khác có liên quan trực tiếp đến quá trình công tác và quan hệ xã hội của cán bộ, công chức.

12. Đối với cán bộ, công chức được bổ nhiệm giữ chức vụ lãnh đạo phải bổ sung đầy đủ các tài liệu có liên quan đến việc bổ nhiệm vào hồ sơ của cán bộ, công chức đó.

Trong 12 tài liệu hồ sơ của cán bộ công chức, tài liệu hay được sử dụng nhất để đánh giá là sơ yếu lịch lịch cán bộ công chức. Sơ yếu lí lịch là bản tóm lược các thông tin cơ bản nhất của công chức phản ánh trong hồ sơ. Quá trình tin học hóa số hóa được tài liệu này từ năm 2005, theo Đề án “tổng điều tra cán bộ công chức” của Bộ Nội vụ:



Hình 1.1 Thống kê tổng hợp số liệu bằng việc sử dụng phần mềm tin học

Thực hiện việc triển khai theo yêu cầu định hướng này, trong năm 2006 Bộ Tài chính đã tiến hành triển khai phần mềm cập nhật phiếu điều tra cán bộ công chức, sơ yếu lí lịch cán bộ công chức, tại Tổng cục Hải quan và các đơn vị trong ngành Tài chính:

BỘ TÀI CHÍNH	CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM Độc lập - Tự do - Hạnh phúc
Số 474 474/BTC-TCCB V/v: triển khai cài đặt phần mềm và nhập dữ liệu điều tra số lượng, chất lượng công chức	Hà Nội, ngày <u>12</u> tháng <u>01</u> năm 2006
TỔNG CỤC HẢI QUAN	Kính gửi: Các đơn vị, tổ chức hành chính thuộc Bộ
BẾN NGÀY: 12-01-2006	theo yêu cầu của Bộ Tài chính và Bộ Nội vụ (tại Công văn số 21/BNV-CTT4-TĐT ngày 31/10/2005 về việc hướng dẫn bổ sung thực hiện Tổng điều tra số lượng, chất lượng đội ngũ công chức nhà nước; công văn số 01/BNV-CT4-ĐA1 ngày 06/01/2006 về việc nhập phần mềm phiếu điều tra CBCCHC), đề nghị các đơn vị triển khai một số công việc sau đây:
CÔNG VĂN SỐ: CTT4-TĐT	<ol style="list-style-type: none">1. Tổ chức cài đặt, hướng dẫn sử dụng phần mềm tin học phục vụ Tổng điều tra cán bộ, công chức (mẫu 1HC-ĐA1) trong toàn hệ thống dọc (đến các đơn vị cấp Cục và tương đương).2. Thực hiện nhập vào máy, sao dữ liệu kết quả đã nhập, kết xuất các biểu tổng hợp vào 02 đĩa CD gửi về Bộ (Cục Tin học và Thống kê tài chính) chậm nhất là <u>ngày 21/01/2006</u> để kịp tổng hợp báo cáo Bộ Nội vụ.
	Theo yêu cầu của Bộ Nội vụ, các đơn vị chú ý cài đặt chương trình phần mềm theo các tham số quy định sau:

Kết quả triển khai tập hợp được dữ liệu sơ yếu lí lịch của 6978 cán bộ công chức toàn ngành Hải quan (thời điểm năm 2006). Ứng dụng xây dựng chạy trên nền “Oracle Form” và cơ sở dữ liệu ORACLE, bước đầu khai thác các báo cáo thống kê theo mẫu báo cáo Bộ Nội vụ quy định.

1.2.4. Khảo sát thực trạng yêu cầu vấn đề quản lý nguồn nhân lực của cơ quan Hải quan:

a) Nguồn Nhân lực: Năng lực cán bộ quyết định năng lực của tổ chức:

Quản lý tốt nguồn nhân lực nhiều khả năng là vấn đề quan trọng hàng đầu duy nhất ảnh hưởng đến hiệu quả hoạt động của hải quan bất luận bộ máy tổ chức của cơ quan hải quan đó là gì. Nói như vậy không phải là một sự cường điệu bởi tất cả các khía cạnh trong quản lý hải quan và thông quan, kể cả việc ứng dụng và bảo trì hệ thống công nghệ thông tin hiện đại đều đòi hỏi đội ngũ cán bộ phải có đủ trình độ để vận hành hệ thống hiện tại một cách hiệu quả và chuẩn bị sẵn sàng để đón nhận các kỹ thuật và quy trình mới. Để làm được điều đó, đội ngũ cán bộ cần bắt kịp với những tiến bộ chung của chuỗi hậu cần thương mại quốc tế và điều chỉnh để thích nghi với những thay đổi trong các nhiệm vụ trọng tâm của hải quan.

Trước đây, công việc của ngành hải quan bao gồm các lao động thủ công như kiểm tra thực tế hàng hóa, tàu biển và hành khách cũng như tuần tra kiểm soát khu vực ranh giới điểm nhập. Cán bộ lãnh đạo hải quan thì làm việc với tư cách là các quan chức chính phủ cao cấp trong khi nhân viên hải quan lại có trình độ kém, được đào tạo không bài bản và có chế độ đãi ngộ thấp. Tình trạng

này đã bào mòn phong cách làm việc chuyên nghiệp cũng như liêm chính hải quan.

Các dịch vụ do chính quyền cung cấp phải đáp ứng các chuẩn mực ngày càng cao hơn. Những nhiệm vụ mà một nền kinh tế toàn cầu hóa đặt ra đối với ngành hải quan cũng trở nên rõ ràng. Một cơ quan hải quan hiện đại vừa có trọng trách bảo vệ quốc gia trước những nguy cơ bên ngoài vừa đảm nhận nhiệm vụ đại diện cho quốc gia tại biên giới và khu vực cảng phải có một đội ngũ cán bộ chuyên nghiệp sử dụng công nghệ tiên tiến giúp thực hiện được sứ mệnh giao phó. Có thể phân tách quy trình quản lý nguồn nhân lực hải quan thành nhiều giai đoạn:

- xác định các tiêu chí cần có;
- xây dựng quy trình tuyển dụng để đảm bảo hải quan tuyển được các cán bộ đáp ứng được các tiêu chí mong muốn đã được xác định;
- đào tạo cho cán bộ đương nhiệm để duy trì kỹ năng làm việc;
- chế độ đãi ngộ thích đáng nhằm tạo động lực cho cán bộ làm việc cũng như giữ họ ở lại làm việc lâu dài;
- áp dụng những hình thức xử phạt kiên quyết và kịp thời đối với các trường hợp không hoàn thành chức trách được giao và vi phạm liêm chính hải quan.

b) Hồ sơ cán bộ

Một cơ quan hải quan hiện đại cần xác định hồ sơ cho đội ngũ cán bộ mà mình muốn có. Hồ sơ đó phải bao gồm một nền tảng giáo dục tổng thể cao đủ mức để cán bộ hải quan có thể tiếp thu và duy trì các kỹ năng cần có trong ngành hải quan. Cùng với thời gian, các kỹ năng này sẽ thay đổi và đòi hỏi ngày càng nhiều hơn kiến thức chuyên ngành về kế toán, thu thập thông tin tình báo, tài chính, điều tra, phân tích, đào tạo, lập kế hoạch và quản lý nguồn nhân lực. Các thủ tục sử dụng trong các khâu nghiệp vụ này lại phụ thuộc ngày càng nhiều hơn vào công nghệ thông tin. Bởi vậy, cần áp dụng kỹ thuật phân tích tiến trình công việc (workflow analysis) hiện đại vào việc xác định phân bổ nhân sự phù hợp cho các mảng nghiệp vụ khác nhau.

Dưới đây là một số nhiệm vụ chính của hải quan cùng các yêu cầu chuyên môn cần có để hoàn thành các nhiệm vụ đó.

- Thực thi luật và quy định trong nước tại biên giới. Các Luật và quy định trong nước phải phù hợp với tất cả các công ước và chuẩn mực hải quan quốc tế mà quốc gia đó đã gia nhập. Do đó, cán bộ hải quan cần nắm được các bước tiến đạt được trong đàm phán thương mại quốc tế cũng như yêu cầu của toàn cầu hóa. Cán bộ hải quan cũng cần có trình độ chuyên môn nhất định về pháp luật để có thể nội luật hóa những tiến triển trong cộng đồng hải quan và thương mại quốc tế.
- Áp dụng quy trình thông quan hiện đại. Quy trình đánh giá rủi ro hiện đại được dựa trên các thông tin điện tử và kỹ thuật thu thập thông tin tình báo hiện đại nhằm tạo thuận lợi cho thương mại và bắt kịp với các tiến bộ trong lĩnh vực hậu cần thương mại tư nhân.
- Duy trì trao đổi thông tin mở với cộng đồng doanh nghiệp Hải quan phải đảm bảo rằng cộng đồng doanh nghiệp có đầy đủ thông tin về nghĩa vụ của mình. Quan điểm của doanh nghiệp cũng cần được tính đến trong quá trình hoạch định chính sách hải quan. Cán bộ hải quan cần có kỹ năng giao tiếp với doanh nghiệp song cần tránh tiếp xúc trực tiếp giữa cán bộ hải quan và đại diện doanh nghiệp trong quá trình xử lý thủ tục hải quan.
- Thực thi các điều luật về quyền sở hữu trí tuệ, an ninh, vận chuyển ma túy và cuối cùng là lao động và nhân quyền. Tuy hiện tại lao động và nhân quyền chưa phải là ưu tiên quốc gia song trong tương lai có thể sẽ phải thực thi các điều luật về lao động và nhân quyền để thực hiện các thoả thuận đạt được trong các cuộc đàm phán thương mại. Điều này đòi hỏi hải quan phải có năng lực đưa được chương trình nghị sự của các ban ngành chức năng khác vào trong quy trình, thủ tục hải quan.
- Thu thập và phổ biến số liệu thống kê thương mại quốc tế đòi hỏi phải có trình độ chuyên môn về công nghệ thông tin, nhận thức được tầm quan trọng của số liệu thống kê trong việc ra quyết định kinh tế.
- Quản lý nguồn nhân lực hải quan đòi hỏi phải quản lý nguồn nhân lực một cách hợp lý và có kiến thức chuyên sâu về phát triển nguồn nhân lực.

c) Tuyển dụng

Thông thường, việc điều chỉnh đội ngũ cán bộ hiện có thành đội ngũ cán bộ mong muốn là một quá trình cần có nhiều thời gian. Khi cán bộ lớn tuổi nghỉ hưu, cần tuyển dụng cán bộ trẻ không chỉ để thay thế đội ngũ đã về hưu mà còn để đáp ứng nhu cầu mở rộng hoạt động của ngành. Công tác tuyển dụng phải

được tiến hành một cách có hệ thống trong đó có đăng quảng cáo tuyển dụng. Quảng cáo tuyển dụng cần chỉ rõ yêu cầu cần có đối với vị trí cần tuyển, ví dụ như trình độ học vấn, kinh nghiệm công tác và các yêu cầu tương tự. Quy trình tuyển dụng cũng cần được nêu rõ. Đảm bảo minh bạch trong quy trình tuyển dụng là tối quan trọng vì nó giúp tạo ra chuẩn mực nghề nghiệp mới trong ngành hải quan, đồng thời ngăn chặn tình trạng thân quen, chạy chọt hiện vẫn thường phổ biến trong công tác tuyển dụng tại các cơ quan công quyền. Ví dụ, tại Hải quan Hoa Kỳ những năm 1880, có ứng cử viên tiềm năng cho các vị trí công tác trong ngành hải quan phải trả một khoản phí chạy việc đôi lúc gấp nhiều lần lương hàng năm. Tại Bolivia thì trước khi có những cải cách gần đây, có nhiều công chức hải quan “làm công không lương” và bù lại bằng cách đòi tiền hối lộ trong quá trình tiến hành nhiệm vụ. Một thực tế khác được biết đến nhiều là các quan chức cấp cao thường xuyên dùng ảnh hưởng của mình để gây sức ép buộc phải bổ nhiệm người thân hay người cùng dòng tộc của họ. Việc này làm hỏng quy trình tuyển dụng và tạo ra một món nợ phải trả cho những cán bộ mới vào ngành hoàn toàn xa lạ với việc thực thi các chức trách của một công chức hải quan). Nên sử dụng các biện pháp như quảng cáo công khai vị trí tuyển dụng, tham gia vào hội chợ việc làm và cung cấp thông tin tuyển dụng trực tiếp tại các trường đại học để đảm bảo những người có đủ tiêu chuẩn biết về kỳ tuyển dụng và đăng ký dự tuyển. Các ứng viên tiềm năng phải được các thành viên trong Ban tuyển dụng kiểm tra nghiêm ngặt về trình độ, năng lực. Các thành viên trong Ban tuyển dụng cần được đào tạo chuyên về kỹ năng tuyển dụng. Họ cũng có thể phỏng vấn hàng xóm, đồng nghiệp cũng như các chủ tuyển dụng trước đây của ứng viên. Cũng cần tiến hành kiểm tra hồ sơ lưu tại cơ quan công an và tài khoản ngân hàng để đánh giá mức thu nhập và nguồn gốc thu nhập của ứng viên. Cũng cần có kiểm tra các kỹ năng và năng khiếu đặc biệt trong yêu cầu tuyển dụng. Các khâu kiểm tra này có thể do cán bộ phụ trách quản lý nguồn nhân lực trong ngành thực hiện hoặc có thể thuê chuyên gia trong lĩnh vực tuyển dụng. Cũng cần cho cán bộ mới tuyển biết về cơ chế luân chuyển công tác khi làm việc trong ngành hải quan. Điều này nhằm mục đích nâng cao khả năng một cán bộ hải quan có thể thực hiện được nhiều công việc khác nhau, đồng thời tránh phát triển những quan hệ móc ngoặc không chính đáng với cộng đồng doanh nghiệp địa phương. Ví dụ như tại Zambia và Ma-rốc, luân chuyển cán bộ đã là một phần gắn liền với sự nghiệp của công chức hải quan. Các cán bộ mới tuyển chưa có kinh nghiệm công tác trước đó trong lĩnh vực hải quan cần tham dự các lớp đào tạo chuyên sâu và sát hạch nghiêm túc để chuẩn bị cho nhiệm vụ

mới. Nếu thành công, họ còn phải trải qua một giai đoạn tập sự thường kéo dài một năm trước khi chính thức trở thành công chức hải quan. Trong thời gian tập sự, cán bộ mới tuyển này có thể bị sa thải ngay nếu có kết quả tập sự không đạt yêu cầu hoặc vi phạm kỷ luật mà không cần phải qua những thủ tục phức tạp như khiếu nại hay đưa vụ việc ra tòa án hành chính. Sau giai đoạn tập sự, nếu được đánh giá là đạt yêu cầu, cán bộ tập sự sẽ chính thức được tuyển dụng. Sử dụng các tiêu chí hoạt động minh bạch sẽ hỗ trợ lãnh đạo trong công tác quản lý cũng như tăng cường tính minh bạch của quy trình tuyển dụng.

Theo truyền thống, phần lớn các cơ quan hải quan đều dựa vào việc tuyển dụng các cán bộ trẻ tuổi sau đó kết hợp đào tạo cả lý thuyết và thực hành về những thông lệ hải quan tốt nhất. Tuy nhiên, hải quan hiện đại đòi hỏi đội ngũ cán bộ phải có trình độ chuyên môn không dễ gì có được nếu chỉ tiến hành đào tạo trong ngành. Kiến thức chuyên môn về công nghệ thông tin và kế toán ngày càng cần thiết để thực thi công tác kiểm tra sau thông quan chỉ là hai ví dụ. Cần có quy trình tuyển dụng phù hợp và chế độ đãi ngộ thỏa đáng để có thể thu hút được cán bộ có các kỹ năng chuyên sâu này. Khi không có các ứng viên có đủ năng lực, đôi lúc có thể né tránh những hạn chế gặp phải trong tuyển dụng công chức bằng cách ký các hợp đồng tư vấn hấp dẫn. Tuy nhiên, nhược điểm là ký hợp đồng tư vấn sẽ làm nhụt đi tinh thần làm việc của cán bộ hải quan nói chung và có thể đặt ra vấn đề không đảm bảo được tính liên tục về nhân sự trong các công việc chuyên sâu.

Quy trình thải hồi và tuyển dụng nhân sự sẽ mất nhiều thời gian nếu có khoảng cách lớn giữa hồ sơ nhân sự mới cần có và hồ sơ nhân sự hiện tại. Song có thời điểm công tác quản lý đòi hỏi quá trình này phải diễn ra nhanh hơn.

d) Đào tạo

Đào tạo cần là một nhiệm vụ chính của bộ phận quản lý nguồn nhân lực thuộc cơ quan hải quan. Các yêu cầu đặt ra bởi toàn cầu hóa và việc áp dụng nhanh chóng công nghệ thông tin vào hàng loạt các lĩnh vực hoạt động hải quan đã khiến cho đào tạo liên tục trở thành một nhu cầu tất yếu đối với hải quan. Công tác đào tạo trong ngành hải quan cần được tiến hành bởi đội ngũ giảng viên bao gồm các cán bộ hải quan giàu kinh nghiệm và các giảng viên chuyên nghiệp.

Trong hải quan hiện đại, việc tham gia thành công một số chương trình đào tạo nhất định là cơ sở cho nhiều quyết định đề bạt. Tất cả cán bộ trong

ngành đều phải được đào tạo hàng năm theo như thống nhất giữa bộ phận quản lý nguồn nhân lực và thủ trưởng đơn vị. Việc triển khai những nội dung đào tạo đã thống nhất cho cán bộ sẽ là một tiêu chí để đánh giá lãnh đạo. Có thể chỉ định các trường đào tạo chuyên nghiệp để thực hiện công tác đào tạo này trên quy mô quốc gia hoặc khu vực. Cần tận dụng tối đa các hỗ trợ đào tạo cung cấp bởi các tổ chức song phương, ...

e) Đãi ngộ cán bộ

Đãi ngộ cán bộ là một tác nhân quan trọng trong quản lý nguồn nhân lực. Cần có một chế độ đãi ngộ đủ cao để thu hút và giữ chân cán bộ có trình độ cần thiết làm việc cho ngành hải quan. Tuy nhiên, tổng gói lương chi trả cho cán bộ hải quan thường không thỏa đáng. Mức chênh lệch giữa lương thưởng cho cán bộ quản lý và cán bộ cấp dưới không có cách biệt lớn giống như trong khu vực tư nhân. Ví dụ như tại Nepal, lương của một cán bộ hải quan chỉ bằng một phần mức lương trung bình trong khu vực tư nhân; lương cơ bản của Tổng cục trưởng chỉ bằng 150% lương khởi điểm của một cán bộ mới được bổ nhiệm.

Cho dù không phải là động lực duy nhất thúc đẩy các cán bộ hải quan hoàn thành tốt công việc được giao song chắc chắn chế độ đãi ngộ là một động lực rất quan trọng. Ngoài ra còn có các động lực bổ sung, đó là xây dựng tinh thần đồng đội và lòng tự hào nghề nghiệp. Các động lực này thường ít được chú trọng đầy đủ. Ví dụ, trong những năm gần đây, khi chú trọng đặc biệt đến yếu tố này thì có vẻ như hải quan Maroc đã thu được những lợi ích nhất định. Trong phần lớn các trường hợp, hải quan ít có độ linh hoạt trong việc đặt ra mức lương mà buộc phải tuân thủ triệt để thang lương công chức. Thường thì, quy định tài chính chặt chẽ đã khiến cho thang lương trong ngành hải quan kém xa so với thang lương chi trả cho nhân viên có cùng năng lực trong khu vực tư nhân. Tình trạng này không khuyến khích cán bộ làm việc tốt và đẩy họ đến chỗ tìm kiếm các khoản phí tiêu cực khác. Hiện tượng các cán bộ sáng giá nhất bỏ ngành để ra làm đại lý khai thuê hải quan là một hiện tượng không có gì bất thường. Khi làm nhân viên khai thuê hải quan, các cán bộ này với hiểu biết về nội bộ ngành cũng như mối quan hệ quen biết với các cán bộ hải quan, có thể giúp tạo thuận lợi cho quá trình hoàn thành các thủ tục hải quan cho khách hàng của mình song cũng có thể đe dọa vi phạm liêm chính hải quan. Nguy cơ liêm chính đã khiến một số nước ngăn cản cán bộ hải quan cung cấp các dịch vụ khai thuê hải quan một vài năm sau khi kết thúc làm việc trong ngành hải quan.

Giải pháp cục bộ cho quy định cứng nhắc về thang lương là chi trả tiền thưởng cho cán bộ. Mặc dù hải quan nhiều nước áp dụng cơ chế thưởng song rất ít trong số đó cơ chế thưởng giúp nâng cao kết quả làm việc và hiệu quả hoạt động cuối cùng. Để cơ chế thưởng có tác dụng thì mức thưởng phải đủ cao để có thể lấp đầy khoảng cách với thu nhập trong khu vực tư nhân (có khấu trừ tính ổn định khi làm việc trong khu vực nhà nước) và thỏa mãn một số các tiêu chí khắt khe. Các khoản thưởng phải đảm bảo tính chính đáng trong nội bộ ngành hải quan cũng như ở ngoài ngành, phải khách quan, minh bạch và dễ quản lý. Ngoài ra, các khoản thưởng cũng phải đáp ứng tiêu chí SMART, tức là cụ thể (Specific), có thể đo lường được (Measurable), có thể đạt được (Achievable) và đúng thời điểm (Timed). Tuy nhiên, không dễ gì khi đảm bảo rằng đánh giá kết quả làm việc cung cấp đủ cơ sở để có thể phân loại cán bộ thỏa đáng (Ví dụ như tại Maroc, cứ nơi nào áp dụng chế độ thưởng 100% lương cơ bản cho cán bộ có kết quả làm việc đạt yêu cầu thì nơi đó đại đa số các cán bộ sẽ được đánh giá là đủ tiêu chuẩn để nhận mức thưởng lớn nhất).

Các khoản thưởng được các đối tượng trong ngành coi là chính đáng khi hệ thống thưởng được phân phối một cách công bằng và minh bạch, không thiên vị và có cơ chế khiếu nại về quyết định thưởng. Các khoản thưởng được các đối tượng ngoài ngành coi là chính đáng khi được chấp nhận ở ngoài ngành hải quan và có ngân sách cho các khoản thưởng đó. Hệ thống thưởng nếu không đảm bảo tính chính đáng dù là trong ngành hay ngoài ngành thì sẽ không bền vững (Ví dụ tại Ga-na, không thể duy trì được mức lương cao hơn cho cán bộ hải quan theo cơ chế Cơ quan Thu thuế Độc lập (ARA) trước sự phản đối của cán bộ công chức ở các ngành khác). Các điều kiện nghiêm ngặt cần đáp ứng để đảm bảo cơ chế thưởng đủ cao để bù đắp được mức lương thấp của cán bộ hải quan cũng đặt ra nhiều thách thức trong quá trình thiết kế và thực thi. Điều này khiến cho các hệ thống kiểu này trở nên hết sức rủi ro và cần được giám sát chặt chẽ.

Một phương án khác có thể thay thế cho việc chi trả mức thưởng cao là áp dụng thang lương cao hơn cho các cán bộ làm trong ngành hải quan. Lý do biện minh cho giải pháp này là việc huy động nguồn thu ngân sách đóng một vai trò quan trọng đối với vận hành của chính phủ. Lương bổng tốt hơn cũng phần nào giúp bảo vệ cán bộ hải quan trước cám dỗ của việc nhận hối lộ từ doanh nghiệp. Cần kết hợp biện pháp áp dụng thang lương cao hơn với cải cách hải quan tổng thể nhằm đảm bảo tăng cường kết quả làm việc cũng như hiệu quả hoạt động.

Hải quan một số nước áp dụng giải pháp mạnh tay hơn để giải quyết tình trạng lương bổng không thỏa đáng bằng cách thiết lập các ARA. Trong thời gian đầu, các ARA đã trả cho cán bộ hải quan một mức lương mang tính cạnh tranh so với mức lương trả trong khu vực tư nhân hoặc mức lương cao nhất trả cho công chức.

Tăng lương và thăng chức dựa trên kết quả làm việc cũng là hai nhân tố tạo động lực làm việc quan trọng cho cán bộ. Tuy nhiên, cả hai lại thường bị hạn chế bởi các chính sách đề bạt cứng nhắc áp dụng cho tất cả các công chức mà theo đó quyết định thăng chức hay tăng lương phụ thuộc nhiều vào thâm niên. Phải thừa nhận là vẫn có những trường hợp ngoại lệ. Ví dụ, ở Mô-dăm-bích, kết quả làm việc của cán bộ được đánh giá hàng quý trong hai năm đào tạo tập sự và sau đó được thanh tra ngành theo dõi chặt chẽ. Hải quan Ăng-gô-la lại áp dụng một hệ thống đánh giá công chức thường niên nhằm mục đích đối chiếu kỹ năng làm việc với mô tả công việc, từ đó xác định một cách chính xác các ứng cử viên để bổ nhiệm lên vị trí cao hơn.

Tình trạng cơ sở hạ tầng thiếu thốn trong đó có văn phòng làm việc cũng như nhà công vụ, nhất là tại các trụ sở hải quan vùng sâu vùng xa, cũng tác động đến đạo đức nghề nghiệp và tinh thần làm việc của cán bộ hải quan. Hạ tầng ở những nơi như vậy thường bị bỏ bê trong nhiều năm vì thiếu ngân sách. Tình trạng nhà ở công vụ tồi tàn ở nhiều điểm kiểm soát biên giới đã khiến cho việc chuyển đến làm việc tại những nơi này bị coi như một hình phạt hay bị đẩy vào chỗ khó khăn một cách bất công. Cải cách hải quan tại Dăm-bia và Tanzania bao gồm cả mảng nâng cấp cơ sở hạ tầng, một điểm được đánh giá cao bởi cán bộ hải quan.

f) Liêm chính và các hình thức kỷ luật

Thông quan hải quan hiện đại với việc áp dụng cao độ công nghệ thông tin và cung cấp chế độ đãi ngộ thỏa đáng cho cán bộ sẽ đóng vai trò quyết định trong cuộc đấu tranh chống lại các vấn đề liêm chính. Tuy nhiên, kinh nghiệm cho thấy như thế vẫn chưa đủ để triệt tiêu hoàn toàn nạn tham nhũng. Vì vậy, bất kỳ chính sách nguồn nhân lực nào cũng cần chỉ rõ cách thức giải quyết những vấn đề. Cán bộ hải quan cần phải nhận thức đầy đủ rằng không có chỗ cho các hành vi tham nhũng và tặc trách. Có thể nêu rõ các hình thức kỷ luật trong sổ tay nhân sự. Có nhiều hình thức kỷ luật ở các mức độ khác nhau từ khiển trách, không tăng lương, cắt thưởng cho đến sa thải. Song cần lưu ý rằng không được áp dụng hình thức sa thải cán bộ một cách khinh suất. Cũng cần có các biện

pháp bảo hộ để đảm bảo tính công bằng trong thi hành kỷ luật. Cần có phản hồi chính thức một cách nhanh chóng đối với các cáo buộc tham nhũng hoặc để kịp thời xóa bỏ những cáo buộc thiếu căn cứ hoặc để tránh tình trạng trì hoãn kéo dài giữa thời điểm vi phạm và thời điểm xử phạt. Các hình thức kỷ luật trong đó có sa thải sẽ trở thành một công cụ mạnh mẽ giúp thực thi pháp luật trong bối cảnh chế độ đãi ngộ cán bộ tốt và tỷ lệ thất nghiệp cao.

g) Cơ cấu Tổ chức và bố trí cán bộ

Bất kỳ tổ chức nào đều có thể vận hành tốt và vượt qua vô vàn các trở ngại nếu có một đội ngũ cán bộ năng động và có năng lực. Song lực lượng cán bộ cũng cần được hỗ trợ bởi một cơ cấu tổ chức phù hợp. Nhưng ngay cả khi có một mô hình tổ chức hoàn hảo mà thiếu vắng các nhân viên có năng lực, có trình độ, năng động và được đào tạo thì tổ chức đó cũng không thể tồn tại được. Một tổ chức lý tưởng là một tổ chức ở trạng thái động. Các tổ chức ở khu vực công và khu vực tư liên tục điều chỉnh cơ cấu của mình để đối mặt với các thách thức, biến đổi trong khối lượng công việc, mở rộng địa bàn hoạt động, trong cạnh tranh cũng như trong việc áp dụng công nghệ mới và các sáng kiến mới. Hải quan cũng không phải là trường hợp ngoại lệ. Cơ quan hải quan cũng thường phải vật lộn để tìm ra một mô hình tổ chức lý tưởng đáp ứng được yêu cầu của một môi trường hải quan liên tục biến đổi. Tuy vậy, tái cơ cấu không phải là một phương thuốc chữa bách bệnh. Đôi lúc, nó chỉ là một cái cớ để che đậy cho một thực tế là không tìm ra được căn nguyên dẫn đến tình trạng vận hành kém hiệu quả của cơ quan hải quan. Tiến hành tái cơ cấu tổ chức còn có thể làm xáo trộn hay gián đoạn hoạt động thường nhật của cơ quan hải quan.

h) Cơ cấu tổ chức nội bộ

Theo truyền thống, các cơ quan hải quan được cơ cấu như một bộ phận của Bộ Tài chính và chịu trách nhiệm trước Bộ Tài chính về các hoạt động và kết quả đạt được. Nhiệm vụ bao trùm của ngành hải quan là tăng thu thuế theo quy định của ngân sách. Và trong quá trình thực hiện nhiệm vụ thu ngân sách, hải quan cũng phải đảm bảo áp dụng các thủ tục và chính sách một cách nhất quán tại tất cả các cửa khẩu nhập và cho tất cả các loại hình vận tải. Đôi lúc, do có trách nhiệm bảo vệ biên giới và xử lý các trường hợp không tuân thủ và buôn lậu nên cơ quan hải quan có quyền tự quyết phần nào lớn hơn các ban ngành và cơ quan ngang bộ khác quản lý tại biên giới. Đặc thù chức trách của cơ quan hải quan quyết định hình thức tổ chức phân cấp trong ngành hải quan. Theo đó, cơ

cấu tổ chức hải quan bao gồm cơ quan tổng cục, các hải quan vùng và các cục hải quan địa phương. Tổng cục chịu trách nhiệm xây dựng các quy trình và chính sách hoạt động, trong đó có việc ứng dụng công nghệ thông tin – nhằm đạt được kết quả và hiệu quả vận hành cũng như đảm bảo tuân thủ các hiệp định quốc tế liên quan đến Tổ chức Thương mại Thế giới. Tổng cục cũng chịu trách nhiệm giám sát hoạt động được phân cấp xuống các văn phòng vùng và địa phương, về chính sách nhân sự bao gồm tuyển dụng, đãi ngộ, đào tạo và thực thi. Các văn phòng vùng giám sát hoạt động của văn phòng địa phương trong phạm vi mình quản lý trong khi các văn phòng địa phương thường là điểm liên hệ của cộng đồng doanh nghiệp quốc tế và hải quan các nước khác. Cán bộ tại các hải quan địa phương quyết định mức kiểm tra cần áp dụng trong xử lý tờ khai và thông quan hàng hóa cũng như mức độ hiệu quả của quá trình làm thủ tục hải quan.

Bên cạnh các phòng ban truyền thống (pháp chế, thủ tục, định giá, công nghệ thông tin, chống lậu, hoạt động thực địa, hợp tác quốc tế, dịch vụ, nhân sự và kiểm tra sau thông quan), hải quan nhiều nước đã bổ sung thêm một vụ phụ trách việc duy trì quan hệ với khu vực tư nhân chuyên giải đáp các lo ngại, giải thích về quy trình, thủ tục, và tiến hành thanh tra khi có tranh cãi với các doanh nghiệp nước ngoài. Các doanh nghiệp nhập khẩu, xuất khẩu, các hãng vận chuyển, giao nhận hay các đại lý khai thuế hải quan cũng yêu cầu có một cơ chế khiếu nại độc lập để họ có thể tiến hành khiếu nại đối với những quyết định mà họ cho là sai lầm hay không phù hợp với thông lệ quốc tế. Một cơ chế mở như vậy đối với khu vực tư có thể giúp xây dựng được lòng tin cần thiết giữa khu vực tư nhân và hải quan, một lòng tin sẽ đem lại lợi ích cho tất cả các bên liên quan.

Khi nhận ra một thực tế là một phần lớn số thuế nhập khẩu được thu từ số lượng khá ít các doanh nghiệp nhập khẩu với quy mô lớn, hải quan một số nước đã điều chỉnh hoạt động của mình để cung cấp những dịch vụ đặc biệt cho các doanh nghiệp nhập khẩu lớn này. Việc này phù hợp với Công ước Kyoto sửa đổi cho phép các doanh nghiệp nhập khẩu “được uỷ quyền” (authorized) (được ưu tiên) được thông quan hàng hóa nhanh hơn, tức là những doanh nghiệp đáp ứng một số tiêu chí về minh bạch và trung thực sẽ được hưởng các thủ tục thông quan đơn giản hơn với điều kiện là có thể tiến hành kiểm tra sau thông quan. Ví dụ như Trung tâm Thuế và Hải quan Mẫu tại Ai cập (Model Customs và Tax Center) được thiết lập để xử lý tờ khai thuế và hải quan cho một nhóm các

doanh nghiệp lớn được chọn lọc (khoảng 200 doanh nghiệp vào thời điểm cuối năm 2003). Nhóm doanh nghiệp này chiếm một tỷ phần lớn trong khối lượng thương mại và tổng số thuế thu được. Trên thực tế, trung tâm này hoạt động như văn phòng một cửa áp dụng các thủ tục mới và hiệu quả.

i) Phương pháp đánh giá nguồn nhân lực:

Chỉ số đánh giá

Phương pháp đánh giá nguồn nhân lực là những cách thức, biện pháp khác nhau được sử dụng trên cơ sở những chỉ tiêu, chỉ số nhất định để từ đó có những nhận xét, kết luận về số lượng, chất lượng, cơ cấu nguồn nhân lực hiện có cũng như khả năng sẽ có trong tương lai dự định. Theo phương pháp quản lý nguồn nhân lực lấy con người làm trung tâm thì để đánh giá nguồn nhân lực dựa trên hệ các chỉ số như:

- Mục tiêu, mức độ đạt được mục tiêu của tổ chức cũng như của các cá nhân thành viên của tổ chức và các biện pháp, cách thức đã thực hiện để đạt mục tiêu đề ra. Chỉ số này gọi theo khoa học là chỉ số mục tiêu (index of objective point). Theo chỉ số này các tổ chức phải đặt ra mục tiêu phát triển và mỗi cá nhân trong tổ chức cũng đặt ra mục tiêu phát triển và thực hiện để trên cơ sở đó thực hiện việc đánh giá;

- Chỉ số công việc (index of job). Chỉ số này được hình thành trên cơ sở sự phân tích công việc (work analysis) theo đó có bảng mô tả công việc với các chỉ số căn bản như nhiệm vụ (task), chức trách (responsibility), yêu cầu của công việc (demand of job). Khi đánh giá ta sẽ sử dụng các chỉ số để đánh giá mức độ hoàn thành công việc của nhân viên từ đó đưa ra kết luận;

- Chỉ số bổ sung (additional index). Các chỉ số này bao gồm tinh thần trách nhiệm, tính chấp hành kỷ luật, phong cách hành động.v.v... nói chung so với các chỉ số mục tiêu và công việc thì các chỉ số này tình xác định hạn chế hơn do vậy chúng được xếp vào hệ các chỉ số bổ sung.

Yêu cầu của công việc đánh giá

Ngoài các chỉ số đánh giá, khi đánh giá nguồn nhân lực còn phải bảo đảm các yêu cầu của việc đánh giá đó là:

- Tính phù hợp. Tính phù hợp thể hiện trên nhiều phương diện khác nhau như sự phù hợp giữa các chỉ số đánh giá với mục tiêu trong mỗi tổ chức hoặc

mối liên hệ giữa công việc đã được xác định thông qua phân tích công việc với các chỉ số đánh giá được thiết kế trong phiếu đánh giá.v.v...;

- Tính nhạy cảm. Tính nhạy cảm đòi hỏi hệ thống đánh giá phải có những công cụ đo lường chuẩn xác các mức độ khác nhau của sự hoàn thành công việc hay không hoàn thành công việc, tức là đạt được hay không đạt được mục tiêu;

- Tính tin cậy. Thể hiện sự nhất quán của các đánh giá trong toàn bộ quá trình đánh giá cho dù chúng được thực hiện với phương pháp nào. Tức là hệ thống đánh giá phải bảo đảm sao cho đối với mỗi đối tượng đánh giá thì kết quả đánh giá của các chủ thể đánh giá khác nhau phải thống nhất về cơ bản;

- Tính được chấp nhận. Tính này thể hiện và cũng đặt ra đòi hỏi với hệ thống đánh giá mà trong đó các chỉ số đánh giá phải được đối tượng đánh giá chấp nhận, tức là thuyết phục được họ;

- Tính thực tiễn. Thể hiện ở việc các phương pháp đánh giá phải khả thi với những công cụ đơn giản, dễ hiểu, dễ thực hiện không chỉ với đối tượng đánh giá mà cả với thủ trưởng đơn vị thực hiện đánh giá;

- Tính không lỗi. Đánh giá nguồn nhân lực là hoạt động của con người do vậy thường hay gặp phải các lỗi như: thiên vị, xu hướng trung bình, thái cực, định kiến, ảnh hưởng của các sự kiện gần nhất.v.v.... để tránh các lỗi này đạt ra nghiên cứu và đưa ra yêu cầu về tính không lỗi của hệ thống đánh giá.

Như vậy, thông qua các chỉ số đánh giá và nội dung yêu cầu của công việc đánh giá nguồn lực mang tính khoa học đã trình bày ở trên cho chúng ta thấy đánh giá nguồn nhân lực là một trong số những vấn đề quan trọng trong phát triển nguồn nhân lực. Đánh giá đúng, có những kế hoạch tốt thì sẽ phát triển có kết quả nguồn nhân lực, ngược lại sẽ cho kết quả "âm tính".

k)Kết luận khảo sát thực trạng yêu cầu về vấn đề quản lý nguồn nhân lực của cơ quan Hải quan:

Quản lý tốt nguồn nhân lực là chìa khoá cho quản lý hải quan hiệu quả. Tuy nhiên, việc này thường bị bỏ qua dẫn đến tình trạng vấn đề liên chính kéo dài gây hậu quả xấu cho việc cung cấp dịch vụ cho khách hàng ở mọi khía cạnh. Công tác quản lý nguồn nhân lực là một quá trình nhiều mặt bao gồm tuyển dụng, đào tạo, chính sách đãi ngộ và đề bạt cán bộ cũng như thực thi các quy định của pháp luật. Không có việc nào trong số này là dễ thực hiện và thường phải thực hiện trong một môi trường đầy khó khăn như ngân sách hạn hẹp hay bị bó buộc bởi các quy định cho cán bộ công chức. Song không vì thế mà không

đầu tư suy nghĩ để tìm ra những sáng kiến mới. Nghiên cứu thực địa cho thấy trong điều kiện còn khó khăn, việc chú trọng nhiều hơn đến các vấn đề nguồn nhân lực sẽ thu được những lợi ích đáng kể. Tăng cường phòng ban quản lý nguồn nhân lực là một cách bắt đầu tốt. Quan điểm này được phổ biến rộng rãi cho toàn thể cán bộ công chức trong ngành hải quan quán triệt tiếp thu và đề cùng nghiên cứu thực hiện.

1.2.5. Thực trạng dữ liệu, đề xuất yêu cầu cần đạt được và kiến trúc của hệ thống:

a) Dữ liệu hiện có:

Cơ sở dữ liệu nhân sự được tập hợp các thông tin cá nhân của một cán bộ công chức. Đây là Cơ sở dữ liệu nhân sự thử nghiệm tập hợp lưu trữ hồ sơ lý lịch của 6978 cán bộ, công chức. Bảng hồ sơ lý lịch được lưu trữ trong bảng chính HC_EMP. Bảng dữ liệu này bao gồm các thông tin được nêu trong hồ sơ biểu mẫu hồ sơ lý lịch 2C/TCTW-98 của Bộ Nội vụ, sau đó được bổ sung thêm một số thông tin theo yêu cầu quản lý riêng gọi là Hồ sơ cán bộ. Thông tin được lưu trữ trên bảng dữ liệu chính có tên HC_EMP với khóa của bảng để đảm bảo sự phân biệt giữa các hồ sơ và bảng này sử dụng khóa để tham chiếu đến các bảng dữ liệu tham chiếu quản lý thông tin về quá trình lương, quá trình đào tạo.. Chi tiết xin mô tả như sau:

** Thông tin bao gồm 64 thông tin quản lý trên bảng dữ liệu chính (HC_EMP):*

1. Các thông tin cơ bản gồm 25 thông tin
2. Quá trình tham gia quân đội 8 thông tin
3. Các thông tin khác 27 thông tin: hoàn cảnh kinh tế gia đình
4. Đặc điểm lịch sử bản thân (2 thông tin)
5. Công việc đảm nhận được ghi nhận trong trường cur_work của bảng chính HC_EMP
6. Chức vụ đảng (được lưu trong bảng chính HC_EMP trong trường Party_pos)

** 13 Thông tin tham chiếu:*

1. Quá trình công tác (tham chiếu tới bảng HC_work_pro)
2. Quá trình tham gia đảng (chưa quản lý nên không có dữ liệu hiện tại được lưu trong bảng dữ liệu chính HC_EMP)

3. Quá trình lương (được lưu trữ trong bảng HC_SAL_PRO được lưu trữ trong bảng HC_SAL_PRO)
4. Quá trình phụ cấp: phụ cấp chức vụ, phụ cấp khác (được lưu trữ trong bảng HC_SAL_PRO được đánh dấu bằng trường phân biệt lương và phụ cấp)
5. Quá trình đào tạo: đào tạo về chuyên môn, Chứng chỉ bồi dưỡng nghiệp vụ, Lý luận chính trị quản lý nhà nước, quản lý kinh tế, tin học ngoại ngữ. (được lưu trữ trong bảng tham chiếu HC_EDU_TRA_PRO)
6. Danh hiệu học hàm (được tham chiếu đến bảng HC_EMP_SCALE bảng này lại tham chiếu đến CLA_SCA_CODE)
7. Quá trình khen thưởng (được lưu trữ trong bảng HC_REWARD)
8. Quá trình kỷ luật (Bảng HC_DISCIPLINE lưu trữ thông tin)
9. Quan hệ gia đình: Bên vợ, bên chồng, anh chị em ruột (tham chiếu tới bảng HC_RELATIVE)
10. Chức vụ đảng (chưa có dữ liệu)
11. Đi công tác nước ngoài (chưa có dữ liệu)
12. Kết quả đánh giá hàng năm (chưa có dữ liệu)
13. Quan hệ với tổ chức nước ngoài (chưa có dữ liệu)

**Minh họa một tham chiếu từ bảng chính đến bảng quan hệ:*

Ví dụ trong bảng dữ liệu HC_EMP có một trường hợp có mã đơn vị là Node_id=110 để có thông tin ý nghĩa của mã này sẽ phải tham chiếu đến bảng dữ liệu DBList chứa tên đơn vị:

	NODE_ID	STATUS	NAME
1	110002	0	Cục Hai quan Bà Rịa - Vũng Tàu
2	110004	0	Cục Hai quan Bình Định
3	110	1	Tổng cục Hai quan
4	110029	0	Cục Hai quan Quang Ngãi
5	110003	0	Cục Hai quan Bình Dương
6	110010	0	Cục Hai quan Đắk Lắk
7	110012	0	Cục Hai quan Đồng Nai

Hình 1.2 Ví dụ bảng tham chiếu

b) Vấn đề khai thác dữ liệu này:

Hệ thống quản trị cơ sở dữ liệu quan hệ (Relational Database Management System - RDBMS) Microsoft SQL Server quản lý các bảng dữ liệu này và quan hệ của chúng. Chính Hệ thống quản trị cơ sở dữ liệu này đã tạo điều kiện thuận lợi để phát triển những phần mềm ứng dụng có những chức năng liên quan đến dữ liệu phục vụ các yêu cầu quản lý. Một các cụ thể hơn, **Dữ liệu dạng "thô"**, có nghĩa là mức độ tổng quát, trừu tượng của dữ liệu này rất thấp. Ví dụ dữ liệu của RDBMS được lưu giữ trong các trường dạng số - number, chữ - text, thời gian - date/time. Muốn chuyển các dữ liệu này thành dạng có ý nghĩa sử dụng hơn đối với người dùng (để chúng "biết nói"), cần tiếp tục phân tích và lập báo cáo. Tính năng tiện ích hướng tới xây dựng khai thác hiện tại từ cơ sở dữ liệu nhân sự này:

- Quản lý được thông tin cơ bản của cá nhân để tổng hợp việc phân tích số liệu như số lượng đội ngũ con người, giới tính, độ tuổi, nơi sinh, quê quán, nơi ở hiện tại, gia đình xuất thân, dân tộc, tôn giáo, danh hiệu Nhà nước phong, diện hưởng chính sách...

- Về quá trình đào tạo phân tích chất lượng cán bộ công chức hiện có; nhu cầu đào tạo thời gian tiếp theo; quá trình phấn đấu nâng cao trình độ cán bộ công chức; bố trí sử dụng cán bộ công chức đúng trình độ đào tạo.

- Quá trình công tác, ghi nhận kinh nghiệm cá nhân trong các lĩnh vực trước và sau khi tuyển dụng.

- Quản lý diễn biến lương, theo dõi việc thực hiện chính sách tiền lương.

- Về công tác Đảng, phục vụ yêu cầu báo cáo cấp trên và bố trí sử dụng. Về khen thưởng kỷ luật, sử dụng trong đánh giá và khi bố trí sử dụng nhân lực cán bộ.

- Đánh giá cán bộ công chức viên chức cần thiết để thực hiện quy chế đánh giá cán bộ công chức viên chức hàng năm theo quyết định số 11/1998/QĐ-TCCP-CCVC ngày 05/12/1998 của Ban tổ chức cán bộ chính phủ này là Bộ Nội vụ và thực hiện đánh giá và ghi nhận các ý kiến đánh giá cán bộ công chức viên chức tại từng thời điểm theo yêu cầu.

- Về quan hệ gia đình, bố trí và sử dụng đặc biệt lưu ý đến truyền thống gia đình.

- Về sức khỏe, cần thiết khi bố trí công tác nhất là các ngành đặc biệt liên quan đến điều kiện sức khỏe.

Với một dữ liệu triển khai lưu trữ hồ sơ lý lịch chưa được chuẩn hóa để nâng cao hiệu quả khai thác thì việc nghiên cứu của luận văn đề ra phương pháp khai thác theo hướng mô hình hóa thông tin liên quan đến dữ liệu và định hướng chi tiết yêu cầu của từng thông tin trong dữ liệu được lưu trữ là một giải pháp tích cực phục vụ tốt hơn công tác. Tiếp theo Luận văn đề xuất một kiến trúc tổng thể phù hợp mang tính khả thi cho hệ thống quản lý nguồn nhân lực trong tương lai.

c) Kiến trúc tổng thể của hệ thống:



Hình 1.3 Đề xuất kiến trúc tổng thể của hệ thống quản lý nhân sự

Từ các yêu cầu chung, yêu cầu cụ thể và các thông tin cần quản lý và đề xuất từ thực trạng dữ liệu nhân sự hiện có hình thành một kiến trúc tổng thể của Hệ thống quản lý nhân sự [hình 1.3]. Các ô hình chữ nhật nhỏ như phân tích công việc, mô tả công việc.. là các chức năng của hệ thống. Hệ quản trị cơ sở dữ liệu SQL Server là nơi quản lý, lưu trữ cơ sở dữ liệu nhân sự, bao gồm hồ sơ lý lịch, bảo hiểm, thông tin công việc.... Khai thác thông tin từ cơ sở dữ liệu nhân sự là thông tin đầu vào cho các chức năng của hệ thống.

d) Yêu cầu đặt ra trong tâm nghiên cứu của Luận văn:

Vấn đề đặt ra, làm thế nào hình thành thông tin đầu vào cho từng chức năng, muốn hình thành thông tin đầu vào thông tin chức năng thì cần phải khai thác dữ liệu nhân sự từ bảng dữ liệu chính HC_EMP cho hợp lý, khoa học đáp ứng đầy

đủ yêu cầu chức năng. Cụ thể hơn công cụ nào có thể truy vấn vào cơ sở dữ liệu để lấy thông tin, công nghệ nào phát hiện dữ liệu tiềm ẩn và kỹ thuật nào thích hợp với khai thác dữ liệu. Tìm được công nghệ, kỹ thuật khai thác rồi thì công cụ hỗ trợ nào sẽ phù hợp cho sự triển khai áp dụng. Với giới hạn nghiên cứu của luận văn, sau đây luận văn sẽ đi tiếp kỹ thuật, công nghệ, thuật toán công cụ liên quan để khai thác hiệu quả cơ sở dữ liệu nhân sự hỗ trợ các yêu cầu cụ thể của công tác quản lý nguồn nhân lực.

Kết luận chương 1

Chương 1 là chương Tổng quan đã trình bày những lý do hình thành hướng nghiên cứu qua việc nghiên cứu thực tế các yêu cầu đề ra kiến trúc của mô hình sử dụng.

Chi tiết các yêu cầu đưa ra hoàn chỉnh bước đầu, lưu trữ đầy đủ thông tin cơ bản của hồ sơ liên quan đến cán bộ và cơ sở dữ liệu thiết kế có cấu trúc mở, dễ liên lạc với các hệ thống cơ sở dữ liệu khác liên quan đã được thực hiện bằng việc triển khai cơ sở dữ liệu và phần mềm quản lý cán bộ theo đề án của Bộ Nội vụ. Yêu cầu đặt ra nghiên cứu trong luận văn là quản lý được nguồn lực trong việc đáp ứng được hỗ trợ xây dựng quy trình nhân sự như: tuyển dụng, quy trình đánh giá nguồn nhân lực, đào tạo, giám sát bổ sung biên chế..., giám sát được quá trình làm việc của nhân viên, giám sát số liệu cập nhật trong hệ thống (dữ liệu luân chuyển, dữ liệu bổ sung từ các đơn vị cấp dưới..), hỗ trợ công tác hoạch định mô hình tổ chức, hỗ trợ tuyển lựa và kết quả thu được mà các báo cáo thống kê từ chương trình không có được.

Với yêu cầu đặt ra như vậy, luận văn tiếp theo sẽ nghiên cứu công nghệ, kỹ thuật và công cụ phục vụ yêu cầu.

CHƯƠNG 2. NGHIÊN CỨU CÔNG NGHỆ, KỸ THUẬT VÀ CÔNG CỤ PHỤC VỤ YÊU CẦU

Theo kết quả đánh giá của J.Han, M.Kamber, Fayyad và Piatetsky-Sapiro Smyth là những nhà nghiên cứu phát triển lâu năm trong lĩnh vực khai phá dữ liệu và phát hiện tri thức trong cơ sở dữ liệu thì công nghệ khai phá dữ liệu được biết như là một dạng tiến hóa mới của công nghệ cơ sở dữ liệu hay giải quyết tình trạng tràn ngập thông tin mà thiếu thốn tri thức. Xét yêu cầu phát hiện những thông tin tiềm ẩn của dữ liệu hồ sơ nhân sự thì công nghệ khai phá dữ liệu là một lựa chọn.

Về kỹ thuật của công nghệ khai phá dữ liệu thì có nhiều kỹ thuật áp dụng chỉ phụ thuộc vào yêu cầu của bài toán cụ thể. Luận văn đi từng thuật toán liên quan để có một lựa chọn phù hợp với bài toán.

Về công cụ (phần mềm) phục vụ công nghệ và kỹ thuật khai phá dữ liệu, phải xét tới yêu cầu chuẩn bị dữ liệu: có một kết nối trực tiếp vào cơ sở dữ liệu để có thể phát triển phân tích trực tuyến cũng như việc thuận tiện cho xây dựng mô hình mà không mất thời gian chuẩn bị dữ liệu.

Tiếp theo đây luận văn sẽ đề cập tới những vấn đề liên quan để đưa ra những lựa chọn.

2.1. Khai phá dữ liệu:

2.1.1. Khái niệm:

Khai phá dữ liệu được định nghĩa là: quá trình trích xuất các thông tin có giá trị tiềm ẩn bên trong lượng lớn dữ liệu được lưu trữ trong các cơ sở dữ liệu, kho dữ liệu... Hiện nay, ngoài thuật ngữ khai phá dữ liệu, người ta còn dùng một số thuật ngữ khác có ý nghĩa tương tự như: khai phá tri thức từ cơ sở dữ liệu (knowledge mining from databases), trích lọc dữ liệu (knowledge extraction), phân tích dữ liệu/mẫu (data/pattern analysis), khảo cổ dữ liệu (data archaeology), nạo vét dữ liệu (data dredging). Nhiều người coi khai phá dữ liệu và một thuật ngữ thông dụng khác là khám phá tri thức trong cơ sở dữ liệu (Knowledge Discovery in Databases – KDD) là như nhau. Tuy nhiên trên thực tế, khai phá dữ liệu chỉ là một bước thiết yếu trong quá trình Khám phá tri thức trong cơ sở dữ liệu. Quá trình này bao gồm các bước sau:

Bước 1) **Làm sạch dữ liệu (data cleaning)**: loại bỏ nhiễu hoặc các dữ liệu không thích hợp;

Bước 2) **Tích hợp dữ liệu** (*data integration*): tích hợp dữ liệu từ các nguồn khác nhau như: Cơ sở dữ liệu, Kho dữ liệu, file text...;

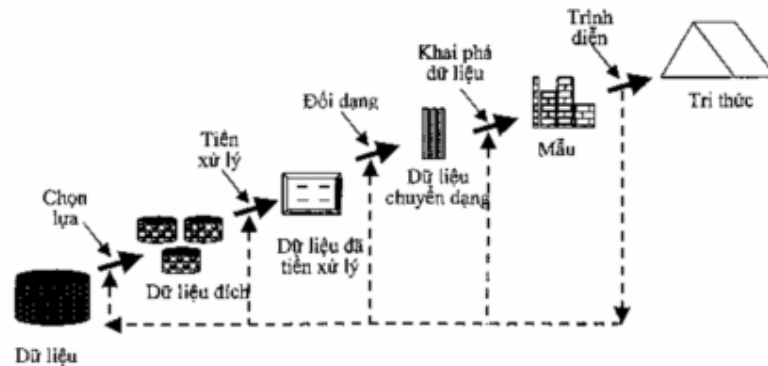
Bước 3) **Chọn dữ liệu** (*data selection*): ở bước này, những dữ liệu liên quan trực tiếp đến nhiệm vụ sẽ được thu thập từ các nguồn dữ liệu ban đầu;

Bước 4) **Chuyển đổi dữ liệu** (*data transformation*): trong bước này, dữ liệu sẽ được chuyển đổi về dạng phù hợp cho việc khai phá bằng cách thực hiện các thao tác nhóm hoặc tập hợp;

Bước 5) **Khai phá dữ liệu** (*data mining*): là giai đoạn thiết yếu, trong đó các phương pháp thông minh sẽ được áp dụng để trích xuất ra các mẫu dữ liệu;

Bước 6) **Đánh giá mẫu** (*pattern evaluation*): đánh giá sự hữu ích của các mẫu biểu diễn tri thức dựa vào một số phép đo;

Bước 7) **Trình diễn dữ liệu** (*knowledge presentation*): sử dụng các kỹ thuật trình diễn và trực quan hoá dữ liệu để biểu diễn tri thức khai phá được cho người sử dụng.



Hình 2.1 Quá trình phát hiện tri thức trong cơ sở dữ liệu

Khai phá dữ liệu và phát hiện tri thức trong các cơ sở dữ liệu đã cuốn hút các phương pháp, thuật toán và kỹ thuật từ nhiều chuyên ngành nghiên cứu khác nhau như học máy, thu nhận mẫu, cơ sở dữ liệu, thống kê, trí tuệ nhân tạo, thu nhận tri thức trong hệ chuyên gia...cùng hướng tới mục tiêu thống nhất là trích lọc ra được các tri thức từ dữ liệu trong các cơ sở dữ liệu khổng lồ. Song so với các phương pháp khác, khai phá dữ liệu có một số ưu thế rõ rệt[1].

2.1.2. Ưu thế khai phá dữ liệu:

Khai phá dữ liệu có nhiều ứng dụng và một số ưu thế rõ rệt được xem xét dưới đây:

+ So với phương pháp học máy, khai phá dữ liệu có lợi thế hơn ở chỗ, khai phá dữ liệu có thể sử dụng với các cơ sở dữ liệu chứa nhiều nhiều, dữ liệu không đầy đủ hoặc biến đổi liên tục. Trong khi đó phương pháp học máy chủ yếu được áp dụng trong các cơ sở dữ liệu đầy đủ, ít biến động và tập dữ liệu không quá lớn;

+ Phương pháp hệ chuyên gia: phương pháp này khác với khai phá dữ liệu ở chỗ các ví dụ của chuyên gia thường ở mức chất lượng cao hơn nhiều so với các dữ liệu trong cơ sở dữ liệu, và chúng thường chỉ bao hàm được các trường hợp quan trọng. Hơn nữa các chuyên gia sẽ xác nhận giá trị và tính hữu ích của các mẫu phát hiện được;

+ Phương pháp thống kê là một trong những nền tảng lý thuyết của Khai phá dữ liệu, nhưng khi so sánh hai phương pháp với nhau ta có thể thấy các phương pháp thống kê còn tồn tại một số điểm yếu mà Khai phá dữ liệu đã khắc phục được:

- ◇ Các phương pháp thống kê chuẩn không phù hợp với các kiểu dữ liệu có cấu trúc trong rất nhiều cơ sở dữ liệu;
- ◇ Các phương pháp thống kê hoạt động hoàn toàn theo dữ liệu, nó không sử dụng tri thức sẵn có về lĩnh vực;
- ◇ Kết quả phân tích của thống kê có thể sẽ rất nhiều và khó có thể làm rõ được;
- ◇ Phương pháp thống kê cần có sự hướng dẫn của người dùng để xác định phân tích dữ liệu như thế nào và ở đâu.

Với những ưu điểm đó, khai phá dữ liệu đang được áp dụng khai phá dữ liệu nhân sự để đáp ứng tính thường xuyên thay đổi, tăng trưởng của dữ liệu. Tìm kiếm những thông tin tiềm ẩn trong dữ liệu mà bằng phương pháp khác không phát hiện được[1].

2.2. Các kỹ thuật khai phá dữ liệu:

Các kỹ thuật khai phá dữ liệu [3] thường được chia thành 2 nhóm chính:

- Kỹ thuật khai phá dữ liệu mô tả: có nhiệm vụ mô tả về các tính chất hoặc các đặc tính chung của dữ liệu trong cơ sở dữ liệu hiện có. Các kỹ thuật này có thể liệt kê: phân cụm (*clustering*), tóm tắt (*summerization*), trực quan hóa (*visualization*), phân tích sự phá hiện biến đổi và độ lệch, phân tích luật kết hợp (*association rules*)...;

- Kỹ thuật khai phá dữ liệu dự đoán: có nhiệm vụ đưa ra các dự đoán dựa

vào các suy diễn trên dữ liệu hiện thời. Các kĩ thuật này gồm có: phân lớp (*classification*), hồi quy (*regression*)...;

3 phương pháp thông dụng nhất trong khai phá dữ liệu là: phân cụm dữ liệu, phân lớp dữ liệu và khai phá luật kết hợp. Ta sẽ xem xét từng phương pháp:

Phân cụm dữ liệu: Mục tiêu chính của phương pháp phân cụm dữ liệu là nhóm các đối tượng tương tự nhau trong tập dữ liệu vào các *cụm* sao cho các đối tượng thuộc cùng một lớp là tương đồng còn các đối tượng thuộc các cụm khác nhau sẽ không tương đồng. Phân cụm dữ liệu là một ví dụ của phương pháp học không có thầy. Không giống như phân lớp dữ liệu, phân cụm dữ liệu không đòi hỏi phải định nghĩa trước các *mẫu dữ liệu huấn luyện*. Vì thế, có thể coi phân cụm dữ liệu là một cách học bằng quan sát (*learning by observation*), trong khi phân lớp dữ liệu là học bằng ví dụ (*learning by example*). Trong phương pháp này bạn sẽ không thể biết kết quả các cụm thu được sẽ như thế nào khi bắt đầu quá trình. Vì vậy, thông thường cần có một chuyên gia về lĩnh vực đó để đánh giá các cụm thu được. Phân cụm dữ liệu được sử dụng nhiều trong các ứng dụng về phân đoạn thị trường, phân đoạn khách hàng, nhận dạng mẫu, phân loại trang Web... Ngoài ra phân cụm dữ liệu còn có thể được sử dụng như một bước tiền xử lí cho các thuật toán khai phá dữ liệu khác.

Khai phá luật kết hợp: mục tiêu của phương pháp này là phát hiện và đưa ra các mối liên hệ giữa các giá trị dữ liệu trong cơ sở dữ liệu. Mẫu đầu ra của giải thuật khai phá dữ liệu là tập luật kết hợp tìm được. Chẳng hạn: phân tích cơ sở dữ liệu bán hàng nhận được thông tin về những khách hàng mua máy tính có khuynh hướng mua phần mềm quản lý tài chính trong cùng lần mua được miêu tả trong luật kết hợp sau:

“Máy tính => Phần mềm quản lý tài chính”

[Độ hỗ trợ: 2%, độ tin cậy: 60%]

Độ hỗ trợ và độ tin cậy là hai độ đo của sự đáng quan tâm của luật. Chúng tương ứng phản ánh sự hữu ích và sự chắc chắn của luật đã khám phá. Độ hỗ trợ 2% có nghĩa là 2% của tất cả các tác vụ đã phân tích chỉ ra rằng máy tính và phần mềm quản lý tài chính là đã được mua cùng nhau. Còn độ tin cậy 60% có nghĩa là 60% các khách hàng mua máy tính cũng mua phần mềm. Khai phá luật kết hợp được thực hiện qua 2 bước:

- Bước 1: tìm tất cả các tập mục phổ biến, một tập mục phổ biến được xác định qua tính độ hỗ trợ và thỏa mãn độ hỗ trợ cực tiểu.

- Bước 2: sinh ra các luật kết hợp mạnh từ tập mục phổ biến, các luật phải thỏa mãn độ hỗ trợ cực tiểu và độ tin cậy cực tiểu.

Phương pháp này được sử dụng rất hiệu quả trong các lĩnh vực như marketing có chủ đích, phân tích quyết định, quản lý kinh doanh, phân tích giỏ thị trường...

Hồi quy: là học một hàm ánh xạ dữ liệu nhằm xác định giá trị thực của một biến. Tình huống ứng dụng hồi quy rất đa dạng, chẳng hạn như dự đoán số lượng sinh vật phát quang trong khu rừng nhờ đo vi sóng các cảm biến (senser) từ xa, hoặc ước lượng xác suất người bệnh có thể chết theo kết quả “test” triệu chứng, hoặc dự báo nhu cầu người tiêu dùng đối với một sản phẩm mới, hoặc dự báo chuỗi thời gian mà các biến đầu vào được coi như bản trễ thời gian của biến dự báo...

Phân lớp dữ liệu: Mục tiêu của phương pháp phân lớp dữ liệu là dự đoán nhãn lớp cho các mẫu dữ liệu. Quá trình phân lớp dữ liệu thường gồm 2 bước: xây dựng mô hình và sử dụng mô hình để phân lớp dữ liệu.

- Bước 1: một mô hình sẽ được xây dựng dựa trên việc phân tích các mẫu dữ liệu sẵn có. Mỗi mẫu tương ứng với một lớp, được quyết định bởi một thuộc tính gọi là thuộc tính lớp. Các mẫu dữ liệu này còn được gọi là *tập dữ liệu huấn luyện (training data set)*. Các nhãn lớp của tập dữ liệu huấn luyện đều phải được xác định trước khi xây dựng mô hình, vì vậy phương pháp này còn được gọi là *học có thầy (supervised learning)* khác với phân cụm dữ liệu là *học không có thầy (unsupervised learning)*.

- Bước 2: sử dụng mô hình để phân lớp dữ liệu. Trước hết chúng ta phải tính độ chính xác của mô hình. Nếu độ chính xác là chấp nhận được, mô hình sẽ được sử dụng để dự đoán nhãn lớp cho các mẫu dữ liệu khác trong tương lai.

Phương pháp hồi qui khác với phân lớp dữ liệu ở chỗ, hồi qui dùng để dự đoán về các giá trị liên tục còn phân lớp dữ liệu thì chỉ dùng để dự đoán về các giá trị rời rạc.

Như vậy, qua việc xem xét 3 phương pháp trên áp vào trong bài toán dự đoán dự đoán của dữ liệu nhân sự ta thấy khai thác dữ liệu nhân sự phần nhiều là những lớp dữ liệu là biết trước dữ liệu huấn luyện. Ví dụ như lựa chọn cán bộ đi đào tạo lớp quản lý nhà nước thì đã có tiêu chuẩn đính kèm, căn cứ vào các tiêu chuẩn này ta dùng phương pháp thống kê truy vấn vào dữ liệu ta đã có tập dữ liệu đầu đủ đáp ứng tiêu chuẩn. Vấn đề còn lại là xem xét thông tin tiềm ẩn

trong dữ liệu đó để dự đoán các mẫu dữ liệu mới. Hơn nữa, dự đoán trong dữ liệu nhân sự là trường dữ liệu có giá trị rời rạc. Do vậy, phân cụm dữ liệu không thích hợp cho bài toán xây dựng mô hình dự liệu dự đoán trên dữ liệu nhân sự.

2.3. Cây quyết định:

Trong phân lớp dữ liệu hình thức trực quan của mô hình là cây quyết định. Sau đây, luận văn sẽ trình bày vai trò, đánh giá về cây quyết định trong khai phá dữ liệu.

2.3.1. Sức mạnh của cây quyết định:

Cây quyết định có các sức mạnh chính sau [6]:

Khả năng sinh ra các quy tắc hiểu được

Cây quyết định có khả năng sinh ra các quy tắc có thể chuyển đổi được sang dạng if..then..else , hoặc các câu lệnh SQL. Đây là ưu điểm nổi bật của kỹ thuật này. Thậm chí với những tập dữ liệu lớn khiến cho hình dáng cây quyết định lớn và phức tạp, việc đi theo bất cứ đường nào trên cây là dễ dàng theo nghĩa phổ biến và rõ ràng. Do vậy sự giải thích cho bất cứ một sự phân lớp hay dự đoán nào đều tương đối minh bạch.

Khả năng xử lý với cả thuộc tính liên tục và thuộc tính rời rạc

Cây quyết định xử lý “tốt” như nhau với thuộc tính liên tục và thuộc tính rời rạc. Tuy rằng với thuộc tính liên tục cần nhiều tài nguyên tính toán hơn. Những thuộc tính rời rạc đã từng gây ra những vấn đề với mạng neural và các kỹ thuật thống kê lại thực sự dễ dàng thao tác với các *tiêu chuẩn phân chia* (splitting criteria) trên cây quyết định: mỗi nhánh tương ứng với từng phân tách tập dữ liệu theo giá trị của thuộc tính được chọn để phát triển tại node đó. Các thuộc tính liên tục cũng dễ dàng phân chia bằng việc chọn ra một số gọi là ngưỡng trong tập các giá trị đã sắp xếp của thuộc tính đó. Sau khi chọn được ngưỡng tốt nhất, tập dữ liệu phân chia theo “test” nhị phân của ngưỡng đó.

Thể hiện rõ ràng những thuộc tính tốt nhất

Các thuật toán xây dựng cây quyết định đưa ra thuộc tính mà phân chia tốt nhất tập dữ liệu đào tạo bắt đầu từ node gốc của cây. Từ đó có thể thấy những thuộc tính nào là quan trọng nhất cho việc dự đoán hay phân lớp.

2.3.2. Nhược điểm của cây quyết định:

Dù có những sức mạnh nổi bật trên, cây quyết định vẫn không tránh khỏi có những điểm yếu. Đó là cây quyết định không thích hợp lắm với những bài toán với mục tiêu là dự đoán giá trị của thuộc tính liên tục như thu nhập, huyết áp hay lãi xuất ngân hàng,... Cây quyết định cũng khó giải quyết với những dữ liệu thời gian liên tục nếu không bỏ ra nhiều công sức cho việc đặt ra sự biểu diễn dữ liệu theo các mẫu liên tục.

Như vậy, từ những ưu điểm và nhược điểm trên cho thấy cây quyết định phù hợp với mô hình dự đoán trên dữ liệu nhân sự bởi những nguyên nhân sau :

- Tốc độ học tương đối nhanh hơn so với những phương pháp phân loại khác;
- Có thể hoán chuyển được thành những luật phân lớp đơn giản và dễ hiểu;
- Có thể dễ dàng chuyển đổi sang câu lệnh SQL sử dụng truy vấn SQL để truy xuất cơ sở dữ liệu một cách hiệu quả;
- Sự chính xác phân lớp có thể so sánh được với những phương pháp khác.

2.4. Các phần mềm công cụ khai phá dữ liệu:

Các phần mềm hỗ trợ khai phá dữ liệu được phát triển nhiều. Luận văn chỉ đề cập tới một số phần mềm đang thông dụng và đang được sử dụng được đánh giá là hiệu quả. Trên cơ sở đó nghiên cứu đưa vào áp dụng trên dữ liệu nhân sự để đánh giá lựa chọn trên các đặc tính: kết quả thu được, tính sử dụng trực quan hiệu quả, khả năng triển khai cao. Các công cụ đã được luận văn nghiên cứu đề cập:

- Phần mềm phân tích thống kê R;
- Phân tích số liệu bằng phần mềm Weka;
- Phân tích số liệu bằng See5/C5.0;
- Phân tích số liệu bằng DTREEG1;
- Phân tích số liệu bằng Microsoft Analysis Service.

2.4.1. Phân tích số liệu bằng R:

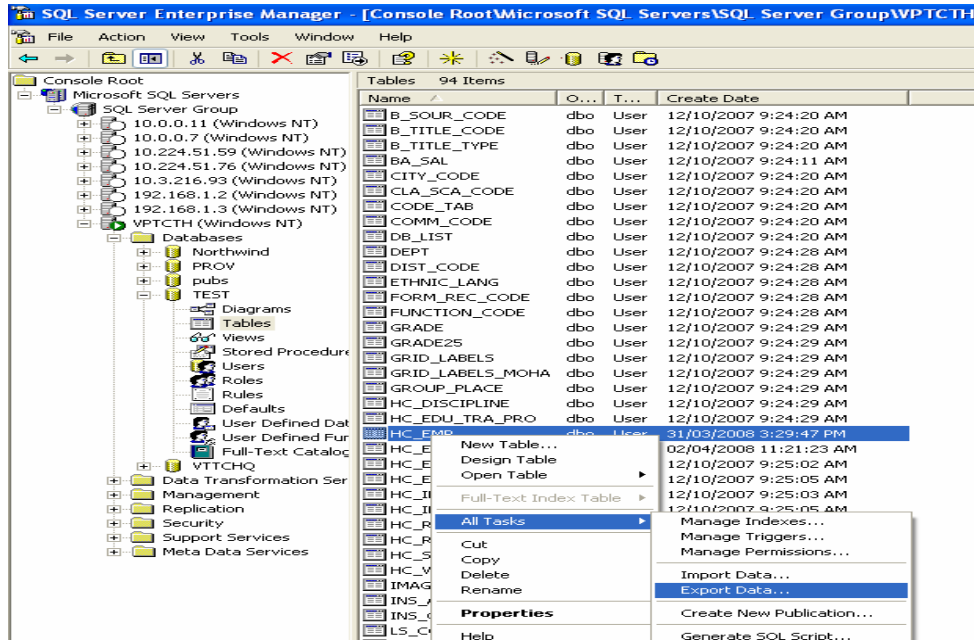
Vậy R là gì? Nói một cách ngắn gọn, R là một phần mềm sử dụng cho phân tích thống kê và vẽ biểu đồ. Thật ra, về bản chất, R là ngôn ngữ máy tính đa năng, có thể sử dụng cho nhiều mục tiêu khác nhau, từ tính toán đơn giản, toán học giải trí (recreational mathematics), tính toán ma trận (matrix),

đến các phân tích thống kê phức tạp. Vì là một ngôn ngữ, cho nên người ta có thể sử dụng R để phát triển thành các phần mềm chuyên môn cho một vấn đề tính toán cá biệt.

File dữ liệu vào cho R:

File dữ liệu dùng trong R là file .csv (file dạng dữ liệu Excel).

Để lấy dữ liệu từ bảng dữ liệu nhân sự ta phải vào dùng chức năng “Export” dữ liệu của Hệ quản trị cơ sở dữ liệu SQL Server



Hình 2.2 Hình mô tả chức năng Export dữ liệu từ dữ liệu nhân sự

Sau khi export ta được file dữ liệu lưu với tên HC_EMP.CSV. Ta dùng file này để thực hiện phân lớp bằng phần mềm R.

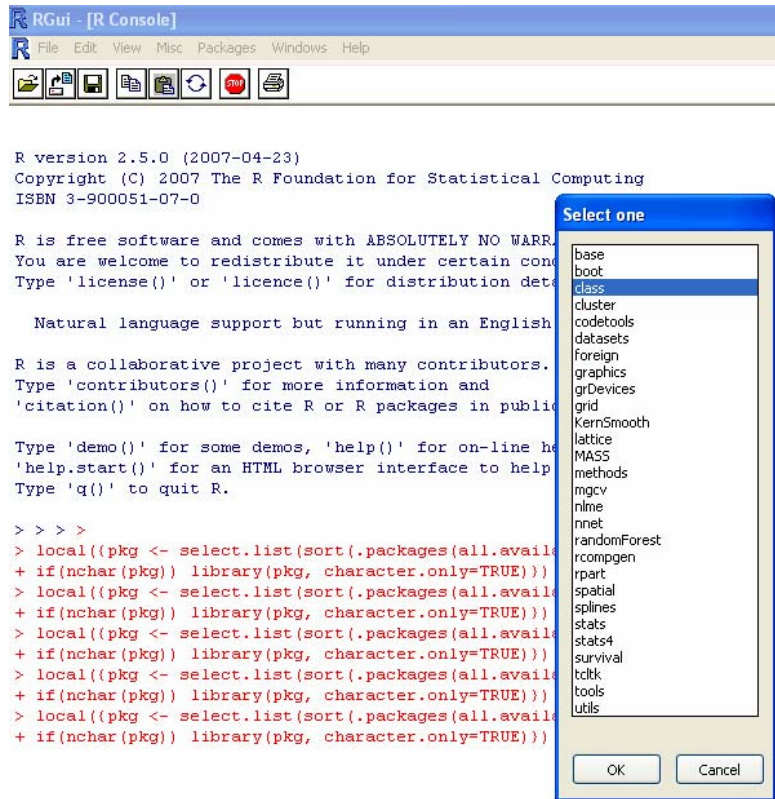
Nhập dữ liệu vào R ta dùng lệnh Read.CSV:

```
> setwd("c:/works/insulin")  
> gh <- read.csv ("HC_EMP.CSV", header=TRUE)
```

Lệnh thứ hai read.csv yêu cầu R đọc số liệu từ “HC_EMP.csv”, dùng dòng thứ nhất là tên cột, và lưu các số liệu này trong một object có tên là gh. Bây giờ chúng ta có thể lưu gh dưới dạng R để xử lí sau này bằng lệnh sau đây:

```
> save(gh, file="gh.rda")
```


Lấy Packages để thực hiện phân lớp chọn Random forest:



Hình 2.3 Minh họa chức năng chọn phân lớp dữ liệu trong R

Đánh giá hỗ trợ của R về mặt phân tích và dự đoán số liệu:

Qua màn hình và một số tư liệu tham khảo cho thấy R là công cụ phân tích thông kê mạnh. Nhưng cũng có một số đánh giá sau:

- Thích hợp mô hình phân tích thống kê và vẽ biểu đồ;
- Giao diện kết quả khó khai thác(đây là ứng dụng cài đặt cho từng máy trạm không phát triển được ứng dụng phân tích trực tuyến);
- Dữ liệu đầu vào cho phân tích dạng bảng theo cấu trúc file csv hoặc file text. Công cụ không có hỗ trợ kết nối trực tiếp vào cơ sở dữ liệu;
- Kết quả đầu ra không trực quan;

Chính vì những lý do này luận văn không tiếp tục nghiên cứu sử dụng R cho bài toán phân tích số liệu nhân sự.

2.4.2. Phân tích số liệu bằng phần mềm weka

Giới thiệu Weka:

Weka là phần mềm khai thác dữ liệu viết bằng ngôn ngữ Java. Weka tập hợp các thuật toán máy học cho các tác vụ khai thác dữ liệu. Weka gồm các công cụ thực hiện: tiền xử lý dữ liệu(data pre-processing), phân lớp (classification), hồi quy (regression), gom cụm (clustering), luật kết hợp (association rules). Chúng ta có thể tìm hiểu và sử dụng nó qua website <http://www.cs.waikato.ac.nz/~ml/weka/index.html>.

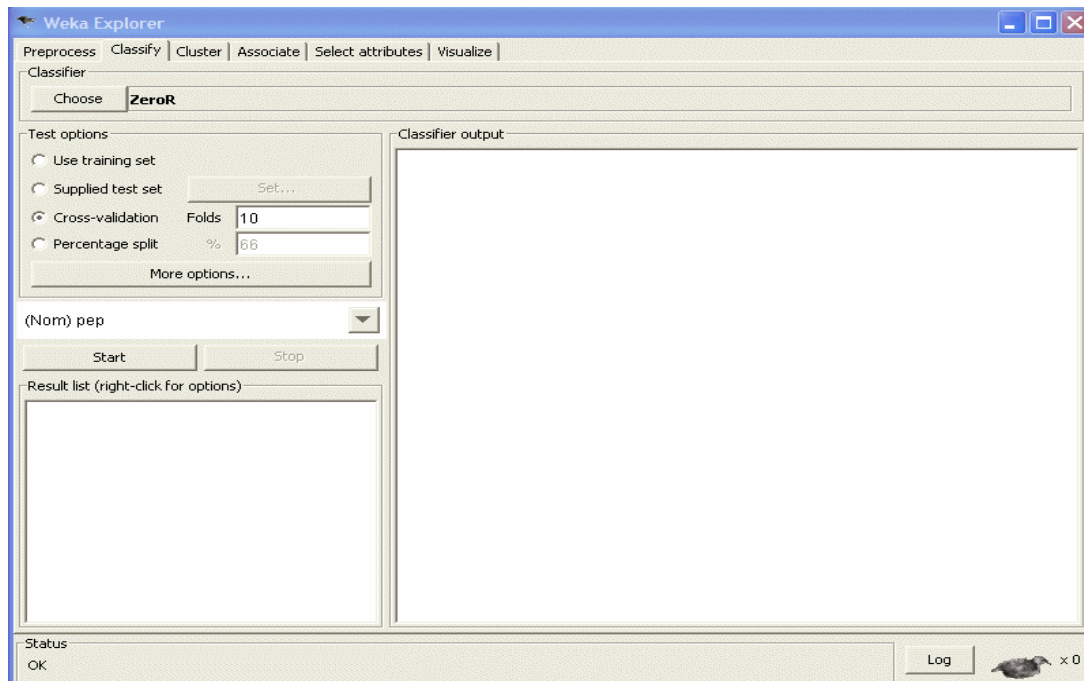
Xử lý file dữ liệu để phân lớp trong Weka:

File dữ liệu dùng trong Weka là file .arff(file này do phần mềm Weka tạo ra) hoặc file .csv (file dạng dữ liệu Excel).

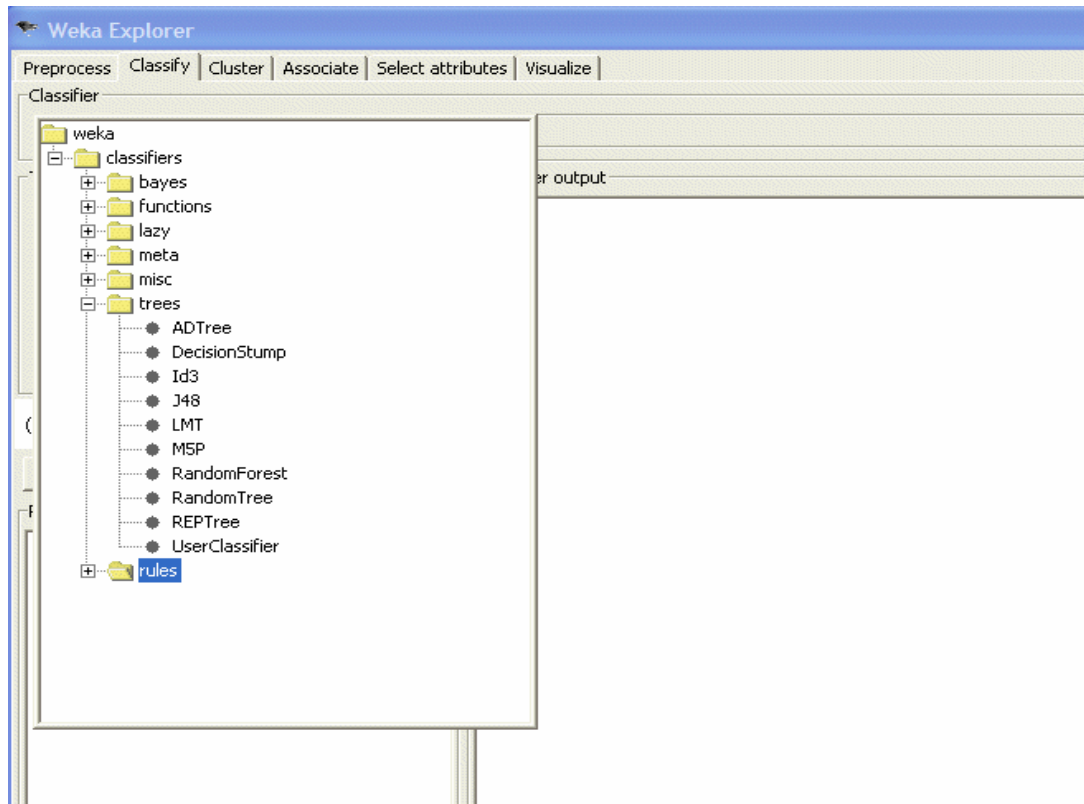
Phân lớp trong Weka

Chúng ta có thể dùng dữ liệu file.csv của cơ sở dữ liệu nhân sự vừa kết xuất (export) bằng chức năng của hệ quản trị Microsoft SQL Server hoặc sử dụng 2 file bank-data.csv và bank.arff do phần mềm cung cấp để kiểm tra chức năng phân lớp của Weka.

- Để thực hiện phân lớp, đầu tiên chúng ta phải chọn file cần phân lớp (Hình 2.4 là hình ảnh sau khi tải file dữ liệu).
- Sau khi mở file, chọn tab **Classify**. Nhấn nút **Choose** để chọn phân lớp theo luật nào: Bayes, C4.5,...(trong đương C4.5 , trong Weka là J48)

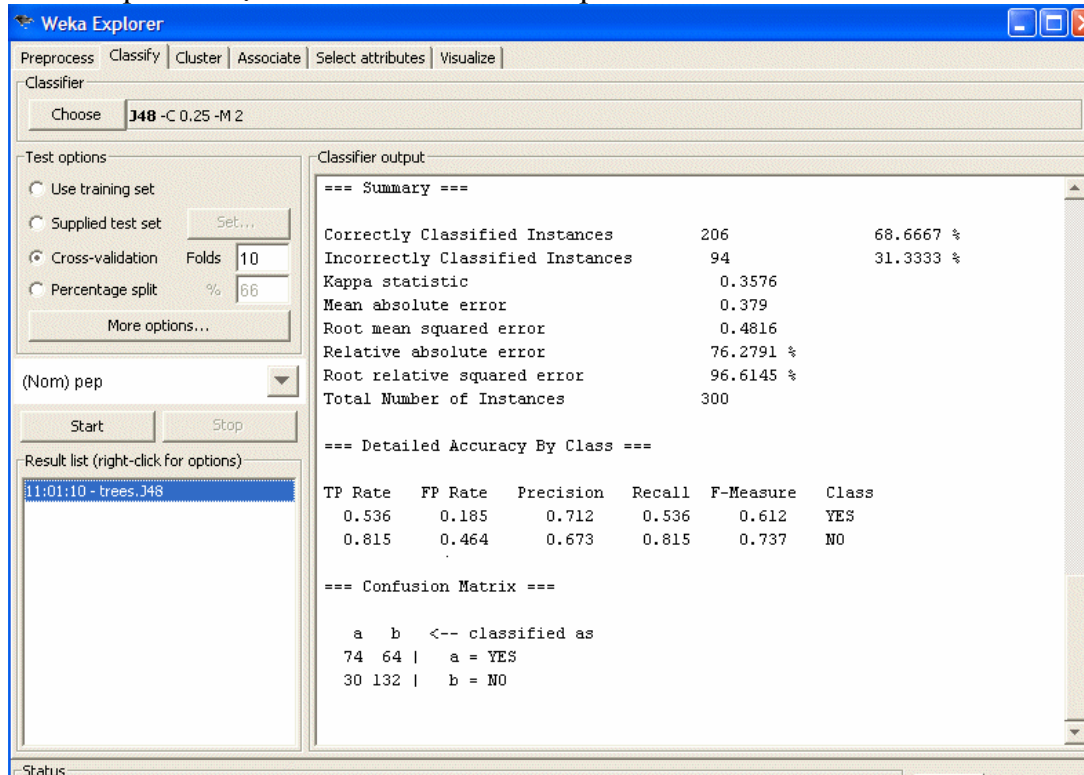


Hình 2.4 Minh họa chọn phân lớp trong weka



Hình 2.5 Liệt kê các phương pháp phân lớp của công cụ

- Kết quả sẽ hiện ra trên màn hình bên phải:



Hình 2.6 Liệt kê kết quả phân lớp

Đánh giá hỗ trợ của weka đối với yêu cầu phân tích trên dữ liệu nhân sự:

- Cơ sở dữ liệu nhân sự dùng Hệ quản trị Microsoft SQL. Vì thế, để có dữ liệu cho phân tích ta phải kết xuất dữ liệu theo định dạng;
- Giao diện kết quả khó khai thác(đây là ứng dụng cài đặt cho từng máy trạm không phát triển được ứng dụng phân tích trực tuyến);
- Dữ liệu đầu vào cho phân tích dạng bảng theo cấu trúc file csv hoặc file text. Không có hỗ trợ kết nối trực tiếp vào cơ sở dữ liệu;
- Kết quả đầu ra không trực quan.

2.4.3 Phân tích số liệu bằng See5/C5.0

“See5 là một dạng nghệ thuật của hệ thống xây dựng sự phân loại trong dạng thức của những cây quyết định và tập luật “. See5 đã được thiết kế và hoạt động trên cơ sở dữ liệu lớn và sự kết hợp đổi mới như là boosting. Kết quả tạo ra bởi See5 và C5.0 là tương tự nhau. Hoạt động trước đây trên Windows95/98/NT của C5.0 là phần hoạt động của nó trên Unix . See 5 và C5.0 là những công cụ khai thác dữ liệu phức tạp cho những mẫu khai phá dữ liệu mà phác họa ra những loại tập hợp chúng thành những đối tượng phân loại và sử dụng chúng để tiên đoán.

Đặc điểm chính của C5.0 là :

- ▶▶ C5.0 được thiết kế để phân tích những cơ sở dữ liệu quan trọng chứa đựng hàng ngàn đến hàng trăm ngàn những records.và hàng chục đến hàng trăm số liệu và hoặc tên trường (field).
- ▶▶ Để tối đa khả năng giải thích , đối tượng phân loại của See5.0 /C5.0 được diễn tả như là cây quyết định hoặc tập của những luật if – then.Dạng thức của nó dễ hiểu hơn so với neutron network.

C5.0 dễ dàng sử dụng do đó không được gọi là kiến thức cao cấp của thống kê và học máy.

Xử lý file dữ liệu

Mỗi bộ dữ liệu dùng trong See5/C5.0 gồm có 3 file:

- Filestem.names: định nghĩa bộ dữ liệu;
- Filestem.data: chứa dữ liệu training, có cấu trúc như sau: mỗi dòng tương ứng với một bản ghi (cases) trong cơ sở dữ liệu. Mỗi dòng một bộ giá trị theo thứ đã định của các thuộc tính định nghĩa trong *filestem.names*. Các giá trị ngăn cách nhau bởi dấu phẩy. Giá trị thiếu (*missing value*) được biểu diễn bằng dấu “?”.

-Filestem.test: chứa dữ liệu test, File này chứa dữ liệu test trên mô hình phân lớp đã được tạo ra từ tập dữ liệu training, và có cấu trúc giống filestem.data.

Đánh giá hỗ trợ của See5/C5.0 về mặt phân tích và dự đoán số liệu:

- Cơ sở dữ liệu nhân sự dùng Hệ quản trị Microsoft SQL. Vì thế, để có dữ liệu cho phân tích ta phải kết xuất dữ liệu theo định dạng;
- Giao diện kết quả khó khai thác theo yêu cầu(đây là ứng dụng cài đặt cho từng máy trạm không phát triển được ứng dụng phân tích trực tuyến);
- Dữ liệu đầu vào cho phân tích dạng phải định nghĩa cấu trúc file. Không có hỗ trợ kết nối trực tiếp vào cơ sở dữ liệu;
- Kết quả đầu ra không trực quan.

2.4.4. Phân tích số liệu bằng DTREG1

Giới thiệu:

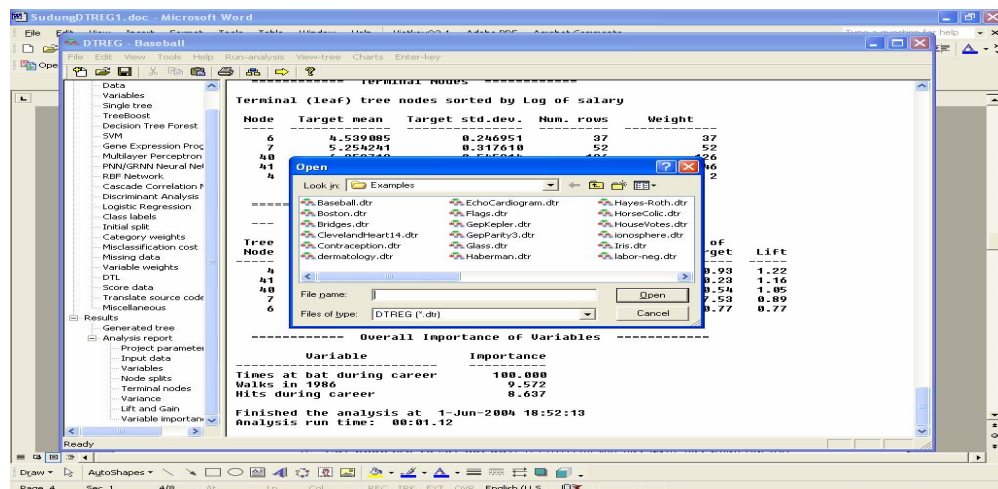
DTREG là chương trình phân tích thống kê mạnh, phát sinh cây quyết định phân lớp, hồi quy và mô hình SVM để mô tả mối liên hệ dữ liệu, có thể sử dụng để dự đoán giá trị cho sự khảo sát tương lai.

File dữ liệu

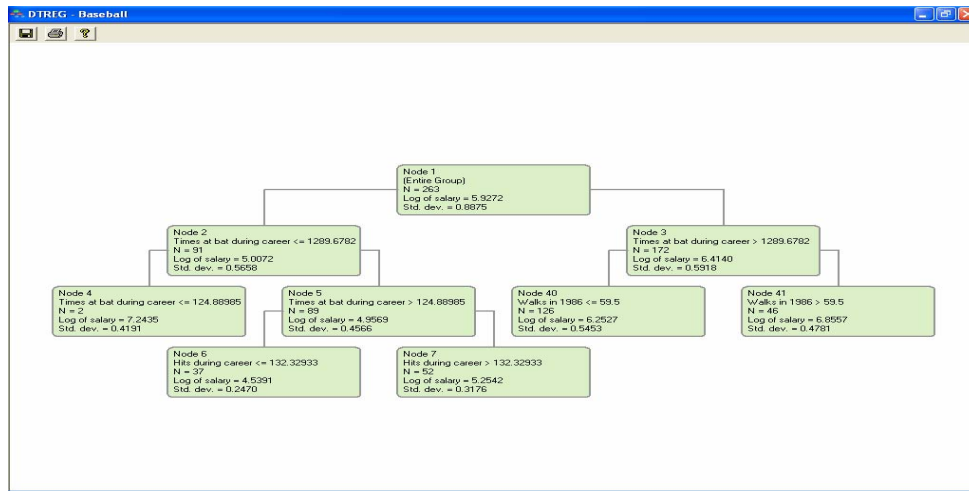
DTREG gồm 2 file:

File .csv (file dạng dữ liệu Excel) chứa dữ liệu nguồn phục vụ cho việc phân tích;

File dtree chứa các kết quả để hiển thị phân tích:



Hình 2.7 Dạng kết quả của DTREE



Hình 2.8 Kết quả cây quyết định

Đánh giá phân tích số liệu DTree

- Giao diện kết quả khó khai thác theo yêu cầu(đây là ứng dụng cài đặt cho từng máy trạm không phát triển được ứng dụng phân tích trực tuyến);
- Dữ liệu đầu vào cho phân tích dạng phải định nghĩa cấu trúc file. Không có hỗ trợ kết nối trực tiếp vào cơ sở dữ liệu.

2.4.5. Phân tích số liệu sử dụng công cụ của Microsoft:

Trong phần này sẽ trình bày cách thức công cụ “Microsoft Analysis Services” được sử dụng để hiện thực mô hình cây quyết định trong phần mềm Microsoft SQL Server 2000 .Chúng ta đề cập đến tạo mô hình cây quyết định với mô hình - một sử dụng những bảng quan hệ chuẩn như là nguồn.

a).Tạo mô hình:

Bước đầu tiên trong hoạt động khai phá dữ liệu là tạo mô hình . Mô hình khai phá dữ liệu được tạo ra khác biệt với các công cụ khác là từ những mẫu tin chứa trong một nguồn dữ liệu (data source) . Một vài nguồn dữ liệu có thể được kết nối thông qua OLE DB có thể được sử dụng để tạo mô hình . Những nguồn này bao gồm cơ sở dữ liệu quan hệ , OLAP cubes, FoxPro tables, text file , hoặc thậm chí Microsoft Excel spread sheets. Chúng ta cũng sẽ tập trung vào cách thức để sử dụng những nguồn dữ liệu này để lưu trữ test case được sử dụng để tạo tiên đoán và cách thức để chứa kết quả của những tiên đoán.

Sản phẩm của Microsoft đi đôi với những tác vụ trong một giới hạn và có thể tiên đoán một số bước. Mining mode wizard sẽ dẫn dắt chúng ta từng bước để tạo một mô hình:

1. Chọn nguồn (Select source);
2. Chọn case table hoặc những bảng cho mô hình khai phá dữ liệu;
3. Chọn kỹ thuật khai phá dữ liệu (giải thuật);
4. Hiệu chỉnh những kết nối của những bảng được chọn như là nguồn trong những bước trước;
5. Chọn cột Case Key;
6. Chọn Input và cột tiên đoán;
7. Kết thúc .

b). Các thuật toán được Microsoft khuyến cáo sử dụng với kỹ thuật thực hiện:

The image shows a matrix titled "Data Mining Algorithms" with columns for Decision Trees, Naïve Bayes, Clustering, Seq. Clustering, Time Series, Association rules, and Neural Network. The rows represent different tasks: Classification, Regression, Segmentation, Assoc. Analysis, Anomaly Detect., Seq. Analysis, and Time series. Green checkmarks indicate a first choice, and orange checkmarks indicate a second choice.

	Decision Trees	Naïve Bayes	Clustering	Seq. Clustering	Time Series	Association rules	Neural Network
Classification	✓	✓	✓	✓		✓	✓
Regression	✓	✓	✓	✓			✓
Segmentation			✓	✓			✓
Assoc. Analysis	✓	✓	✓	✓		✓	✓
Anomaly Detect.			✓	✓			✓
Seq. Analysis				✓			
Time series					✓		

Legend: ✓ - first choice, ✓ - second choice

Hình 2.9 Bảng khuyến cáo lựa chọn thuật toán của Microsoft

Từ bảng cho thấy cây quyết định là lựa chọn số 1 cho các kỹ thuật phân lớp, hồi quy và luật kết hợp. Cây quyết định không có lựa chọn thứ 2.

c). Kết luận về công cụ “Microsoft Analysis service”:

Trong các công cụ trên công cụ phân tích của Microsoft thể hiện được tính ưu việt:

- Dễ dàng kết nối với hệ quản trị cơ sở dữ liệu dùng Microsoft SQL Server;
- Sử dụng máy chủ (Server) phân tích chỉ cần kết nối với máy chủ phân tích có thể làm việc từ bất kỳ đâu không cần cài đặt;

- Công cụ phân tích của Microsoft sử dụng nhiều thuật toán của Datamining. Vì thế mềm dẻo thuận tiện cho khai thác nghiệp vụ thay đổi (Chi tiết về việc sử dụng để phân lớp hình thành cây quyết định sẽ được giới thiệu đầy đủ hơn trong phần công cụ lựa chọn);

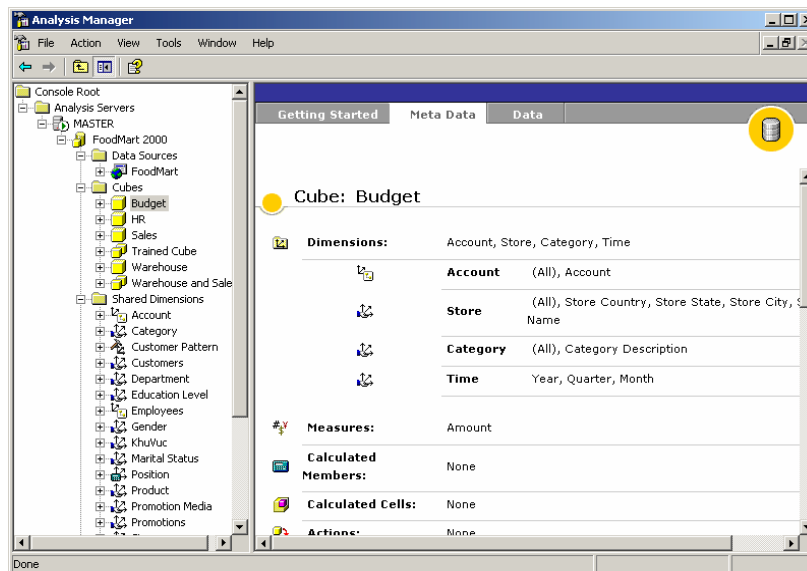
- Khi cơ sở dữ liệu dùng hệ quản trị Microsoft SQL Server việc lựa chọn công cụ để phân tích dữ liệu với yêu cầu trực tuyến thì chỉ có một lựa chọn là sản phẩm của Microsoft: Microsoft Analysis Service. Với sản phẩm này người sử dụng có thể dùng Microsoft Excel để lấy dữ liệu từ cơ sở dữ liệu dùng để phân tích bảng tính.

2.5. Công cụ lựa chọn:

Giới thiệu công cụ:

Analysis Service - Một dịch vụ phân tích dữ liệu rất hay của Microsoft. Hỗ trợ lấy dữ liệu bổ ích từ chứa trong cơ sở dữ liệu. Microsoft cung cấp cho bạn một công cụ rất mạnh giúp cho việc phân tích dữ liệu trở nên dễ dàng và hiệu quả bằng cách dùng khái niệm hình khối nhiều chiều (multi-dimension cubes) và kỹ thuật "khai phá dữ liệu". Phần này giới thiệu về khả năng của công cụ trong việc xây dựng mô hình.

Mô tả màn hình giao diện



Hình 2.10 Màn hình Analysis Manager

- Bên trái màn hình là cây theo dõi cơ sở dữ liệu và các đối tượng trong cơ sở dữ liệu. Bên dưới nút có tên Analysis Server, ta có thể khai báo nhiều cơ sở dữ liệu, hoặc import từ nhiều nguồn khác nhau như : Access, SQL Server,

Oracle,...;

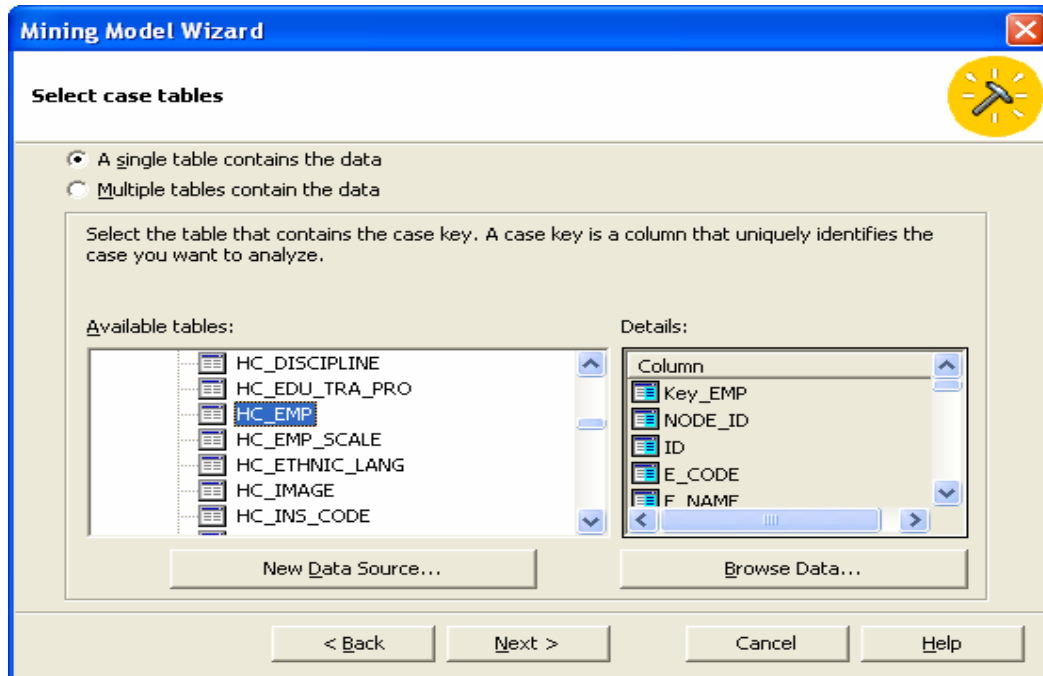
- Bên phải màn hình hiển thị thông tin chi tiết về 1 đối tượng đang chọn trên cây bên trái. Tab 'Meta Data' hiển thị thông tin chung về cấu trúc của đối tượng đó. Tab 'Data' hiển thị dữ liệu đang chứa trong đối tượng đó.

Các chức năng cơ bản trên màn hình Analysis Manager :

- Đăng ký một Analysis Server (máy chủ phân tích dữ liệu);
- Tạo một database (cơ sở dữ liệu phân tích);
- Tạo một khối dữ liệu (dữ liệu tổng hợp-phân tích nhiều chiều);
- Tạo một chiều (dimension);
- Tạo phép đo (measure).
- Hiển thị mô hình (Brow)
- Lấy dữ liệu (process)

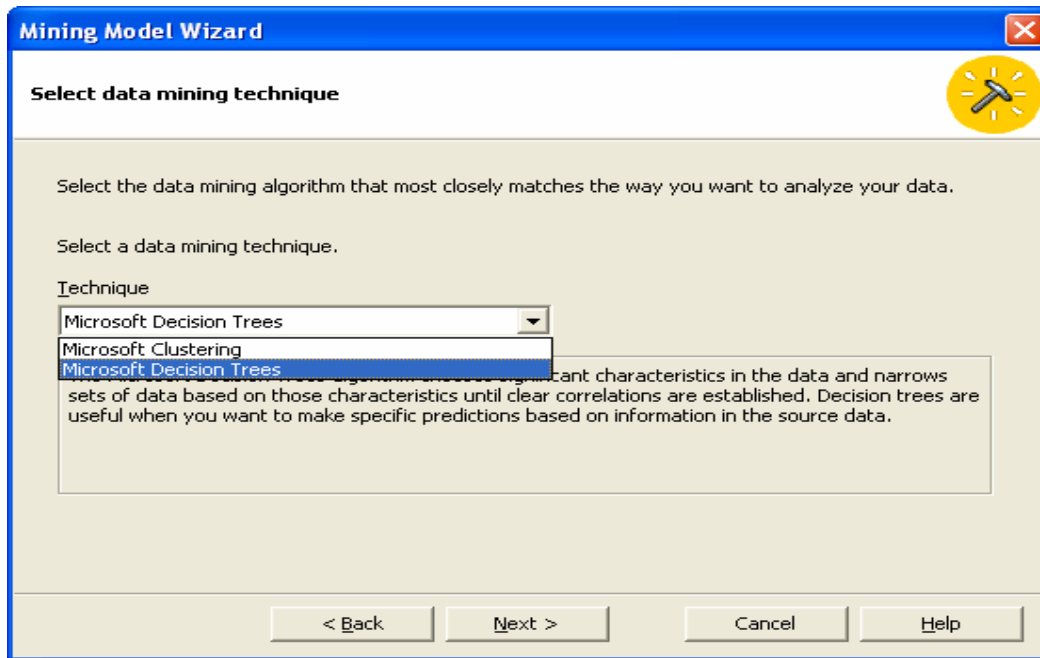
Xây dựng mô hình phân tích kho dữ liệu nhân sự:

Lấy dữ liệu nguồn chọn bảng HC_EMP (bảng chứa thông tin hồ sơ nhân sự)



Hình 2.11 Chọn bảng dữ liệu đầu vào cho mô hình

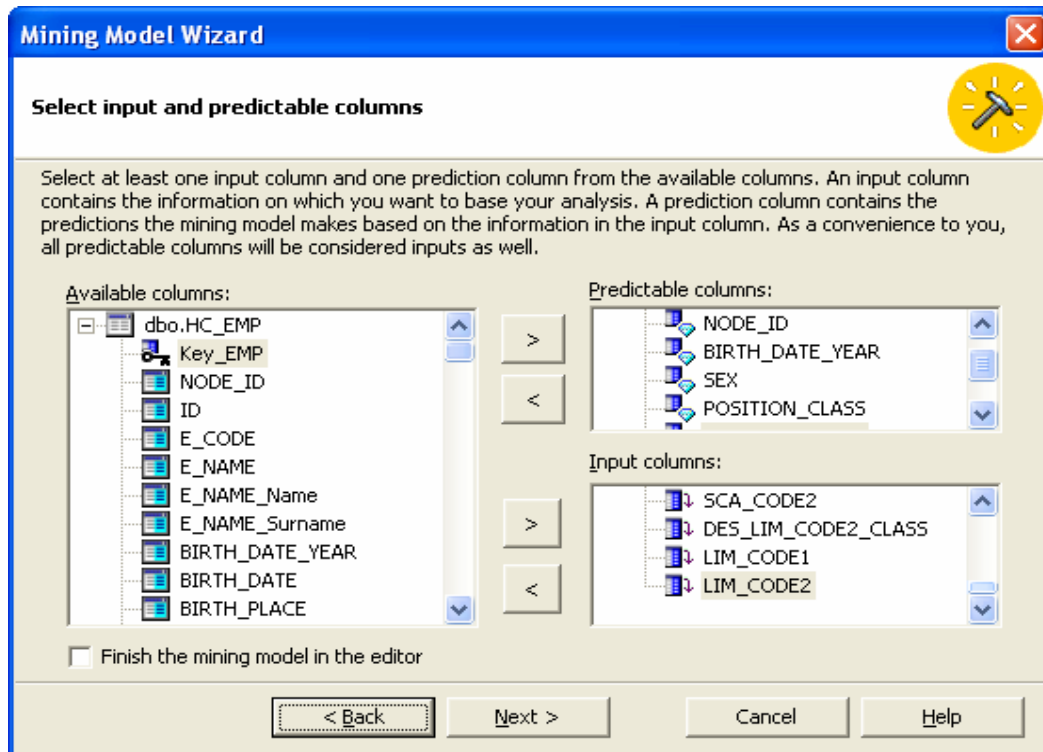
Lựa chọn Datamining technique là Microsoft Decision Trees



Hình 2.12 Chọn kỹ thuật “Data Mining”

Chọn Thuộc tính đầu vào

Chọn thuộc tính cần dự đoán (nhân lớp): Giới tính (Sex); Đơn vị công tác (Node_ID); Năm sinh (Birth_Date_year); Chuyên ngành đào tạo (Des_Lim_Code2_class); Sức khỏe (HEA_CON_CLASS); Chuyên ngành đào tạo (DES_LIM_CODE2_CLASS); Đăng viên (RPOS_DATE_YEAR)



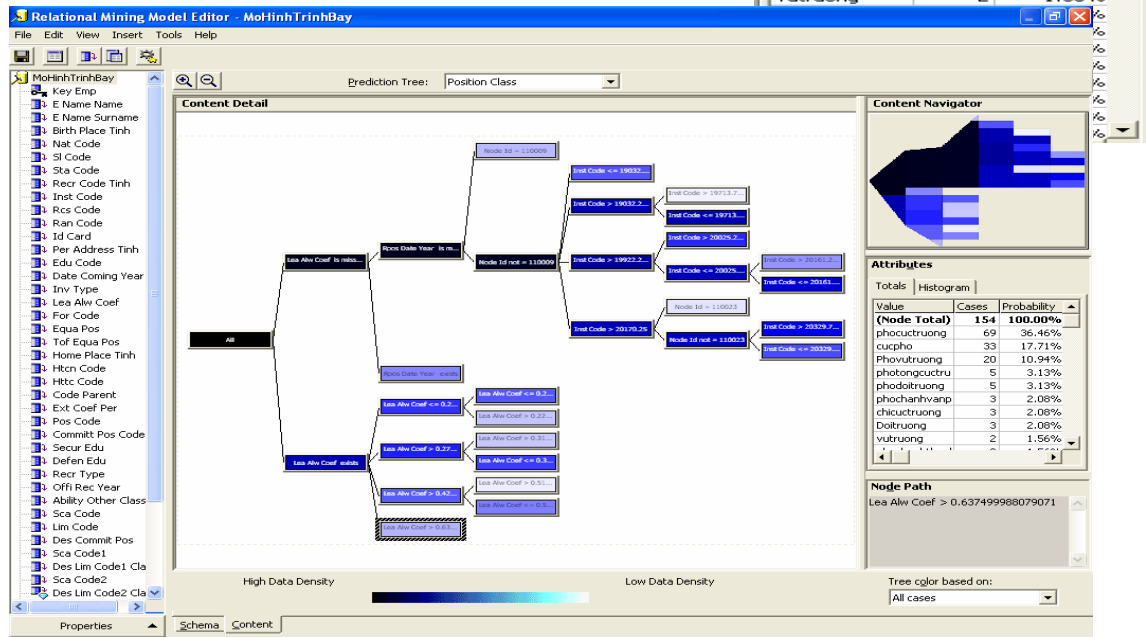
Hình 2.13 Lựa chọn các cột dữ liệu cho mô hình dự đoán

Kết quả mô hình là một cây quyết định, bên trên cùng là các lựa chọn thuộc tính cần dự đoán, ở giữa là cây quyết định tương ứng, bên phải là đánh giá, bên trái là các thuộc tính. Để có thông tin chi tiết của mô hình ta chọn từng lá sẽ có thông tin dự đoán về thuộc tính lựa chọn.

Chọn lá cuối cùng của cây dựa đoán nhân lớp lãnh đạo

Lea Alw Coef > 0.637499988079071 (hệ số phụ cấp lãnh đạo 0.637499988076071 thì thuộc lớp lãnh đạo. Giá trị của lớp này như sau:

Value	Cases	Probability
(Node Total)	154	100.00%
phocuctruong	69	36.46%
cucpho	33	17.71%
Phovutruong	20	10.94%
photongcuctru	5	3.13%
phodoitruong	5	3.13%
phochanhvanp	3	2.08%
chicuctruong	3	2.08%
Doitruong	3	2.08%
vutruong	2	1.56%



Hình 2.14 Kết quả cây quyết định

Khai thác mô hình

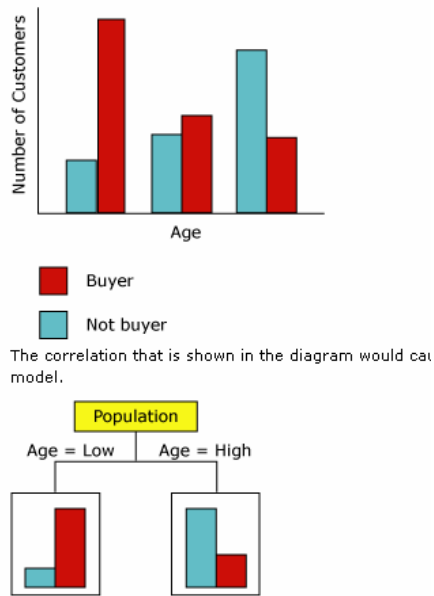
Bây giờ ta có một trường hợp mà trường vị trí là lãnh đạo cấp cao thì trường hệ số phụ cấp lãnh đạo sẽ có giá trị : Lea Alw Coef > 0.637499988079071. Ngược lại nói khác đi nếu một trường hợp có hệ số phụ cấp lãnh đạo Lea_Alw_Coef > 0.637499988079071 thì sẽ là lãnh đạo cấp cao.

Thuật toán áp dụng

Thuật toán Microsoft Decision Tree hỗ trợ cả việc phân lớp và hồi quy , và tạo rất tốt các mô hình dự đoán. Dữ liệu vào đối với mô hình cây quyết định: Cột dữ liệu cần dự đoán, cột dữ liệu vào, một thuộc tính khóa để phân biệt các trường hợp. Sử dụng thuật toán này có thể dự đoán cả các thuộc tính rời rạc và liên tục.

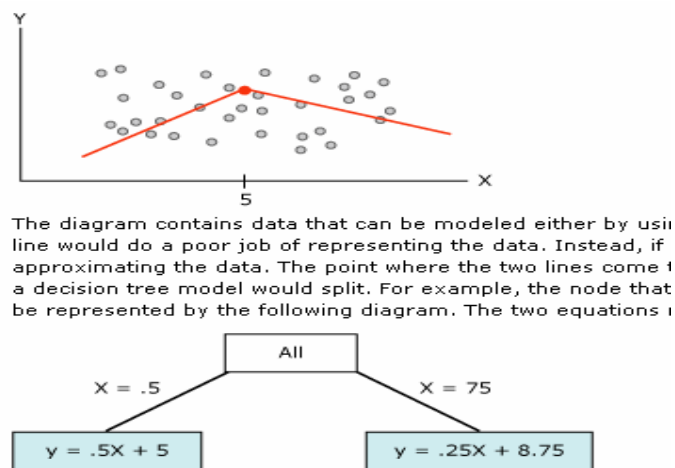
Với thuộc tính rời rạc, thuật toán dự đoán dựa trên mối quan hệ của các thuộc tính vào. Đặc biệt thuật toán phát hiện thuộc tính vào có mối tương đồng với thuộc tính dự đoán. Ví dụ để dự đoán khách hàng thích mua xe đạp, có 9 trong 10 khách hàng mua xe đạp là người ít tuổi, chỉ có 1 khách là người lớn tuổi. Thuật toán sẽ suy luận tuổi dự đoán của khách hàng mua xe đạp, cây quyết định dự đoán dựa trên thiên hướng này.

Với thuộc tính liên tục, thuật toán Microsoft Decision Tree dùng lược đồ:



Hình 2.15 Lược đồ minh họa cho dự đoán thuộc tính liên tục

Với thuộc tính rời rạc thuật toán dùng hồi quy để xác định sự phân chia cây:



Hình 2.16 Minh họa cho dự đoán thuộc tính liên tục

Nếu có nhiều dự đoán (tập dự đoán) thuật toán sẽ xây dựng cây quyết định độc lập cho mỗi cột dự đoán.

Trong việc xây dựng mô hình, thuật toán này sẽ khảo sát sự **ảnh hưởng** của mỗi thuộc tính trong tập dữ liệu và kết quả của thuộc tính dự đoán . Tiếp đến nó sử dụng các thuộc tính **input** (thuộc tính vào với các quan hệ rõ ràng) để tạo thành 1 nhóm **phân hoá** gọi là các node . Khi 1 node mới được thêm vào mô hình , 1 cấu trúc cây sẽ được thiết lập . Node đỉnh của cây sẽ miêu tả sự phân tích (bằng thống kê)của các thuộc tính dự đoán thông qua các mẫu . Mỗi node thêm vào sẽ được tạo ra dựa trên sự sắp xếp các trường của thuộc tính dự đoán, để so sánh với dữ liệu thuộc tính input . Nếu 1 thuộc tính input được coi là nguyên nhân của thuộc tính dự đoán (to favour one state over another), 1 node mới sẽ thêm vào mô hình. Mô hình tiếp tục phát triển cho đến lúc không còn thuộc tính nào, tạo thành 1 sự phân tách(split) để cung cấp một dự báo hoàn chỉnh thông qua các node đã tồn tại . Mô hình đòi hỏi tìm kiếm một sự **kết hợp** giữa các thuộc tính và trường của nó , nhằm thiết lập một sự phân phối không cân xứng giữa các trường trong thuộc tính dự đoán, Vì thế cho phép dự đoán kết quả của thuộc tính dự đoán một cách tốt nhất.

Kết luận chương 2

Chương 2 nghiên cứu các công nghệ kỹ thuật và công cụ sử dụng phù hợp với yêu cầu đặt ra. Kết quả đã cho thấy kỹ thuật phân lớp dữ liệu dựa trên cây quyết định có nhiều ưu biệt và việc sử dụng công cụ “Microsoft Analysis Service” khá thuận tiện, trực quan, mang lại độ chính xác hỗ trợ của mô hình dự đoán. Trong các công cụ trên công cụ phân tích của Microsoft thể hiện được tính ưu việt:

- Dễ dàng kết nối với hệ quản trị cơ sở dữ liệu dùng hệ quản trị Microsoft SQL server;

- Sử dụng hỗ trợ phân tích trên “web” chỉ cần kết nối với máy chủ (server) kết nối là có thể dùng công cụ để làm việc từ bất kỳ đâu không cần cài đặt (Analysis service on web);

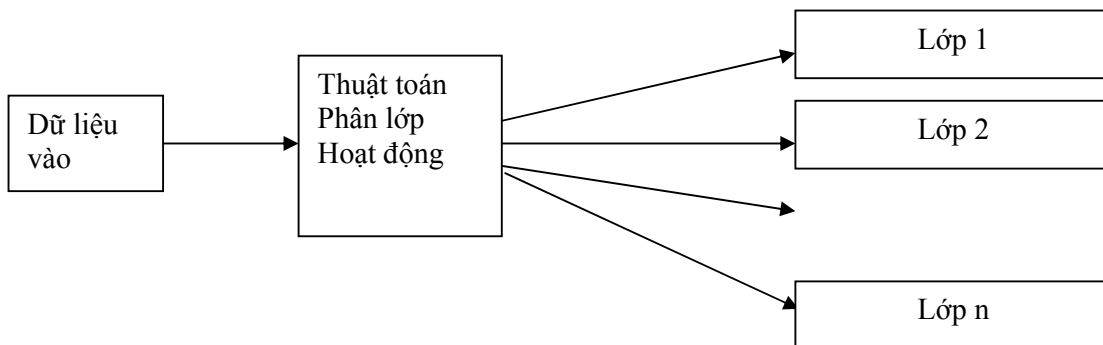
- Công cụ phân tích của Microsoft sử dụng nhiều thuật toán của Data Mining. Vì thế mềm dẻo thuận tiện cho khai thác nghiệp vụ thay đổi.

CHƯƠNG 3. PHÂN LỚP DỮ LIỆU SỬ DỤNG CÂY QUYẾT ĐỊNH

3.1. Tổng quan về phân lớp dữ liệu trong khai phá dữ liệu

3.1.1. Phân lớp dữ liệu

Một trong các nhiệm vụ chính của khai phá dữ liệu là giải quyết bài toán phân lớp. Đầu vào của bài toán phân lớp là một tập các mẫu học đã được phân lớp trước, mỗi mẫu được mô tả bằng một số thuộc tính. Các thuộc tính dùng để mô tả một mẫu gồm hai loại là thuộc tính liên tục và thuộc tính rời rạc. Trong số các thuộc tính rời rạc có một thuộc tính đặc biệt là *phân lớp*, mà các giá trị của nó được gọi là nhãn lớp. Thuộc tính liên tục sẽ nhận các giá trị có thứ tự, ngược lại thuộc tính rời rạc sẽ nhận các giá trị không có thứ tự. Ngoài ra, các thuộc tính có thể nhận giá trị không xác định (chẳng hạn, vì những lý do khách quan ta không thể biết được giá trị của nó). Chú ý rằng nhãn lớp của tất cả các mẫu không được phép nhận giá trị không xác định. Nhiệm vụ của quá trình phân lớp là thiết lập được ánh xạ giữa giá trị của các thuộc tính với các nhãn lớp. Mô hình biểu diễn quan hệ nói trên sau đó sẽ được dùng để xác định nhãn lớp cho các quan sát mới không nằm trong tập mẫu ban đầu.



Hình 3.1 Bài toán phân lớp

Thực tế đặt ra nhu cầu từ một cơ sở dữ liệu với nhiều thông tin ẩn ta có thể trích rút ra các quyết định nghiệp vụ thông minh. Phân lớp và dự đoán là hai dạng của phân tích dữ liệu nhằm trích rút ra một mô hình mô tả các lớp dữ liệu quan trọng hay dự đoán xu hướng dữ liệu tương lai. Phân lớp dự đoán giá trị của những nhãn xác định (*categorical label*) hay những giá trị rời rạc (*discrete value*), có nghĩa là phân lớp thao tác với những đối tượng dữ liệu mà có bộ giá trị là biết trước. Trong khi đó, dự đoán lại xây dựng mô hình với các hàm nhận

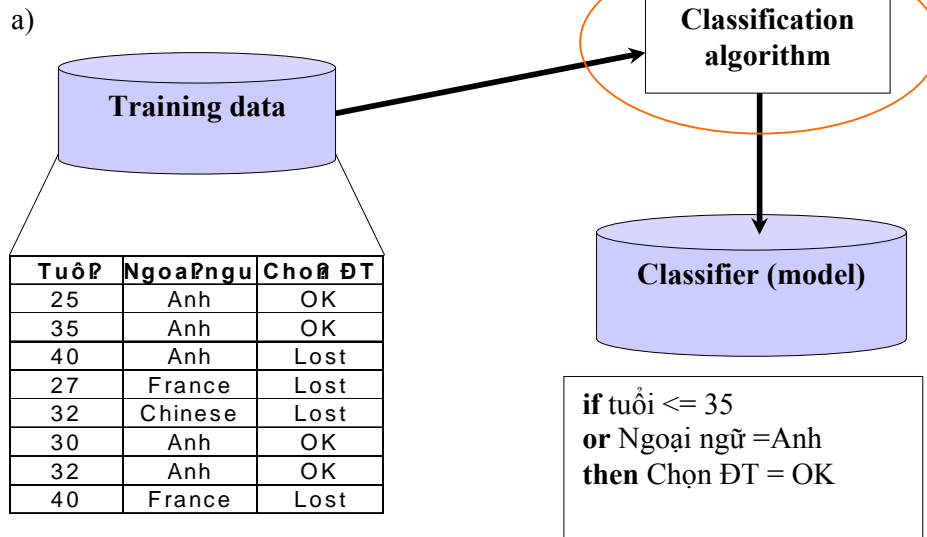
giá trị liên tục. Ví dụ mô hình phân lớp dự báo thời tiết có thể cho biết thời tiết ngày mai là mưa, hay nắng dựa vào những thông số về độ ẩm, sức gió, nhiệt độ,... của ngày hôm nay và các ngày trước đó. Hay nhờ các luật về xu hướng mua hàng của khách hàng trong siêu thị, các nhân viên kinh doanh có thể ra những quyết sách đúng đắn về lượng mặt hàng cũng như chủng loại bày bán... Một mô hình dự đoán có thể dự đoán được lượng tiền tiêu dùng của các khách hàng tiềm năng dựa trên những thông tin về thu nhập và nghề nghiệp của khách hàng. Trong những năm qua, phân lớp dữ liệu đã thu hút sự quan tâm các nhà nghiên cứu trong nhiều lĩnh vực khác nhau như học máy (machine learning), hệ chuyên gia (expert system), thống kê (statistics)... Công nghệ này cũng ứng dụng trong nhiều lĩnh vực khác nhau như: thương mại, nhà băng, marketing, nghiên cứu thị trường, bảo hiểm, y tế, giáo dục...

Quá trình phân lớp dữ liệu gồm hai bước [1]:

- Bước thứ nhất (learning)

Quá trình học nhằm xây dựng một mô hình mô tả một tập các lớp dữ liệu hay các khái niệm định trước. Đầu vào của quá trình này là một tập dữ liệu có cấu trúc được mô tả bằng các thuộc tính và được tạo ra từ tập các bộ giá trị của các thuộc tính đó. Mỗi bộ giá trị được gọi chung là một phần tử dữ liệu (data tuple), có thể là các mẫu (sample), ví dụ (example), đối tượng (object), bản ghi (record) hay trường hợp (case). Luận văn sử dụng các thuật ngữ này với nghĩa tương đương. Trong tập dữ liệu này, mỗi phần tử dữ liệu được giả sử thuộc về một lớp định trước, lớp ở đây là giá trị của một thuộc tính được chọn làm thuộc tính gán nhãn lớp hay thuộc tính phân lớp (class label attribute). Đầu ra của bước này thường là các quy tắc phân lớp dưới dạng luật dạng if-then, cây quyết định, công thức logic, hay mạng nơron. Quá trình này được mô tả như trong hình vẽ:

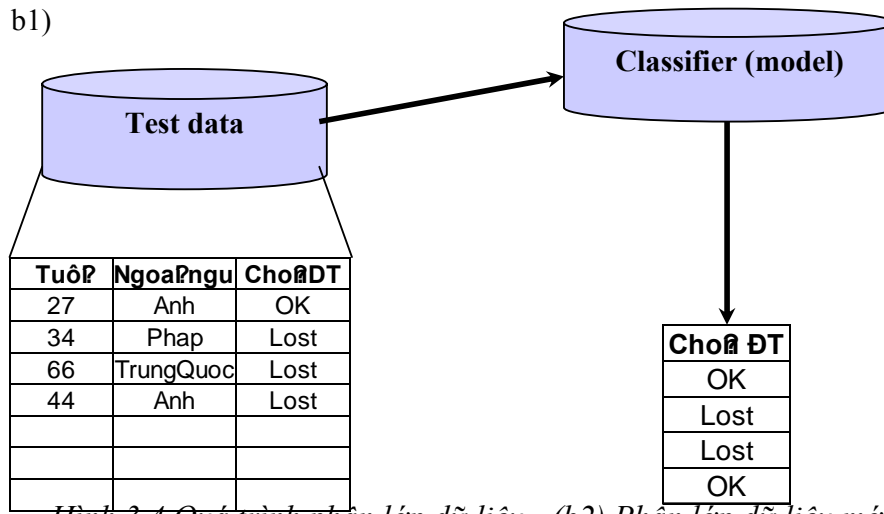
Hình 3.2 Quá trình phân lớp dữ liệu - (a) Bước xây dựng mô hình phân lớp



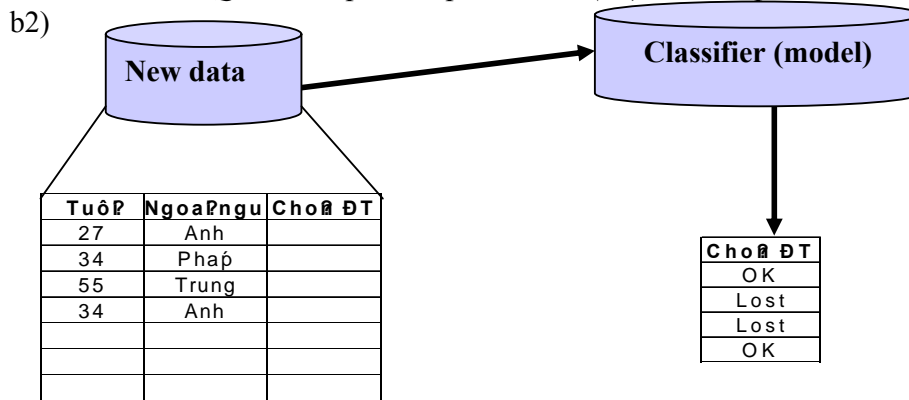
- **Bước thứ hai (classification)**

Bước thứ hai dùng mô hình đã xây dựng ở bước trước để phân lớp dữ liệu mới. Trước tiên độ chính xác mang tính chất dự đoán của mô hình phân lớp vừa tạo ra được ước lượng. Holdout là một kỹ thuật đơn giản để ước lượng độ chính xác đó. Kỹ thuật này sử dụng một tập dữ liệu kiểm tra với các mẫu đã được gán nhãn lớp. Các mẫu này được chọn ngẫu nhiên và độc lập với các mẫu trong tập dữ liệu đào tạo. Độ chính xác của mô hình trên tập dữ liệu kiểm tra đã đưa là tỉ lệ phần trăm các các mẫu trong tập dữ liệu kiểm tra được mô hình phân lớp đúng (so với thực tế). Nếu độ chính xác của mô hình được ước lượng dựa trên tập dữ liệu đào tạo thì kết quả thu được là rất khả quan vì mô hình luôn có xu hướng “quá vừa” dữ liệu. Quá vừa dữ liệu là hiện tượng kết quả phân lớp trùng khít với dữ liệu thực tế vì quá trình xây dựng mô hình phân lớp từ tập dữ liệu đào tạo có thể đã kết hợp những đặc điểm riêng biệt của tập dữ liệu đó. Do vậy, cần sử dụng một tập dữ liệu kiểm tra độc lập với tập dữ liệu đào tạo. Nếu độ chính xác của mô hình là chấp nhận được, thì mô hình được sử dụng để phân lớp những dữ liệu tương lai, hoặc những dữ liệu mà giá trị của thuộc tính phân lớp là chưa biết.

Hình 3.3 Quá trình phân lớp dữ liệu - (b1) Ước lượng độ chính xác của mô hình



Hình 3.4 Quá trình phân lớp dữ liệu - (b2) Phân lớp dữ liệu mới



Trong mô hình phân lớp, thuật toán phân lớp giữ vai trò trung tâm, quyết định tới sự thành công của mô hình phân lớp. Do vậy chìa khóa của vấn đề phân lớp dữ liệu là tìm ra được một thuật toán phân lớp nhanh, hiệu quả, có độ chính xác cao và có khả năng mở rộng được. Trong đó khả năng mở rộng được của thuật toán được đặc biệt chú trọng và phát triển.

3.1.2. Các vấn đề liên quan đến phân lớp dữ liệu

a). Chuẩn bị dữ liệu cho việc phân lớp

Việc tiền xử lý dữ liệu cho quá trình phân lớp là một việc làm không thể thiếu và có vai trò quan trọng quyết định tới sự áp dụng được hay không của mô hình phân lớp. Quá trình tiền xử lý dữ liệu sẽ giúp cải thiện độ chính xác, tính hiệu quả và khả năng mở rộng được của mô hình phân lớp.

Quá trình tiền xử lý dữ liệu gồm có các công việc sau:

Làm sạch dữ liệu

Làm sạch dữ liệu liên quan đến việc xử lý với *noise* và *missing value* trong tập dữ liệu ban đầu. *Noise* là các lỗi ngẫu nhiên hay các giá trị không hợp lệ của các biến trong tập dữ liệu. Để xử lý với loại lỗi này có thể dùng kỹ thuật làm trơn. *Missing value* là những ô không có giá trị của các thuộc tính. Giá trị thiếu có thể do lỗi chủ quan trong quá trình nhập liệu, hoặc trong trường hợp cụ thể giá trị của thuộc tính đó không có, hay không quan trọng. Xử lý giá trị có thể bằng cách thay giá trị thiếu bằng giá trị phổ biến nhất của thuộc tính đó hoặc bằng giá trị có thể xảy ra nhất dựa trên thống kê. Mặc dù phần lớn thuật toán phân lớp đều có cơ chế xử lý với những giá trị thiếu và *noise* trong tập dữ liệu, nhưng bước tiền xử lý này có thể làm giảm sự hỗn độn trong quá trình học (xây dựng mô hình phân lớp).

Phân tích sự cần thiết của dữ liệu

Có rất nhiều thuộc tính trong tập dữ liệu có thể hoàn toàn không cần thiết hay liên quan đến một bài toán phân lớp cụ thể. Ví dụ dữ liệu về ngày trong tuần hoàn toàn không cần thiết đối với ứng dụng phân tích độ rủi ro của các khoản tiền cho vay của ngân hàng, nên thuộc tính này là dư thừa. Phân tích sự cần thiết của dữ liệu nhằm mục đích loại bỏ những thuộc tính không cần thiết, dư thừa khỏi quá trình học vì những thuộc tính đó sẽ làm chậm, phức tạp và gây ra sự hiểu sai trong quá trình học dẫn tới một mô hình phân lớp không dùng được.

Chuyển đổi dữ liệu

Việc khái quát hóa dữ liệu lên mức khái niệm cao hơn đôi khi là cần thiết trong quá trình tiền xử lý. Việc này đặc biệt hữu ích với những *thuộc tính liên tục* (*continuous attribute* hay *numeric attribute*). Ví dụ các giá trị số của thuộc tính *thu nhập* của khách hàng có thể được khái quát hóa thành các dãy giá trị rời rạc: *thấp, trung bình, cao*. Tương tự với những *thuộc tính rời rạc* (*categorical attribute*) như *địa chỉ phố* có thể được khái quát hóa lên thành *thành phố*. Việc khái quát hóa làm cô đọng dữ liệu học nguyên thủy, vì vậy các thao tác vào/ ra liên quan đến quá trình học sẽ giảm.

b). So sánh các mô hình phân lớp

Trong từng ứng dụng cụ thể cần lựa chọn mô hình phân lớp phù hợp. Việc lựa chọn đó căn cứ vào sự so sánh các mô hình phân lớp với nhau, dựa trên các tiêu chuẩn sau:

- ***Độ chính xác dự đoán*** (*predictive accuracy*)

Độ chính xác là khả năng của mô hình để dự đoán chính xác nhãn lớp của dữ liệu mới hay dữ liệu chưa biết.

- **Tốc độ** (*speed*)

Tốc độ là những chi phí tính toán liên quan đến quá trình tạo ra và sử dụng mô hình.

- **Sức mạnh** (*robustness*)

Sức mạnh là khả năng mô hình tạo ra những dự đoán đúng từ những dữ liệu *noise* hay dữ liệu với những giá trị thiếu.

- **Khả năng mở rộng** (*scalability*)

Khả năng mở rộng là khả năng thực thi hiệu quả trên lượng lớn dữ liệu của mô hình đã học.

- **Tính hiểu được** (*interpretability*)

Tính hiểu được là mức độ hiểu và hiểu rõ những kết quả sinh ra bởi mô hình đã học.

- **Tính đơn giản** (*simplicity*)

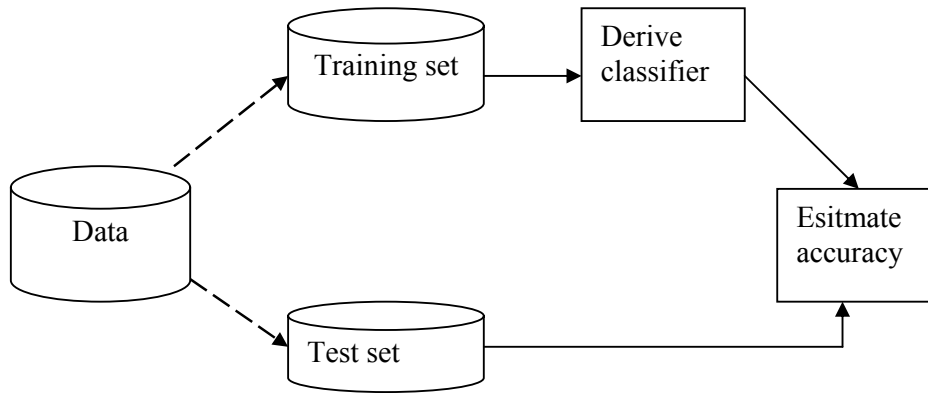
Tính đơn giản liên quan đến kích thước của cây quyết định hay độ cô đọng của các luật.

Trong các tiêu chuẩn trên, khả năng mở rộng của mô hình phân lớp được nhấn mạnh và trú trọng phát triển, đặc biệt với cây quyết định.

3.1.3 Các phương pháp đánh giá độ chính xác của mô hình phân lớp

Ước lượng độ chính xác của bộ phân lớp là quan trọng ở chỗ nó cho phép dự đoán được độ chính xác của các kết quả phân lớp những dữ liệu tương lai. Độ chính xác còn giúp so sánh các mô hình phân lớp khác nhau. Trong luận văn này đề cập đến 2 phương pháp đánh giá phổ biến là *holdout* và *k-fold cross-validation*. Cả 2 kỹ thuật này đều dựa trên các phân hoạch ngẫu nhiên tập dữ liệu ban đầu.

- Trong phương pháp *holdout*, dữ liệu đưa ra được phân chia ngẫu nhiên thành 2 phần là: tập dữ liệu đào tạo và tập dữ liệu kiểm tra. Thông thường 2/3 dữ liệu cấp cho tập dữ liệu đào tạo, phần còn lại cho tập dữ liệu kiểm tra [6].



Hình 3.5 Ước lượng độ chính xác của mô hình phân lớp với phương pháp holdout

- Trong phương pháp *k-fold cross validation* tập dữ liệu ban đầu được chia ngẫu nhiên thành k tập con (*fold*) có kích thước xấp xỉ nhau S_1, S_2, \dots, S_k . Quá trình học và “test” được thực hiện k lần. Tại lần lặp thứ i , S_i là tập dữ liệu kiểm tra, các tập còn lại hợp thành tập dữ liệu đào tạo. Có nghĩa là, đầu tiên việc dạy được thực hiện trên các tập S_2, S_3, \dots, S_k , sau đó test trên tập S_1 ; tiếp tục quá trình dạy được thực hiện trên tập $S_1, S_3, S_4, \dots, S_k$, sau đó test trên tập S_2 ; và cứ thế tiếp tục. Độ chính xác là toàn bộ số phân lớp đúng từ k lần lặp chia cho tổng số mẫu của tập dữ liệu ban đầu.

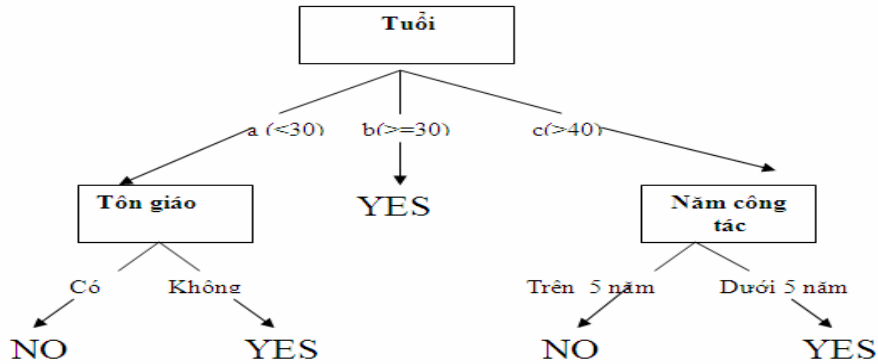
3.2 Cây quyết định ứng dụng trong phân lớp dữ liệu

3.2.1. Định nghĩa:

Trong những năm qua, nhiều mô hình phân lớp dữ liệu đã được các nhà khoa học trong nhiều lĩnh vực khác nhau đề xuất như mạng neuron, mô hình thông kê tuyến tính bậc 2, cây quyết định, mô hình di truyền. Trong số những mô hình đó, cây quyết định với những ưu điểm của mình được đánh giá là một công cụ mạnh, phổ biến và đặc biệt thích hợp cho Data Mining nói chung và phân lớp dữ liệu nói riêng [12]. Có thể kể ra những ưu điểm của cây quyết định như: xây dựng tương đối nhanh; đơn giản, dễ hiểu. Hơn nữa các cây có thể dễ dàng được chuyển đổi sang các câu lệnh SQL để có thể được sử dụng để truy nhập cơ sở dữ liệu một cách hiệu quả. Cuối cùng, việc phân lớp dựa trên cây

quyết định đạt được sự tương tự và đôi khi là chính xác hơn so với các phương pháp phân lớp khác.

Cây quyết định là một flow-chart giống cấu trúc cây, nút bên trong biểu thị một kiểm tra trên một thuộc tính, nhánh biểu diễn đầu ra của kiểm tra, nút lá biểu diễn nhãn lớp hoặc sự phân bố của lớp. Cây quyết định là biểu đồ phát triển có cấu trúc dạng cây, như mô tả trong hình vẽ sau:



Hình 3.6 Ví dụ về cây quyết định

Trong cây quyết định:

- Gốc: là node trên cùng của cây;
- Node trong: biểu diễn một kiểm tra trên một thuộc tính đơn (hình chữ nhật);
- Nhánh: biểu diễn các kết quả của kiểm tra trên node trong (mũi tên);
- Node lá: biểu diễn lớp hay sự phân phối lớp.

Để phân lớp mẫu dữ liệu chưa biết, giá trị các thuộc tính của mẫu được đưa vào kiểm tra trên cây quyết định. Mỗi mẫu tương ứng có một đường đi từ gốc đến lá và lá biểu diễn dự đoán giá trị phân lớp mẫu đó.

3.2.2. Các vấn đề trong khai phá dữ liệu sử dụng cây quyết định

a). Độ lợi thông tin (Information Gain) trong cây quyết định :

Information gain là đại lượng được sử dụng để chọn lựa thuộc tính với information gain lớn nhất. Giả sử có hai lớp, P và N. Cho tập hợp của những ví dụ S chứa p phần tử của lớp P và n phần tử của lớp N. Khối lượng của thông tin, cần để quyết định nếu những mẫu tùy ý trong S thuộc về P hoặc N được định nghĩa như là :

$$I(p,n) = -[p/(p+n)]\log_2 [p/(p+n)] - [n/(p+n)]\log_2 [n/(p+n)]$$

Giả sử rằng sử dụng thuộc tính A một tập hợp S được phân hoạch thành những tập hợp $\{S_1, S_2, \dots, S_v\}$. Nếu S_i chứa những mẫu của P và n_i mẫu của N_i entropy hoặc thông tin mong đợi cần để phân loại những đối tượng trong cây con S_i là :

$$E(A) = \sum_{i=1}^v [(p_i+n_i)/(p+n)] I(p_i, n_i)$$

Thông tin nhận được ở nhánh A là : $\text{Gain}(A) = I(p,n) - E(A)$

b).Tránh “quá vừa” dữ liệu

Thế nào là “quá vừa” dữ liệu? Có thể hiểu đây là hiện tượng cây quyết định chứa một số đặc trưng riêng của tập dữ liệu đào tạo, nếu lấy chính tập training data để test lại mô hình phân lớp thì độ chính xác sẽ rất cao, trong khi đối với những dữ liệu tương lai khác nếu sử dụng cây đó lại không đạt được độ chính xác như vậy.

Quá vừa dữ liệu là một khó khăn đáng kể đối với học bằng cây quyết định và những phương pháp học khác. Đặc biệt khi số lượng ví dụ trong tập dữ liệu đào tạo quá ít, hay có noise trong dữ liệu.

Có hai phương pháp tránh “quá vừa” dữ liệu trong cây quyết định:

- Dừng phát triển cây sớm hơn bình thường, trước khi đạt tới điểm phân lớp hoàn hảo tập dữ liệu đào tạo. Với phương pháp này, một thách thức đặt ra là phải ước lượng chính xác thời điểm dừng phát triển cây;
- Cho phép cây có thể “quá vừa” dữ liệu, sau đó sẽ cắt, tỉa cây.

Mặc dù phương pháp thứ nhất có vẻ trực tiếp hơn, nhưng với phương pháp thứ hai thì cây quyết định được sinh ra được thực nghiệm chứng minh là thành công hơn trong thực tế. Hơn nữa việc cắt tỉa cây quyết định còn giúp cải thiện độ chính xác của mô hình phân lớp. Dù thực hiện phương pháp nào thì vấn đề mấu chốt ở đây là tiêu chuẩn nào được sử dụng để xác định kích thước hợp lý của cây cuối cùng.

c). Thao tác với thuộc tính liên tục

Việc thao tác với thuộc tính liên tục trên cây quyết định hoàn toàn không đơn giản như với thuộc tính rời rạc.

Thuộc tính rời rạc có *tập giá trị (domain)* xác định từ trước và là tập hợp các giá trị rời rạc. Ví dụ *ngạch công chức* là một thuộc tính rời rạc với tập giá trị là: {nhân viên, kiểm tra viên, kiểm tra viên chính, kiểm tra viên cao cấp}. Việc phân chia dữ liệu dựa vào phép kiểm tra giá trị của thuộc tính rời rạc được chọn tại một ví dụ cụ thể có thuộc tập giá trị của thuộc tính đó hay không: $value(A) \in X$ với $X \subset domain(A)$. Đây là phép kiểm tra logic đơn giản, không tốn nhiều tài nguyên tính toán. Trong khi đó, với thuộc tính liên tục (thuộc tính dạng số) thì tập giá trị là không xác định trước. Chính vì vậy, trong quá trình phát triển cây, cần sử dụng kiểm tra dạng nhị phân: $value(A) \leq \theta$. Với θ là *hằng số ngưỡng (threshold)* được lần lượt xác định dựa trên từng giá trị riêng biệt hay từng cặp giá trị liên nhau (theo thứ tự đã sắp xếp) của thuộc tính liên tục đang xem xét trong tập dữ liệu đào tạo. Điều đó có nghĩa là nếu thuộc tính liên tục A trong tập dữ liệu đào tạo có d giá trị phân biệt thì cần thực hiện $d-1$ lần kiểm tra $value(A) \leq \theta_i$ với $i = 1..d-1$ để tìm ra ngưỡng θ_{best} tốt nhất tương ứng với thuộc tính đó. Việc xác định giá trị của θ và tiêu chuẩn tìm θ tốt nhất tùy vào chiến lược của từng thuật toán.

3.2.3. Xây dựng cây quyết định

Bao gồm 2 giai đoạn: tạo cây và tỉa cây. Tạo cây ở thời điểm bắt đầu tất cả những ví dụ huấn luyện là ở gốc sau đó phân chia ví dụ huấn luyện theo cách đệ qui dựa trên thuộc tính được chọn. Việc tỉa cây là xác định và xóa những nhánh mà có phần tử hỗn loạn hoặc những phần tử nằm ngoài (những phần tử không thể phân vào một lớp nào đó).

Thuật toán qui nạp tạo cây quyết định được chia thành các bước như sau:

1. Cây được xây dựng đệ qui từ trên xuống dưới (top-down) và theo cách thức chia để trị (divide-conquer);
2. Ở thời điểm bắt đầu, tất cả những ví dụ huấn luyện ở gốc;
3. Thuộc tính được phân loại (nếu là giá trị liên tục chúng được rời rạc hóa);
4. Những ví dụ huấn luyện được phân chia đệ qui dựa trên thuộc tính mà nó chọn lựa;

5. Kiểm tra những thuộc tính được chọn dựa trên nền tảng của heuristic hoặc của một định lượng thông kê .

Điều kiện để dừng việc phân chia :

1. Tất cả những mẫu huấn luyện đối với một node cho trước thuộc về cùng một lớp;
2. Không còn thuộc tính còn lại nào để phân chia tiếp;
3. Không còn mẫu nào còn lại .

Sơ đồ tổng quát xây dựng cây quyết định

- 1: procedure XayDungCay(n :Nút, D : dữ kiện , LL : phương thức)
- 2: Áp dụng LL trên D để tìm ra chuẩn chia ;
- 3: Chia n bằng cách sử dụng chuẩn chia ở trên ;
- 4: $k \leftarrow$ số các nút con của n ;
- 5: if $k > 0$ then
- 6: Tạo ra k nút con n_1, \dots, n_k của n ;
- 7: Chia D thành D_1, \dots, D_k ;
- 8: for $i \leftarrow 1, k$ do
- 9: XayDungCay(n_i, D_i, LL) ;
- 10: end for
- 11: end if
- 12: end procedure

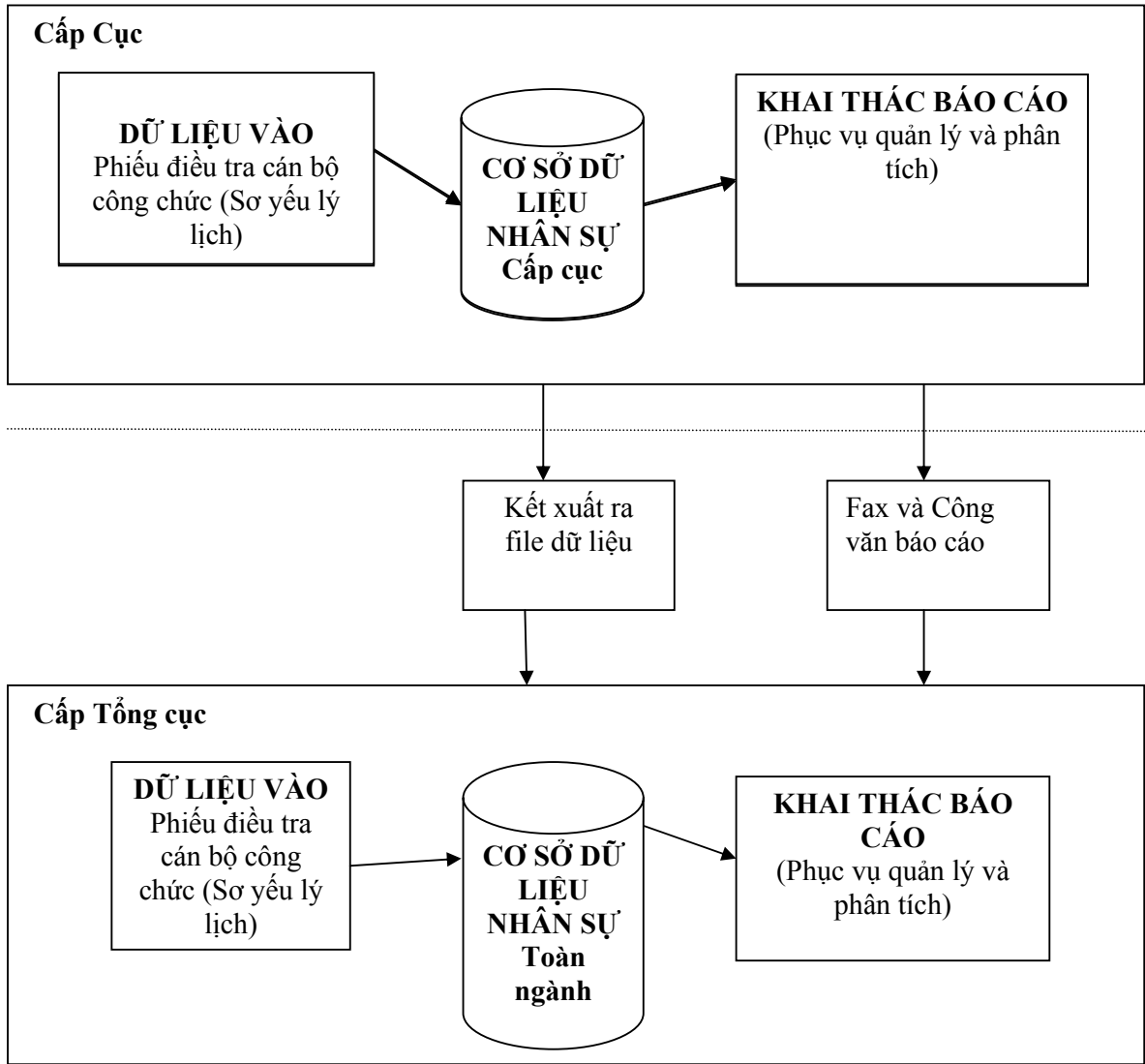
Kết luận chương 3

Chương này nghiên cứu về kỹ thuật phân lớp dữ liệu sử dụng cây quyết định. Luận văn tiếp cận một cách tổng quan bài toán phân lớp đi vào những đánh giá của thuật toán phân lớp. Tiếp theo là những vấn đề sử dụng cây quyết định: quá trình hình thành cây, các vấn đề liên quan và những đặc tính, điểm mạnh, điểm yếu của cây quyết định. Chương này mục đích xem xét kiến thức kỹ thuật nền tảng được sử dụng trong phần thực nghiệm trong luận văn làm cơ sở so sánh đưa ra những kết luận nghiên cứu.

CHƯƠNG 4. THỰC NGHIỆM

4.1. Giới thiệu về mô hình xây dựng:

4.1.1. Sơ đồ luồng dữ liệu thông tin nhân sự:



Hình 4.1 Sơ đồ luồng dữ liệu thông tin nhân sự

Chương trình được triển khai xây dựng ở 2 cấp: cấp Cục và cấp Tổng cục. Cấp Cục cập nhật dữ liệu hồ sơ lý lịch của đơn vị mình lưu trữ tại cơ sở dữ liệu cấp Cục để khai thác báo cáo quản lý và báo cáo theo yêu cầu nghiệp vụ.

Cấp Tổng cục ngoài chức năng cập nhật thông tin sơ yếu lý lịch nhân sự ở Tổng cục như ở cấp Cục còn tiến hành cập nhật dữ liệu nhận về từ các Cục địa phương: thông tin hồ sơ cán bộ mới, bổ sung, hồ sơ luân chuyển từ đơn vị này sang đơn vị khác. Việc khai thác báo cáo từ dữ liệu này phụ thuộc vào chương

trình xây dựng nên có nhiều hạn chế. Các báo cáo của hệ thống hiện tại xây dựng theo phương pháp thống kê để đáp ứng việc làm báo cáo theo biểu mẫu của Bộ Nội vụ (cách làm thủ công). Chương trình cũng hỗ trợ thêm các báo cáo theo yêu cầu riêng của từng đơn vị như báo cáo công chức theo ngạch chức danh, các báo cáo chế độ công chức trong giai đoạn tập sự, hết tập sự... Hay chức năng làm báo cáo động theo cách để người dùng tự chọn các trường dữ liệu trong hồ sơ cần để làm báo cáo. Song việc hỗ trợ như vậy vẫn tồn tại các yêu cầu phải nghiên cứu hỗ trợ như sau:

- + Giám sát được dữ liệu cập nhật từ cấp dưới: Việc này gây mất rất nhiều thời gian đòi hỏi cán bộ thực hiện phải có chuyên môn. Đánh giá về kết quả giám sát không được nhiều.
- + Việc lựa chọn cán bộ theo tiêu chuẩn để đào tạo, quy hoạch như thế đã đúng và đã đủ đối tượng chưa.
- + Việc bổ sung, cắt giảm biên chế hay điều chuyển cán bộ trong của các đơn vị gặp nhiều khó khăn (kết quả báo cáo thì có nhiều dẫn tới khó có thể làm rõ được).

Theo yêu cầu bảo quản thông tin hồ sơ, dữ liệu sử dụng trong luận văn được chạy trên cơ sở dữ liệu thử nghiệm. Cơ sở dữ liệu này được quản lý bằng hệ quản trị SQL Server (Hệ quản trị cơ sở dữ liệu của hồ sơ gốc bằng Oracle 9i).

4.1.2. Giải quyết vấn đề:

Các vấn đề không thể thực hiện được bằng chương trình quản lý hiện tại. Ta giải quyết vấn đề tồn tại này bằng nghiên cứu của luận văn sử dụng công cụ “Microsoft Analysis Services” của Microsoft để tiến hành tạo mô hình phân lớp dữ liệu dựa trên kỹ thuật “Microsoft Decision Tree” – Cây quyết định.

Dữ liệu vào của mô hình là bảng chính của cơ sở dữ liệu sơ yếu lý lịch nhân sự. Cơ sở dữ liệu này bao gồm: Bảng dữ liệu chính, dữ liệu tham chiếu và dữ liệu danh mục việc bảng dữ liệu chính sẽ là bảng dữ liệu vào cho mô hình xây dựng. Bảng dữ liệu được làm việc HC_EMP đây là bảng dữ liệu chính chứa thông tin cần cho mô hình xây dựng. Việc xây dựng mô hình không mấy khó khăn và rất nhanh chóng qua 3 bước xác định:

Bước 1:

Chạy chương trình “Microsoft Analysis Service manager” kết nối vào Máy chủ phân tích (Analysis Server) để có kết nối đến cơ sở dữ liệu nhân sự;

Bước 2:

Xác định thuộc tính cần dự đoán và thuộc tính đầu vào cho mô hình. Ưu điểm của công cụ này không phụ thuộc quá nhiều vào việc chuẩn bị dữ liệu;

Bước 3:

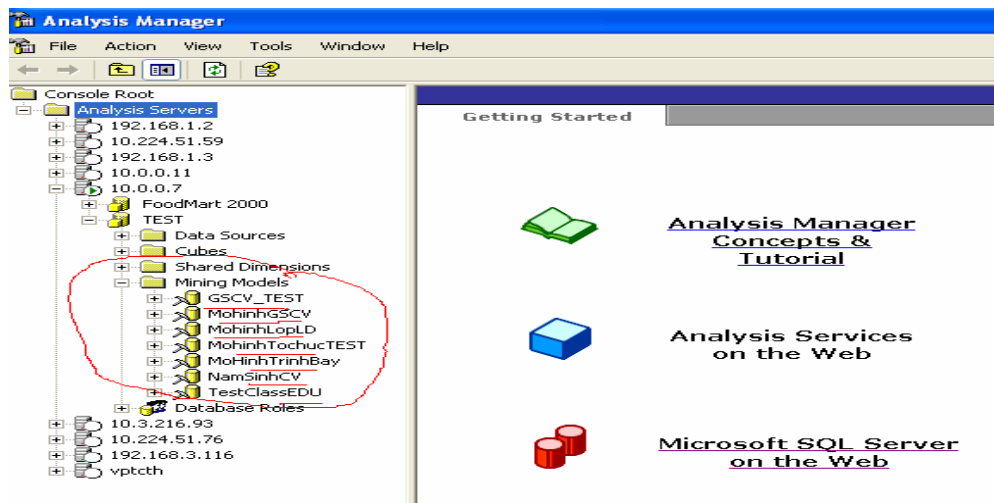
Dùng chức năng xây dựng mô hình để:

+Mô tả bảng dữ liệu với dữ liệu vào và trường dữ liệu cần dự đoán;

+Chọn thuật toán cần sử dụng trong luận văn này ta dùng “Microsoft Decision Tree”.

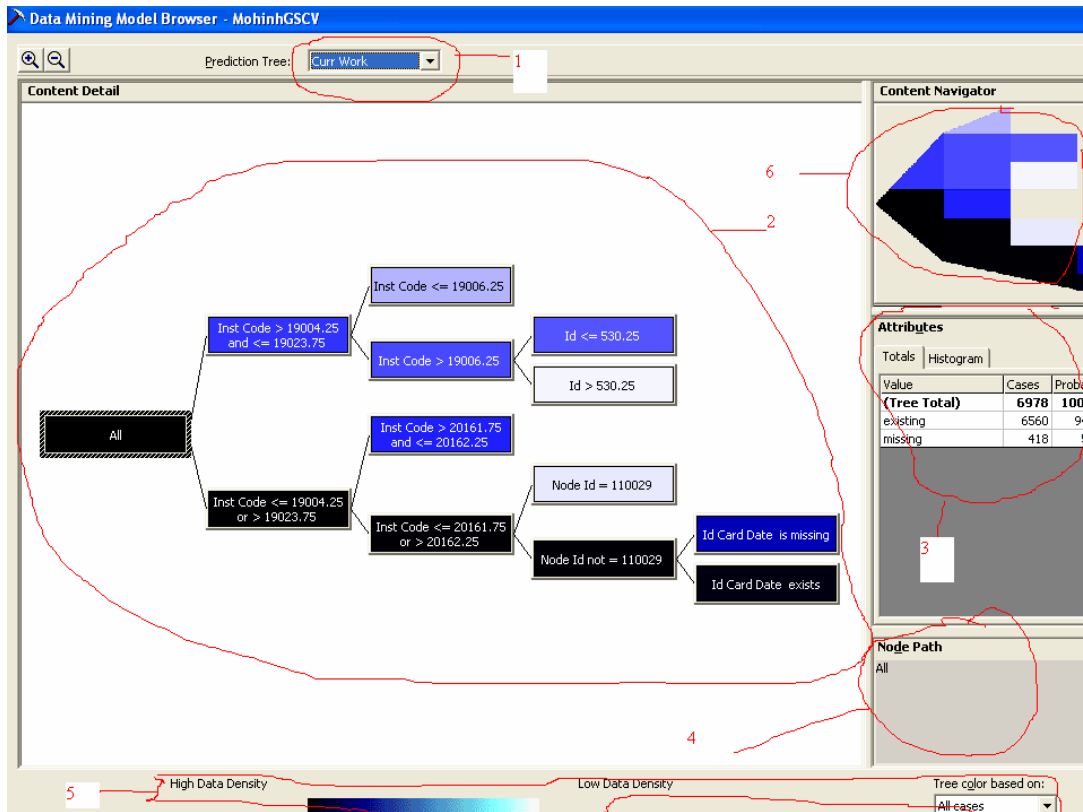
Ta có mô hình cần.

4.1.3.Các mô hình được xây dựng:



Hình 4.2 Hình minh họa là các mô hình được xây dựng hỗ trợ cho công tác

Giao diện hỗ trợ của mô hình xây dựng gồm có 6 thành phần:



Hình 4.3 Hình minh họa các thành phần của giao diện hỗ trợ

1. Thuộc tính cần dự đoán ở đây là trường dữ liệu thông tin về công việc hiện tại
2. Hình ảnh mô hình cây thu được
3. Con số thống kê và khả năng dự đoán cho từng node
4. Mô tả mối quan hệ giữa các nút hay luật ví dụ:
(Node Id not = 110029 and Inst Code <= 19004.25 or (> 19023.75 and <= 20161.75) or > 20162.25 and Id Card Date exists)
5. Chỉ dẫn màu sắc biểu hiện mật độ phân bố dữ liệu được đánh dấu hiển thị theo màu sắc và lựa chọn các trường hợp (trên hình vẽ là All cases nghĩa là tất cả các trường hợp). Công cụ hỗ trợ người dùng chi tiết từng trường hợp.
6. Hình ảnh tổng thể cùng mật độ phân bố dữ liệu của cây mô hình thu được.

Vậy thì, các mô hình có giá trị như thế nào cho các công việc đang vướng mắc đã đề cập ở trên. Ta sẽ thấy rõ hơn ở phần minh họa các kết quả thu được từ các mô hình.

4.2. Minh họa kết quả hỗ trợ thu được từ mô hình xây dựng

4.2.1 Minh họa hỗ trợ công tác tuyển lựa và đào tạo-cây lựa chọn cán bộ đào tạo quản lý nhà nước

Đơn vị nhận trách nhiệm đào tạo (sau đây gọi là Vụ Tổ chức cán bộ) phải làm công văn thông báo tới các đơn vị trong ngành biết về khóa đào tạo: Chương trình học, địa điểm, thời gian đào tạo và đối tượng đào tạo. Dưới đây là một hình minh họa công văn thông báo đào tạo:

<input type="checkbox"/>	Ø Số ghi nhận	Ngày ký	Trích yếu
<input checked="" type="checkbox"/>	Ø 670/CV-ĐUy	03/04/2008	Triệu tập Hội nghị BCH mở rộng
<input checked="" type="checkbox"/>	Ø 03/VP	03/04/2008	Tài liệu giao ban quý I/2008
<input checked="" type="checkbox"/>	Ø 03.4-TCCB	03/04/2008	Thông báo tin buồn
<input checked="" type="checkbox"/>	Ø 62-TTDT-NVTQ	01/04/2008	Tổ chức khoá đào tạo về kỹ năng phát hiện chứng từ giả trong bộ s
<input checked="" type="checkbox"/>	Ø 01.4-TB-TCCB	01/04/2008	Thông báo khai giảng lớp học ngoại ngữ Tiếng Anh
<input checked="" type="checkbox"/>	Ø 31/TCCB	31/03/2008	THÔNG BÁO TIN BUỒN
<input checked="" type="checkbox"/>	Ø 31.03-TTBTDCCHQ	31/03/2008	Ủy quyền điều hành công việc
<input checked="" type="checkbox"/>	Ø 180/NHNN	31/03/2008	Thông báo tỷ giá tính chéo
<input checked="" type="checkbox"/>	Ø 28-03-HĐH	28/03/2008	Tài liệu Hội thảo Crown Agents
<input checked="" type="checkbox"/>	Ø 28.3.2008	28/03/2008	DANH MỤC DỮ LIỆU GIÁ
<input checked="" type="checkbox"/>	Ø 27-03-HTQT	27/03/2008	Tham gia ý kiến về quy chế phối hợp của vụ HTQT
<input checked="" type="checkbox"/>	Ø 27/VP-TVQT	27/03/2008	THÔNG BÁO KHÔNG PHỤC VỤ CƠM TRƯA
<input checked="" type="checkbox"/>	Ø 25/VP	25/03/2008	Xin ý kiến tham gia đề án Net office
<input checked="" type="checkbox"/>	Ø 21/PTPLMT	25/03/2008	Thư cảm tạ
<input checked="" type="checkbox"/>	Ø 24/TCCB	24/03/2008	Thông báo tin buồn
<input checked="" type="checkbox"/>	Ø 24.03-TCCB	24/03/2008	Cử cán bộ đi học QLNN chương trình Chuyên viên chính
<input checked="" type="checkbox"/>	Ø 49-TTDT-KTSTQ	21/03/2008	Mở lớp Kế toán và Kiểm toán cho lực lượng KTSTQ
<input checked="" type="checkbox"/>	Ø 21.03-TCCB	21/03/2008	Thông báo tin buồn
<input checked="" type="checkbox"/>	Ø 165/NHNN	20/03/2008	Thông báo tỷ giá tính chéo
<input checked="" type="checkbox"/>	Ø 19/VP	19/03/2008	Xin ý kiến về chi bổ sung thu nhập 0.2 lần 6 tháng cuối năm 2007

Hình 4.4 Minh họa công văn đào tạo

BỘ TÀI CHÍNH
Số : 3404/BTC-TCCB
V/v cử cán bộ tham dự lớp
Bồi dưỡng kiến thức QLNN
Chương trình CVC

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc
Hà Nội, ngày 19 tháng 3 năm 2008

Kính gửi: Các đơn vị thuộc Bộ

TỔNG CỤC HẢI QUAN
ĐẾN NGÀY: 20-03-2008
CÔNG VĂN SỐ:

Thực hiện triển khai kế hoạch đào tạo, bồi dưỡng nâng cao kiến thức về quản lý nhà nước cho công chức, viên chức năm 2008, Trường Bồi dưỡng cán bộ tài chính phối hợp với Vụ Tổ chức cán bộ mở 02 lớp Bồi dưỡng kiến thức quản lý nhà nước chương trình chuyên viên chính tại Hà Nội và TP Hồ Chí Minh:

Đối tượng : Công chức có chứng chỉ học lớp QLNN ngạch chuyên viên thuộc các đối tượng sau:

- Lãnh đạo cấp Cục, Vụ, Viện, Phòng và tương đương trong các đơn vị thuộc khối cơ quan Bộ;
- Lãnh đạo cấp Ban và tương đương trong các đơn vị Tổng cục Thuế, Tổng cục Hải quan, Kho bạc nhà nước, Cục Dự trữ Quốc gia, Ủy ban chứng khoán nhà nước;
- Lãnh đạo cấp Cục và tương đương tại các Cục địa phương (Cục Thuế, Cục Hải quan, Kho bạc địa phương, Dự trữ Quốc gia khu vực) của TP Hà Nội, TP Hồ Chí Minh và một số tỉnh lân cận.
- Chuyên viên chính có mã ngạch công chức 01.002, và những cán bộ công chức có mức lương tương đương chuyên viên chính.

2. Thời gian và địa điểm mở lớp:

2.1- Lớp tại Hà Nội: dành cho các đơn vị thuộc khối cơ quan Bộ; các đơn vị cấp Tổng cục (Tổng cục Thuế, Tổng cục Hải quan, Kho bạc nhà nước, Cục Dự trữ Quốc gia, Ủy ban chứng khoán nhà nước) và lãnh đạo cấp Cục địa phương của các tỉnh, thành phố: TP Hà Nội, Hà Tây, Bắc Ninh.

- **Thời gian:** 03 tháng, khai giảng ngày 01/4/2008
- **Địa điểm:** Số 8 Kim Mã Thượng, Quận Ba Đình – Hà Nội
(Học viên tự lo chỗ ăn, nghỉ)

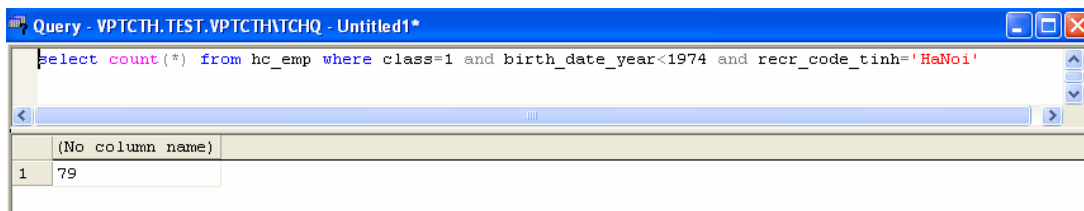
Hình 4.5 Minh họa yêu cầu đối tượng đào tạo

Sau khi Vụ Tổ chức cán bộ nhận được danh sách đăng ký tham gia lớp đào tạo từ các đơn vị. Vụ Tổ chức tiến hành: tổng hợp, kiểm tra danh sách (đối chiếu với tiêu chuẩn). Kết quả kiểm tra sẽ ra Quyết định cử cán bộ đi đào tạo. Vấn đề hỗ trợ Quyết định được trình bày chi tiết dưới đây:

Đối tượng đào tạo (tiêu chuẩn):

- Là lãnh đạo cấp phòng trở lên (**class=1**)
- 35 tuổi trở lên (birth_date_year<1974)
- Nơi làm việc ở Hà Nội(recr_code_tinh='HaNoi')

Kết quả truy vấn trực tiếp từ cơ sở dữ liệu nhân sự ta có 79 trường hợp đạt yêu cầu. Kết quả này sẽ dùng để so sánh độ chính xác với mô hình dự đoán:



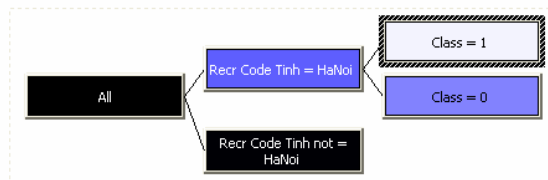
Dữ liệu vào và dữ liệu cần dự đoán:

Trường dữ liệu vào là Class, Recr_code_Tinh

Trường dữ liệu cần dự đoán là Class_Edu

Tên mô hình trong công cụ là "TestClassEdu"

Cây quyết định thu được như sau:



Value	Cases	Probability
(Node Total)	84	100.00%
0	5	6.90%
1	79	91.95%
missing	0	1.15%

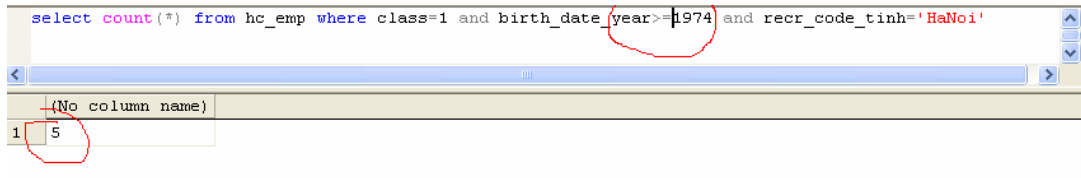
Node Path: Class = 1 and Recr Code Tinh = HaNoi

Hình 4.6 Cây phân lớp cán bộ, nhân viên

Giải thích cây

Mọi người (ALL) Nơi làm việc ở Hà nội(Recr Code Tinh=HaNoi) được xem xét trong mô hình.

Ở đây ta xem xét những người làm việc ở Hà nội (Recr Code Tinh=HaNoi) có vị trí công tác lãnh đạo cấp phòng trở lên (Class=1). Kết quả dự đoán cho thấy có 79 người đạt tiêu chuẩn, 5 người không đạt tiêu chuẩn là do là lãnh đạo nhưng chưa đủ tuổi 35. Điều này ta có thể kiểm chứng bằng cây lệnh truy vấn trực tiếp từ cơ sở dữ liệu nhân sự như sau:



Với cây quyết định phân lớp này ta có số lượng từ hồ sơ đem so sánh với danh sách trình sẽ có cơ sở lựa chọn chính xác hơn. Chẳng hạn, nếu số lượng vượt quá 79 người điều đó có nghĩa danh sách này bao gồm cả những người không đủ tiêu chuẩn, ngược lại ít hơn 79 người điều đó có nghĩa danh sách chưa đủ số lượng người đủ tiêu chuẩn như thông tin hỗ trợ phản ánh. Bước còn lại là chỉ còn việc kiểm tra chi tiết thông tin của từng người trong danh sách.

Tính trực quan của mô hình còn biểu hiện ở chỗ ta bấm vào bất kỳ nút nào cũng có con số thống kê cụ thể. Nếu ta bấm vào nút `Rec_code_tinh=HaNoi` sẽ có con số thống kê trên cửa sổ Attributes:

Value	Cases	Probability
(Node Total)	312	100.00%
0	233	74.29%
1	79	25.40%
missing	0	0.32%

Hình 4.7 Minh họa thông tin một node của cây

Số liệu thống kê cho ta biết tổng số hồ sơ của những người làm việc tại Hà Nội là 312 người, trong đó:

- + là lãnh đạo 79 người khả năng đoán nhận là 25.40%,
- + không giữ chức vụ lãnh đạo 233 khả năng đoán nhận 74.29%,
- + không xác định chỗ làm việc có khả năng chiếm 0.32%. Cụ thể hơn một số trường hợp đủ tiêu chuẩn nhưng không có thông tin về nơi làm việc nên cũng có khả năng xem xét . Ta có thể kiểm chứng bằng câu lệnh truy vấn sau:

```
select count(*) from hc_emp where recr_code_tinh='' and class=1 and birth_date_year<1974
```

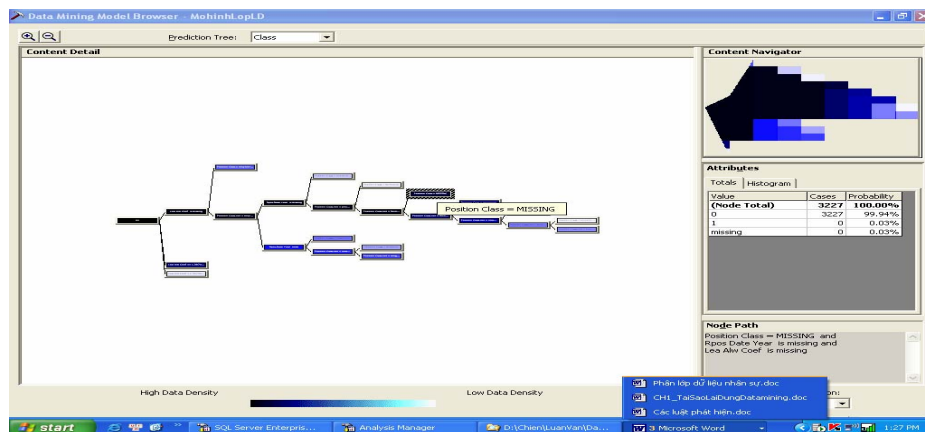
(No column name)	
1	424

Kết quả truy vấn trực tiếp xác định có 424 người đủ tiêu chuẩn là Lãnh đạo tuổi trên 35 nhưng không khai báo thông tin nơi làm việc nên mô hình đưa ra vẫn phải xem xét, tỷ lệ dự đoán khả năng này là 0.32%.

Như thế, mô hình đã hỗ trợ ngoài tính chính xác còn thấy được khả năng dự đoán của mô hình khuyến cáo đưa ra. Ngoài ra, kết quả thu được cây quyết định được công cụ sử dụng hỗ trợ người sử dụng có thông tin về phân bố mật độ dữ liệu dựa trên màu sắc, màu sắc đậm thể hiện dữ liệu nhiều, màu sắc nhạt dữ liệu ít. Thông tin hỗ trợ trong công tác đào tạo thường phải ứng dụng nhiều việc áp dụng công cụ và kỹ thuật này sẽ làm tăng khả năng đào tạo cũng như tuyển lựa cán bộ.

4.2.2. Minh họa công tác kiểm tra thông tin hồ sơ nhập máy-cây phân lớp vị trí công tác (lãnh đạo, nhân viên)

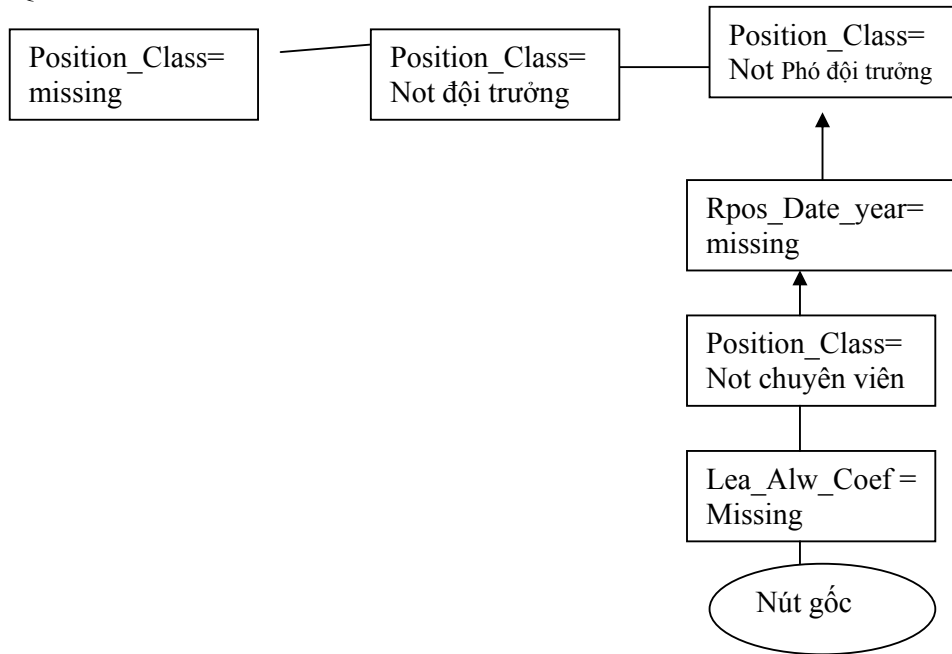
Đây là bộ dữ liệu đã phân lớp những người giữ vị trí lãnh đạo có chức danh: Phó Trưởng phòng, Trưởng phòng, Phó cục trưởng, Cục trưởng.. được gọi là lãnh đạo trường phân lớp Class =1. Không có chức danh: Chuyên viên, nhân viên, cán bộ.. được gọi là lớp không giữ chức vụ lãnh đạo trường phân lớp Class=0. Mô hình phân lớp cây quyết định trên bảng chính của dữ liệu hồ sơ (HC_EMP) có trường dữ liệu cần dự đoán là Class và 37 thuộc tính đầu vào là các thuộc tính được coi là có giá trị của bảng chính dữ liệu hồ sơ. Mô hình cây quyết định thu được (tên mô hình trong công cụ là mô hình lớp lãnh đạo):



Hình 4.8 mô hình phân lớp vị trí công tác

Các giá trị lấy được từ mô hình:

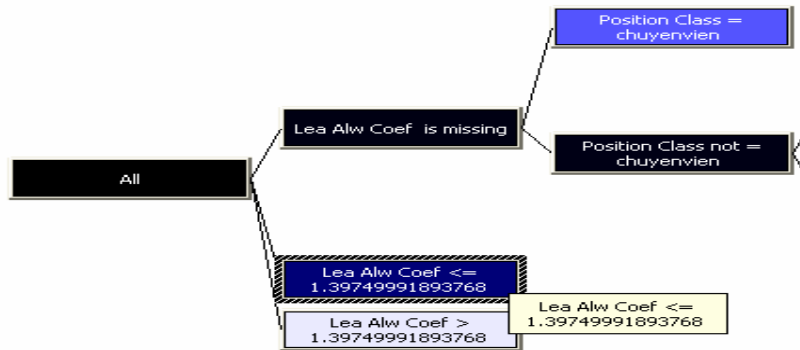
Nhìn hình vẽ mô hình tồn tại một lá có giá trị Position_Class= missing,
Quan hệ các nút như sau:



Hình 4.9 Quan hệ giữa các nút

Thông tin từ nhánh trên cây quyết định này cho thấy những trường hợp dữ liệu không xác định được chức danh là do nhập dữ liệu thiếu. Dữ liệu một số thông tin trường Rpos_date_year,lea_alw_coef và dữ liệu thiếu xảy ra trên hồ sơ của các cán bộ có chức danh không phải là chuyên viên, không phải là đội trưởng, không phải là phó đội trưởng (hình minh họa). Trên cơ sở đó ta có thể thống kê các hồ sơ thiếu thông tin theo chỉ dẫn trên. Đề nghị bổ sung thông tin và yêu cầu đối tác xây dựng chương trình phần mềm sửa chữa rằng buộc bắt người nhập liệu phải nhập dữ liệu trên 2 trường Rpos_date_year và riêng Lea_Alw_Coef với những người có hệ số phụ cấp chức vụ.

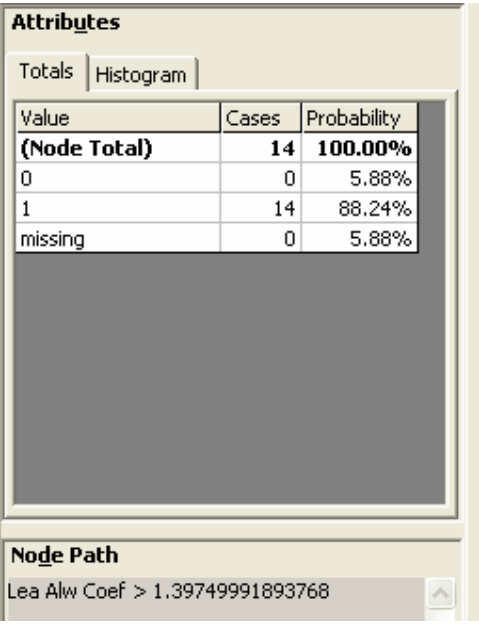
Một trường hợp khác cũng liên quan đến việc theo dõi, phát hiện điều chỉnh số liệu.



Hình 4.10 Cây phân lớp chức vụ

Hình vẽ trên là mô hình cây quyết định phân lớp chức vụ này tồn tại một quy luật nếu là lãnh đạo thì hệ số Lea_Alw_Coef sẽ có giá trị khác missing và trong 2 giá trị của trường Lea_Alw coef có giá trị thống kê như sau:

Giá trị cột	Thông tin thống kê tương ứng															
Lea Alw Coef <= 1.39749991893768	<div><p>Attributes</p><p>Totals Histogram</p><table border="1"><thead><tr><th>Value</th><th>Cases</th><th>Probability</th></tr></thead><tbody><tr><td>(Node Total)</td><td>1537</td><td>100.00%</td></tr><tr><td>0</td><td>11</td><td>0.78%</td></tr><tr><td>1</td><td>1526</td><td>99.16%</td></tr><tr><td>missing</td><td>0</td><td>0.06%</td></tr></tbody></table><p>Node Path</p><p>Lea Alw Coef <= 1.39749991893768</p></div>	Value	Cases	Probability	(Node Total)	1537	100.00%	0	11	0.78%	1	1526	99.16%	missing	0	0.06%
Value	Cases	Probability														
(Node Total)	1537	100.00%														
0	11	0.78%														
1	1526	99.16%														
missing	0	0.06%														

Giá trị cột	Thông tin thống kê tương ứng															
Lea Alw Coef > 1.39749991893768	 <table border="1"><thead><tr><th>Value</th><th>Cases</th><th>Probability</th></tr></thead><tbody><tr><td>(Node Total)</td><td>14</td><td>100.00%</td></tr><tr><td>0</td><td>0</td><td>5.88%</td></tr><tr><td>1</td><td>14</td><td>88.24%</td></tr><tr><td>missing</td><td>0</td><td>5.88%</td></tr></tbody></table> <p>Node Path Lea Alw Coef > 1.39749991893768</p>	Value	Cases	Probability	(Node Total)	14	100.00%	0	0	5.88%	1	14	88.24%	missing	0	5.88%
Value	Cases	Probability														
(Node Total)	14	100.00%														
0	0	5.88%														
1	14	88.24%														
missing	0	5.88%														

Dữ liệu hồ sơ tồn tại các hồ sơ trường Lea Alw Coef \leq 1.39749991893768 nhập không đúng dẫn đến kết quả phân lớp 11 trường hợp không giữ chức danh lãnh đạo trường Lea_Alw_coef vẫn tồn tại giá trị. Căn cứ vào kết quả này ta có cơ sở để thống kê các hồ sơ có trường dữ liệu lea_alw_coef \leq 1.39749991893768 để các đơn vị liên quan điều chỉnh lại hồ sơ cho chính xác.

Qua các minh chứng trên cho thấy, thông tin từ cây quyết định của mô hình rất trực quan cho việc theo dõi số liệu hỗ trợ việc điều chỉnh bổ sung dữ liệu sát hơn, đầy đủ hơn. Điều này không thể phát hiện nếu dùng phương pháp thống kê thông thường.

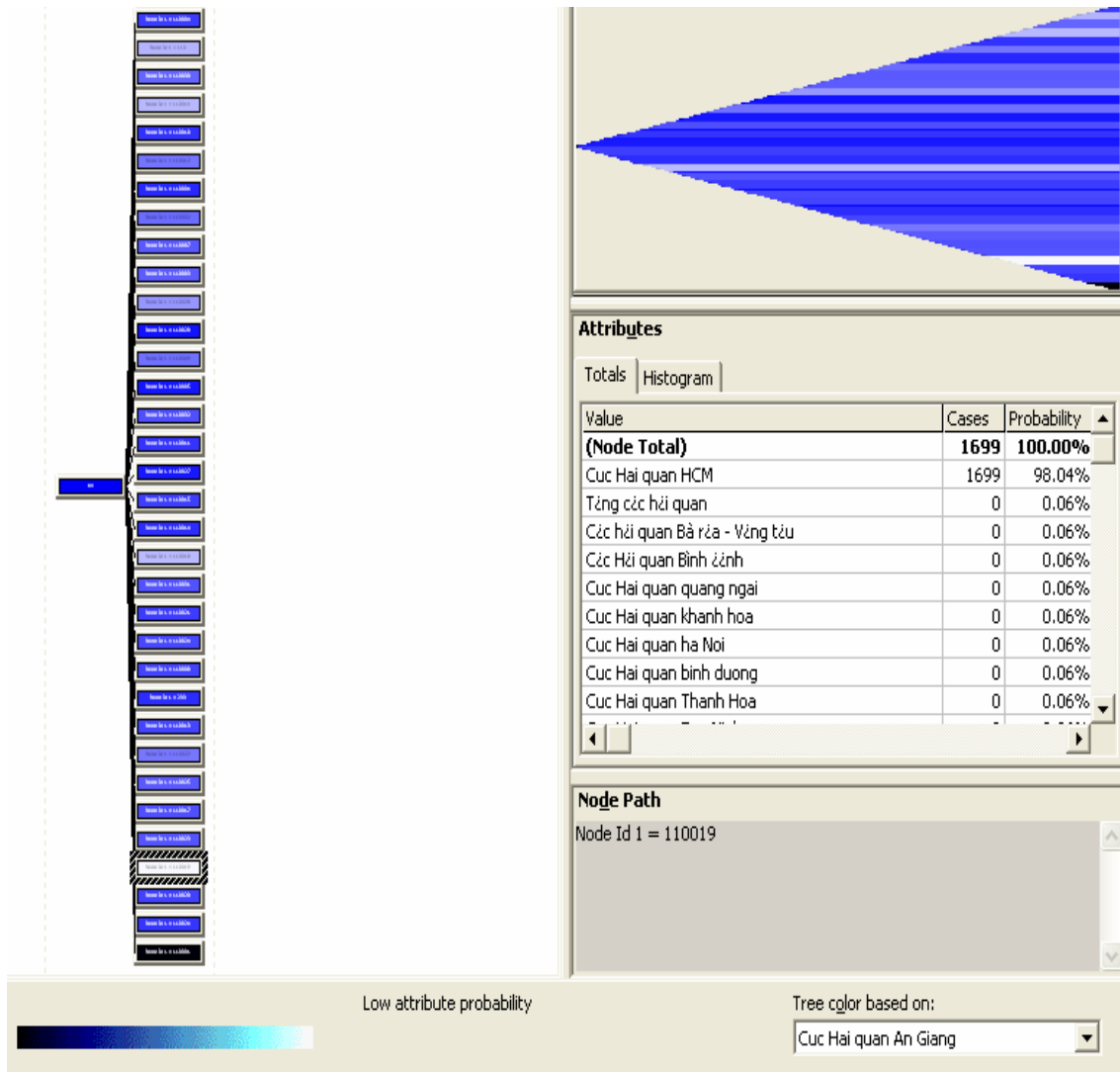
4.2.3. Minh họa công tác hoạch định mô hình tổ chức

Phần này giới thiệu kết quả hỗ trợ từ cây quyết định phân lớp. Hình thành mô hình từ thông tin phân bổ và mối quan hệ giữa dữ liệu nhân sự với trường dữ liệu đơn vị thể hiện trên mô hình. Ta cần xem xét sự phân bổ nhân sự trong các đơn vị để thực hiện kế hoạch tuyển dụng hay điều chuyển nhân sự dựa vào số nhân sự phân bổ trong từng đơn vị thể hiện bằng màu sắc trên mô hình. Ở đây ta xem xét các đơn vị có phân bổ nhân sự tại Cục Hải quan An Giang để xem các đơn vị khác có cùng nhu cầu phân bổ nhân sự giống cục Hải quan An Giang và cần phải cắt giảm biên chế dựa trên màu sắc phân bổ trên mô hình.

Trường dữ liệu vào là các thuộc tính của nhân sự lưu trong hồ sơ

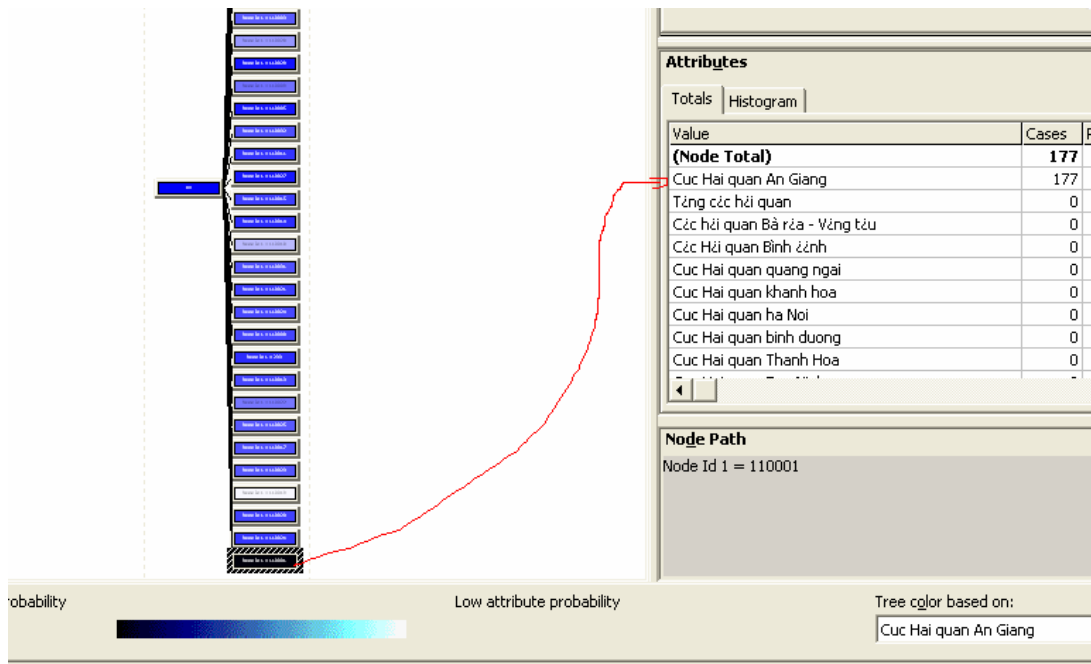
Trường dữ liệu cần dự đoán là tên đơn vị công tác của nhân sự

Cây quyết định theo mô hình tổ chức của các đơn vị như sau (tên mô hình trong công cụ là “MohinhTochucTEST”):



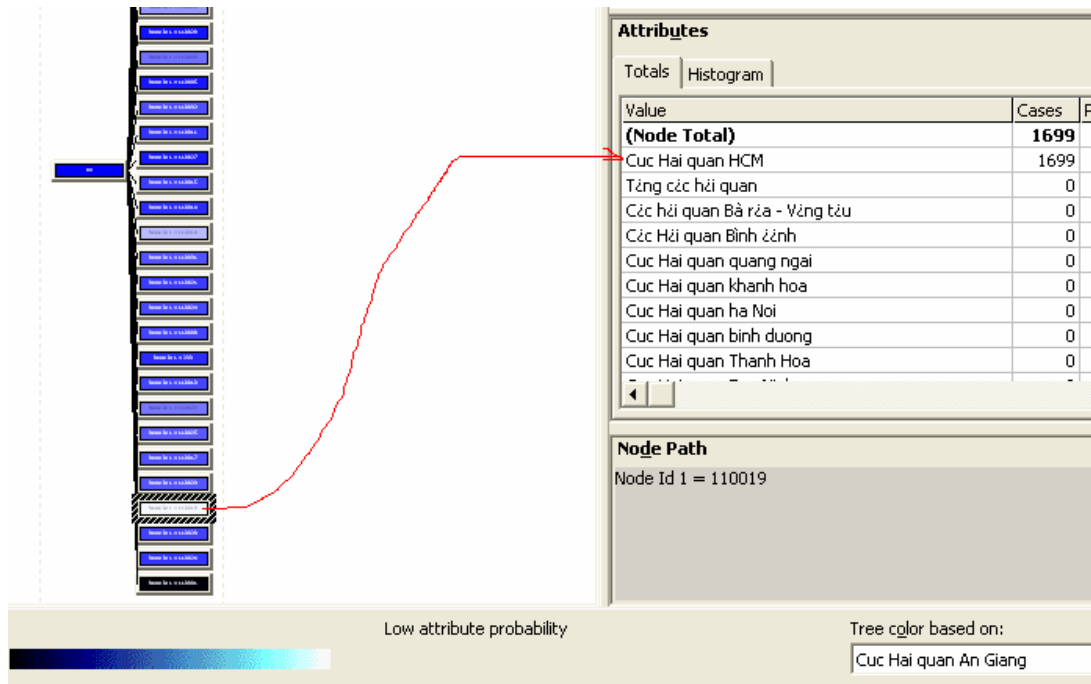
Hình 4.11 Cây hoạch định mô hình tổ chức

Bên phải cây là các thông tin thống kê về từng nút của cây. Đánh giá của mô hình biểu hiện của màu sắc càng đậm thể hiện ở khả năng phải xem xét bổ sung nhân sự, như trong mô hình là nút có màu đen có giá trị Node Id 1 = 110001 là Cục Hải quan An Giang có 177 nhân sự cần xem xét bổ sung nhân sự tại đây:



Hình 4.12 Thông tin nút đơn vị cần bổ sung nhân sự

Nút có màu trắng bên trên có giá trị Node Id=110019 là Cục Hải quan Thành phố Hồ Chí Minh có 1699 nhân sự:



Hình 4.13 Thông tin đơn vị không cần điều chỉnh bổ sung nhân sự

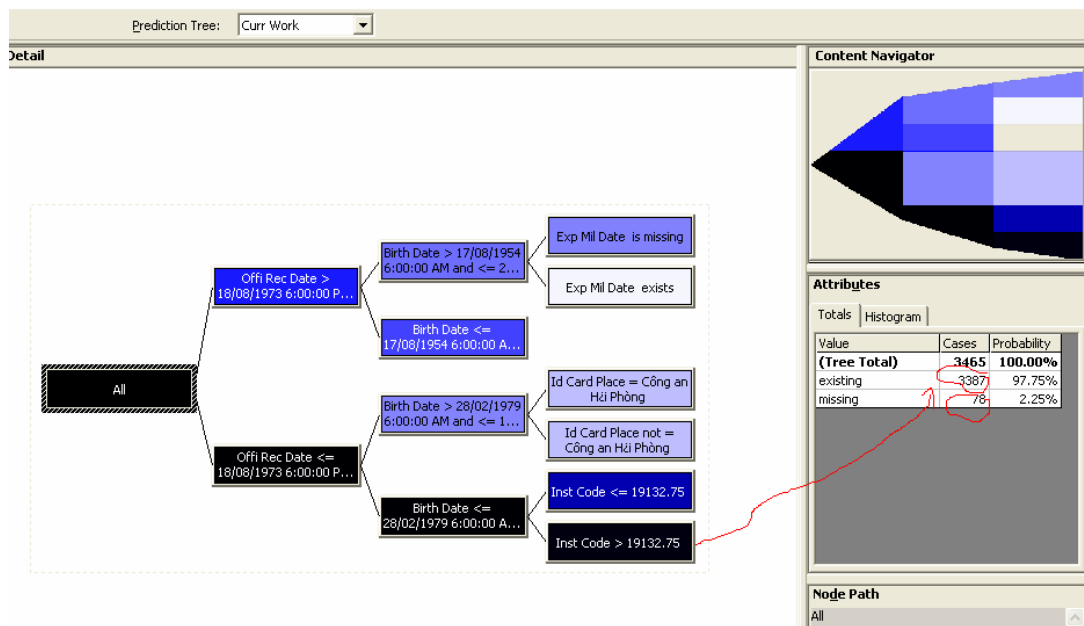
Biểu hiện về màu sắc cho thấy không cần xem xét bổ sung nhân sự tại đơn vị này. Thông tin này cũng khuyến cáo cần phải xem thêm thông tin mô tả chức danh công việc để có thêm thông tin điều chuyển hay cắt giảm biên chế của đơn vị này.

Như vậy, 2 kết quả minh chứng trên cây quyết định hỗ trợ thông tin hoạch định mô hình tổ chức trong việc bổ sung hay cắt giảm nhân sự của các đơn vị trong tổ chức. Mô hình biểu hiện tính trực quan kết nối trực tiếp với dữ liệu khi có thay đổi kết quả cũng sẽ thay đổi theo. Điều này rất có ý nghĩa cho việc kiểm thử nâng cao chất lượng dự đoán của mô hình và cũng cho thấy khả năng đáp ứng của công cụ cũng như việc phân lớp dựa trên cây quyết định trong việc hoạch định mô hình tổ chức.

4.2.4. Minh họa mô hình giám sát theo dõi công việc phục vụ công tác đánh giá:

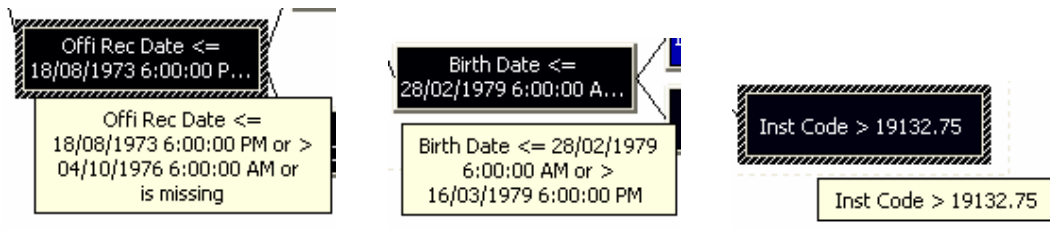
Mô hình biểu hiện mối quan hệ giữa công việc với nhân sự và đơn vị liên quan đến công việc đó (tên mô hình trong công cụ là GSCV_TEST). Trường dữ liệu vào là thông tin hồ sơ nhân sự. Trường dữ liệu cần dự đoán là công việc hiện tại (Curr_work) mã đơn vị công tác của nhân viên đó (name) và nhân viên đó (Key_emp).

Cây quyết định của công việc hiện tại:



Hình 4.14 Cây mô tả công việc hiện tại

Nhìn vào cây ta thấy mật độ nhân viên có công việc cụ thể ở nhánh màu sẫm ở nhánh này có 3387 nhân viên thông tin mô tả công việc cụ thể và 78 nhân viên không có thông tin trong trường công việc cụ thể. Chi tiết thông tin của nhánh này như sau:



Birth Date <= 28/02/1979 > 16/03/1979

Inst Code > 19132 - 19175

Offi Rec Date <= 18/08/1973 or > 04/10/1976

Những người công tác trong những đơn vị có mã trong khoảng 19132-19175 và sinh trong khoảng 28/02/1975, 04/10/1976 và

Vào ngành trước năm 18/08/1973 hay sau 04/10/1976 thì có thông tin mô tả công việc cụ thể

Điều này chứng tỏ các đơn vị có mã trong khoảng 19132 và 19175 có mô tả công việc cụ thể cho cán bộ trong đơn vị. Đây là cơ sở quan trọng để giám sát công việc và đánh giá nhân viên, ta chỉ cần đưa khoảng mã đơn vị này vào truy vấn sẽ được tên đơn vị có tên đơn vị tương ứng. Ý nghĩa của mô hình thể hiện ở tính trực quan nhanh chóng phân loại bước đầu được đơn vị nào giám sát được nhân viên tốt (do thời điểm này đang trong quá trình xây dựng mô tả chức danh công việc của từng nhân viên và thời gian hoàn thành công việc của nhân viên nên chỉ có thể đánh giá bước đầu tới đơn vị mà chưa đánh giá chi tiết đến nhân viên trong đơn vị đó).

4.2.5. Minh họa hỗ trợ xây dựng các quy trình quản lý nguồn nhân lực:

Phần này minh họa đặc tính hỗ trợ của mô hình xây dựng đối với quy trình công việc quản lý nguồn nhân lực: tuyển lựa, tuyển dụng, đào tạo... Quy trình thủ công cũ, phụ thuộc chủ yếu vào đề xuất của người đứng đầu đơn vị (đề xuất của người đứng đầu chuyển đến bộ phận quản lý nhân sự là được phê duyệt). Việc kiểm tra giám sát thông tin đề xuất gặp nhiều khó khăn trong việc tra cứu kiểm tra thông tin (thông tin nhiều nhưng khó mà làm rõ được). Chúng ta sẽ thấy cụ thể hơn bởi quy trình tổng quát sau:

Người thực hiện	Trình tự thực hiện	Giải thích
Thủ trưởng đơn vị	Xác định và đề xuất nhu cầu	
Bộ phận nhân sự (đơn vị quản lý hồ sơ cán bộ)	Tiếp nhận nhu cầu và đề xuất	
Bộ phận nhân sự (đơn vị quản lý hồ sơ cán bộ)	Kiểm tra, xem xét	
Lãnh đạo bộ phận quản lý nhân sự	Phê duyệt chủ trương	
Bộ phận nhân sự (đơn vị quản lý hồ sơ cán bộ)	Chuẩn bị hồ sơ	
Bộ phận nhân sự (đơn vị quản lý hồ sơ cán bộ)	Kiểm tra hồ sơ của đơn vị	
Lãnh đạo Tổng cục	Phê duyệt	
Lãnh đạo Tổng cục	Trao đổi với đơn vị liên quan	
Lãnh đạo Tổng cục	Phê duyệt	
Lãnh đạo Tổng cục, Vụ TCCB và Đơn vị liên quan	Quyết định và triển khai thực hiện	

Với quy trình trên mô hình hỗ trợ sẽ tăng được tính chính xác khi phê duyệt, ra quyết định chính xác hơn. Công tác tuyển lựa đúng và đầy đủ đối tượng kiểm soát được đối tượng lựa chọn, đề xuất người đứng đầu đơn vị, giảm được tiêu cực trong công tác phụ thuộc vào đề xuất người đứng đầu. Kết quả hỗ trợ mở ra một khả năng tiến tới khi dữ liệu được hóa đầy đủ thông tin hơn thì đề xuất của người đứng đầu trong quy trình trên sẽ dần mất đi.

Kết luận chương 4

Các mô hình đưa ra minh chứng ở đây minh họa một số đặc điểm mà phương pháp thống kê thông thường không thể làm được cũng như tính ưu việt thể hiện trong đặc tính hỗ trợ quyết định của công cụ và phương pháp. Qua mô hình lựa chọn cán bộ đào tạo ta sẽ thấy được đặc tính hỗ trợ cho việc lựa chọn chính xác, đủ đối tượng cần chọn (đào tạo, tuyển dụng..). Tính trực quan trong việc theo dõi số liệu hỗ trợ việc điều chỉnh bổ sung dữ liệu sát hơn, đầy đủ hơn trong minh họa phân lớp vị trí công tác. Thấy được đặc tính hỗ trợ thông tin hoạch định mô hình tổ chức trong việc bổ sung, điều chuyển hay cắt giảm nhân sự của các đơn vị trong tổ chức. Mô hình giám sát theo dõi công việc trình bày khả năng hỗ trợ giám sát công việc phục vụ công tác đánh giá.

Như vậy, kết quả đánh giá sử dụng thông tin thu được từ các đặc tính cung cấp của các mô hình phân lớp dựa trên cây quyết định sử dụng công cụ phân tích Microsoft Analysis mang lại những đặc điểm tiện ích sau:

Tìm kiếm những dữ liệu tiềm ẩn trong dữ liệu, bằng phương pháp thống kê thông thường không phát hiện được. Phương pháp thống kê truyền thống đáp ứng đều phải định nghĩa mục đích trước, sau đó sử dụng phương pháp thích hợp để có được thông tin chúng ta cần, để có được những thông tin quý giá, đáng quan tâm, chúng ta phải tốn rất nhiều công sức để tìm kiếm. Việc tìm kiếm không phải dễ dàng nếu không hiểu biết cấu trúc và ý nghĩa thông tin của bảng dữ liệu;

Việc sử dụng công cụ bớt được rất nhiều công sức để chuẩn bị dữ liệu. Công cụ hỗ trợ rất nhiều trong việc phân tích trên một bảng hoặc các bảng có mối quan hệ với nhau;

Dữ liệu nhân sự là dữ liệu thường xuyên thay đổi, tăng trưởng. Việc dùng mô hình phân lớp trên cây quyết định đã đáp ứng yêu cầu giám sát dữ liệu. Khẳng định giám sát dữ liệu, phát hiện sự sai sót trong quá trình nhập liệu chỉ có thể sử dụng phương pháp khai phá dữ liệu;

Kết quả còn biểu hiện tính trực quan trong phân tích. Từ những đặc điểm đó đã mở ra một cách nhìn mới về công tác thống kê báo cáo.

KẾT LUẬN

Thông qua thu thập thông tin và khảo sát yêu cầu, luận văn đã đề cập một quan điểm quản lý nguồn nhân lực trong thời điểm mới áp dụng cho cơ quan Hải quan. Tiếp cận bước đầu của luận văn tiến hành khảo sát được thực trạng khai thác cơ sở dữ liệu hồ sơ nhân sự hiện có để đưa ra phương pháp khai thác mới khoa học hiệu quả hơn mang đặc tính mô hình hỗ trợ ra quyết định cho một phần công việc trọng tâm của công tác quản lý nguồn nhân lực. Nâng cao hiệu quả việc khai thác dữ liệu sơ yếu lý lịch đã được lưu trữ chưa được chuẩn hóa là mang tính tích cực, rất có ý nghĩa phục vụ công tác. Các mô hình đưa ra minh chứng ở đây minh họa một số đặc điểm mà phương pháp thống kê thông thường cũng như các báo cáo mẫu đưa ra được xây dựng theo phương pháp thông không thể làm được cũng như tính ưu việt thể hiện trong đặc tính hỗ trợ quyết định của công cụ và phương pháp. Qua mô hình phân thực nghiệm thấy được nhiều đặc tính hỗ trợ cho việc phân tích trực tuyến, tính trực quan của mô hình, đặc tính giám sát công việc, hoạch định mô hình tổ chức, nâng cao chất lượng cho quy trình quyết định... Sử dụng phân lớp dữ liệu dựa trên cây quyết định để dự đoán các thông tin quan trọng của 1 hồ sơ nhân sự là cơ sở đánh giá.

Về mặt nghiên cứu công cụ, phân lớp dữ liệu nhân sự dựa trên quyết định và sử dụng công cụ Analysis Service của Microsoft đã khám phá ra thuộc tính ẩn trong dữ liệu và đáp ứng được yêu cầu phân tích trên dữ liệu nhân sự luôn tăng trưởng ,thay đổi thường xuyên phù hợp cho việc triển khai áp dụng việc phân tích, xây dựng mô hình(giảm thời gian công sức chuẩn bị dữ liệu xây dựng mô hình phân lớp).

Hướng mở của luận văn là dữ liệu hồ sơ nhân sự hiện có đang ở giai đoạn sơ khởi nhưng kết quả minh họa cho thấy tính hữu dụng của phương pháp áp dụng và cũng đặt ra một giải pháp phân tích hỗ trợ cho những bộ dữ liệu nghiệp vụ khác. Luận văn đã đề xuất một mô hình tổng thể áp dụng công nghệ thông tin cho quản lý nguồn nhân lực cho cơ quan hải quan với những chức năng cần có dựa trên quan điểm xây dựng về quản lý nguồn nhân lực và khảo sát yêu cầu hướng tới của việc khai thác dữ liệu hồ sơ nhân sự hiện có. Hình thành chức năng khai thác phân tích thống kê và những kết quả hỗ trợ từ công nghệ khai phá dữ liệu. Vì đề xuất đưa ra phụ thuộc quá nhiều vào việc xây dựng các hệ thống thông tin quản lý nguồn nhân lực, dữ liệu hồ sơ lý lịch nhân sự hiện có còn chưa đủ thông tin nên nghiên cứu của Luận văn chỉ đưa ra nghiên cứu giải pháp khai thác dữ liệu mới minh chứng tính hiệu quả để phục vụ phân tích đánh

giá một số công việc cần thiết của công tác quản lý nguồn nhân lực. Mong muốn hướng phát triển tiếp theo của đề tài được phát triển xây dựng chuẩn hóa kho dữ liệu quản lý nguồn nhân lực, hoàn thiện những đề xuất đã đưa ra nhằm giải quyết vấn đề bức bối chưa giải quyết được trong công tác quản lý nguồn nhân lực hiện nay.



TÀI LIỆU THAM KHẢO

Tiếng Việt

- [1] Hà Quang Thụy, “*Bài giảng cao học môn học Khai phá web*”, Trường Đại học Công nghệ, 2007.
- [2] Quyết định số: 14/2006/QĐ-BNV ngày 06/11/2006 của Bộ trưởng Bộ Nội vụ về việc ban hành quy chế quản lý hồ sơ cán bộ, công chức
- [3] Học viện bưu chính viễn thông, “*Ứng dụng các kỹ thuật khai phá dữ liệu vào lĩnh vực viễn thông*”
<http://www.uit.edu.vn/forum/index.php?act=Attach&type=post&id=22124>
- [4] “*Pháp lệnh cán bộ công chức*” - 2003
- [5] “*Tạp chí Tổ chức nhà nước*” số tháng 1+2/2008, bài 21 trang 65

Tiếng Anh

- [6] Jiawei Han, Micheline Kamber. “*Datamining- Concept and Techniques*”, Chapter 7-*Classification and Prediction* - Morgan Kaufman Publishers, 2001
- [7] [Wikinews: U.S. Army intelligence detection of 9/11 terrorists before attack](#)
- [8] C.Heckerman-*Decision trees* (classification/regression):
<http://research.microsoft.com/~dmax/publications/dmart-final.pdf>
- [9] msdn.microsoft.com (search “data mining”)
- [10] <http://proquest.safaribooksonline.com>

PHỤ LỤC 1: MÔ TẢ BẢNG DỮ LIỆU SỬ DỤNG

Stt	Tên trường	Kiểu dữ liệu	Ý nghĩa
1	Key_emp	Int(số)	Khóa chính của bảng
2	E_CODE	varchar(chữ)	Mã cán bộ
3	E_NAME_NAME	varchar(chữ)	Tên cán bộ
4	BIRTH_DATE_YEAR	datetime(ngày)	Năm sinh
5	BIRTH_PLACE_TINH	varchar(chữ)	Nơi sinh
6	SEX	varchar(chữ)	Giới tính
7	STA_CODE	varchar(chữ)	Mã ngạch
8	RECR_CODE_TINH	varchar(chữ)	Tên tỉnh trụ sở làm việc
9	INST_CODE	numeric(số)	Mã đơn vị công tác
10	POSITION_CLASS	varchar(chữ)	Lớp chức vụ
11	ID_CARD	varchar(chữ)	Số chứng minh
12	PER_ADDRESS_TINH		Tên tỉnh của chỗ ở
13	EDU_CODE	numeric(số)	Mã đào tạo
14	DATE_COMING_YEAR	int(số)	Năm vào ngành
15	RPOS_DATE_YEAR	int(số)	Năm vào đảng
16	INV_TYPE	varchar(chữ)	Mã tham chiếu->nhóm máu
17	HEA_CON_CLASS	varchar(chữ)	Phân lớp sức khỏe
18	LEA_ALW_COEF	numeric(số)	Hệ số phụ cấp lãnh đạo
19	HOME_PLACE_TINH	varchar(chữ)	Tên tỉnh của quê quán
20	CODE_PARENT	varchar(chữ)	Mã tham chiếu->Mã đơn vị cấp trên
21	CURR_WORK	Varchar(số)	Tên công việc hiện tại
22	OFFI_REC_YEAR	datetime(ngày)	Ngày vào ngành
23	ABILITY_OTHER_CLASS	varchar(chữ)	Phân lớp chứng chỉ khác
24	DES LIM CODE2 CLASS	int(số)	Phân lớp Chuyên ngành đào tạo
25	EDU_NV	int(số)	Có chứng chỉ đào tạo nghiệp vụ
26	Class_Edu	Int(số)	Phân lớp: 0 chưa đào tạo quản lý nhà nước; 1 là đã đi đào tạo
27	Class	Int(số)	Phân lớp chức vụ lãnh đạo cấp phòng trở lên

PHỤ LỤC 2: THÔNG TIN VỀ QUẢN LÝ QUÁ TRÌNH CÁ NHÂN CỦA CHỨC NĂNG HỒ SƠ HIỆN TẠI

Quá trình công tác

Tên thông tin	Giá trị
Từ ngày	Ngày bắt đầu quá trình công tác tại đơn vị công tác (MM/YYYY)
Đến ngày	Ngày kết thúc quá trình công tác tại đơn vị công tác (MM/YYYY)
Đơn vị công tác	Đơn vị công tác trong khoảng thời gian <i>Từ ngày - Đến ngày</i> (có thể tự gõ khi đơn vị công tác nằm ngoài hệ thống hoặc chọn từ danh sách chọn bằng cách ấn <i>F9</i> khi đơn vị công tác nằm trong hệ thống quản lý)
Ngành	Tự cập nhật khi đơn vị công tác nằm trong hệ thống quản lý hoặc là Null khi đơn vị nằm ngoài hệ thống (có thể chọn từ danh sách chọn bằng cách ấn <i>F9</i>)
Chức vụ	Mã chức vụ thực tế theo khối (có thể chọn từ danh sách chọn bằng cách ấn <i>F9</i>)
Công việc	Mô tả các công việc đảm nhiệm trong quá trình công tác
Hình thức tuyển dụng	Hình thức tuyển dụng vào cơ quan (có thể chọn từ danh sách chọn bằng cách ấn <i>F9</i>)
Sự nghiệp	Sự nghiệp được chọn theo đơn vị công tác của cán bộ (có thể chọn từ danh sách chọn bằng cách ấn <i>F9</i>)
Số quyết định	Lưu trữ số quyết định công tác của cán bộ

Quá trình công việc

Tên thông tin	Giá trị
Từ ngày	Ngày bắt đầu công việc (MM/YYYY)
Đến ngày	Ngày kết thúc công việc (MM/YYYY)
Công việc đảm nhiệm	ông việc chuyên ngành đảm nhiệm trong khoảng thời gian <i>Từ ngày - Đến ngày</i> dựa theo Ngành trong quá trình công tác (có thể chọn từ danh sách chọn bằng cách ấn <i>F9</i>)

Quá trình công tác nước ngoài

Tên thông tin	Giá trị
Từ ngày	Ngày bắt đầu đi công tác (MM/YYYY)
Đến ngày	Ngày kết thúc chuyên công tác (MM/YYYY)
Quốc gia công tác	Quốc gia đến công tác (có thể chọn từ danh sách chọn bằng cách ấn <i>F9</i>)
Hình thức công tác	Hình thức đi nước ngoài (có thể chọn từ danh sách chọn bằng cách ấn <i>F9</i>)
Lý do	Lý do đi công tác (ví dụ như đi học, hội thảo...)
Loại hộ chiếu	Chọn từ danh sách chọn với ba loại: PT-Phổ thông, CV-Công vụ, NG-Ngoại giao

Tên thông tin	Giá trị
Đơn vị tổ chức	Đơn vị tổ chức chuyển đi công tác

Quá trình tham gia Đảng

Tên thông tin	Giá trị
Ngày vào Đảng	Ngày vào Đảng (DD/MM/YYYY)
Ngày chính thức vào Đảng	Ngày chính thức vào Đảng (DD/MM/YYYY)
Ngày ra	Ngày ra khỏi Đảng hoặc bị khai trừ (DD/MM/YYYY)
Nơi vào Đảng	Nơi kết nạp Đảng
Số thẻ Đảng	Số thẻ Đảng viên
Ngày cấp thẻ	Ngày cấp thẻ Đảng (DD/MM/YYYY)
Tại Đảng bộ	Đảng bộ cấp thẻ Đảng

Quá trình chức vụ Đảng

Tên thông tin	Giá trị
Từ ngày	Ngày bắt đầu chức vụ Đảng (MM/YYYY)
Đến ngày	Ngày kết thúc chức vụ Đảng (MM/YYYY)
Chức vụ Đảng	Chức vụ Đảng (có thể chọn từ danh sách chọn bằng cách ấn <i>F9</i>)
Đảng bộ	Đảng bộ nơi sinh hoạt

Quá trình lương

Tên thông tin	Giá trị
Từ ngày	Ngày bắt đầu hưởng mức lương (MM/YYYY)
Đến ngày	Ngày kết thúc hưởng mức lương (MM/YYYY). Thông tin này sẽ tự cập nhật khi nhập quá trình lương mới sao cho các quá trình lương là liên tiếp
Ngạch công chức	Ngạch công chức (có thể chọn từ danh sách chọn bằng cách ấn <i>F9</i> hoặc nhập trực tiếp)
Bậc lương	Bậc lương tương ứng với ngạch công chức (chỉ xác định khi Ngạch công chức được chọn từ danh sách chọn)
Hệ số lương	Hệ số lương tương ứng với Ngạch công chức và Bậc lương
Phần trăm hưởng	Phần trăm hưởng lương (mặc định là 100%). Giá trị của trường này nhỏ hơn 100% khi tập sự và lớn hơn 100% khi vượt khung
Ngày xét NL sau	Ngày xét nâng lương lần sau (DD/MM/YYYY). Mặc định bằng ngày bắt đầu hưởng mức lương

Quá trình phụ cấp chức vụ

Tên thông tin	Giá trị
Từ ngày	Ngày bắt đầu hưởng phụ cấp chức vụ (MM/YYYY)
Đến ngày	Ngày kết thúc hưởng phụ cấp chức vụ (MM/YYYY)
Chức vụ	Chức vụ thực tế theo khối (có thể chọn từ danh sách chọn bằng cách ấn <i>F9</i>)
Hệ số	Hệ số phụ cấp chức vụ

Quá trình phụ cấp khác

Tên thông tin	Giá trị
Từ ngày	Ngày bắt đầu hưởng phụ cấp (MM/YYYY)
Đến ngày	Ngày kết thúc hưởng phụ cấp (MM/YYYY)
Loại phụ cấp	Loại phụ cấp được hưởng (có thể chọn từ danh sách chọn bằng cách ấn F9)
Cách tính	Chọn cách tính từ danh sách chọn (theo 4 cách)
Hệ số	Hệ số hưởng phụ cấp
Thành tiền	Số tiền phụ cấp thực hưởng (tự cập nhật theo Cách tính đã chọn hoặc tự nhập)

Chuyên môn - Nghiệp vụ, Tin học - Ngoại ngữ

Tên thông tin	Giá trị
Từ ngày	Ngày bắt đầu quá trình đào tạo (MM/YYYY)
Đến ngày	Ngày kết thúc quá trình đào tạo (MM/YYYY)
Chuyên ngành	Ngành nghề đào tạo (có thể chọn từ danh sách chọn bằng cách ấn F9 hoặc tự nhập)
Văn bằng	Loại văn bằng (có thể chọn từ danh sách chọn bằng cách ấn F9)
Hạng	Hạng văn bằng (có thể chọn từ danh sách chọn bằng cách ấn F9)
Ngày – Số VB	Ngày cấp bằng và số của văn bằng được cấp
Hình thức	Hình thức đào tạo (có thể chọn từ danh sách chọn bằng cách ấn F9)
Nơi đào tạo	Nơi đào tạo (có thể chọn từ danh sách chọn bằng cách ấn F9)
Quốc gia	Quốc gia nơi đào tạo (có thể chọn từ danh sách chọn bằng cách ấn F9). Mặc định là Việt Nam

Lý luận chính trị - Quản lý nhà nước - Quản lý kinh tế

Tên thông tin	Giá trị
Từ ngày	Ngày bắt đầu quá trình đào tạo (MM/YYYY)
Đến ngày	Ngày kết thúc quá trình đào tạo (MM/YYYY)
Văn bằng	Loại văn bằng (có thể chọn từ danh sách chọn bằng cách ấn F9)
Hạng	Hạng văn bằng (có thể chọn từ danh sách chọn bằng cách ấn F9)
Ngày – Số VB	Ngày cấp bằng và số của văn bằng được cấp
Hình thức	Hình thức đào tạo (có thể chọn từ danh sách chọn bằng cách ấn F9)
Nơi đào tạo	Nơi đào tạo (có thể chọn từ danh sách chọn bằng cách ấn F9)
Quốc gia	Quốc gia nơi đào tạo (có thể chọn từ danh sách chọn bằng cách ấn F9). Mặc định là Việt Nam

Quá trình danh hiệu, học hàm

Tên thông tin	Giá trị
----------------------	----------------

Từ ngày	Ngày bắt đầu được nhận danh hiệu hoặc học hàm (MM/YYYY)
Danh hiệu, học hàm	Danh hiệu, học hàm được nhận (có thể chọn từ danh sách chọn bằng cách ấn F9)

Quá trình khen thưởng

Tên thông tin	Giá trị
Năm	Năm khen thưởng (YYYY)
Hình thức khen thưởng	Hình thức khen thưởng (có thể chọn từ danh sách chọn bằng cách ấn F9)
Lý do	Lý do khen thưởng
Số quyết định	Số của quyết định khen thưởng
Đơn vị khen thưởng	Đơn vị ký quyết định khen thưởng
Cấp khen thưởng	Cấp khen thưởng (có thể chọn từ danh sách chọn bằng cách ấn F9)

Quá trình kỷ luật

Tên thông tin	Giá trị
Từ ngày	Ngày bắt đầu nhận hình thức kỷ luật (MM/YYYY)
Đến ngày	Ngày kết thúc nhận hình thức kỷ luật (MM/YYYY)
Hình thức kỷ luật	Hình thức kỷ luật (có thể chọn từ danh sách chọn bằng cách ấn F9)
Lý do	Lý do kỷ luật (có thể chọn từ danh sách chọn bằng cách ấn F9)
Số quyết định	Số của quyết định kỷ luật
Đơn vị kỷ luật	Đơn vị ký quyết định kỷ luật
Cấp kỷ luật	Cấp kỷ luật (có thể chọn từ danh sách chọn bằng cách ấn F9)

Quan hệ gia đình

Tên thông tin	Giá trị
Quan hệ	Loại quan hệ gia đình (có thể chọn từ danh sách chọn bằng cách ấn F9)
Họ và tên	Họ và tên người có quan hệ với công chức
Năm sinh	Năm sinh của người có quan hệ với công chức (YYYY)
Công việc	Mô tả công việc của người có quan hệ với công chức
Nơi ở	Nơi ở của người có quan hệ với công chức
Tình trạng	Tình trạng của người có quan hệ với công chức

Đánh giá hàng năm

Tên thông tin	Giá trị
Năm	Năm đánh giá (YYYY)
Kết quả	Kết quả đánh giá được chọn từ danh sách chọn
Ghi chú	Ghi chú