

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG

NGUYỄN ĐĂNG HIỆP

NGHIÊN CỨU HỆ THỐNG TRỢ GIÚP QUYẾT ĐỊNH
TRONG TƯ VẤN NGHỀ NGHIỆP CHO HỌC SINH
TRUNG HỌC PHỔ THÔNG

LUẬN VĂN THẠC SĨ CÔNG NGHỆ THÔNG TIN

CHUYÊN NGÀNH: HỆ THỐNG THÔNG TIN

MÃ SỐ: 60480104

NGƯỜI HƯỚNG DẪN KHOA HỌC:

PGS.TS. ĐỖ TRUNG TUẤN

Hải Phòng, 10/2017

LỜI CẢM ƠN

Luận văn “Nghiên cứu hệ thống trợ giúp ra quyết định trong tư vấn nghề nghiệp cho học sinh trung học phổ thông” được hoàn thành không những nhờ sự nỗ lực của cá nhân tác giả mà còn có sự trợ giúp, giúp đỡ từ rất nhiều tập thể và cá nhân.

Trước hết, tác giả xin chân thành cảm ơn tất cả các Giáo sư, Phó giáo sư, tiến sĩ, các thầy giáo cô giáo của Khoa công nghệ thông tin trường Đại học Dân lập Hải Phòng đã nhiệt tình giảng dạy, tạo điều kiện thuận lợi cho tác giả trong quá trình học tập, nghiên cứu, hoàn thành chương trình học tập của khóa học.

Đặc biệt, tác giả bày tỏ lòng biết ơn sâu sắc đến thầy PGS.TS. Đỗ Trung Tuấn, các thầy cô trong hội đồng khoa học đã giành thời gian chỉ bảo tận tình giúp em hoàn thành luận văn.

Tác giả xin chân thành cảm ơn Sở GD và ĐT Hải Phòng, Ban giám hiệu và giáo viên trường THPT Lê Ích Mộc huyện Thủy Nguyên thành phố Hải Phòng đã quan tâm giúp đỡ tạo mọi điều kiện thuận lợi cho tác giả trong suốt quá trình học tập, nghiên cứu và hoàn thành luận văn.

Tác giả xin cảm ơn gia đình, bạn bè, đồng nghiệp, đã động viên tiếp thêm nghị lực để tác giả hoàn thành khóa học và luận văn.

Mặc dù đã có nhiều cố gắng, song luận văn khó tránh khỏi những thiếu sót. Tác giả rất mong sự chỉ bảo, góp ý của các nhà khoa học, các thầy cô giáo và đồng nghiệp.

Xin trân trọng cảm ơn!

Hải Phòng, tháng 10 năm 2017

Nguyễn Đăng Hiệp

MỤC LỤC

LỜI CẢM ƠN.....	ii
MỤC LỤC	iii
DANH SÁCH NHỮNG TỪ VIẾT TẮT	vi
Danh mục các hình vẽ và bảng biểu	vii
MỞ ĐẦU	1
1. Lý do chọn đề tài	1
2. Mục tiêu nghiên cứu	2
3. Nội dung nghiên cứu.....	2
4. Phương pháp nghiên cứu	2
5. Cấu trúc luận văn	2
Chương 1.....	3
TỔNG QUAN VỀ HỆ THỐNG TRỢ GIÚP QUYẾT ĐỊNH.....	3
1.1. Hệ thống thông tin	3
1.2. Hệ trợ giúp ra quyết định.....	4
1.2.1. Quyết định.....	4
1.2.2. Quá trình ra quyết định	5
1.3. Hệ trợ giúp quyết định.....	7
1.3.1. Khái niệm hệ trợ giúp quyết định	7
1.3.2. Các thành phần của hệ trợ giúp quyết định	7
1.3.3. Mô hình ra quyết định.....	8
1.3.4. Phân loại hệ trợ giúp ra quyết định.....	9
1.3.5. Năng lực của hệ trợ giúp quyết định.....	11
1.3.6. Phân tích “What-if”	13
1.4. Quá trình ứng dụng tin học trong các bài toán phi cấu trúc	14
1.4.1. Bài toán phi cấu trúc	14
1.4.2. Loại bài toán phi cấu trúc	14
1.5. Cây quyết định.....	16
1.5.1 Khái niệm về cây quyết định	16
1.5.2 Một số vấn đề khi sử dụng cây quyết định	17
1.5.3. Đánh giá cây quyết định trong lĩnh vực khai phá dữ liệu.....	19
1.6. Thuật toán C4.5.....	22

1.6.1 Giới thiệu về thuật toán cây quyết định.....	22
1.6.2 Giới thiệu thuật toán C4.5	22
1.6.3 Thuật toán C4.5 dùng Gain-entropy làm độ đo lựa chọn thuộc tính “tốt nhất”	23
1.6.4 Thuật toán C4.5 với cơ chế riêng trong xử lý những giá trị thiếu.....	26
1.6.5 Tránh “quá vừa” dữ liệu	26
1.6.6 Chuyển đổi từ cây quyết định sang luật	27
1.6.7 Nhận xét về thuật toán C4.5	28
1.7. Kết luận.....	28
Chương 2.....	29
NHU CẦU RA QUYẾT ĐỊNH ĐỐI VỚI VIỆC LỰA CHỌN NGHỀ NGHIỆP CỦA HỌC SINH TRUNG HỌC PHỔ THÔNG	29
2.1 Thực trạng của việc chọn nghề của thanh niên.....	29
2.1.1. Thực trạng chung của thanh niên.....	29
2.1.2. Thực trạng của lựa chọn nghề nghiệp của học sinh.....	31
2.2.3 Nhu cầu về nguồn nhân lực	33
2.2. Yêu cầu của nhà trường về trợ giúp quyết định.....	35
2.2.1. Về ngành nghề lao động	35
2.2.2. Ngành nghề đông người chọn trong mười năm tới	40
2.3. Định hướng nghề tại cơ sở giáo dục phổ thông	43
2.3.1. Tư vấn nghề tại cơ sở giáo dục	43
2.3.2. Tư vấn nhờ hệ thống công nghệ thông tin	46
2.4. Kết luận.....	47
Chương 3.....	48
THỬ NGHIỆM TRỢ GIÚP QUYẾT ĐỊNH LỰA CHỌN NGHỀ NGHIỆP TẠI TRƯỜNG TRUNG HỌC PHỔ THÔNG	48
3.1. Đặt vấn đề	48
3.1.1. Tổng quan về phần mềm Weka	48
3.1.2. Xây dựng dữ liệu đưa vào phần mềm.....	50
3.2. Chuẩn bị dữ liệu thử nghiệm	52
3.2.1 Trích chọn thuộc tính.....	52
3.2.2 Trích chọn thuộc tính bằng phần mềm Weka.....	53
3.2.3 Chuẩn hóa dữ liệu	54
3.2.4 Lựa chọn thuật toán J48 để xây dựng cây quyết định	56

3.2.5	Đánh giá hiệu quả phân lớp của thuật toán.....	58
3.3.	Đánh giá kết quả sử dụng cây quyết định.....	67
3.3.1.	Giải pháp xây dựng hệ thống.....	67
3.3.2.	Giao diện phần mềm.....	67
3.4.	Kết luận chương.....	68
KẾT LUẬN.....		69
1.	Những kết quả đạt được của luận văn	69
2.	Hướng phát triển	69
TÀI LIỆU THAM KHẢO		71

DANH SÁCH NHỮNG TỪ VIẾT TẮT

ACM	Hội máy tính Hoa Kỳ
AEC	Cộng đồng kinh tế ASEAN
ANN	Mạng thần kinh nhân tạo
BCHTW	Ban chấp hành trung ương
CBGV	Vật tư giáo viên
CBQL	Vật tư quản lý
CNTT	Công nghệ thông tin
CSDL	Cơ sở dữ liệu
DB	Database Management
DBMS	Database Management System
DD	Data Dictionary
DGMS	Dialog Generation and Management System
DSS	Decision Support Systems
EIS	Hệ thống thông tin khai thác
ES	Hệ chuyên gia
GDSS	Hệ trợ giúp quyết định theo nhóm
GV	Giảng viên
GVCN	Giáo viên chủ nhiệm
HHTQD	Hệ trợ giúp quyết định
HSSV	Sinh viên sinh viên
MBMS	Model Base Management System
MSS	Hệ trợ giúp quản lý
What if	Câu hỏi trong DSS

Danh mục các hình vẽ và bảng biểu

Hình 1.1. Hệ thống thông tin xí nghiệp	4
Hình 1.2. Các giai đoạn của quá trình ra quyết định	6
Hình 1.3. Hệ thống ra quyết định và môi trường của nó	7
Hình 1.4. Cấu trúc chung của mô hình định lượng.....	8
Hình 1.5. Mô hình khái niệm của DSS	10
Hình 1.6. Thí dụ về DSS và EIS	13
Hình 1.7 Thí dụ về chức năng what-if để phân tích dữ liệu và phân tích nhân quả trong phần mềm Excel	13
Hình 1.8. Chức năng quản lý	14
Hình 1.9 Ví dụ về cây quyết định	17
Hình 1.10. Thuật toán cây quyết định.....	22
Hình 1.11. Giả mã của thuật toán C4.5.....	23
Bảng 1.1 Bảng dữ liệu tập huấn luyện với thuộc tính phân lớp là buys_computer	24
.....	
Hình 2.1. Tư vấn chọn nghề tại Hội chợ việc làm thanh niên	30
Bảng 2.1. Lựa chọn của học sinh THPT sau khi tốt nghiệp (nguồn: tự điều tra) ...	30
Bảng 2.2 Những thuận lợi của học sinh THPT khi lựa chọn.....	31
Bảng 2.3. Những khó khăn của học sinh THPT khi lựa chọn	31
Hình 2.2. Nhu cầu về nhân lực về công nghệ thông tin ngày càng tăng	33
Hình 2.3. Tư vấn việc làm tại Học viện ngân hàng	36
Hình 2.4. Khám chữa bệnh	38
Hình 2.5. Công nghệ thông tin thời đại số hóa.....	39
Hình 2.6. Nhu cầu tuyển dụng theo trình độ, tháng 3-4 năm 2015	40
Hình 2.7. Các nước khối AEC	41
Hình 2.8. Định hướng nghề tại cơ sở giáo dục	44
Hình 2.9. Tư vấn nhờ máy tính.....	46
Hình 3.1. Công cụ của Weka	48
Hình 3.2. Môi trường chính của Weka	49
Hình 3.3. Dữ liệu đầu vào thông dụng trong weka.....	50
Bảng 3.1. Nghề lựa chọn thử nghiệm của luận văn.....	51
Hình 3.4. Dữ liệu huấn luyện để đưa vào Weka.....	51

Hình 3.5. Lược đồ thử nghiệm trong luận văn	52
Hình 3.6: Trích chọn thuộc tính bằng phần mềm Weka.....	53
Hình 3.7: Lựa chọn thuật toán để chuẩn hóa dữ liệu bằng Weka.....	54
Hình 3.7: Chuẩn dữ liệu cho các các thuộc tính	55
Hình 3.4: Bảng dữ liệu sau khi đã chuẩn hóa.....	56
Hình 3.8: Lựa chọn thuật toán J48 với phần mềm Weka	57
Hình 3.9: Đánh giá hiệu quả của thuật toán bằng phương pháp cross-validation ..	58
Hình 3.10: Cây quyết định.....	60
Hình 3.11. Màn hình Weka.....	62
Hình 3.12. Đánh giá kết quả	62
Hình 3.13. Màn hình Weka lần hai	63
Hình 3.14. Đánh giá đối với lần hai.....	63
Hình 3.15. Màn hình đối với lần ba.....	64
Hình 3.16. Đánh giá đối với lần ba.....	64
Hình 3.17. Màn hình đối với lần thứ tư	65
Hình 3.18. Đánh giá đối với lần thứ tư	65
Hình 3.18. Màn hình đối với lần thứ năm	66
Hình 3.19. Đánh giá đối với lần thứ năm	66
Hình 3.20. Giao diện chương trình hỗ trợ ra quyết định trong tư vấn nghề nghiệp	67
Hình 3.12. Giao diện nhập các thông tin để được tư vấn	67
Hình 3.13. Giao diện ra quyết định tư vấn	68

MỞ ĐẦU

1. Lý do chọn đề tài

Trong thời đại ngày nay, trước sự đa dạng và phức tạp của hoàn cảnh khách quan cùng với sự phát triển như vũ bão của công nghệ thông tin, nhà quản lý cần có những tính toán đảm bảo được tính chính xác, kịp thời công việc quản lý của mình. Các ứng dụng máy tính cho quản lý ngày càng nhiều. Cách mạng về máy vi tính khiến máy vi tính giúp nhiều cho các nhà quản lý. Các nhà quản lý có thể truy cập đến hàng ngàn cơ sở dữ liệu trong nhiều nước. Hầu hết các tổ chức, tư nhân hay tập thể, đều dùng phân tích có tính toán trong quyết định của mình.

Các công ty đang phát triển các hệ thống phân tán cho phép khai thác dễ dàng các dữ liệu tại nhiều địa điểm. Các hệ thống thông tin đa dạng có thể được tích hợp với các hệ thống thông tin khác. Nhà quản lý dễ ra quyết định hơn do họ có thông tin chính xác hơn. Việc dùng các hệ thống trợ giúp nhờ máy tính sẽ làm thay đổi cung cách quản lý, tổ chức và công nghệ. Hơn nữa nó giúp tạo ra tiềm năng trợ giúp quản lý.

Tuy nhiên cho dù có nhiều phát triển về công nghệ, nhiều nhà quản lý không dùng máy tính, hay chỉ dùng với các quyết định đơn giản. Hệ trợ giúp quyết định (DSS), Hệ trợ giúp quyết định theo nhóm (GDSS), hệ thống thông tin khai thác (EIS), Hệ chuyên gia (ES), và mạng thần kinh nhân tạo (ANN) là các công nghệ chính để thay đổi tình trạng này. Các công nghệ được nghiên cứu được gọi chung là các hệ trợ giúp quản lý (MSS). DSS và ES cũng là giáo trình được hội ACM khuyến cáo trong chương trình khung của các hệ thống thông tin máy tính.

Cùng với sự phát triển của khoa học kỹ thuật, nền kinh tế thế giới cũng đang phát triển theo hướng nền kinh tế tri thức, với sự phát triển của những ngành công nghệ cao. Việt Nam cũng nằm trong xu hướng chung đó. Trong tình hình chung của đất nước, xã hội hóa giáo dục là một trong những hướng đi cần thiết. Trên thực tế thì hàng nghìn sinh viên ra trường không có việc làm, hoặc sau khi tốt nghiệp phải làm trái ngành, không đúng với chuyên môn được đào tạo, vì vậy vấn đề lựa chọn sao cho đúng, định hướng nghề nghiệp ngay từ trên ghế nhà trường là vô cùng cấp thiết.

Chính vì vậy, với tư cách là một người trực tiếp giảng dạy trong nhà trường Trung học phổ thông (THPT) tôi quyết định chọn đề tài "*Nghiên cứu hệ thống trợ giúp quyết định trong tư vấn nghề nghiệp cho học sinh trung học phổ thông*" nhằm hỗ trợ cho việc lựa chọn nghề nghiệp cho các em học sinh ngay từ khi ngồi trên ghế nhà trường, xác định nghề nghiệp cho tương lai.

2. Mục tiêu nghiên cứu

Tìm hiểu về hệ thống trợ giúp quyết định trong quản lý nói chung và sử dụng các công cụ trợ giúp quyết định để trợ giúp công tác định hướng nghề nghiệp cho học sinh THPT nói riêng.

3. Nội dung nghiên cứu

Nội dung nghiên cứu của luận văn là:

- Lý thuyết về tâm lý chọn nghề, tư chất và năng lực con người, xu hướng chọn nghề.
- Các kỹ thuật ra quyết định.
- Lý thuyết về định hướng nghề nghiệp dựa trên thực tế địa phương

4. Phương pháp nghiên cứu

Để hoàn thành các nội dung nghiên cứu đề tài, học viên lựa chọn sử dụng các phương pháp sau:

- Thu thập dữ liệu;
- Tìm hiểu vấn đề;
- Thử nghiệm;
- Viết báo cáo.

5. Cấu trúc luận văn

Luận văn được chia thành các chương :

1. Chương đầu trình bày tổng quan về các hệ thống DSS, một số khía cạnh liên quan đến hệ thống thông tin quản lý và nhu cầu của quá trình ra quyết định trong công tác quản lý;
2. Chương 2 về nhu cầu ra quyết định cho công tác lựa chọn nghề cho học sinh THPT
3. Chương 3 Thử nghiệm trợ giúp quyết định lựa chọn nghề nghiệp cho học sinh THPT

Cuối luận văn là phần kết luận và danh sách các tài liệu tham khảo, sử dụng trong luận văn.

Chương 1

TỔNG QUAN VỀ HỆ THỐNG TRỢ GIÚP QUYẾT ĐỊNH

Hiện nay nước ta đang diễn ra sự ứng dụng các thành tựu của Công nghệ thông tin vào cuộc sống. Có những ứng dụng đã và đang đem lại những hiệu quả to lớn cho xã hội như Internet, Trí tuệ nhân tạo, Hệ chuyên gia, Cơ sở dữ liệu, ... tuy nhiên, có những ứng dụng rất mới và cần được sự quan tâm và đóng góp của mọi người như Hệ trợ giúp quyết định. Người đóng góp đầu tiên cho DSS là nhà tin học Scott- Morton từ những năm 70 của thế kỷ 20. Ngay từ ban đầu DSS được coi như một hệ thống tương tác máy tính, giúp người ra quyết định dùng dữ liệu và các mô hình để giải quyết bài toán không có cấu trúc. Ban đầu DSS được phát triển và ứng dụng vào hầu hết các lĩnh vực của cuộc sống. DSS càng ngày càng trở thành một trợ lý, trợ lý máy tính, với những đặc tính vô cùng quý báu như chính xác, tính toán nhanh, khả năng ghi nhớ lớn, trung thành và không bao giờ biết mệt mỏi, phù hợp với một xã hội biến động như ngày nay.

Hệ trợ giúp quyết định DSS giúp người lãnh đạo trong việc quyết định các chủ trương kịp thời, đúng đắn, khả thi và có hiệu quả. Thông thường, việc trợ giúp nói trên dựa trên cơ sở các phương pháp tối ưu hoá toán học (lý thuyết quy hoạch, xác suất thống kê, quá trình ngẫu nhiên...) và các kinh nghiệm của các chuyên gia, đồng thời tận dụng thế mạnh của các phương tiện. Tin học trong việc phân tích, đánh giá, đề xuất phương án, lựa chọn tiêu chuẩn tối ưu và tìm phương án tối ưu cũng như dự báo, dự đoán các tình huống có thể xảy đến và các hướng dự phòng, ứng phó hiệu quả.

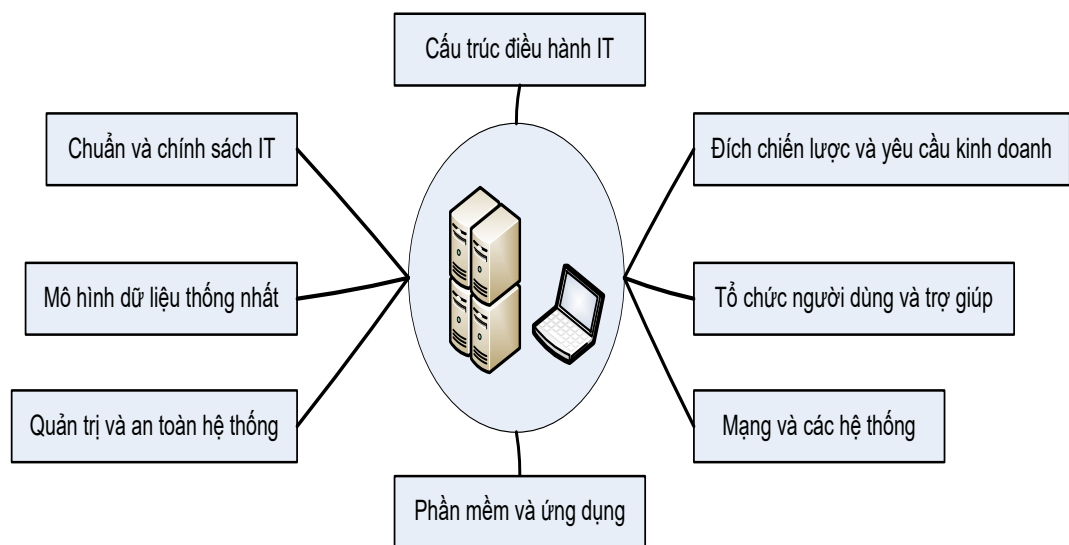
Mặc dù DSS kết hợp trong mình những thành tựu của nhiều chuyên ngành Tin học tiên tiến như Trí tuệ nhân tạo, Quản trị hệ thống, Hệ chuyên gia, khoa học về quản lý, ... nhưng DSS chỉ làm một nhiệm vụ là trợ giúp con người. Nói cách khác, nhà quản lý có thể hỏi Hệ trợ giúp quyết định để nhận được các thông tin định hướng cho việc ra quyết định còn quyền quyết định cuối cùng vẫn thuộc về con người – người ra quyết định.

1.1. Hệ thống thông tin

Hệ thống thông tin là một hệ thống mà mục tiêu tồn tại của nó là cung cấp thông tin phục vụ cho hoạt động của con người trong một tổ chức nào đó.

Hệ thống thông tin có một vai trò quan trọng cho việc ra quyết định quản lý

hệ thống. Trong thời đại ngày nay sự bùng nổ của công nghệ thông tin làm cho nhu cầu có được thông tin chính xác, đầy đủ kịp thời cho việc ra quyết định quản lý không thể thiếu được. Trợ giúp quyết định quản lý thuộc phạm trù các công nghệ trợ giúp quản lý. Do đó tầm quan trọng của việc trợ giúp quản lý gắn liền với tầm quan trọng của hệ trợ giúp quyết định. Một quyết định kịp thời, đúng đắn sẽ đẩy công tác quản lý nói chung tiến triển tốt. Do đó việc nghiên cứu và tìm ra một phương pháp có tính chất tương đối phổ biến cho các dạng quyết định là công việc của hệ trợ giúp quyết định mà ta sẽ đề cập tới trong đề tài này.



Hình 1.1. Hệ thống thông tin xí nghiệp

1.2. Hệ trợ giúp ra quyết định

1.2.1. Quyết định

1.2.1.1. Khái niệm về quyết định

Đó là một lựa chọn về “đường lối hành động” (Simon 1960; Costello & Zalkind 1963; Churchman 1968), hay “chiến lược hành động” (Fishburn 1964) dẫn đến một mục tiêu mong muốn” (Churchman 1968).

“Một quá trình lựa chọn có ý thức giữa hai hay nhiều phương án để chọn ra một phương án tạo ra được kết quả mong muốn trong các điều kiện ràng buộc đã biết”.

1.2.1.2. Hiểu rõ thêm về ra quyết định

Việc đưa ra quyết định đối với một vấn đề xuất hiện trong khắp các lĩnh vực,

hoạt động của đời sống mà đôi khi chúng ta không nhận ra. Từ những việc đơn giản như chọn một bộ quần áo để đi dự tiệc cho đến các công việc lớn lao như phân bổ ngân sách vào các chương trình của quốc gia đều là các công việc đưa ra quyết định.

Vậy đưa ra quyết định chính là chọn ra trong các giải pháp khả thi một giải pháp mà theo người đưa ra quyết định là phù hợp nhất.

1.2.2. Quá trình ra quyết định

1.2.2.1. Phân loại quyết định

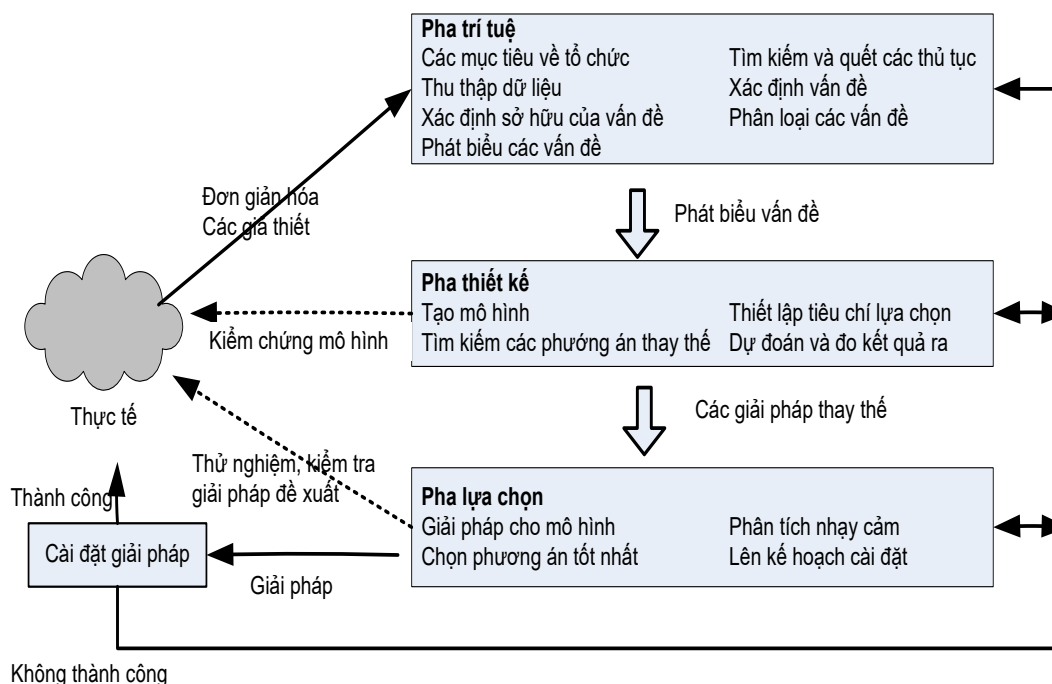
Có thể phân ra bốn loại quyết định như sau [2]:

1. Quyết định có cấu trúc (Structured Decision): Các quyết định mà người ra quyết định biết chắc chắn đúng. Ví dụ: Bài toán quyết định thưởng, phạt Nhân viên.
2. Quyết định không có cấu trúc (NonStructured Decision): Các quyết định mà người ra quyết định biết là có nhiều câu trả lời gần đúng và không có cách nào để tìm ra câu trả lời chính xác nhất. Ví dụ: Bài toán quyết định chiến lược phát triển của Nhà trường.
3. Quyết định đệ quy (Recurring Decision): Các quyết định lặp đi lặp lại.
4. Quyết định không đệ quy (Nonrecurring Decision): Các quyết định không xảy ra thường xuyên.

1.2.2.2. Các giai đoạn của quá trình ra quyết định

Theo Simon [5], quá trình ra quyết định và quan hệ giữa chúng được giới thiệu ở hình dưới đây:

- Giai đoạn thứ nhất là nhận định (Intelligence): Tìm kiếm các tình huống dẫn đến việc phải ra quyết định, nhận dạng các vấn đề, nhu cầu, cơ hội, rủi ro.



Hình 1.2. Các giai đoạn của quá trình ra quyết định

- Giai đoạn thứ hai là thiết kế (Design): Phân tích các hướng tiếp cận để giải quyết vấn đề, đáp ứng các nhu cầu, tận dụng các cơ hội, hạn chế các rủi ro.
- Giai đoạn thứ ba là lựa chọn (Choice): Cân nhắc và đánh giá từng giải pháp và chọn giải pháp tối ưu.
- Cuối cùng là tiến hành ra quyết định (Implementation): Thực hiện giải pháp được chọn, theo dõi kết quả và điều chỉnh khi thấy cần thiết.

1.2.2.3. Tìm kiếm và đánh giá các lựa chọn một phần rất quan trọng trong trợ giúp ra quyết định.

Giai đoạn lựa chọn (Choice Phase) là giai đoạn quan trọng nhất của quá trình ra quyết định. Giai đoạn này bao gồm ba bước chính sau đây:

1. Tìm kiếm lựa chọn
2. Đánh giá lựa chọn
3. Giới thiệu lựa chọn

Trong trường hợp này người ra quyết định muốn sử dụng mô hình quy chuẩn để tìm kiếm một lựa chọn tối ưu, thì Hệ trợ giúp có thể sử dụng phương pháp vét cạn để duyệt hết các lựa chọn hay mô hình toán học để phân tích.

Đối với mô hình mô tả, ta có thể sử dụng phương pháp kinh nghiệm (Heuristic search) để duyệt các lựa chọn dựa trên các quy luật rút ra từ thử và sai hay kinh nghiệm.

Phương pháp đánh giá các lựa chọn được quy định khác nhau trong bài toán một mục tiêu và bài toán đa mục tiêu. Bài toán một mục tiêu có thể được mô hình hóa bằng bảng ra quyết định hay cây quyết định.

Một trong các phương pháp hiệu quả để giải quyết đa mục tiêu là đo lường trọng số của các ưu tiên ra quyết định (Analytical Hierarechy press of Expert choice). Một phương pháp khác là tối ưu hóa dựa trên các mô hình toán học tuyến tính (Microsoft Excel, Lingo...). Một phương pháp khác là lập trình kinh nghiệm sử dụng Heuristics như là tabu search, giải thuật di truyền.

1.3. Hệ trợ giúp quyết định

1.3.1. Khái niệm hệ trợ giúp quyết định

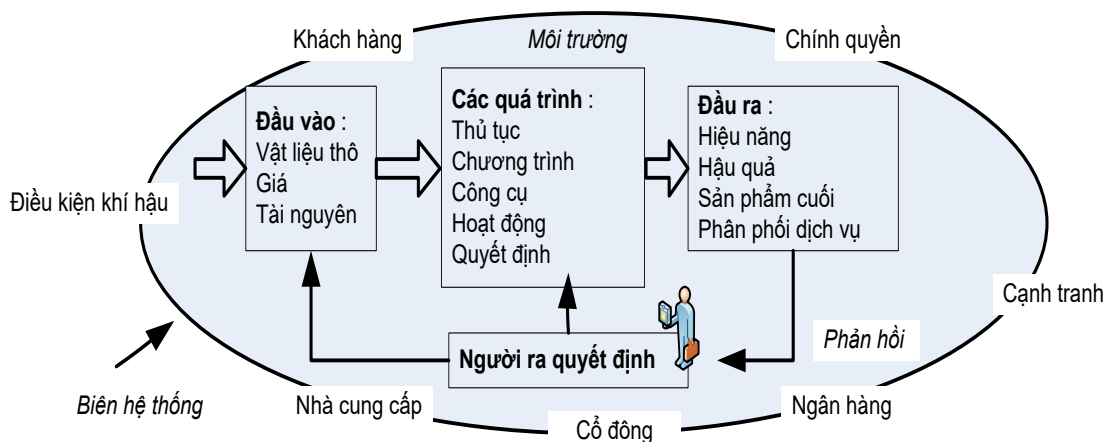
Trong thập niên 1970, Scott Norton đưa những khái niệm đầu tiên về hệ trợ giúp quyết định (Decision Support System, DSS). Ông định nghĩa “DSS là các hệ dựa trên máy tính, có tính tương tác, giúp các nhà ra quyết định dùng dữ liệu và mô hình để giải các bài toán phi cấu trúc, những bài toán mờ, phức tạp với lời giải không hoàn chỉnh”.

Cho đến nay chưa có một định nghĩa thống nhất về DSS. Tuy nhiên tất cả đều đồng ý mục đích cơ bản nhất của DSS là trợ giúp và cải tiến việc ra quyết định.

1.3.2. Các thành phần của hệ trợ giúp quyết định

Một hệ trợ giúp quyết định gồm có ba thành phần chính:

1. Quản lý dữ liệu
2. Quản lý mô hình
3. Quản lý giao diện người dùng



Hình 1.3. Hệ thống ra quyết định và môi trường của nó

Quản lý dữ liệu (Data Management): Thực hiện công việc lưu trữ các thông tin của hệ và phục vụ cho việc lưu trữ, cập nhật, truy vấn thông tin.

Quản lý mô hình (Model Management) hay còn gọi là hệ quản trị cơ sở mô hình (MBMS, model base management system): bao gồm các mô hình ra quyết định (DSS models) và việc quản lý các mô hình này. Một số ví dụ của các mô hình này bao gồm: mô hình nếu thì, mô hình tối ưu, mô hình tìm kiếm mục đích, mô hình thống kê.

Quản lý giao diện người dùng giúp người sử dụng giao tiếp với và ra lệnh cho hệ thống. Các thành phần vừa kể trên tạo nên HHTQĐ, có thể kết nối với intranet/extranet của tổ chức hay kết nối trực tiếp với Internet.

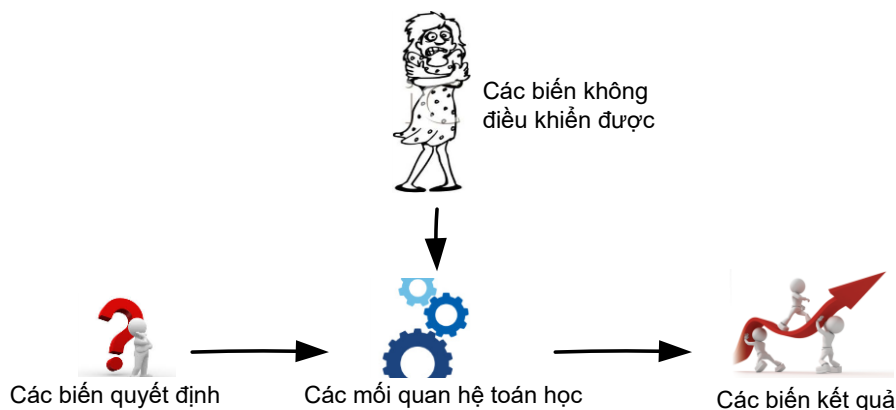
1.3.3. Mô hình ra quyết định

Một đặc trưng cơ bản của hệ trợ giúp ra quyết định là phải có ít nhất một mô hình trợ giúp ra quyết định. Việc chọn lựa và xây dựng mô hình nằm trong giai đoạn thứ 2 (Design Phase) của quá trình ra quyết định.

Mô hình là một khái quát hóa hay trừu tượng hóa các vấn đề thực tế thành các mô hình định tính hay định lượng. Đó là một quy trình kết hợp cả khoa học (sự chính xác, logic) và nghệ thuật (sự sáng tạo).

Một mô hình gồm ba thành phần cơ bản:

1. *Decision Variables*: Đây là các lựa chọn xác định bởi người ra quyết định. Chẳng hạn trong bài toán quyết định thưởng phạt nhân viên ...
2. *Uncontrollable Variables*: Đây là các biến không nằm trong sự kiểm soát của người ra quyết định (bị tác động bởi các yếu tố bên ngoài). Chẳng hạn trong bài toán trên thì đây là...
3. *Result Variables*: Đây là biến kết quả của mô hình. Chẳng hạn trong bài toán trên đây...



Hình 1.4. Cấu trúc chung của mô hình định lượng

Khi lựa chọn quyết định cuối cùng, người ra quyết định có thể muốn có một quyết định tối ưu hay một quyết định thỏa đáng, phần tối ưu. Do vậy có thể chia ra hai loại mô hình trợ giúp ra quyết định.

1. Mô hình quy chuẩn (Normative Model): Mô hình này xem xét tất cả các phương án và chọn ra phương án tối ưu.
2. Mô hình mô tả (Descriptive Model): Mô hình xem xét một tập hợp các điều kiện theo ý người dùng và xem xét các phương án theo các điều kiện này và đưa ra một kết quả thỏa đáng. Vì mô hình này không xem xét hết tất cả các phương án nên kết quả cuối cùng chỉ gần tối ưu.

Mô hình quy chuẩn thường được sử dụng trong bài toán tối ưu hóa một mục tiêu. Mô hình mô tả thường được sử dụng trong bài toán tối ưu hóa đa mục tiêu khi các mục tiêu này có thể mâu thuẫn nhau.

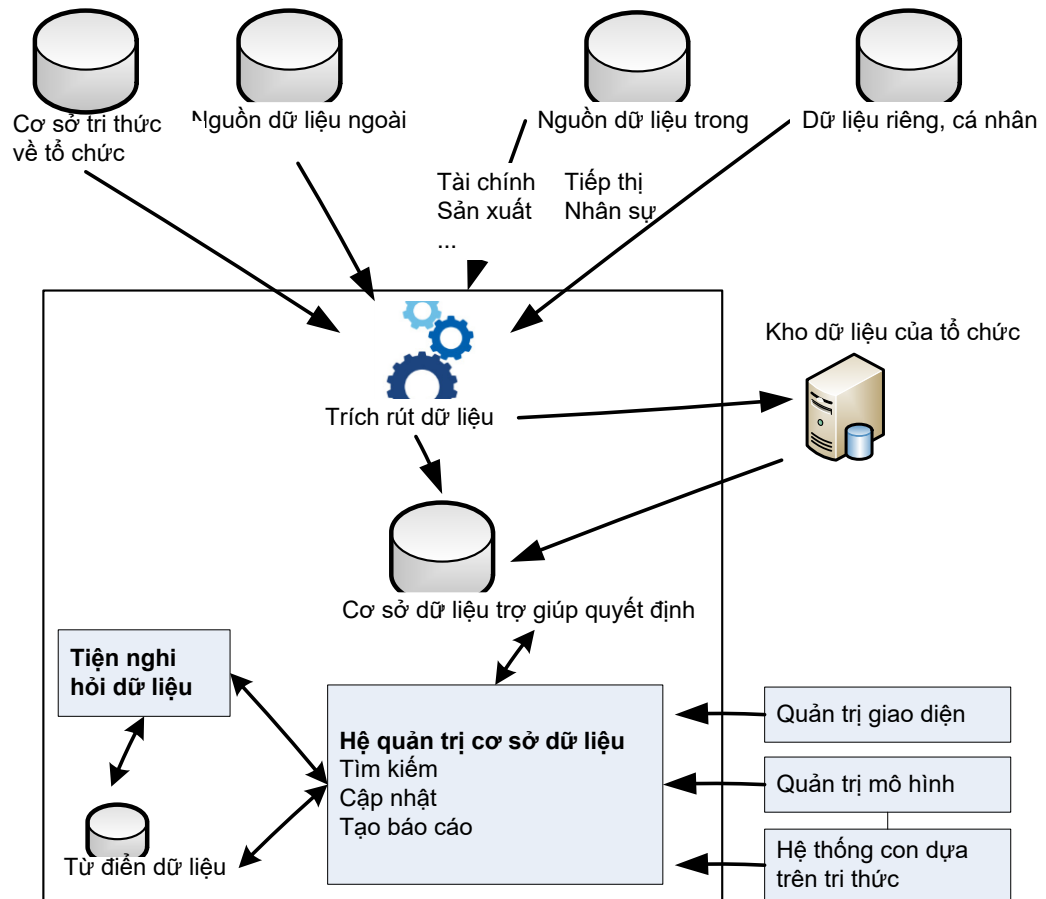
1.3.4. Phân loại hệ trợ giúp ra quyết định

Hệ trợ giúp ra quyết định được phân loại dựa trên nhiều tiêu chí. Hiện nay, vẫn chưa có cách phân loại thống nhất. Sau đây là hai cách phổ biến nhất.

Theo [2] có tất cả năm loại hệ trợ giúp ra quyết định

1. Hướng giao tiếp (Communication, Driven DSS);
2. Hướng dữ liệu (Data, Driven DSS);
3. Hướng tài liệu (Document, Driven DSS);
4. Hướng tri thức (Knowledge, Driven DSS);
5. Hướng mô hình (Model, Driven DSS).

Hướng giao tiếp : Hệ trợ giúp ra quyết định sử dụng mạng và công nghệ viễn thông để liên lạc và cộng tác. Công nghệ viễn thông bao gồm mạng cục bộ (LAN), mạng diện rộng (WAN), Internet, ISDN, mạng riêng ảo.... là then chốt trong việc trợ giúp ra quyết định. Các ứng dụng của Hệ trợ giúp ra quyết định hướng giao tiếp là Phần mềm nhóm (Group ware), hội thảo từ xa (Videoconferencing), bản tin (Bulletin Boards) ...



Hình 1.5. Mô hình khái niệm của DSS

Hướng dữ liệu, Hệ trợ giúp ra quyết định dựa trên truy xuất và xử lý dữ liệu. Phiên bản đầu tiên được gọi là Hệ chỉ dành cho việc truy xuất dữ liệu (Retrieval, Only DSS). Kho dữ liệu (Data warehouse) là cơ sở dữ liệu tập trung chứa thông tin từ nhiều nguồn đồng thời sẵn sàng cung cấp thông tin cần thiết cho việc ra quyết định. OLAP có nhiều tính năng cao cấp. Ví dụ dữ liệu vật tư cần phải phân cấp theo nhiều chiều như theo trình độ học vấn, số năm công tác, theo người làm việc...

Hướng tài liệu, Hệ trợ giúp ra quyết định dựa trên việc truy xuất và phân tích các văn bản, tài liệu... Trong một đơn vị, có rất nhiều văn bản như các công văn đi, đến, nội bộ, giấy tờ... Internet cho phép truy xuất các kho tài liệu lớn như kho văn bản, hình ảnh, âm thanh.. Một công cụ tìm kiếm hiệu quả là phần quan trọng đối với Hệ trợ giúp ra quyết định dạng này.

Hướng tri thức, Hệ trợ giúp ra quyết định có thể đề nghị và đưa ra những tư vấn cho người ra quyết định. Những hệ này là các hệ chuyên gia với một kiến thức chuyên ngành cụ thể, nắm vững các vấn đề trong chuyên ngành đó và có kỹ năng để giải quyết những vấn đề này. Các công cụ khai mở dữ liệu có thể dùng để tạo ra các hệ dạng này.

Theo Holsapple và Whinston (1996) [6] phân ra 6 loại Hệ trợ giúp ra quyết định.

1. Hướng văn bản (Text, Oriented DSS);
2. Hướng cơ sở dữ liệu (Database, Oriented DSS);
3. Hướng bảng tính (Spreasheet, Oriented DSS);
4. Hướng người giải quyết (Solver, Oriented DSS);
5. Hướng luật (Rule, Oriented DSS);
6. Hướng kết hợp (Compound DSS).

Hướng văn bản, Thông tin (bao gồm dữ liệu và kiến thức) được lưu trữ dưới dạng văn bản. Vì vậy hệ thống đòi hỏi lưu trữ và xử lý văn bản một cách hiệu quả. Các công nghệ mới như quản lý văn bản một cách hiệu quả. Các công nghệ mới như hệ quản lý văn bản dựa trên Web, Interlligent Agents có thể được sử dụng cùng với hệ này.

Hướng cơ sở dữ liệu, Cơ sở dữ liệu đóng vai trò chủ yếu trong hệ này. Thông tin trong cơ sở dữ liệu thường có cấu trúc chặt chẽ, các mô tả rõ ràng. Hệ này cho phép người dùng truy vấn thông tin dễ dàng và rất mạnh về báo cáo.

Hướng bảng tính, Một bản tính là một mô hình để cho phép người dùng thực hiện việc phân tích trước khi ra quyết định. Bản tính bao gồm nhiều mô hình thống kê, lập trình tuyến tính... Bản tính phổ biến nhất Microsoft Excel. Hệ này được dùng rộng rãi trong các hệ liên quan tới người dùng cuối.

Hướng người giải quyết, Một trợ giúp là một giải thuật hay chương trình để giải quyết một vấn đề cụ thể chẳng hạn như tính lượng hàng đặt tối ưu hay tính toán xu hướng bán hàng. Một số trợ giúp khác phức tạp như tối ưu hóa đa mục tiêu. Hệ này bao gồm nhiều trợ giúp như vậy.

Hướng luật, Kiến thức của hệ này được mô tả các quy luật thủ tục hay lý lẽ. Hệ này gọi là hệ chuyên gia. Các quy luật này có thể định tính hay định lượng. Các ví dụ của hệ này như là hướng dẫn không lưu, hướng dẫn giao thông trên biển, trên bộ...

Hướng kết hợp, Một hệ tổng hợp có thể kết hợp hai hay nhiều trong số năm kể trên.

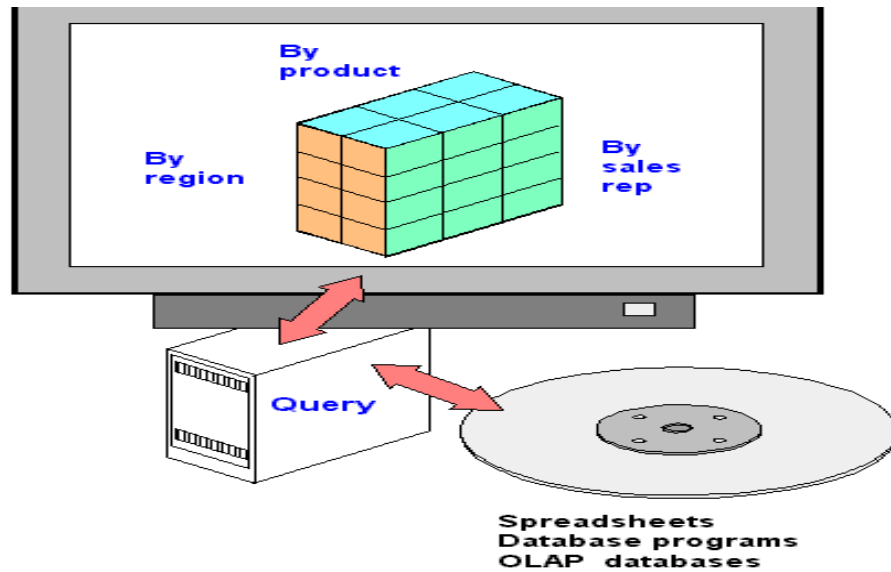
1.3.5. Năng lực của hệ trợ giúp quyết định

Năng lực của DSS, người ta thấy:

- Cung cấp trợ giúp cho người ra quyết định trong những tình huống không cấu trúc và nửa cấu trúc. Những tình huống này không thể

giải quyết bằng các hệ thống tính toán khác.

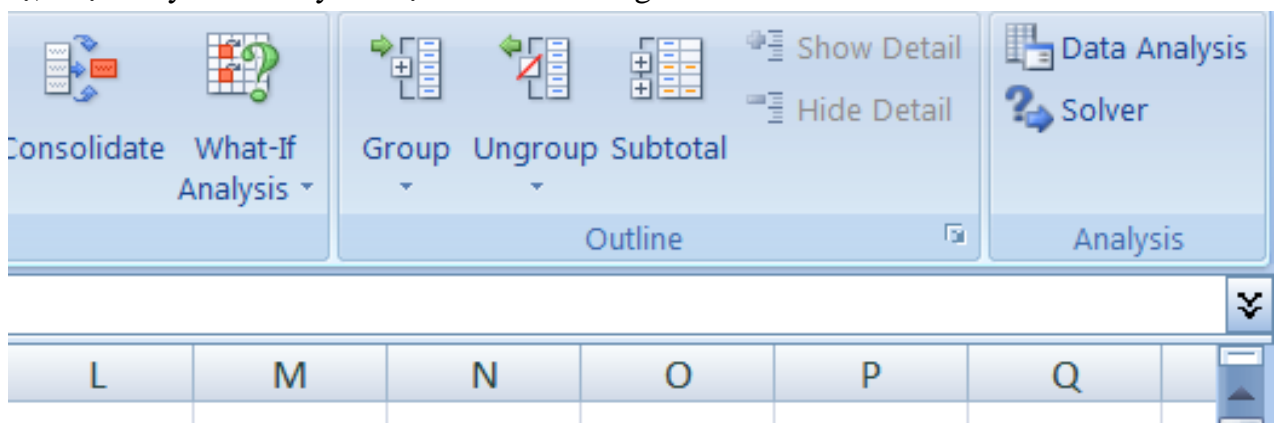
- Sự trợ giúp được cung cấp cho các mức quản lý khác nhau từ người thực thi đến các nhà quản lý.
- Sự trợ giúp cho cá nhân và cho cả nhóm
- DSS trợ giúp cho các giai đoạn của quá trình ra quyết định: Giai đoạn trí tuệ, thiết kế, lựa chọn và cài đặt.
- DSS trợ giúp cho sự đa dạng của quá trình ra quyết định và các kiểu quyết định. Có sự phù hợp giữa DSS và tính cách của cá nhân người ra quyết định, như từ vựng và kiểu ra quyết định.
- DSS thích nghi và mềm dẻo. Do vậy người dùng có thêm xóa, kết hợp, thay đổi hoặc sắp đặt lại các phần tử cơ bản để DSS có thể cung cấp sự trả lời nhanh chóng cho những tình huống không mong đợi.
- DSS dễ sử dụng, người dùng cảm thấy thoải mái đối với hệ thống do DSS thân thiện dùng, mềm, dẻo, những khả năng đồ họa mạnh và có ngôn ngữ giao diện người và máy thích hợp.
- DSS cố gắng nâng cao hiệu quả của quá trình ra quyết định, chẳng hạn như đúng đắn, chính xác, thời gian và chất lượng...
- Người ra quyết định điều khiển toàn bộ các bước của quá trình ra quyết định trong việc giải quyết các bài toán. DSS hướng vào sự trợ giúp chứ không thay thế những người ra quyết định. Người ra quyết định có thể bỏ qua lời khuyên của máy tính vào bất kỳ giai đoạn nào trong quá trình xử lý.
- DSS thường sử dụng các mô hình cho sự phân tích các tình huống ra quyết định. Khả năng mô hình hóa cho phép thí nghiệm với những chiến lược khác nhau và với những cấu hình khác nhau.
- DSS ở mức cao được trang bị thành phần trí thức, do vậy nó cho phép những giải pháp tiềm năng và hiệu quả để giải quyết những bài toán khó.



Hình 1.6. Thí dụ về DSS và EIS

1.3.6. Phân tích “What-if”

Một người làm mô hình tạo ra những dự đoán và những giả định để đánh giá dữ liệu vào. Công việc này nhiều khi để đánh giá tương lai không chắc chắn. Khi mô hình được giải quyết, các kết quả tất nhiên phụ thuộc vào những dữ liệu này. Phân tích nhạy cảm cố gắng kiểm tra sự tác động của những sự thay đổi của dữ liệu vào trên những giải pháp được đề nghị (các biến kết quả). Kiểu của phân tích nhạy cảm được gọi là phân tích “What - if”, bởi vì nó được cấu trúc như là “Điều gì xảy ra cho giải pháp nếu biến vào, giả thiết, hoặc giá trị của tham số được thay đổi”. Nếu giao diện người sử dụng phù hợp thì các nhà quản lý dễ hỏi máy tính những câu hỏi kiểu như thế này. Hơn nữa họ có thể nhắc lại những câu hỏi và thay đổi tỷ lệ, hoặc thay đổi bất kỳ dữ liệu nào khác trong câu hỏi.



Hình 1.7 Thí dụ về chức năng what-if để phân tích dữ liệu và phân tích nhân quả trong phần mềm Excel

1.4. Quá trình ứng dụng tin học trong các bài toán phi cấu trúc

1.4.1. Bài toán phi cấu trúc

Bài toán phi cấu trúc là bài toán không có thuật toán (tức là không có lời giải chính xác trên máy tính điện tử) mà dùng máy tính điện tử để trợ giúp một số phần việc.

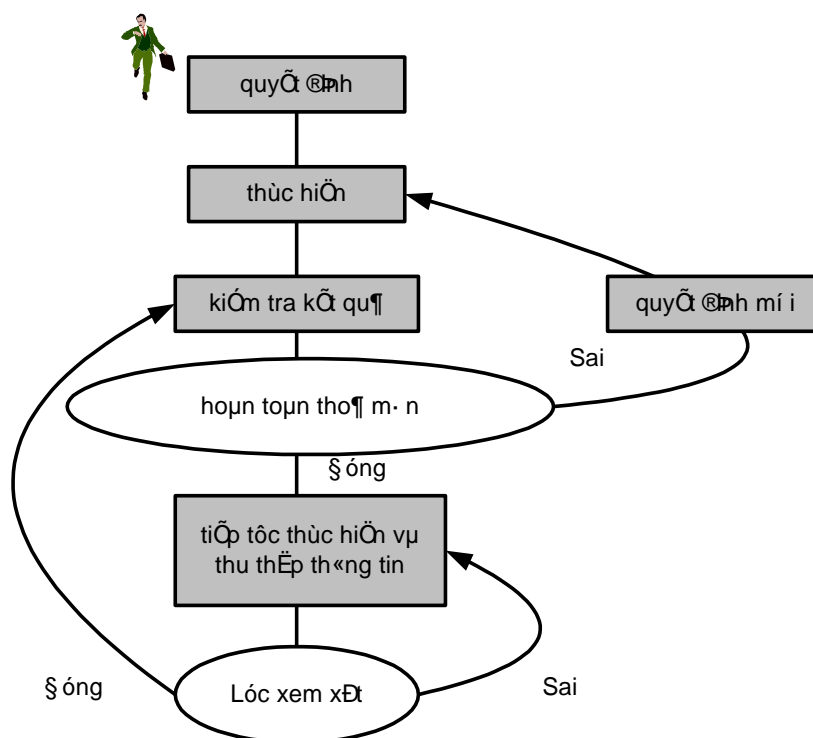
Hệ trợ giúp quyết định cơ bản trợ giúp các nhà ra quyết định trong các tình huống nửa cấu trúc và phi cấu trúc bằng cách kết hợp phán xử của con người và xử lý thông tin bằng máy tính. Các bài toán như vậy không thể/không thuận tiện giải quyết được chỉ bằng các công cụ máy tính hóa hay các phương pháp định lượng.

1.4.2. Loại bài toán phi cấu trúc

Loại bài toán đó là loại bài toán quản lý và ra quyết định

1.4.2.1. Quản lý

Công việc quản lý về cơ bản là quá trình ra quyết định, tức lựa chọn các phương án .



Hình 1.8. Chức năng quản lý

Theo Herbert A.Simon, 1960, quản lý là ra quyết định. Quản lý được mô tả như chu trình liên tục các quyết định và kiểm tra, xem xét.

Đối với khoa học người ta có khái niệm: Khoa học quản lý gồm việc áp dụng

phương pháp khoa học để giải quyết vấn đề của hệ thống người - máy, dùng đo định lượng có hiệu quả. Giải pháp gồm các hoạt động được liệt kê.

Một số lĩnh vực được kể đến, liên quan đến DSS: (i) tối ưu; (ii) điều khiển tác nghiệp; (iii) trợ giúp quyết định.

1.4.2.2. Ra quyết định

Ra quyết định ở một quá trình lựa chọn có ý thức giữa hai hoặc nhiều phương án để chọn ra một phương án này sẽ tạo ra một kết quả mong muốn trong các điều kiện ràng buộc đã biết.

Các loại ra quyết định trong quản lý gồm có hai loại:

Ra quyết định theo chương trình : Nhằm giải quyết các bài toán cấu trúc tốt, lặp đi lặp lại các phương án hầu như có sẵn, lời giải thường dựa trên các kinh nghiệm. Thường để giải quyết bài toán dạng này, các nhà quản lý lập ra các quy trình, luật hay chính sách.

- Quy trình (Procedure): Bao gồm một chuỗi các bước có liên quan nhau mà người ra quyết định có thể sử dụng để xử lý các bài toán cấu trúc tốt.
- Luật (Rule): Là phát biểu cụ thể hướng dẫn người ra quyết định nên làm điều gì và không nên làm điều gì.
- Chính sách (Policy): Là các hướng dẫn để định hướng cho người ra quyết định trong việc giải quyết vấn đề. Khác với luật, chính sách thường là những khái niệm chung chung để người ra quyết định tham khảo hơn là những điều buộc người ra quyết định phải làm.

Ra quyết định không theo chương trình :

- Nhằm giải quyết các bài toán cấu trúc kém, các vấn đề mới, đơn chiếc không lặp đi lặp lại, thông tin không rõ ràng.
- Trong thực tế có nhiều bài toán ở dạng trung gian giữa hai loại vấn đề trên.
- Ra quyết định theo tính chất của vấn đề

Theo tính chất của vấn đề, có thể chia quyết định làm ba loại:

1. Ra quyết định trong điều kiện chắc chắn (certainty): Khi ra quyết định đã biết chắc chắn trạng thái nào sẽ xảy ra, do đó sẽ dễ dàng và nhanh chóng ra quyết định.
2. Ra quyết định trong điều kiện rủi ro (risk): Khi ra quyết định đã

biết được xác suất xảy ra ở mỗi trạng thái.

3. Ra quyết định trong điều kiện không chắc chắn (uncertainty): Khi ra quyết định, không biết được xác suất xảy ra của mỗi trạng thái hoặc không biết được các dữ liệu liên quan đến các vấn đề cần giải quyết.

Quá trình ra quyết định: Quá trình ra quyết định gồm có một số vấn đề sau:

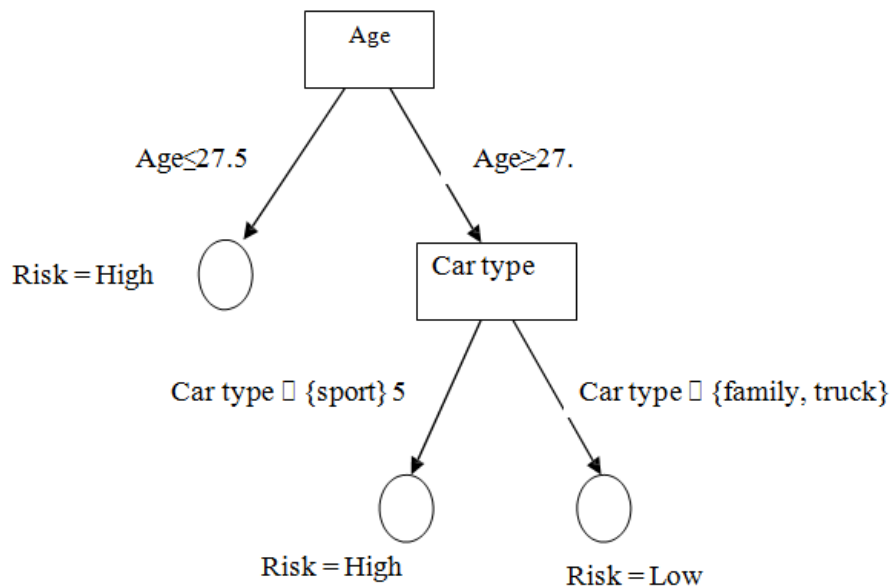
- Hệ thống mục tiêu và môi trường;
- Dự kiến giải pháp;
- Các kĩ thuật tìm giải pháp;
- Các hoạt động sau khi có giải pháp.

1.5. Cây quyết định

1.5.1 Khái niệm về cây quyết định

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, hay bậc hai, 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. 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 biểu đồ phát triển có cấu trúc dạng cây, như mô tả trong hình vẽ sau:



Hình 1.9 Ví dụ về cây quyết định

Trong cây quyết định:

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

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.

1.5.2 Một số vấn đề khi sử dụng cây quyết định

Các vấn đề đặc thù trong khi học hay phân lớp dữ liệu bằng cây quyết định gồm: xác định độ sâu để phát triển cây quyết định, xử lý với những thuộc tính liên tục, chọn phép đo lựa chọn thuộc tính thích hợp, sử dụng tập dữ liệu đào tạo với những giá trị thuộc tính bị thiếu, sử dụng các thuộc tính với những chi phí khác nhau, và cải thiện hiệu năng tính toán. Sau đây sẽ đề cập đến những vấn đề chính đã được giải quyết trong các thuật toán phân lớp dựa trên cây quyết định.

1.5.2.1 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 dữ liệu huấn luyện để thử 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 nhiều dữ liệu.

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

1. 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.
2. 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 tổng quát hóa, và 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.

1.5.2.2. 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 loại ô tô là một thuộc tính rời rạc với tập giá trị là: {xe tải, xe khách, xe con, taxi}. 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 [13][1]. Trong thuật toán C4.5, θ_i được chọn là

giá trị trung bình của hai giá trị liền kề nhau trong dãy giá trị đã sắp xếp.

Ngoài ra còn một số vấn đề liên quan đến sinh tập luật, xử lý với giá trị thiếu sẽ được trình bày cụ thể trong phần thuật toán C4.5.

1.5.3. Đánh giá cây quyết định trong lĩnh vực khai phá dữ liệu

1.5.3.1 Sức mạnh của cây quyết định

Cây quyết định có 5 sức mạnh chính sau [5]:

1. *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 tiếng Anh, 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.
2. *Khả năng thực thi trong những lĩnh vực hướng quy tắc.* Điều này có nghe có vẻ hiển nhiên, nhưng quy tắc quy nạp nói chung và cây quyết định nói riêng là lựa chọn hoàn hảo cho những lĩnh vực thực sự là các quy tắc. Rất nhiều lĩnh vực từ di truyền tới các quá trình công nghiệp thực sự chứa các quy tắc ẩn, không rõ ràng do khá phức tạp và tối nghĩa bởi những dữ liệu nhiễu loạn. Cây quyết định là một sự lựa chọn tự nhiên khi chúng ta nghi ngờ sự tồn tại của các quy tắc ẩn, không rõ ràng.
3. *Dễ dàng tính toán trong khi phân lớp.* Mặc dù như chúng ta đã biết, cây quyết định có thể chứa nhiều định dạng, nhưng trong thực tế, các thuật toán sử dụng để tạo ra cây quyết định thường tạo ra những cây với số phân nhánh thấp và các thử đơn giản tại từng nút. Những thử điển hình là: so sánh số, xem xét phần tử của một tập hợp, và các phép nối đơn giản. Khi thực thi trên máy tính, những thử này chuyển thành các toán hàm logic và số nguyên là những toán hạng thực thi nhanh và không đắt. Đây là một ưu điểm quan trọng bởi trong môi trường thương mại, các mô hình dự đoán thường được sử dụng để phân lớp hàng triệu thậm chí hàng tỉ bản ghi.
4. *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 nơ ron nhân tạo và các kỹ thuật thống kê lại thực sự dễ dàng thao tác với các tiêu chuẩn phân chia 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 nút 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 thử nhị phân của ngưỡng đó.

5. 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 nút gốc của cây. Từ đó có thể thấy những thuộc tính nào là quan trọng nhất cho việc dự đoán hay phân lớp.

1.5.3.2. 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.

1. *Dễ xảy ra lỗi khi có quá nhiều lớp.* Một số cây quyết định chỉ thao tác với những lớp giá trị nhị phân dạng *yes/no* hay *accept/reject*. Số khác lại có thể chỉ định các bản ghi vào một số lớp bất kỳ, nhưng dễ xảy ra lỗi khi số ví dụ đào tạo ứng với một lớp là nhỏ. Điều này xảy ra càng nhanh hơn với cây mà có nhiều tầng hay có nhiều nhánh trên một nút.
2. *Chi phí tính toán đắt để đào tạo.* Điều này nghe có vẻ mâu thuẫn với khẳng định ưu điểm của cây quyết định ở trên. Nhưng quá trình phát triển cây quyết định đắt về mặt tính toán. Vì cây quyết định có rất nhiều nút trong trước khi đi đến lá cuối cùng. Tại từng nút, cần tính một *độ đo* (hay *tiêu chuẩn phân chia*) trên từng thuộc tính, với thuộc tính liên tục phải thêm thao tác sắp xếp lại tập dữ liệu theo thứ tự giá trị của thuộc tính đó. Sau đó mới có thể chọn được một thuộc tính phát triển và tương ứng là một phân chia tốt nhất. Một vài thuật toán sử dụng tổ hợp các thuộc tính kết hợp với nhau có trọng số để phát triển cây quyết định. Quá trình cắt cụt cây cũng “đắt” vì nhiều cây con ứng cử phải được tạo ra và so sánh.

1.5.3.3 Xây dựng cây quyết định

Quá trình xây dựng cây quyết định gồm hai giai đoạn:

1. Giai đoạn thứ nhất phát triển cây quyết định: Giai đoạn này phát triển bắt đầu từ gốc, đến từng nhánh và phát triển quy nạp theo cách thức chia để trị cho tới khi đạt được cây quyết định với tất cả các lá được gán nhãn lớp.
2. Giai đoạn thứ hai cắt, tỉa bớt các cành nhánh trên cây quyết định. Giai đoạn này nhằm mục đích đơn giản hóa và khái quát hóa từ đó làm tăng độ chính xác của cây quyết định bằng cách loại bỏ sự phụ thuộc vào mức độ lỗi (noise) của dữ liệu đào tạo mang tính chất thống kê, hay những sự biến đổi mà có thể là đặc tính riêng biệt của dữ liệu đào tạo. Giai đoạn này chỉ truy cập dữ liệu trên cây quyết định đã được phát triển trong giai đoạn trước và quá trình thực nghiệm cho thấy giai đoạn này không tốn nhiều tài nguyên tính toán, như với phần lớn các thuật toán, giai đoạn này chiếm khoảng dưới 1% tổng thời gian xây dựng mô hình phân lớp [7][1].

Do vậy, ở đây chỉ tập trung vào nghiên cứu giai đoạn phát triển cây quyết định. Dưới đây là khung công việc của giai đoạn này:

1. Chọn thuộc tính “tốt” nhất bằng một độ đo đã định trước
2. Phát triển cây bằng việc thêm các nhánh tương ứng với từng giá trị của thuộc tính đã chọn
3. Sắp xếp, phân chia tập dữ liệu đào tạo tới nút con
4. Nếu các ví dụ được phân lớp rõ ràng thì dừng. Ngược lại: lặp lại bước 1 tới bước 4 cho từng nút con

1.5.3.4. Thuật toán xây dựng cây quyết định

Tư tưởng chung: Phần lớn các thuật toán phân lớp dữ liệu dựa trên cây quyết định có mã giả như sau:

Make Tree (Training Data T)

```
{
Partition (T)
}
```

Partition (Data S)

```
{
if (all points in S are in the same class) then return
for each attribute A do
evaluate splits on attribute A;
```

```

use best split found to partition S into S1, S2, ..., Sk
Partition (S1) Partition (S2)
...
Partition (Sk)
}

```

Hình 1.10. Thuật toán cây quyết định

Các thuật toán phân lớp như C4.5 (Quinlan, 1993), CDP (Agrawal và các tác giả khác, 1993), SLIQ (Mehta và các tác giả khác, 1996) và SPRINT (Shafer và các tác giả khác, 1996) đều sử dụng phương pháp của Hunt làm tư tưởng chủ đạo. Phương pháp này được Hunt và các đồng sự nghĩ ra vào những năm cuối thập kỷ 50 đầu thập kỷ 60 [5].

1.6. Thuật toán C4.5

1.6.1 Giới thiệu về thuật toán cây quyết định

Theo [5] C4.5 là sự kế thừa của của thuật toán học máy bằng cây quyết định dựa trên nền tảng là kết quả nghiên cứu của HUNT và các cộng sự của ông trong nửa cuối thập kỷ 50 và nửa đầu những năm 60 (Hunt 1962). Phiên bản đầu tiên ra đời là ID3 (Quinlan, 1979), hệ thống đơn giản ban đầu chứa khoảng 600 dòng lệnh Pascal, và tiếp theo là C4 (Quinlan 1987). Năm 1993, J. Ross Quinlan đã kế thừa các kết quả đó phát triển thành C4.5 với 9000 dòng lệnh C chứa trong một đĩa mềm. Mặc dù đã có phiên bản phát triển từ C4.5 là C5.0 - một hệ thống tạo ra lợi nhuận từ Rule Quest Research, nhưng nhiều tranh luận, nghiên cứu vẫn tập trung vào C4.5 vì mã nguồn của nó là sẵn dùng.

Trong các thuật toán phân lớp dữ liệu dựa trên cây quyết định, C4.5 và thuật toán tiêu biểu cho hai phạm vi ứng dụng khác nhau. C4.5 là thuật toán hiệu quả và được dùng rộng rãi nhất trong các ứng dụng phân lớp với lượng dữ liệu nhỏ cỡ vài trăm nghìn bản ghi.

1.6.2 Giới thiệu thuật toán C4.5

C4.5 là thuật toán phân lớp dữ liệu dựa trên cây quyết định hiệu quả và phổ biến trong những ứng dụng khai phá cơ sở dữ liệu có kích thước nhỏ. C4.5 sử dụng cơ chế lưu trữ dữ liệu thường trú trong bộ nhớ, chính đặc điểm này làm C4.5 chỉ thích hợp với những cơ sở dữ liệu nhỏ, và cơ chế sắp xếp lại dữ liệu tại mỗi nút trong quá trình phát triển cây quyết định. C4.5 còn chứa một kỹ thuật cho phép biểu diễn lại cây quyết định dưới dạng một danh sách sắp thứ tự các luật if-

then (một dạng quy tắc phân lớp dễ hiểu). Kỹ thuật này cho phép làm giảm bớt kích thước tập luật và đơn giản hóa các luật mà độ chính xác so với nhánh tương ứng cây quyết định là tương đương.

Tư tưởng phát triển cây quyết định của C4.5 là phương pháp HUNT đã nghiên cứu ở trên. Chiến lược phát triển theo độ sâu (*depth-first strategy*) được áp dụng cho C4.5.

Giả mã của thuật toán C4.5:

```
ComputerClassFrequency (T);
if OneClass or FewCases return a leaf;
Create a decision node N;
ForEach Attribute A ComputeGain (A);
N.test=AttributeWithBestGain;
if N.test is continuous find Threshold;
ForEach T' in the splitting of T
if T' is Empty Child of N is a leaf else
Child of N=FormTree (T');
ComputeErrors of N;
return N
```

Hình 1.11. Giả mã của thuật toán C4.5

1.6.3 Thuật toán C4.5 dùng Gain-entropy làm độ đo lựa chọn thuộc tính “tốt nhất”

Phần lớn các hệ thống học máy đều cố gắng để tạo ra 1 cây càng nhỏ càng tốt, vì những cây nhỏ hơn thì dễ hiểu hơn và dễ đạt được độ chính xác dự đoán cao hơn.

Do không thể đảm bảo được sự cực tiểu của cây quyết định, C4.5 dựa vào nghiên cứu tối ưu hóa, và sự lựa chọn cách phân chia mà có *độ đo lựa chọn thuộc tính* đạt giá trị cực đại. Hai độ đo được sử dụng trong C4.5 là *information gain* và *gain ratio*. $RF(C_j, S)$ biểu diễn tần suất (*Relative Frequency*) các trường hợp trong S thuộc về lớp C_j . $RF(C_j, S) = |S_j| / |S|$, với $|S_j|$ là kích thước tập các trường hợp có giá trị phân lớp là C_j . $|S|$ là kích thước tập dữ liệu đào tạo.

Chỉ số thông tin cần thiết cho sự phân lớp: $I(S)$ với S là tập cần xét sự phân phối lớp được tính bằng:

$$I(S) = - \sum_{j=1}^x RF(C_j, S) \log(RF(C_j, S)).$$

Sau khi S được phân chia thành các tập con S₁, S₂, ..., S_t bởi thử B thì *information gain* được tính bằng:

$$G(S, B) = I(S) - \sum_{i=1}^t \frac{|S_i|}{|S|} I(S_i).$$

Thử B sẽ được chọn nếu có G(S, B) đạt giá trị lớn nhất. Tuy nhiên có một vấn đề khi sử dụng G(S, B) ưu tiên thử có số lượng lớn kết quả, ví dụ G(S, B) đạt cực đại với thử mà từng S_i chỉ chứa một *trường hợp* đơn. Tiêu chuẩn *gain ratio* giải quyết được vấn đề này bằng việc đưa vào thông tin tiềm năng (*potential information*) của bản thân mỗi phân hoạch

$$P(S, B) = - \sum_{i=1}^t \frac{|S_i|}{|S|} \log \left(\frac{|S_i|}{|S|} \right).$$

Thử B sẽ được chọn nếu có tỉ số giá trị *gain ratio* = G(S, B) / P(S, B) lớn nhất. Trong mô hình phân lớp C4.5 thế hệ 8, có thể dùng một trong hai loại chỉ số *Information Gain* hay *Gain ratio* để xác định thuộc tính tốt nhất. Trong đó *Gain ratio* là lựa chọn mặc định.

1.6.3.1. Với thuộc tính rời rạc

Bảng 1.1 Bảng dữ liệu tập huấn luyện với thuộc tính phân lớp là `buys_computer`

Số	Tuổi	Thu nhập	Sinh viên	Mức thanh toán	Mua máy tính
1	<30	high	no	fair	no
2	<30	high	no	excellent	no
3	30-40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	30-40	low	yes	excellent	yes
8	<30	medium	no	fair	no
9	<30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<30	medium	yes	excellent	yes
12	30-40	medium	no	excellent	yes
13	30-40	high	yes	fair	yes
14	>40	medium	no	excellent	no

Thuộc tính phân loại

Các thuộc tính quyết định

Trong tập dữ liệu trên: s_1 là tập những bản ghi có giá trị phân lớp là *yes*, s_2 là tập những bản ghi có giá trị phân lớp là *no*. Khi đó:

$$I(S) = I(s_1, s_2) = I(9, 5) = -9/14 \cdot \log_2 9/14 - 5/14 \cdot \log_2 5/14 = 0.940$$

Tính $G(S, A)$ với A lần lượt là từng thuộc tính:

$A = \text{age}$. Thuộc tính age đã được rời rạc hóa thành các giá trị <30 , $30-40$, và >40 .

- Với $\text{age} = "<30"$: $I(S_1) = I(s_{11}, s_{21}) = -2/5 \log_2 2/5 - 3/5 \log_2 3/5 = 0.971$
- Với $\text{age} = "30-40"$: $I(S_2) = I(s_{12}, s_{22}) = 0$
- Với $\text{age} = ">40"$: $I(S_3) = I(s_{13}, s_{23}) = 0.971$

$$\sum |S_i| / |S| \cdot I(S_i) = 5/14 \cdot I(S_1) + 4/14 \cdot I(S_2) + 5/14 \cdot I(S_3) = 0.694$$

$$\text{Gain}(S, \text{age}) = I(s_1, s_2) - \sum |S_i| / |S| \cdot I(S_i) = 0.246$$

Tính tương tự với các thuộc tính khác ta được:

- $A = \text{income}$: $\text{Gain}(S, \text{income}) = 0.029$
- $A = \text{student}$: $\text{Gain}(S, \text{student}) = 0.151$
- $A = \text{credit_rating}$: $\text{Gain}(S, \text{credit_rating}) = 0.048$

Thuộc tính age là thuộc tính có độ đo *Information Gain* lớn nhất. Do vậy age được chọn làm thuộc tính phát triển tại nút đang xét.

1.6.3.2. Với thuộc tính liên tục

Xử lý thuộc tính liên tục đòi hỏi nhiều tài nguyên tính toán hơn thuộc tính rời rạc. Gồm các bước sau:

- Kỹ thuật *Quick sort* được sử dụng để sắp xếp các trường hợp trong tập dữ liệu đào tạo theo thứ tự tăng dần hoặc giảm dần các giá trị của thuộc tính liên tục V đang xét. Được tập giá trị $V = \{v_1, v_2, \dots, v_m\}$
- Chia tập dữ liệu thành hai tập con theo ngưỡng $\theta_i = (v_i + v_{i+1}) / 2$ nằm giữa hai giá trị liên kế nhau v_i và v_{i+1} . Thử để phân chia dữ liệu là thử nhị phân dạng $V \leq \theta_i$ hay $V > \theta_i$. Thực thi thử đó ta được hai tập dữ liệu con: $V_1 = \{v_1, v_2, \dots, v_i\}$ và $V_2 = \{v_{i+1}, v_{i+2}, \dots, v_m\}$.
- Xét $(m-1)$ ngưỡng θ_i có thể có ứng với m giá trị của thuộc tính V bằng cách tính *Information gain* hay *Gain ratio* với từng ngưỡng đó. Ngưỡng có giá trị của *Information gain* hay *Gain ratio* lớn nhất sẽ được chọn làm ngưỡng phân chia của thuộc tính đó.

Việc tìm ngưỡng (theo cách tuyến tính như trên) và sắp xếp tập training theo thuộc tính liên tục đang xem xét đôi khi gây ra thất cổ chai vì tốn nhiều tài nguyên tính toán.

1.6.4 Thuật toán C4.5 với cơ chế riêng trong xử lý những giá trị thiếu

Giá trị thiếu của thuộc tính là hiện tượng phổ biến trong dữ liệu, có thể do lỗi khi nhập các bản ghi vào cơ sở dữ liệu, cũng có thể do giá trị thuộc tính đó được đánh giá là không cần thiết đối với *trường hợp* cụ thể.

Trong quá trình xây dựng cây từ tập dữ liệu đào tạo S , B là thử dựa trên thuộc tính A_a với các giá trị đầu ra là b_1, b_2, \dots, b_t . Tập S_0 là tập con các *trường hợp* trong S mà có giá trị thuộc tính A_a không biết và S_i biểu diễn các *trường hợp* với đầu ra là b_i trong thử B . Khi đó độ đo *information gain* của thử B giảm vì chúng ta không học được gì từ các *trường hợp* trong S_0 .

$$G(S, B) = \frac{|S - S_0|}{|S|} G(S - S_0, B).$$

Tương ứng với $G(S, B)$, $P(S, B)$ cũng thay đổi,

$$P(S, B) = -\frac{|S_0|}{|S|} \log\left(\frac{|S_0|}{|S|}\right) - \sum_{i=1}^t \frac{|S_i|}{|S|} \log\left(\frac{|S_i|}{|S|}\right).$$

Hai thay đổi này làm giảm giá trị của thử liên quan đến thuộc tính có tỉ lệ giá trị thiếu cao. Nếu thử B được chọn, C4.5 không tạo một nhánh riêng trên cây quyết định cho S_0 . Thay vào đó, thuật toán có cơ chế phân chia các trường hợp trong S_0 về các tập con S_i là tập con mà có giá trị thuộc tính thử xác định theo trong số $|S_i|/|S - S_0|$.

1.6.5 Tránh “quá vừa” dữ liệu

“Quá vừa” dữ liệu là một khó khăn đáng kể đối với học bằng cây quyết định và những phương pháp học khác. Quá vừa dữ liệu là hiện tượng: nếu không có các trường hợp xung đột (là những trường hợp mà giá trị cho mọi thuộc tính là giống nhau nhưng giá trị của lớp lại khác nhau) thì cây quyết định sẽ phân lớp chính xác toàn bộ các trường hợp trong tập dữ liệu đào tạo. Đôi khi dữ liệu đào tạo lại chứa những đặc tính cụ thể, nên khi áp dụng cây quyết định đó cho những tập dữ liệu khác thì độ chính xác không còn cao như trước.

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

1. 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.

2. Cho phép cây có thể “quá vừa” dữ liệu, sau đó sẽ cắt, tia cây

Mặc dù phương pháp thứ nhất có vẻ trực quan hơn, nhưng với phương pháp thứ hai thì cây quyết định được sinh ra được thử nghiệm chứng minh là thành công hơn trong thực tế, vì nó cho phép các tương tác tiềm năng giữa các thuộc tính được khám phá trước khi quyết định xem kết quả nào đáng giữ lại. C4.5 sử dụng kỹ thuật thứ hai để tránh “quá vừa” dữ liệu.

1.6.6 Chuyển đổi từ cây quyết định sang luật

Việc chuyển đổi từ cây quyết định sang luật sản xuất (*production rules*) dạng if-then tạo ra những quy tắc phân lớp dễ hiểu, dễ áp dụng. Các mô hình phân lớp biểu diễn các khái niệm dưới dạng các luật sản xuất đã được chứng minh là hữu ích trong nhiều lĩnh vực khác nhau, với các đòi hỏi về cả độ chính xác và tính hiệu được của mô hình phân lớp. Dạng kết quả đầu ra tập luật sản xuất là sự lựa chọn “khôn ngoan”. Tuy nhiên, tài nguyên tính toán dùng cho việc tạo ra tập luật từ tập dữ liệu đào tạo có kích thước lớn và nhiều giá trị sai là vô cùng lớn [12]. Khẳng định này sẽ được chứng minh qua kết quả thực nghiệm trên mô hình phân lớp C4.5

Giai đoạn chuyển đổi từ cây quyết định sang luật bao gồm 4 bước:

1. *Cắt tia*: Luật khởi tạo ban đầu là đường đi từ gốc đến lá của cây quyết định. Một cây quyết định có l lá thì tương ứng tập luật sản xuất sẽ có l luật khởi tạo. Từng điều kiện trong luật được xem xét và loại bỏ nếu không ảnh hưởng tới độ chính xác của luật đó. Sau đó, các luật đã cắt tia được thêm vào tập luật trung gian nếu nó không trùng với những luật đã có.
2. *Lựa chọn*: Các luật đã cắt tia được nhóm lại theo giá trị phân lớp, tạo nên các tập con chứa các luật theo lớp. Sẽ có k tập luật con nếu tập training có k giá trị phân lớp. Từng tập con trên được xem xét để chọn ra một tập con các luật mà tối ưu hóa độ chính xác dự đoán của lớp gắn với tập luật đó.
3. *Sắp xếp*: Sắp xếp K tập luật đã tạo ra từ trên bước theo tần số lỗi. Lớp mặc định được tạo ra bằng cách xác định các trường hợp trong tập training không chứa trong các luật hiện tại và chọn lớp phổ biến nhất trong các trường hợp đó làm lớp mặc định.
4. *Ước lượng, đánh giá*: Tập luật được đem ước lượng lại trên toàn bộ tập training, nhằm mục đích xác định xem liệu có luật nào làm giảm độ chính xác của sự phân lớp. Nếu có, luật đó bị loại bỏ và quá trình ước lượng được lặp cho đến khi không thể cải tiến thêm.

1.6.7 Nhận xét về thuật toán C4.5

Thuật toán C4.5 là một thuật toán hiệu quả cho những tập dữ liệu vừa và nhỏ.

C4.5 có cơ chế sinh cây quyết định hiệu quả và chặt chẽ bằng việc sử dụng độ đo lựa chọn thuộc tính tốt nhất là *information-gain*. Các cơ chế xử lý với giá trị lỗi, thiếu và chống “quá vừa” dữ liệu của C4.5 cùng với cơ chế cắt tỉa cây đã tạo nên sức mạnh của C4.5.

Ngoài ra, mô hình phân lớp C4.5 còn có phần chuyển đổi từ cây quyết định sang luật dạng *if-then*, làm tăng độ chính xác và tính dễ hiểu của kết quả phân lớp. Đây là tiện ích rất có ý nghĩa đối với người sử dụng.

1.7. Kết luận

Hệ trợ giúp quyết định DSS là hệ thống cơ sở máy tính được thiết kế nâng cao hiệu quả của những người làm quyết định từng công việc cụ thể. Nói cách khác, DSS là hệ thống dựa trên tương tác máy tính, giúp người ra quyết định dùng dữ liệu và các mô hình để giải quyết bài toán nửa cấu trúc hoặc phi cấu trúc. DSS được sử dụng để trợ giúp công tác quản lý. Hệ DSS hoàn chỉnh có thể đưa ra những thông tin trợ giúp cho sự giải quyết các vấn đề.

Với những công việc như vậy người ra quyết định không thể nắm bắt được toàn bộ thông tin chính xác và cụ thể về các vấn đề như: vấn đề các bài toán, cơ hội, sự lựa chọn các cách giải quyết sao cho tối ưu, hoặc là quyết định các vấn đề theo tiêu chuẩn hoặc giá trị trong sự lựa chọn.

DSS được thiết kế như là một tổ hợp với mục đích của cá nhân, tổ chức và với hoàn cảnh chung của môi trường cạnh tranh, và đề cao quá trình ra quyết định của cá nhân, tổ chức. DSS dùng công nghệ để trợ giúp quá trình của nhóm, như hệ thống trợ giúp quyết định nhóm tăng cường công tác quản lý của các cá nhân.

Chương 2

NHU CẦU RA QUYẾT ĐỊNH ĐỐI VỚI VIỆC LỰA CHỌN NGHỀ NGHIỆP CỦA HỌC SINH TRUNG HỌC PHỔ THÔNG

Lao động, việc làm và nghề nghiệp là một vấn đề được rất nhiều nhà khoa học quan tâm nghiên cứu. Mặt khác, khi nghiên cứu vấn đề này, các tác giả thường hay đặt mục tiêu tìm hiểu về định hướng nghề nghiệp, về những dự định việc làm nghề nghiệp, và nói chung, về hiện trạng lao động - việc làm - nghề nghiệp xã hội của giới trẻ.

Trên bình diện định hướng việc làm nghề nghiệp ở thanh niên, nhiều tác giả đặc biệt quan tâm tới đối tượng là những học sinh sắp kết thúc trường THPT. Các tác giả thường nhấn mạnh giá trị việc làm, bên cạnh nhiều giá trị khác của xã hội mà thanh niên cần hướng tới, hay những yếu tố khác như nơi làm việc, cơ quan, khu vực làm việc...

Tuy nhiên, khi ngồi trên ghế nhà trường thì các em học sinh đa phần chưa định hướng được mình sau này sẽ làm gì, học sinh, chuẩn bị những gì cho những công việc sau này mình lựa chọn,

2.1 Thực trạng của việc chọn nghề của thanh niên

2.1.1. Thực trạng chung của thanh niên

Chọn nghề là một nhiệm vụ quan trọng, chi phối phần lớn suy nghĩ và hoạt động của các em học sinh cuối cấp THPT. Tuy nhiên hoạt động giáo dục hướng nghiệp ở các trường vẫn còn không ít hạn chế. Hầu hết các trường THPT đều đặc cách các giáo viên thiếu tiết đảm nhiệm công tác này, cho nên quá trình chuẩn bị thông tin, kiến thức cho công tác tư vấn hướng nghiệp trong nhà trường của các giáo viên còn mang tính tự phát, chưa có hệ thống. Công tác tư vấn hướng nghiệp cho học sinh lớp 12 chưa đúng và chung chung: Ở các trường THPT chỉ khi đến kỳ học sinh chuẩn bị làm hồ sơ thi vào các trường chuyên nghiệp, các trường mới công bố các thông tin hướng nghiệp trên báo chí, trên Internet ở bảng thông báo để học sinh tham khảo, sau đó hướng dẫn các em ghi hồ sơ tuyển sinh, và cho đó là cách làm của tư vấn hướng nghiệp. Hàng năm, khi gần đến kỳ thi tuyển sinh Đại học, các trường THPT mới bắt đầu liên hệ với một số trường Đại học, Cao đẳng về làm tư vấn tuyển sinh cho học sinh trường mình. Khi tư vấn tuyển sinh như vậy thì các trường chỉ tổ chức gói gọn trong một buổi, chủ yếu các

em học sinh chỉ được nghe thông tin về chỉ tiêu tuyển sinh, về điểm chuẩn, về các ngành đào tạo, về mức học phí của trường đến làm tư vấn tuyển sinh. Khi các em có những băn khoăn khác nhau về sự phù hợp nghề, về thị trường lao động, khả năng phát triển của nghề đã chọn trong tương lai... thì hầu như không có thời gian để hỏi, hoặc không được giải thích một cách thoả đáng. Chính vì vậy đã dẫn đến tình trạng thừa thầy thiếu thợ và hàng năm có rất nhiều sinh viên tốt nghiệp ở nhiều ngành nghề khác nhau nhưng các em rất khó tham gia vào thị trường lao động trong tỉnh, một phần do các em không đáp ứng được yêu cầu của nghề nghiệp, một phần do ngành nghề của các em chọn không có trong nhu cầu của thị trường lao động.



Hình 2.1. Tư vấn chọn nghề tại Hội chợ việc làm thanh niên

Qua điều tra thấy rằng, tỷ lệ học sinh dự định sẽ thi ĐH, CĐ sau khi tốt nghiệp THPT chiếm tỷ lệ khá cao (78, 32%). Thực tế cho thấy rằng ở Hà Nội hiện nay, những gia đình khá giả, có điều kiện thường đầu tư cho con em đi học tiếp ở nước ngoài sau khi tốt nghiệp THPT. Và có thể nói đây là một trong những con đường tốt nhất để con em họ có được những ngành nghề vững vàng, ổn định trong tương lai. Do đó số học sinh dự định đi du học chiếm tỷ lệ không nhỏ (9, 09%). Trong khi đó, 1 số ít học sinh khác thay cho dự định sẽ chọn ngành nghề cho mình thông qua các trường ĐH, CĐ thì lại quyết tâm đi làm ngay sau khi tốt nghiệp. Số này chiếm 2, 79%.

Bảng 2.1. Lựa chọn của học sinh THPT sau khi tốt nghiệp (nguồn: tự điều tra)

Học ĐH	Đi du học	Đi làm	Học nghề	Chưa rõ	Phương án khác
78, 32%	9, 09%	2, 79%	0, 72%	4, 19%	4, 89%

Qua đây, chúng ta có thể thấy rằng trong định hướng nghề nghiệp của học

sinh THPT, việc thi vào ĐH, CĐ không phải là con đường duy nhất để họ có được 1 ngành nghề tốt như mong đợi. Thế hệ trẻ ngày nay đã linh hoạt hơn trong việc chọn ngành, chọn nghề. Họ biết lựa chọn cho mình 1 hướng đi phù hợp với hoàn cảnh, khả năng của bản thân, gia đình và xã hội. Bên cạnh đó cũng có 1 số ít học sinh dự định học nghề sau khi tốt nghiệp (0, 92%); hay có 4, 19% các em vẫn chưa xác định rõ ràng nghề nghiệp trong tương lai của mình. Tuy nhiên không phải tất cả học sinh THPT đều chọn những dự định trên, họ còn có những phương án khác như: thi ĐH nếu trượt thì đi du học, lấy vợ, lấy chồng, buôn bán nhỏ...

2.1.2. Thực trạng của lựa chọn nghề nghiệp của học sinh

Một vấn đề không kém phần quan trọng trong việc lựa chọn của học sinh THPT, đó là việc các em nhận thức được những khó khăn và thuận lợi khi lựa chọn. Trong tổng số lượng phiếu thu được thì có 95, 8% học sinh đã nhận thức được khó khăn và thuận lợi và một phần nhỏ còn lại là 4, 2% thì không nhận thức được. Điều này chứng tỏ học sinh đã có ý thức tự giác với việc định hướng nghề nghiệp của bản thân, đã xác định được những thuận lợi, khó khăn nảy sinh khi mình lựa chọn.

Theo thống kê của chúng tôi thì các em đã có những thuận lợi sau:

Bảng 2.2 Những thuận lợi của học sinh THPT khi lựa chọn

Có lòng say mê, ham thích	56, 5%
Được gia đình ủng hộ	28, 9%
Có kinh nghiệm từ gia đình, họ hàng, người quen	10, 5%
Phương tiện học tập thuận lợi (xe cộ đi lại, trang thiết bị...)	4, 1%

Như vậy học sinh THPT đã đưa ra được những nhân tố thúc đẩy sự lựa chọn nghề nghiệp của mình. Lòng say mê, ham thích chính là động lực lớn nhất mà phần đông học sinh lựa chọn. Bên cạnh đó là sự tác động không nhỏ từ phía gia đình, thông tin đại chúng và những cơ sở vật chất phục vụ cho việc học tập, cho nghề nghiệp lựa chọn. Bên cạnh thuận lợi, các em cũng đã nhận thức được những khó khăn mà mình gặp phải.

Bảng 2.3. Những khó khăn của học sinh THPT khi lựa chọn

Không hứng thú mấy	14%
Gia đình không đồng ý	8, 4%
Không có những hiểu biết, thông tin, kinh nghiệm về nghề nghiệp đó	35%

Thiếu phương tiện học tập	4, 9%
Không có khó khăn	26, 5%
Phương án khác	11, 2%

Như vậy với những số liệu trên, chúng ta đều nhận thấy rằng sự thiếu hiểu biết về trường học, ngành nghề là khó khăn lớn nhất đối với hầu hết các học sinh THPT. Tuy là đã chọn trường, chọn nghề nhưng thực tế nhiều em không có những thông tin cụ thể về trường và ngành đó ra sao? Đây được coi là tình trạng phổ biến của học sinh THPT. Đây không chỉ là nỗi lo của riêng học sinh, phụ huynh học sinh mà còn đặt ra cho toàn xã hội : làm sao để học sinh THPT hiểu hết được những định hướng tương lai của mình?

Bên cạnh đó, các yếu tố : gia đình, hứng thú, phương tiện học tập cũng là những khó khăn trở ngại cho việc lựa chọn nghề nghiệp của học sinh THPT.

Với thực tế nước ta hàng năm số học sinh tốt nghiệp lớp 12 tăng lên nhanh chóng trong khi đó chỉ tiêu tuyển sinh vào ĐH, CĐ tăng không đáng kể, số dư trong nỗi thất vọng tìm kiếm một nghề nghiệp của lớp trẻ càng nhiều hơn sau mỗi năm. Vì vậy, việc xác định trước những khó khăn thuận lợi là vô cùng cần thiết vì điều đó sẽ phần nào giúp học sinh có định hướng rõ ràng, nâng cao sự tự giác bản thân với nghề nghiệp.

Nếu như những khó khăn như không hiểu biết về trường, về ngành mình lựa chọn, không hứng thú với ngành đó thì đây thuộc về lỗi chủ quan. Những học sinh này, theo chúng tôi cần phải chủ động hơn nữa để tìm kiếm những thông tin về ngành nghề mà mình đã lựa chọn, có thể tìm kiếm trên báo, đài hay từ bạn bè, người quen... Tuy nhiên cũng có những nguyên nhân khách quan khiến cho học sinh chưa khắc phục được như cha mẹ không đồng ý, thiếu phương tiện học tập...Điều này rất cần sự quan tâm của cha mẹ, xã hội...

Bên cạnh đó, có 15, 38% học sinh đã khắc phục được khó khăn và 27, 3% không có quan niệm gì về vấn đề này. Đáng mừng cho những học sinh đã khắc phục được khó khăn, từ đó các em có thể vững tin hơn để đi theo ngành nghề mà mình đã lựa chọn. Thế nhưng, câu hỏi đặt ra là tại sao số lượng học sinh không có ý kiến lại khá cao như vậy? Phải chăng những em này không thật sự quyết tâm với ngành nghề mà mình lựa chọn, các em không chịu nỗ lực, cố gắng? Vì thế, theo chúng tôi các em này rất cần sự giúp đỡ từ người thân, bạn bè... giúp các em có những định hướng rõ ràng hơn trong việc lựa chọn ngành nghề.

2.2.3 Nhu cầu về nguồn nhân lực

Riêng về nhân lực công nghệ thông tin: Mỏng số lượng, yếu chất lượng. Ngành công nghệ thông tin (CNTT) đang bước vào thời kỳ tăng trưởng nóng nhưng “lực lượng làm CNTT của Việt Nam còn rất mỏng về số lượng, yếu về chất lượng”, như nhận xét của Phó thủ tướng Vũ Đức Đam, đang là thực trạng báo động... [6].

1.2.3.1. Nhà nhà “khát” nhân lực

Theo [6], “Để mở rộng hoạt động sản xuất, trong đợt 1 năm 2017, chúng tôi tuyển dụng hàng ngàn kỹ sư, cử nhân đã/sẽ tốt nghiệp trước 31/12/2017”. Đó là thông báo tuyển dụng mới nhất của Samsung Việt Nam trong tháng 3/2017.

Hiện Samsung có 8 Trung tâm, Nhà máy và đang có tới hơn 150.000 nhân lực làm việc, nhưng với kế hoạch phát triển mới của mình, lượng lao động cần tuyển mới hàng năm lên tới hàng chục ngàn lao động. Nhu cầu nhân sự ngành CNTT đang ở mức cao nhất trong lịch sử với gần 15.000 việc làm được tuyển dụng trong năm 2016.



Hình 2.2. Nhu cầu về nhân lực về công nghệ thông tin ngày càng tăng

Tại FPT Software, đơn vị này vừa tiếp nhận nhân sự thứ 10.000, đưa tổng số nhân sự của mình lên chiếm 10% tổng nhân lực CNTT mảng phần mềm. Để đáp ứng nhu cầu tăng trưởng trong giai đoạn 2017 – 2020, ông FPT Hoàng Nam Tiến, Chủ tịch Công ty FPT Software cho biết, cần tuyển 20.000 nhân sự ở tất cả các vị trí, từ kiểm thử, lập trình viên, kỹ sư cầu nối, biên dịch (comtor) đến quản trị dự án. Trong đó nhiều vị trí có mức thu nhập hấp dẫn và cơ hội làm việc tại nước ngoài, như cơ hội làm việc dài hạn tại Nhật Bản với mức lương từ 2.800 – 3.000USD/tháng.

“Chúng tôi đang có lợi thế về năng lực trong dịch vụ chuyển đổi công nghệ số (Digital transformation) và khả năng mở rộng quy mô nhân lực một cách nhanh

chóng với tốc độ 30-40%/ năm”, ông Tiên cho biết.

Tính chung cả Tập đoàn FPT, nhu cầu tuyển dụng từ nay tới năm 2020 là khoảng 50.000 nhân sự ở nhiều vị trí, từ sinh viên thực tập, quản lý tới lãnh đạo cấp cao trong các khối ngành kỹ thuật, công nghệ, kinh tế.

Nhu cầu về nhân sự CNTT tại VNPT, Viettel, CMC và hàng loạt các công ty CNTT phần mềm, các startup trong nước, các tập đoàn đa quốc gia hoạt động tại Việt Nam như Samsung, LG, Intel, IBM... cũng đang gia tăng nhanh chóng để phục vụ nhu cầu mở rộng sản xuất, đáp ứng các đơn hàng mới...

1.2.3.2. Vấn đề “cung” thì lớn mà “cầu” rất hạn chế

Báo cáo mới nhất về ngành CNTT Việt Nam 2017 của Vietnamworks cho hay, nhu cầu nhân sự ngành CNTT đang ở mức cao nhất trong lịch sử với gần 15.000 việc làm được tuyển dụng trong năm 2016. Theo dự báo của Vietnamworks, với gần 80.000 nhân lực CNTT sẽ được các trường cho “ra lò” trong hai năm, 2017 và 2018, so với nhu cầu tính đến cuối năm 2018, Việt Nam sẽ thiếu khoảng 70.000 nhân lực về CNTT.

Việt Nam cần khoảng 1, 2 triệu nhân lực ngành CNTT vào năm 2020, nhưng theo tính toán, số nhân lực thiếu hụt lên tới trên 500.000 người.

Còn nếu tính tới năm 2020, số nhân lực thiếu hụt sẽ lên tới hơn 500.000 người.

Báo cáo của Bộ Thông tin và Truyền thông cho hay, tổng số nhân lực trong ngành công nghiệp CNTT hiện nay là hơn 600.000 người, trong đó số lao động đang làm việc trong các ngành công nghiệp phần cứng - điện tử là khoảng 300.000 người. Số còn lại thuộc về lĩnh vực công nghiệp phần mềm và công nghiệp nội dung số. Cũng theo Báo cáo này, Việt Nam cần khoảng 1, 2 triệu nhân lực CNTT vào năm 2020.

1.2.3.3. Yếu về chất lượng

Không chỉ thiếu nhiều về số lượng, chất lượng nguồn nhân lực CNTT cũng là trực trạng quan ngại.

Tại nhà máy của LG ở Hải Phòng, kỹ sư CNTT làm việc tại Trung tâm R&D LG phải đào tạo - chuyển giao thêm trong 3 năm, nhân lực đảm nhận công việc liên quan đến thử nghiệm chất lượng hoặc bảo hành sản phẩm phải đào tạo giám sát thêm từ 4 tháng đến một năm, còn công nhân dây chuyền lắp ráp sẽ làm việc được sau khi được đào tạo một tháng.

Ở FPT, kỹ sư CNTT tốt nghiệp đại học cũng phải đào tạo lại một năm.

Để đối phó với thực tế này, Samsung đã đầu tư hơn 38 tỷ đồng hỗ trợ các trường đại học Việt Nam phát triển nguồn nhân lực CNTT chất lượng cao, còn Huawei chỉ 2 triệu USD cho các chương trình xã hội, trong đó chủ yếu tập trung vào các hoạt động hỗ trợ đào tạo và phát triển nguồn nhân lực...

Tại buổi làm việc với Đại học FPT mới đây, Phó thủ tướng Vũ Đức Đam đã chỉ ra rằng, nhu cầu nhân lực ngành CNTT tăng rất cao, trong khi đó, nguồn sinh viên tốt nghiệp ngành này từ các trường ĐH, CĐ còn rất ít, không đảm bảo yêu cầu của doanh nghiệp.

Một nguyên nhân quan trọng được Phó thủ tướng chỉ ra là “môi trường giáo dục đại học chưa đáp ứng được yêu cầu của doanh nghiệp, của nền kinh tế”.

Chia sẻ thực tế này, ông Nguyễn Thành Nam, Phó chủ tịch Hội đồng quản trị ĐH FPT cũng cho hay, ngành CNTT đang phát triển, thay đổi hàng ngày, hàng giờ, vì vậy cần phải có phương án đào tạo hợp lý, trong thời gian ngắn nhất để những người dân bình thường cũng có thể tham gia vào cuộc cách mạng Công nghiệp lần thứ 4.

Khẳng định sẽ tập trung mọi nguồn lực để hỗ trợ, đào tạo, phát triển nguồn nhân lực chuyên ngành điện tử, viễn thông, CNTT và an toàn thông tin, Bộ trưởng Bộ Thông tin và Truyền thông Trương Minh Tuấn cũng cho hay, mục tiêu không chỉ tập trung vào số lượng mà còn đảm bảo được chất lượng chuyên môn, và đặc biệt phải có các kỹ năng mềm cần thiết và khả năng ngoại ngữ tốt để có thể hội nhập sâu hơn với thị trường công nghệ toàn cầu.

Theo [6] “Trong năm 2017, Bộ sẽ đẩy mạnh phối hợp với Bộ Giáo dục và Đào tạo để việc đào tạo nguồn nhân lực Viễn thông, CNTT sát với nhu cầu thị trường và xu hướng phát triển công nghệ trên thế giới, gắn kết chặt chẽ giữa hoạt động đào tạo, nghiên cứu khoa học với hoạt động và nhu cầu của các doanh nghiệp trong lĩnh vực viễn thông, CNTT”, Bộ trưởng Trương Minh Tuấn cho biết.

2.2. Yêu cầu của nhà trường về trợ giúp quyết định

2.2.1. Về ngành nghề lao động

Theo nguồn <https://giasutriviet.edu.vn>, người ta thấy có một số ngành nghề dễ xin việc trong thời đại số hóa.

Th7 01

Những bạn học sinh may mắn có thể vượt qua kỳ thi tuyển sinh Đại học, Cao đẳng luôn có được những sự lựa chọn ngành nghề theo nguyện vọng, ước mơ của bản thân. Vậy những bạn không may mắn thì nên có những sự lựa chọn và định hướng như thế nào để không bỏ lỡ mất con đường đi đến thành công.

Hiện nay vô cùng nhiều ngành nghề, đa dạng về hình thức làm việc trong ngành công nghiệp hiện đại. Nhu cầu phát triển của xã hội ngày càng tăng thì nhu cầu việc làm cũng tăng theo. Vậy trong xã hội hiện nay có những nghề nghiệp nào được xem là hấp dẫn và giúp các bạn sinh viên sau khi tốt nghiệp về chuyên ngành đó có thể dễ dàng xin việc ?

Đầu tiên, nên xét theo lợi nhuận của doanh nghiệp thu về cùng thu nhập cá nhân của ngành nghề đó mang lại. Nhu cầu của con người trong xã hội về ngành đó có cao hay không ? Đặc biệt là những ngành nghề đó là phát triển lâu dài hay nhất thời.



Hình 2.3. Tư vấn việc làm tại Học viện ngân hàng

2.2.1.1. Quản trị kinh doanh

Đây được coi là ngành nghề luôn hấp dẫn cho các bạn sinh viên lựa chọn bởi vì nhu cầu tuyển dụng của các công ty, xí nghiệp, nhà máy... luôn ở tình trạng “Cần nhân viên liên tục”.

Về ngành quá trình kinh doanh, cơ bản khi học ở các trường đại học thì các sinh viên sẽ được học những kiến thức cơ bản và chuyên sâu về “Quản trị”. Có những bạn chọn Quản trị nhân sự, Quản trị tài chính hay có thể là Quản trị doanh nghiệp... là chuyên ngành chính để học phục vụ cho công việc sau này khi ra trường.

Có thể nhìn vào tất cả các Công ty, doanh nghiệp hiện nay họ đều luôn có nhu cầu tuyển dụng về các ngành nghề ở trên để phục vụ cho công tác sản xuất tạo ra lợi nhuận. Họ mong muốn có được những nhân viên làm việc hiệu quả nên sẵn sàng bỏ tiền ra thuê nhân viên với mức lương cao.

Khi học về chuyên ngành này thì các sinh viên nên học được các kỹ năng: Kiến thức nền tảng, chuyên sâu về quản trị doanh nghiệp, quản trị tiếp thị, quản trị tài chính ngân hàng cũng như được học về Quản trị và đàm phán đa văn hóa, dịch vụ cung ứng, Giao tiếp trong kinh doanh, Thương mại điện tử... Để sau này có thể đảm nhận những công việc như: Phòng kinh doanh, phòng kế hoạch, quản lý, điều hành ở các công ty, tập đoàn trong và ngoài nước...

2.2.1.2. Quản trị nhà hàng khách sạn

Ai cũng biết du lịch bây giờ là phát triển với mức khá cao. Các doanh nghiệp luôn tập trung khai thác tài nguyên du lịch một cách tối đa. Vậy nên nguồn lực cung cấp cho ngành này luôn được các nhà tuyển dụng yêu cầu với số lượng lớn.

Các sinh viên khi đã chọn ngành này thì chắc chắn tư tưởng đã nghĩ đến việc không bao giờ ngành này lỗi thời. Yêu cầu cho ngành học này cũng liên quan đến Quản trị nhân sự, Quản lý tài chính... Nhưng ngành này yêu cầu thêm một hình thức nữa đó là ngoại hình, ngoài ra những bạn sinh viên chọn ngành này ít nhất phải có khả năng giao tiếp tốt để sau này có thể giải quyết được tình huống bất ngờ của khách hàng.

Chuyên ngành quá trình này yêu cầu kỹ năng Giao tiếp, đàm phán, xử lý tình huống cũng như quản trị, điều hành nhân lực để đạt hiệu quả cao trong công việc. Sau khi tốt nghiệp các bạn có thể xin việc ở những vị trí sau: Quản lý bộ phận tiếp tân, quản lý phục vụ, quản lý tiền sảnh Nhà hàng, khách sạn, quản lý bộ phận phòng... trong các Nhà hàng, khách sạn cao cấp; hoặc trở thành nhân viên nghiên cứu và quản lý

Những sinh viên có yếu tố năng động, tự tin, có khả năng tổ chức sau này cộng thêm tấm bằng tốt nghiệp loại tốt chắc chắn sẽ có được công việc khá dễ dàng và đạt được kết quả cao trong công việc.

2.2.1.3. Tài chính ngân hàng

Đây một ngành học khá rộng, nó liên quan đến các dịch vụ giao dịch tài chính, vận hành tiền tệ.

Các sinh viên khi theo học ngành này sẽ được học những kiến thức chuyên sâu về lĩnh vực tài chính, phát hành cổ phiếu, trái phiếu, huy động vốn cho doanh nghiệp... Nhu cầu về nhân lực của ngành này ở mức khá cao nên sinh viên sau khi ra trường có thể đảm nhận được những vị trí như: chuyên viên quản lý tiền, tài trợ thương mại, chuyên viên tín dụng, thẩm định... Mức lương khoảng từ 8 – 10 triệu/tháng.

2.2.1.4. Kỹ thuật điện tử

Nhu cầu cuộc sống hiện nay, không thể thiếu các thiết bị điện và điện tử. Cũng vì vậy ngành kỹ thuật là một ngành quan trọng và được khá nhiều các sinh viên theo học.

Ngành này là ngành học nghiên cứu và áp dụng tất cả các vấn đề liên quan đến điện và điện tử. Học những ngành này, sinh viên phải có khả năng thiết kế, xây dựng, vận hành, bảo trì các thiết bị...

Sinh viên sau khi tốt nghiệp các ngành kỹ thuật điện có thể xin việc và đảm nhận các vị trí như: chuyên viên kỹ thuật, tư vấn thiết kế, vận hành, bảo trì mạng lưới điện... tại các doanh nghiệp như: công ty điện lực, khu chế xuất, các khu dân cư...

2.2.1.5. Bác sĩ

Đây là ngành dịch vụ, nó giữ vai trò chủ lực trong việc khám chữa bệnh phục vụ cho người dân, vô cùng quan trọng với xã hội, nếu không có ngành nghề này thì mọi hoạt động hay những khó khăn trong sức khỏe của mỗi người sẽ không được giải quyết và dẫn đến hậu quả xấu.



Hình 2.4. Khám chữa bệnh

Các sinh viên khi chọn ngành nghề này để học thường là những bạn có học lực từ giỏi đến xuất sắc. Các lĩnh vực trong y khoa luôn là những lĩnh vực cần những kỹ năng cũng như kiến thức vô cùng phức tạp và đòi hỏi trình độ tiếp thu nhanh và xử lý tình huống nhạy bén.

Các chuyên ngành trong y khoa không bao giờ là dư thừa đối với nhu cầu

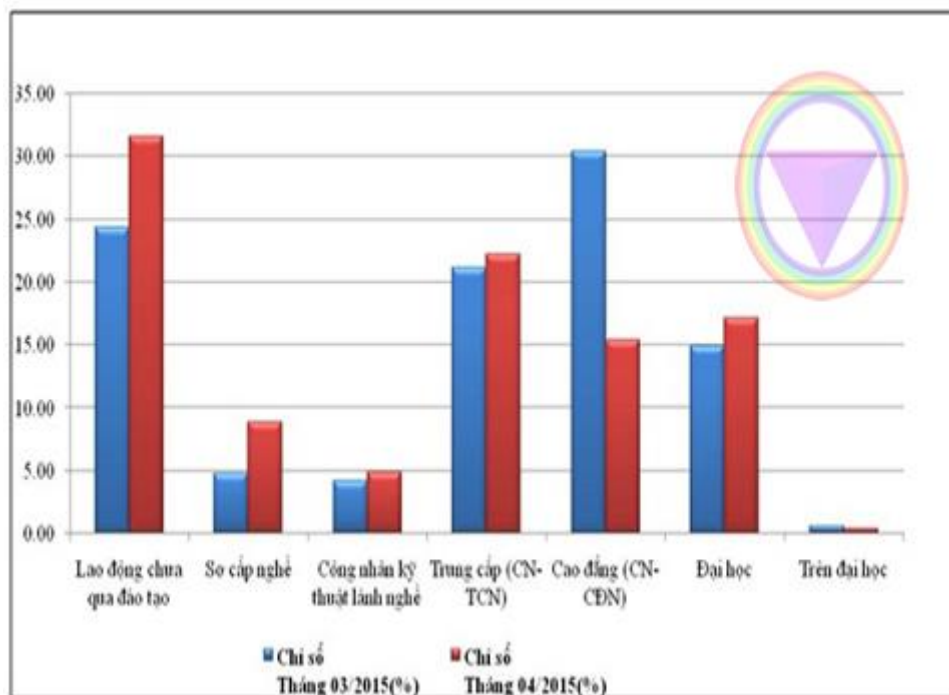
Các sinh viên theo học ngành này đều sẽ được chọn học các chuyên ngành phù hợp với năng lực cũng như sở thích nghiêng về ngành nghề nào. Yêu cầu của ngành nghề này là sinh viên phải tiếp nhận những kiến thức liên quan đến nghiên cứu phát triển, gia công hay ứng dụng hệ thống phần mềm, kiến thức về thiết kế, xây dựng, cài đặt cũng như bảo trì phần cứng và phần mềm.

Các sinh viên sau khi ra trường nếu đảm bảo khả năng thực hành nghề nghiệp tốt và đáp ứng nhu cầu công việc ở từng lĩnh vực cụ thể ở mức thành thạo, kinh nghiệm chuyên môn cao thì sẽ xin việc một cách dễ dàng ở các công ty doanh nghiệp.

2.2.2. Ngành nghề đông người chọn trong mười năm tới

Theo [6] giới trẻ hiện nay đang đối mặt với nhiều thách thức, trong đó thách thức lớn nhất là việc làm. Hàng triệu thanh niên được đào tạo tốt và có năng lực nhưng vẫn thất nghiệp vì chọn nghề không phù hợp.

Theo bản tin cập nhật thị trường lao động Việt Nam của Bộ Lao động thương binh và xã hội đến tháng 10/2015, cả nước có 1.128, 7 nghìn người trong độ tuổi lao động bị thất nghiệp trong đó thanh niên (nhóm từ 15-24 tuổi) bị thất nghiệp chiếm 7, 30%. Đặc biệt, có 225, 5 nghìn người (20 %) lao động trình độ từ đại học trở lên bị thất nghiệp. Qua đó chúng ta thấy rõ thực trạng không phải cứ học Đại học là có được một công việc tốt vì tốt nghiệp Đại học mà vẫn thất nghiệp là một thực trạng khá phổ biến.



Hình 2.6. Nhu cầu tuyển dụng theo trình độ, tháng 3-4 năm 2015

2.2.2.1. Ngành dịch vụ

Việt Nam đã và đang tham gia một cách tích cực vào tiến trình hội nhập quốc tế thông qua việc ký kết các hiệp định đối tác kinh tế với các nước trong khu vực như Hiệp định Đối tác kinh tế chiến lược xuyên Thái Bình Dương (TPP) đã ký kết trong năm 2015 hay Cộng đồng Kinh tế ASEAN (AEC) hình thành vào cuối năm 2015. Theo đó các doanh nghiệp xuất khẩu sẽ mở rộng hoạt động sản xuất kinh doanh, các doanh nghiệp mới sẽ ra đời. Trong xu hướng đó, các ngành thuộc lĩnh vực tiêu dùng nhanh (FMCG), công nghệ thông tin, công nghiệp chế biến và phụ trợ, dệt may, xây dựng, vận tải và nông sản xuất khẩu sẽ hưởng lợi. Những lĩnh vực Cơ khí, xây dựng, công nghệ thông tin, dịch vụ, y tế, du lịch, điện tử, điện - điện công nghiệp - điện lạnh... sẽ là những ngành “hot”. Ứng viên có ngoại ngữ tốt sẽ có nhiều cơ hội lớn hơn, và lĩnh vực giáo dục đào tạo như dạy ngoại ngữ, nhất là tiếng Anh cũng được hưởng lợi từ xu hướng này.

Đổi lại, những nghề làm trong lĩnh vực tài chính – ngân hàng sẽ là đối tượng ít được chú ý đến. Nguyên nhân có lẽ xuất phát từ quá trình mua bán, sáp nhập và tái cấu trúc của ngành. Tuy nhiên, điều này lại kích cầu đối với nhân sự ngành Luật nói chung và Luật kinh tế nói riêng.



Hình 2.7. Các nước khối AEC

AEC sẽ tạo thêm 14 triệu việc làm ở khu vực ASEAN, trong đó Việt Nam chiếm 1/6 tổng lực lượng lao động của khu vực. Có 8 ngành nghề mà lao động có kỹ năng tay nghề cao được phép di chuyển trong khu vực Hiệp định công nhận kỹ năng nghề trong 8 ngành này giữa các nước trong khu vực ASEAN. Đó là, kiểm toán, kiến trúc, kỹ sư, nha sĩ, bác sĩ, y tá, điều tra viên và du lịch. Điểm yếu của lao động Việt Nam chính là thiếu kỹ năng mềm như làm việc theo nhóm, giao tiếp, ngoại ngữ

và thứ hai là kỹ năng nghề nói chung là thấp...

Đặc biệt, một ngành rất mới mẻ ở Việt Nam đang thiếu nhân lực là ngành điện hạt nhân. Theo báo cáo tháng 2/2016 trong hội thảo về phát triển dự án điện hạt nhân tại Ninh Thuận được Thủ tướng Chính phủ phê duyệt tại QĐ số 584/QĐ-TTg ngày 11/4/2013, Việt Nam đang thiếu trầm trọng những nhân lực có chất lượng. Như vậy, có thể nói nhu cầu nhân lực tay nghề cao đang tăng nhanh trong 5 năm tới. Thị trường lao động Việt Nam đang ở thời kỳ cao điểm trong năm nay khi vừa đón một loạt nhân lực mới tham gia vào từ các trường đại học, cao đẳng, trung cấp và các trường nghề trong cả nước. Lực lượng lao động này sẽ phải đối mặt với những thách thức trên. Đồng thời, đội ngũ lao động kế cận sẽ phải lựa chọn hướng đi để phù hợp với xu thế khi gia nhập vào thị trường này trong vài năm tới.

2.2.2.2. *Nhu cầu về tư vấn hướng nghiệp*

Làm thế nào để chọn ngành phù hợp ? là câu hỏi được khá nhiều bạn trẻ tốt nghiệp trung học phổ thông băn khoăn trong khi cũng có nhiều bạn khác chưa hề chuẩn bị hành trang gì để tham gia thị trường lao động trong những năm tới.

Một xu hướng gần đây được nhiều học sinh cuối trung học phổ thông có xu hướng lựa chọn là học đại học và đi du học. Tuy nhiên, trong quá trình chia sẻ, AUM nhận thấy khá nhiều quyết định kiểu chạy theo đám đông vì suy nghĩ đơn giản là học cao thì mới có công việc tốt và thu nhập cao.

Vì vậy nhưng trước khi quyết định chọn ngành nào, học trường nào, nên phân tích kỹ những yếu tố sau:

1. Những biến động của thị trường Lao động Việt Nam hiện tại và trong 5-10 năm;
2. Xác định nhóm công việc mà bạn ưa thích;
3. Lựa chọn 2 đến 3 trường bạn muốn học và phù hợp với năng lực của mình sau đó tìm hiểu kỹ và so sánh về chất lượng đào tạo, cơ sở vật chất, học phí của trường;
4. Ưu tiên những trường, ngành có hợp tác với các tập đoàn lớn như Samsung, Sông Đà, VinGroup... hay những ngành đang được nhà nước quy hoạch để đào tạo cán bộ nguồn để đảm bảo đầu ra.

Cao đẳng là một trong những lựa chọn tốt các bạn nên cân nhắc vì những lợi ích như thời gian đào tạo ngắn và có thể tham gia thị trường lao động sớm hơn; kiến thức về thực hành thực tế nhiều hơn, kinh phí học tập cũng tương đối thấp. Hiện nay đã có một vài trường nắm bắt được xu thế này và tiên phong trong việc hỗ trợ việc làm đầu ra cũng như tạo điều kiện cho sinh viên từ năm thứ 2 thực tập có

lượng, như Trường Cao đẳng Công nghệ và Thương mại Hà Nội (HTT). Trường Cao đẳng Công nghệ và Thương mại Hà Nội có vốn đầu tư nước ngoài từ Nhật và Úc, có hỗ trợ sinh viên ra trường làm việc tại Singapore. Đây là một điểm cộng rất lớn vì trong giai đoạn kinh tế hội nhập việc dịch chuyển nhân lực trong khu vực Đông Nam Á là rất phát triển. Cao đẳng nghề Công Thương Việt Nam cũng là một trong những trường khá cập nhật với xu hướng thị trường lao động tương lai khi mở ra hàng loạt ngành đào tạo dịch vụ có tính thực dụng cao như chăm sóc sắc đẹp, thiết kế tạo mẫu tóc, lắp đặt điện nước, phát triển ứng dụng di động, ...

Với ngành điện hạt nhân, Đại học Đà Lạt, Đại học Khoa học Hà Nội, Đại học Khoa học và Công nghệ Hà Nội là ba trường đang được Bộ GD&ĐT lưu ý để đào tạo cán bộ nguồn cho ngành này. Vì vậy, đây cũng là thời điểm rất tốt để các bạn "gieo" mầm cho nghề nghiệp tương lai. Tỷ lệ có việc làm có thể lên tới 99% nếu bạn đủ năng lực học ngành này bởi có quá ít sự cạnh tranh và quá thiếu đầu vào để lựa chọn.

Có thể nói, chọn trường, chọn ngành là quyết định quan trọng cho cuộc sống tương lai mỗi người. Với mong muốn chia sẻ được những thông tin quan trọng giúp các bạn có cái nhìn đa chiều để xác định con đường tương lai, AUM Việt Nam đã tổng hợp nhu cầu thị trường lao động và phân tích các yếu tố ảnh hưởng ở trên các bạn tham khảo. AUM chúc các bạn có những quyết định sáng suốt để có thể có một công việc phù hợp với sở thích mà vẫn đáp ứng nhu cầu nhân lực hiện tại cũng như 5-10 năm tới của thị trường lao động.

2.3. Định hướng nghề tại cơ sở giáo dục phổ thông

2.3.1. Tư vấn nghề tại cơ sở giáo dục

2.3.1.1. Nhìn nhận chung về công tác hướng nghiệp

Theo <http://www.nhandan.com.vn/giaoduc/item/13791702-.html>, Giáo dục hướng nghiệp là một bộ phận của giáo dục toàn diện giúp mỗi học sinh có sự hiểu biết về tính chất của ngành nghề mà mình hướng tới, biết phân tích thị trường hoạt động và tháo gỡ vướng mắc hoặc rèn luyện bản thân từ đó, mỗi học sinh tự xác định được đâu là nghề nghiệp phù hợp hoặc không phù hợp với mình. Trong hoạt động giáo dục hướng nghiệp ở trường THPT, hoạt động tư vấn nghề có liên quan tới hiệu trưởng, ban hướng nghiệp, giáo viên chủ nhiệm, giáo viên bộ môn, thư viện, y tế...

Hiệu trưởng là người phụ trách chung về các hoạt động hướng nghiệp trong nhà trường, trong đó có hoạt động tư vấn. Hiệu trưởng có trách nhiệm thông qua và ký các quyết định về kế hoạch tiến hành các hoạt động tư vấn trong và ngoài trường. Ban hướng nghiệp chịu trách nhiệm thu thập xử lý những thông tin do các

bộ phận cung cấp, đưa ra những nhận định, đánh giá sơ bộ về xu hướng nghề của học sinh. Những thông tin sau xử lý do ban hướng nghiệp thực hiện sẽ là những tài liệu bổ ích cho cán bộ làm công tác tư vấn khi tiến hành hoạt động này, làm cho nội dung tư vấn có tính sát thực, đáp ứng đúng nhu cầu định hướng nghề của đối tượng tư vấn.

Ban hướng nghiệp còn chịu trách nhiệm tham mưu cho hiệu trưởng đề xuất kế hoạch và tổ chức các hoạt động tư vấn về nhân lực, cơ sở vật chất, ... phù hợp kế hoạch năm học của nhà trường trên từng loại đối tượng cụ thể. Giáo viên bộ môn thu thập và cung cấp những thông tin có liên quan thái độ, năng lực học tập của từng học sinh đối với những môn học cụ thể. Giáo viên chủ nhiệm cung cấp những thông tin phản ánh trình độ nhận thức xã hội, phẩm chất đạo đức, kỹ năng hòa nhập với cộng đồng của mỗi học sinh do mình phụ trách. Mỗi giáo viên chủ nhiệm còn có trách nhiệm tập hợp những thông tin do những bộ phận khác cung cấp để thiết lập các phiếu đánh giá về xu hướng nghề đối với từng học sinh trong lớp làm cơ sở cho hoạt động tư vấn. Đoàn Thanh niên thu thập và cung cấp những thông tin về năng lực hoạt động xã hội, tập thể, về ý thức, thái độ, lối sống của mỗi thành viên trong tổ chức.

Đáng chú ý, học sinh là đối tượng của hoạt động tư vấn đồng thời là chủ thể của quá trình tiếp nhận thông tin nghề do hoạt động tư vấn mang lại học sinh không chỉ có nhiệm vụ tiếp thu thông tin do chủ thể tư vấn cung cấp mà cùng với nó là quá trình lựa chọn những thông tin hữu ích phù hợp với năng lực, sở trường, tình trạng sức khỏe và nhu cầu lựa chọn nghề nghiệp của bản thân.



Hình 2.8. Định hướng nghề tại cơ sở giáo dục

Vì vậy, công tác hướng nghiệp cần giúp cho học sinh hiểu được hệ thống nghề nghiệp trong xã hội, phương hướng phát triển nền kinh tế quốc dân nói chung và của địa phương nói riêng nhằm xác định cho bản thân trách nhiệm, nghĩa vụ sẵn sàng tham gia vào lao động sản xuất. Trên cơ sở của sự hiểu biết nghề nghiệp và nền kinh tế quốc dân, của địa phương, những đòi hỏi khách quan của hoàn cảnh, biết đối chiếu với sự phát triển, năng lực, sở trường, tình trạng tâm sinh lý sức khỏe của bản thân để điều chỉnh động cơ lựa chọn nghề. Tạo ra những điều kiện cần thiết về cơ sở vật chất, về các mối quan hệ xã hội và ý thức cầu tiến bộ của học sinh để các em tích cực tham gia các hình thức lao động kỹ thuật do nhà trường tổ chức, nâng cao ý thức và thái độ lao động, có dịp thử sức mình trong hoàn cảnh thực tiễn, từ đó kết luận về sự phù hợp nghề nghiệp của bản thân. Phải làm cho mỗi học sinh có được tính chủ động trong lựa chọn nghề, có khả năng tự quyết định được con đường nghề nghiệp tương lai của mình.

2.3.1.2. Vai trò của giáo viên chủ nhiệm

Có thể nói, công tác hướng nghiệp nhằm phân luồng học sinh sau tốt nghiệp và định hướng nghề nghiệp cho các em rất quan trọng, đặc biệt là ở bậc THPT. Vậy, người giáo viên nhất là giáo viên chủ nhiệm (GVCN) có vai trò hết sức quan trọng trong việc định hướng nghề nghiệp của các em học sinh

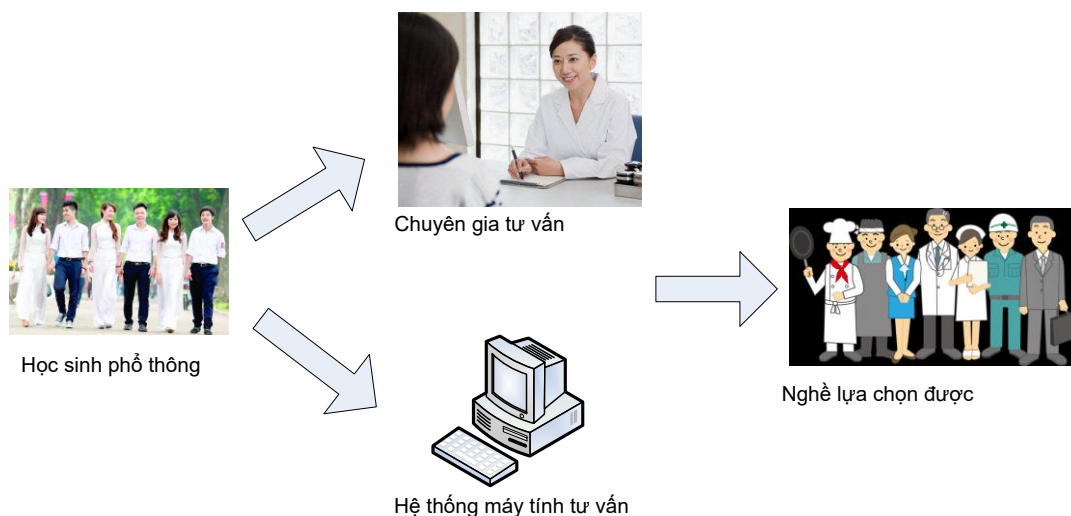
Theo thống kê của Bộ LĐ-TB&XH, có khoảng 60% sinh viên ra trường làm việc trái với ngành đã được đào tạo, trong số đó có 47, 6% làm trái ngành vì không thích ngành mình đã học, điều này chứng tỏ công tác hướng nghiệp đã bị “bỏ quên” hoặc không mang lại hiệu quả như mong đợi. Đã có không ít những chương trình hướng nghiệp cho học sinh như tư vấn tại sân trường, tư vấn online hay phòng tư vấn học đường... nhưng có lẽ, hiệu quả của việc định hướng nghề nghiệp chưa cao, đôi khi các chương trình đang có sự lẫn lộn giữa tư vấn hướng nghiệp và tư vấn tuyển sinh, dẫn đến hệ quả là các em học sinh nghe tư vấn xong vẫn không biết mình phù hợp với ngành nào, nghề nào.

Bên cạnh các chương trình tư vấn hướng nghiệp được tổ chức hàng năm, GVCN có vai trò rất quan trọng trong việc định hướng nghề nghiệp cho học sinh. Hơn ai hết, GVCN là người tiếp xúc với các em hàng ngày, biết được khá rõ lực học của các em, biết được ít nhiều tính cách, hoàn cảnh gia đình của mỗi em, do đó, họ là những người sâu sát các em nhất. Nhưng trong thực tế GVCN không được đào tạo để hướng nghiệp, cũng không được học một khóa tập huấn hướng nghiệp nào bài bản? Kênh nào là kênh chính thống trong hướng nghiệp để GVCN có thể tham khảo cũng như giới thiệu đến phụ huynh và học sinh? Câu hỏi này có lẽ cần sự trợ

giúp của nhiều cấp và nhiều ngành trong tương lai để giảm tình trạng học sinh học nhầm ngành, người lao động làm nhầm nghề...

2.3.2. Tư vấn nhờ hệ thống công nghệ thông tin

Với sự cần thiết của việc quyết định lựa chọn nghề nghiệp của các em học sinh cùng với sự phát triển, sự cần thiết của công nghệ thông tin trong các lĩnh vực hiện nay, việc ra quyết định nhờ máy tính sẽ trợ giúp rất nhiều trong việc đưa ra những quyết định chính xác, tiết kiệm thời gian trong công việc và là xu hướng phát triển hiện nay.



Hình 2.9. Tư vấn nhờ máy tính

Khi cơ sở giáo dục phổ thông sử dụng hệ thống tư vấn với trợ giúp của công nghệ thông tin, người ta thấy có đánh giá sau :

- Hệ thống máy tính cho phép tư vấn tự động, tiện cho bất kì thời gian rảnh rỗi nào của các em học sinh;
- Việc tư vấn với trợ giúp của công nghệ thông tin sẽ thực hiện thuật toán như nhau; chỉ yêu cầu thông tin vừa đủ;
- Tuy nhiên, hệ thống công nghệ thông tin không được thân thiện như chuyên gia tư vấn. Điều này đòi hỏi hệ thống tư vấn cần cải thiện về giao diện người dùng;
- Hệ thống tư vấn với máy tính làm việc tại bất kì nơi đâu; trong khi phòng tư vấn của nhà trường chỉ hoạt động trong thời gian định sẵn.

Việc sử dụng hệ thống tư vấn với trợ giúp của công nghệ thông tin, theo mô hình trợ giúp quyết định là xu thế trong thời đại số hóa. Đặc biệt với công nghệ 4.0, người ta có thể sử dụng các sản phẩm đồ họa, để đưa ra nhiều khả năng lựa chọn cho các em học sinh.

2.4. Kết luận

Chương này đã nêu tình hình chung và thực trạng của việc lựa chọn nghề nghiệp và định hướng tương lai cho của học sinh. Một số ngành nghề hấp dẫn thanh niên được đề cập trong chương.

Trước những dẫn đo và vướng mắc trong khi lựa chọn học gì và làm gì cho tương lai, hệ thống trợ giúp quyết định sử dụng máy tính là giải pháp phù hợp với thời đại mạng máy tính và thông tin rộng khắp. Ứng dụng công nghệ thông tin và truyền thông trong việc trợ giúp quyết định nghề nghiệp cho học sinh là hết sức cần thiết.

Chương 3

THỬ NGHIỆM TRỢ GIÚP QUYẾT ĐỊNH LỰA CHỌN NGHỀ NGHIỆP TẠI TRƯỜNG TRUNG HỌC PHỔ THÔNG

3.1. Đặt vấn đề

Dựa trên tình hình cũng như số liệu về học lực, cơ hội việc làm, nhu cầu của địa phương, ứng dụng quá trình ra quyết định vào việc lựa chọn nghề.

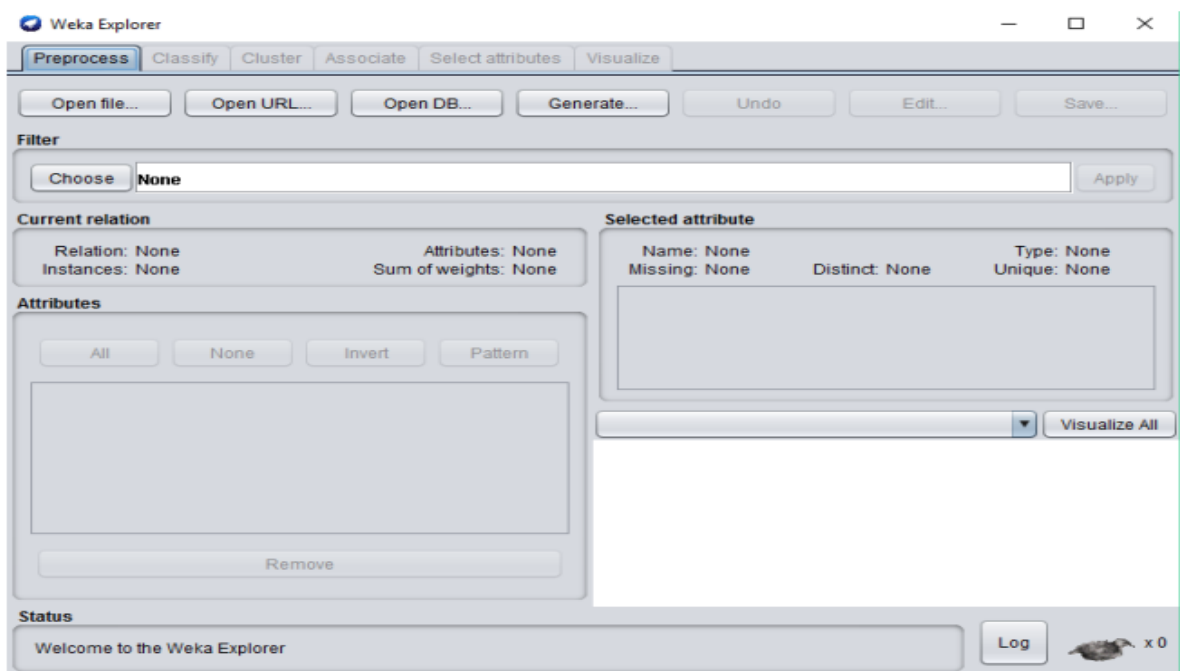
Trên hệ thống thông tin với cơ sở dữ liệu vật tư đã được tạo nên, học sinh có thể ra các quyết định chọn nghề nghiệp cho phù hợp.

Để giải quyết vấn đề này, em sẽ xây dựng mô hình phân lớp (Classification Model) bằng cây quyết định trong weka.

3.1.1. Tổng quan về phần mềm Weka

Weka – Waikato Environment for Knowledge Analysis, là bộ phần mềm học máy, mã nguồn mở, do đại học Waikato phát triển bằng Java, nhằm phục vụ cho các nhiệm vụ chuyên về khai phá dữ liệu.

Weka chứa các công cụ phục vụ cho tiền xử lý dữ liệu, phân loại, hồi quy, phân cụm, các luật liên quan và trực quan hóa. Nó cũng phù hợp cho việc phát triển, xây dựng các mô hình học máy và có khả năng chạy được trên nhiều hệ điều hành khác nhau như Windows, Mac, Linux.



Hình 3.1. Công cụ của Weka

3.1.1.1. Các tính năng chính

Những tính năng vượt trội trong Weka có thể kể đến là:

- Mã nguồn mở;
- Hỗ trợ các thuật toán học máy (**machine learning**) và khai phá dữ liệu;
- Trực quan hóa, dễ dàng xây dựng các ứng dụng thực nghiệm;
- Do sử dụng JVM nên Weka độc lập với môi trường.

Kiến trúc trong thư viện Weka bao gồm hơn 600 class và được tổ chức thành 10 package. Chính vì thế, người sử dụng có thể dùng trực tiếp trên phần mềm hoặc sử dụng những class này làm bộ thư viện để phát triển các ứng dụng của riêng mình.



Hình 3.2. Môi trường chính của Weka

Môi trường chính của Weka bao gồm:

1. Explorer: cho phép sử dụng các tính năng của Weka để khai phá dữ liệu;
2. Experimenter: cho phép tiến hành các thí nghiệm và thực hiện các bài kiểm tra thống kê giữa các mô hình học máy;
3. KnowledgeFlow: môi trường cho phép tương tác đồ họa để thiết kế các thành phần của một thí nghiệm;
4. SimpleCLI: giao diện dòng lệnh.

3.1.1.2 Định dạng dữ liệu

Định dạng dữ liệu đầu vào thông dụng trong Weka là định dạng **.arff** và **.csv**

```
@relation bank-data-final
@attribute age {0_34,35_51,52_max}
@attribute sex {FEMALE,MALE}
@attribute region {INNER_CITY,TOWN,RURAL,SUBURBAN}
@attribute income {0_24386,24387_43758,43759_max}
@attribute married {NO,YES}
@attribute children {0,1,2,3}
@attribute car {NO,YES}
@attribute save_act {NO,YES}
@attribute current_act {NO,YES}
@attribute mortgage {NO,YES}
@attribute pep {YES,NO}

@data
35_51,FEMALE,INNER_CITY,0_24386,NO,1,NO,NO,NO,NO,YES
35_51,MALE,TOWN,24387_43758,YES,3,YES,NO,YES,YES,NO
52_max,FEMALE,INNER_CITY,0_24386,YES,0,YES,YES,YES,NO,NO
0_34,FEMALE,TOWN,0_24386,YES,3,NO,NO,YES,NO,NO
52_max,FEMALE,RURAL,43759_max,YES,0,NO,YES,NO,NO,NO
52_max,FEMALE,TOWN,24387_43758,YES,2,NO,YES,YES,NO,YES
0_34,MALE,RURAL,0_24386,NO,0,NO,NO,YES,NO,YES
52_max,MALE,TOWN,24387_43758,YES,0,YES,YES,YES,NO,NO
35_51,FEMALE,SUBURBAN,24387_43758,YES,2,YES,NO,NO,NO,NO
52_max,MALE,TOWN,0_24386,YES,2,YES,YES,YES,NO,NO
52_max,FEMALE,TOWN,43759_max,YES,0,NO,YES,YES,NO,NO
52_max,FEMALE,INNER_CITY,24387_43758,NO,0,YES,YES,YES,YES,NO
35_51,FEMALE,TOWN,0_24386,YES,1,NO,YES,YES,YES,YES
```

Hình 3.3. Dữ liệu đầu vào thông dụng trong weka

Tuy nhiên, với những định dạng khác, để có thể sử dụng trong Weka, chúng ta cần phải thực hiện việc chuyển đổi dữ liệu sang định dạng **.csv** hoặc những định dạng mà Weka có thể xử lý được.

Với những tính năng vượt trội và mạnh mẽ, đặc biệt là phần mềm mã nguồn mở, giờ đây người dùng có thể nghiên cứu các giải thuật trong khai phá dữ liệu hay học máy một cách dễ dàng và trực quan hơn, đặc biệt là những người mới tập tành nghiên cứu, tìm hiểu lĩnh vực này. Ngoài ra, với những chuyên gia, chúng ta có thể sử dụng các class có sẵn để làm thư viện và tùy biến, chỉnh sửa thành những chức năng phục vụ cho nghiên cứu chuyên sâu.

3.1.2. Xây dựng dữ liệu đưa vào phần mềm

Dữ liệu được dùng trong ví dụ này là file nghe.CSV thực hiện trên Excel có 5 thuộc tính Học lực (HL), Nhu cầu xã hội (nhucauXH), Sở thích (sothich), Năng khiếu (nangkhiếu), Điều kiện gia đình (Dieukien) và thuộc tính phân loại là ngành nghề lựa chọn (NNLC) có 10 giá trị:

Bảng 3.1. Nghề lựa chọn thử nghiệm của luận văn

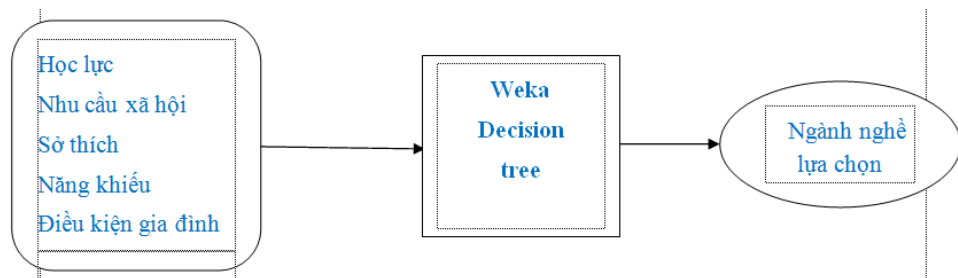
STT	Nghề lựa chọn	STT	Nghề lựa chọn
1	Công an (Congan)	6	Giáo viên (Giaovien)
2	Bồ đội (Bodoi)	7	Công nhân (Congnhan)
3	Bác sĩ (Bacsi)	8	Công nghệ thông tin (CNTT)
4	Ca sĩ (Casi)	9	NV bán hàng (Nvbanhang)
5	Nghệ thuật (Nghethuat)	10	NV văn phòng (Nvvanphong)

Bảng dữ liệu huấn luyện xây dựng trên file Excel có tên là Nghe.CSV:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	HL	NhucauXH	Sothich	Nangkhieu	Dieukiengiadinh	Lcnghe							
2	G	Cao	no	Khong	K	Congan							
3	G	Cao	yes	Khong	K	Bodoi							
4	G	Cao	yes	Khong	C	Bacsi							
5	TB	trungbinh	yes	Co	K	Casi							
6	TB	trungbinh	yes	Co	K	Nghethuat							
7	G	Thap	yes	Co	C	giaovien							
8	TB	Thap	yes	Co	K	Congnhan							
9	K	Cao	no	Co	K	CNTT							
10	K	Thap	yes	Co	K	Nvbanhang							
11	G	Cao	yes	Co	C	Congan							
12	G	Cao	yes	Co	C	Bodoi							
13	G	Cao	yes	Co	C	Bacsi							
14	G	Cao	no	Co	C	Nghethuat							
15	G	trungbinh	yes	Co	C	Nvvanphong							
16	TB	Cao	yes	Khong	C	Nvbanhang							
17	G	Cao	no	Khong	K	Congan							
18	G	Cao	yes	Khong	K	Bodoi							
19	G	Cao	yes	Khong	C	Bacsi							
20	TB	trungbinh	yes	Co	K	Casi							
21	TB	trungbinh	yes	Co	K	Nghethuat							
22	G	Thap	yes	Co	C	giaovien							
23	TB	Thap	yes	Co	K	Congnhan							
24	K	Cao	no	Co	K	CNTT							
25	K	Thap	yes	Co	K	Nvbanhang							
26	G	Cao	yes	Co	C	Congan							
27	G	Cao	yes	Co	C	Bodoi							
28	G	Cao	yes	Co	C	Bacsi							
29	G	Cao	no	Co	C	Nghethuat							
30	G	trungbinh	yes	Co	C	Nvvanphong							
31	TB	Cao	yes	Khong	C	Nvbanhang							
32	G	Cao	no	Khong	K	Congan							
33	G	Cao	no	Khong	K	Bodoi							
34	G	Cao	yes	Khong	C	Bacsi							
35	TB	trungbinh	yes	Co	K	Casi							
36	TB	trungbinh	yes	Co	K	Nghethuat							
37	G	Thap	yes	Co	C	giaovien							
38	TB	Thap	yes	Co	K	Congnhan							
39	K	Cao	no	Co	K	CNTT							
40	K	Thap	yes	Co	K	Nvbanhang							
41	G	Cao	yes	Co	C	Congan							

Hình 3.4. Dữ liệu huấn luyện để đưa vào Weka

Bằng cách sử dụng phần mềm WEKA để thực thi một cây quyết định, chúng ta có thể xây dựng một công cụ trợ giúp được các yêu cầu trên dựa vào những dữ liệu khách quan thu thập trước đó.



Hình 3.5. Lược đồ thử nghiệm trong luận văn

3.2. Chuẩn bị dữ liệu thử nghiệm

Trong quá trình khai phá dữ liệu, công việc xử lý dữ liệu trước khi đưa vào các mô hình là rất cần thiết, bước này làm cho dữ liệu có được ban đầu qua thu thập dữ liệu (gọi là dữ liệu gốc ordinal data) có thể áp dụng được (thích hợp) với các mô hình khai phá dữ liệu (data mining model) cụ thể. Các công việc cụ thể của tiền xử lý dữ liệu bao gồm những công việc như:

1. Filtering Attributes: Chọn các thuộc tính phù hợp với mô hình;
2. Filtering samples: Lọc các mẫu (instances, patterns) dữ liệu cho mô hình;
3. Clean data: Làm sạch dữ liệu như xóa bỏ các dữ liệu bất thường (Outlier);
4. Transformation: Chuyển đổi dữ liệu cho phù hợp với các mô hình như chuyển đổi dữ liệu từ numeric qua nominal hay ordinal;
5. Discretization (rời rạc hóa dữ liệu): Nếu bạn có dữ liệu liên tục nhưng một vài mô hình chỉ áp dụng cho các dữ liệu rời rạc (như luật kết hợp chẵn lẻ) thì bạn phải thực hiện việc rời rạc hóa dữ liệu.

3.2.1 Trích chọn thuộc tính

Trích chọn thuộc tính là nhiệm vụ rất quan trọng giai đoạn tiền xử lý dữ liệu khi triển khai các mô hình khai phá dữ liệu. Một vấn đề gặp phải là các dataset dùng để xây dựng các Data mining Models thường chứa nhiều thông tin không cần thiết cho việc xây dựng mô hình. Chẳng hạn, một dataset gồm hàng trăm thuộc tính dùng để mô tả về khách hàng của một doanh nghiệp được thu thập, tuy nhiên khi xây dựng một Data mining model nào đó chỉ cần khoảng 50 thuộc tính từ hàng trăm thuộc tính đó. Nếu ta sử dụng tất cả các thuộc tính (hàng trăm) của khách hàng để xây dựng mô hình thì ta cần nhiều CPU, nhiều bộ nhớ trong quá trình Training model, thậm chí các thuộc tính không cần thiết

đó làm giảm độ chính xác của mô hình và gây khó khăn trong việc phát hiện tri thức.

Các phương pháp trích chọn thuộc tính thường tính trọng số (score) của các thuộc tính và sau đó chỉ chọn các thuộc tính có trọng số tốt nhất để sử dụng cho mô hình. Các phương pháp này cho phép bạn hiệu chỉnh ngưỡng (threshold) để lấy ra các thuộc tính có Score trên ngưỡng cho phép. Quá trình trích chọn thuộc tính luôn được thực hiện trước quá trình Training Model.

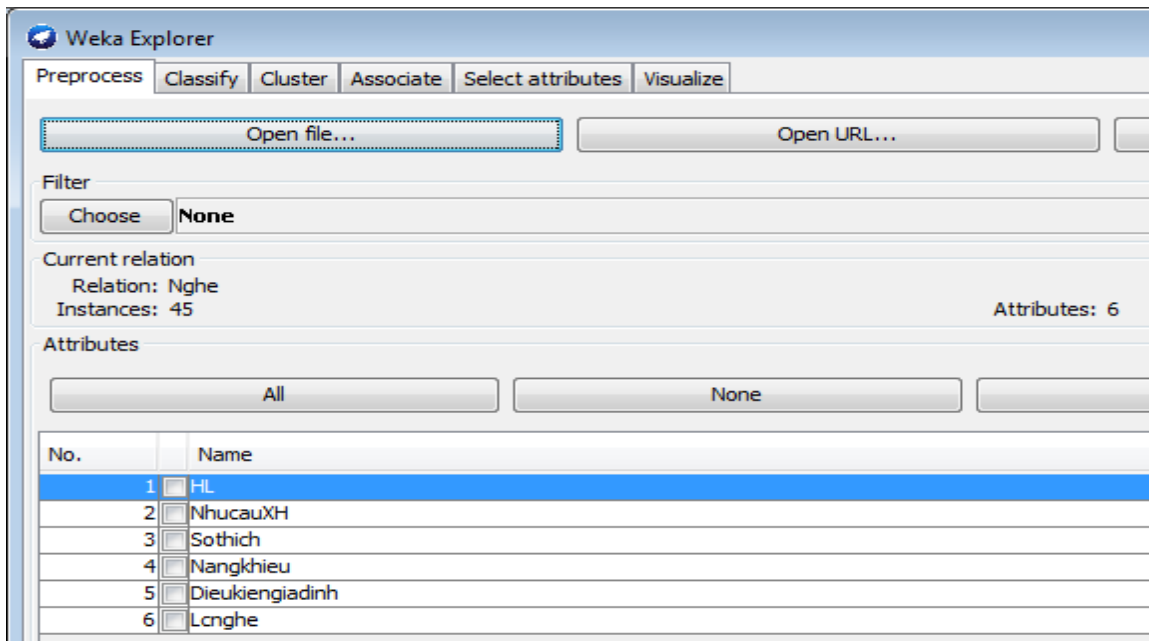
Có rất nhiều phương pháp để lựa chọn thuộc tính tùy thuộc vào cấu trúc của dữ liệu dùng cho mô hình và thuật toán được dùng để xây dựng mô hình. Sau đây là một số phương pháp phổ biến dùng trong trích chọn thuộc tính:

1. Interestingness score: Được sử dụng để xếp hạng (rank) các thuộc tính đối với các thuộc tính có kiểu dữ liệu liên tục (continuous).
2. Shannon's Entropy: Được sử dụng đối với các dữ liệu kiểu rời rạc (discretized data).
3. Ngoài ra còn có một số phương pháp khác cũng thường được sử dụng trong lựa chọn thuộc tính như Bayesian with K2 Prior, Bayesian Dirichlet Equivalent with Uniform Prior.

3.2.2 Trích chọn thuộc tính bằng phần mềm Weka

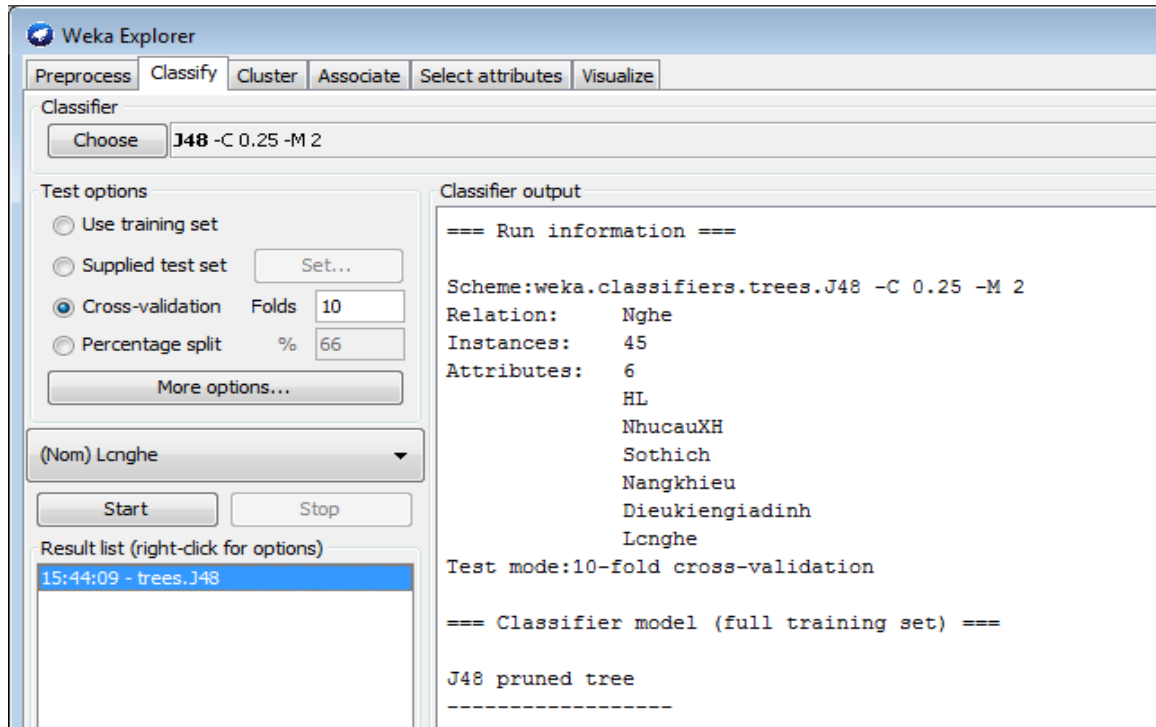
Dataset được dùng là file định dạng của weka nghe.csv

Khởi động Weka > Chọn Explorer > Chọn Open file > Chọn Dataset “nghe.csv“ kết quả như sau:



Hình 3.6: Trích chọn thuộc tính bằng phần mềm Weka

- Chọn Tab “Select attributes”.
- Trong mục Attribute Evaluator chọn WrapperSubsetEval.
- Trong mục Search Method chọn GreedyStepwise .
- Chọn Tab “Classify”: Trong mục classifier chọn tree, chọn j48
- *Bấm Start để thực hiện, kết quả như sau:*



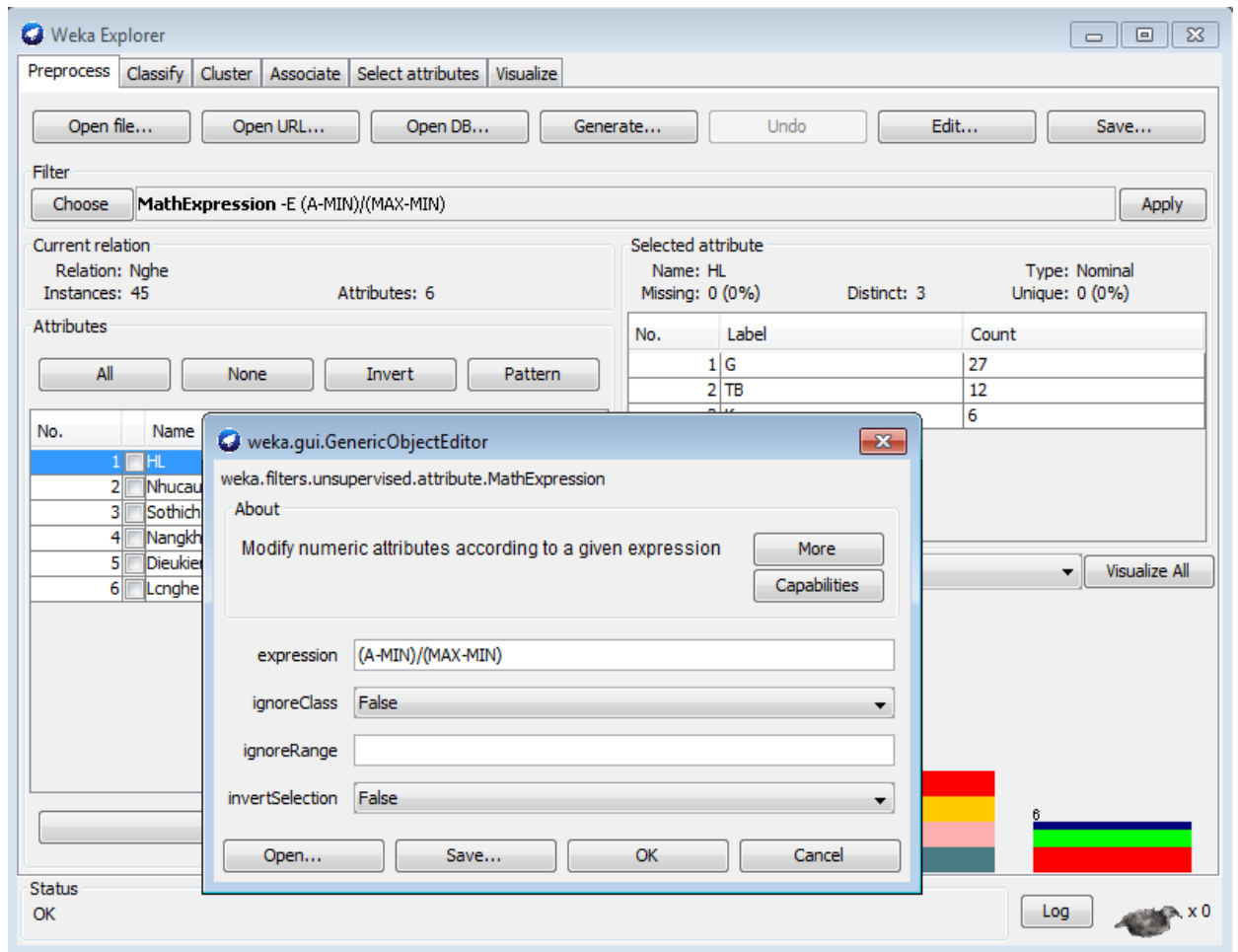
Hình 3.7: Lựa chọn thuật toán để chuẩn hóa dữ liệu bằng Weka

Vậy 5 thuộc tính được chọn đó là: HL, NhucauXH, Sothich, Nangkhieu, Dieukiengiadin, Lnghe (Học lực, nhu cầu xã hội, sở thích, năng khiếu, năng khiếu, điều kiện gia đình, lựa chọn nghề).

3.2.3 Chuẩn hóa dữ liệu

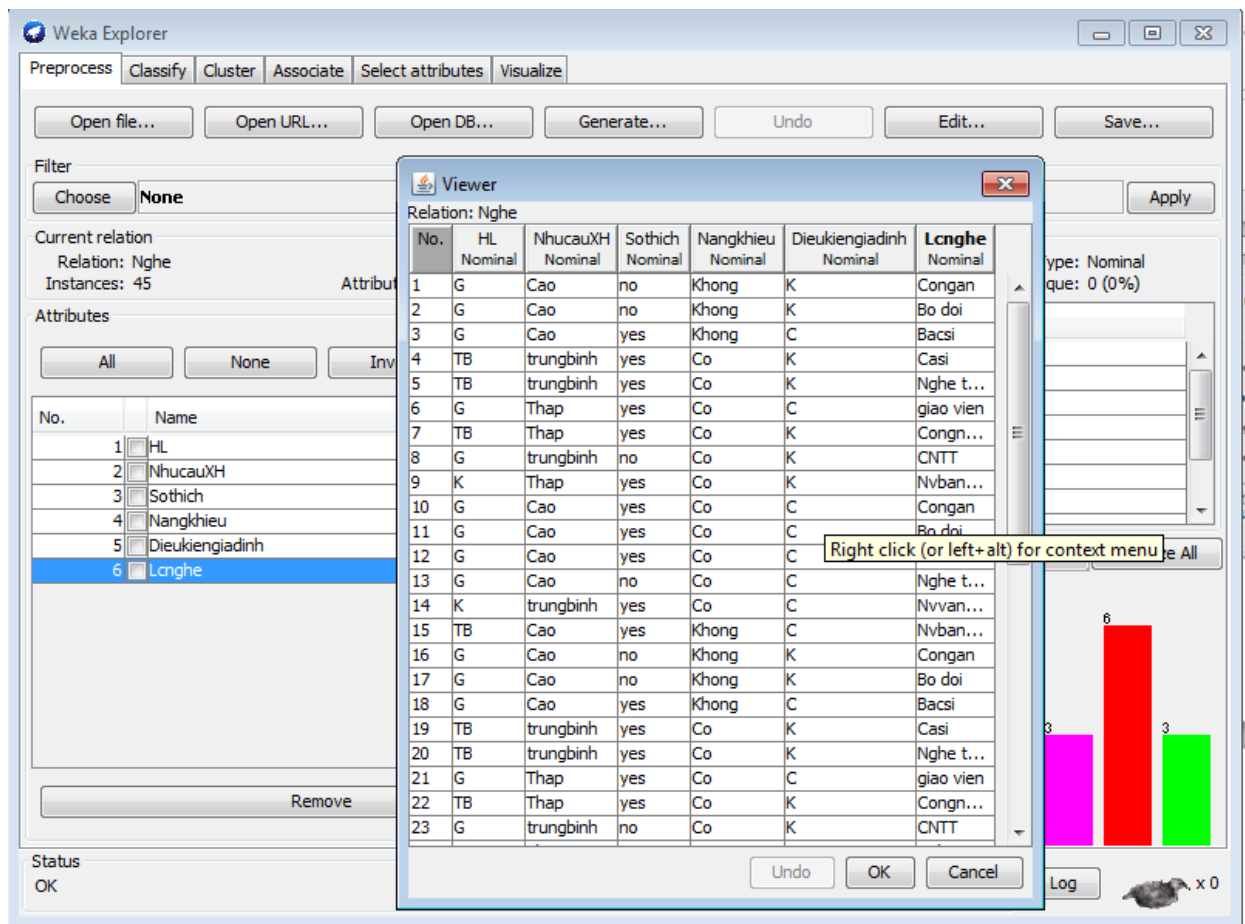
Luận văn thực hiện :

- Chuyển kiểu dữ liệu của thuộc tính **HL** thành kiểu Nominal với các giá trị tương ứng G (giỏi), K (Khá), Tb (trung bình).
- Trong Filter chọn Choose trong cây thư mục chọn MathExpression.



Hình 3.7: Chuẩn dữ liệu cho các thuộc tính

- Nhập biểu thức lọc vào nhấn OK để chấp nhận. Chọn Apply để áp đổi với từng thuộc tính.
- Sau khi tiến hành quá trình Preprocess ta thu được bảng dữ liệu chỉ toàn kiểu Nominal như sau:



Hình 3.4: Bảng dữ liệu sau khi đã chuẩn hóa

3.2.4 Lựa chọn thuật toán J48 để xây dựng cây quyết định

3.2.4.1 Cây quyết định trong thử nghiệm

Giới thiệu về cây quyết định như đã giới thiệu trong chương 1: Cây quyết định (Decision Tree) là một cây phân cấp có cấu trúc được dùng để phân lớp các đối tượng dựa vào dãy các luật (series of rules). Các thuộc tính của đối tượng (ngoại trừ thuộc tính phân lớp – Category attribute) có thể thuộc các kiểu dữ liệu khác nhau (Binary, Nominal, ordinal, quantitative values) trong khi đó thuộc tính phân lớp phải có kiểu dữ liệu là Binary hoặc Ordinal. Tóm lại, cho dữ liệu về các đối tượng gồm các thuộc tính cùng với lớp (classes) của nó, cây quyết định sẽ sinh ra các luật để dự đoán lớp của các đối tượng chưa biết (unseen data).

3.2.4.2. Quy trình huấn luyện và thử bộ phân loại

Một số qui định thử nghiệm :

- Dữ liệu để xây dựng mô hình: dữ liệu gốc (original dataset), dữ

liệu này phải có thuộc tính phân lớp gọi là thuộc tính phạm trù.

- Dữ liệu gốc sẽ được chia thành hai phần là (i) tập huấn luyện, để xây dựng mô hình, và (ii) tập thử nghiệm, để kiểm định mô hình;
- Cuối cùng là tính toán lỗi để đánh giá mô hình.

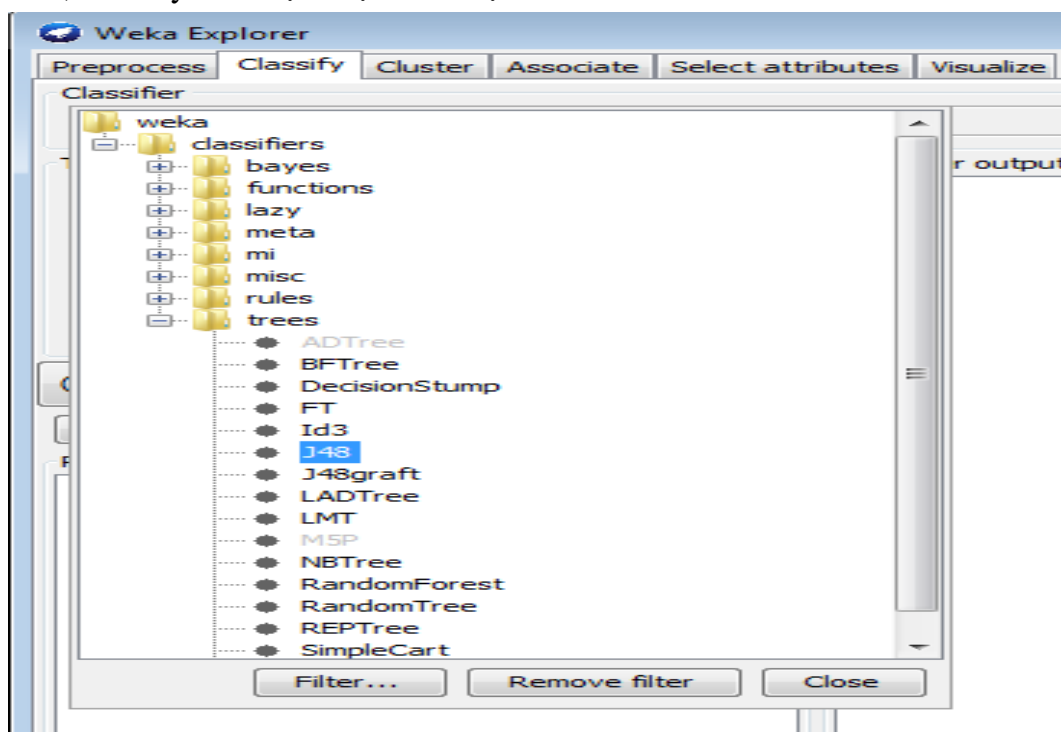
3.2.4.3. Lí do sử dụng thuật toán này

Có rất nhiều thuật toán phân lớp như ID3, J48, C4.5, CART (Classification and Regression Tree), ... Việc chọn thuật toán nào để có hiệu quả phân lớp cao tuy thuộc vào rất nhiều yếu tố, trong đó cấu trúc dữ liệu ảnh hưởng rất lớn đến kết quả của các thuật toán.

Với thuật toán ID3 và CART cho hiệu quả phân lớp rất cao đối với các trường dữ liệu số (quantitative value) trong khi đó các thuật toán như J48, C4.5 có hiệu quả hơn đối với các dữ liệu Qualitative value (ordinal, Binary, nominal). Sau khi đã chuẩn hóa dữ liệu thì được bảng dữ liệu chỉ toàn kiểu Nominal, vì vậy ta sử dụng thuật toán J48 để đạt hiệu quả phân lớp cao.

3.2.4.4. Sử dụng thuật toán với phần mềm Weka

Nhấn vào tab Classify chọn thuật toán sử dụng bằng cách nhấn vào nút Choose; khi cây thư mục hiện thư mục Trees/J48:



Hình 3.8: Lựa chọn thuật toán J48 với phần mềm Weka

3.2.5 Đánh giá hiệu quả phân lớp của thuật toán

Đánh giá hiệu quả phân lớp của thuật toán đối với tập dữ liệu được cho theo hai phương pháp sau.

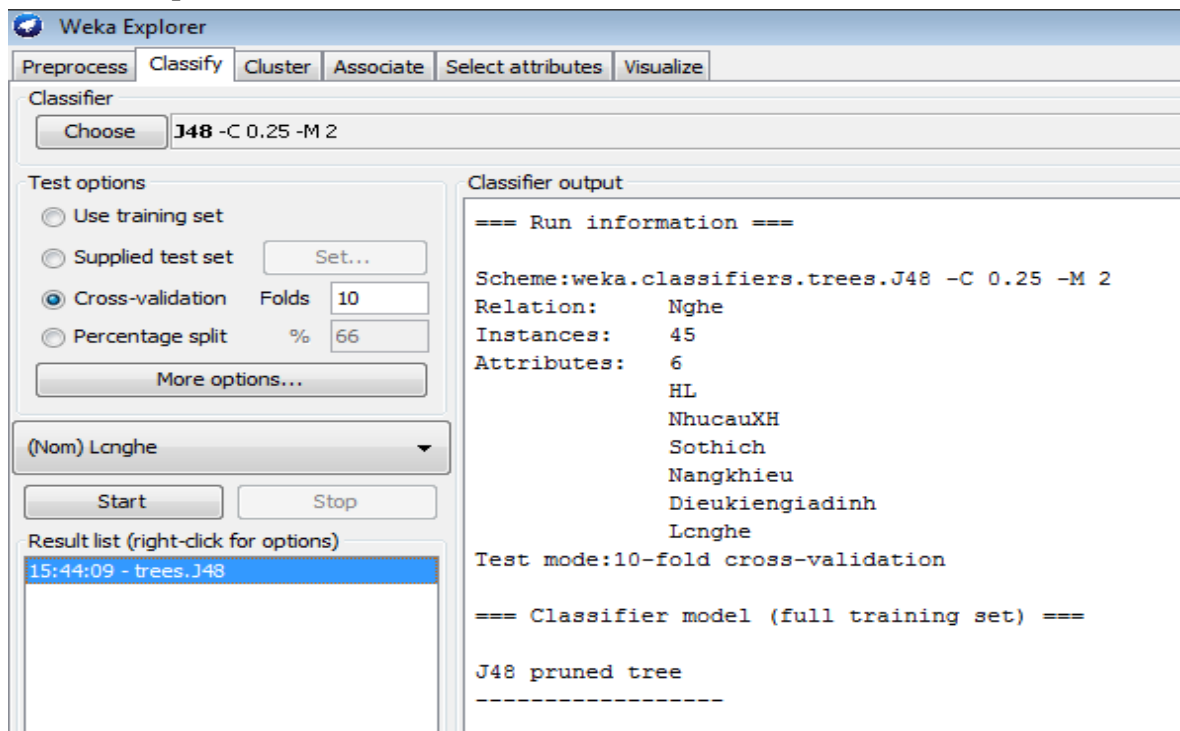
3.2.5.1 Phương pháp Cross-validation

Tập dữ liệu sẽ được chia đều thành k tập (folds) có kích thước xấp xỉ nhau, và bộ phân loại học được sẽ được đánh giá bởi phương pháp *cross-validation*.

Đây là kỹ thuật chủ yếu được sử dụng trong xây dựng predictive Model. Trong đó dữ liệu gốc sẽ được chia thành n phần bằng nhau (n-fold), và quá trình Train/Test Model thực hiện lặp lại n lần. Tại mỗi lần Train/Test Model, 1 phần dữ liệu dùng để Thử và (n-1) phần còn lại dùng để Train. (Người ta đã chứng minh 10-fold Cross –Validation là tối ưu)

Với phương pháp này ta thu được kết quả hiển thị ở khung Classify.

Kết quả đầu ra như sau:



Hình 3.9: Đánh giá hiệu quả của thuật toán bằng phương pháp cross-validation

Nội dung kết quả: Kết quả được trả về theo ba vùng dữ liệu:

- Vùng Run Information: Cho biết thông tin về dữ liệu nguồn.
- Đề án sử dụng: weka.classifiers.trees.J48 -C 0.25 -M 2
- Cơ sở dữ liệu: Nghe

- Số trường: 45
- Số thuộc tính: 6
- Chế độ kiểm tra: 10-fold cross-validation

```

=== Run information ===
Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:  Nghe
Instances:  45
Attributes:  6
HL
NhucauXH
Sothich
Nangkhieu
Dieukiengiadinh
Lcnghe
Test mode:10-fold cross-validation
=== Classifier model (full training set) ===

```

Vùng hiển thị kết quả training:

- Chế độ phân lớp: toàn bộ dữ liệu
- Cây J48 sau khi tiến hành training:
- Số lượng lá: 12
- Kích thước cây: 19
- Thời gian tiến hành: 0.01 giây
-

```

J48 pruned tree
-----
Number of Leaves :      12
Size of the tree :     19
Time taken to build model: 0.01 seconds
=== Stratified cross-validation ===

```

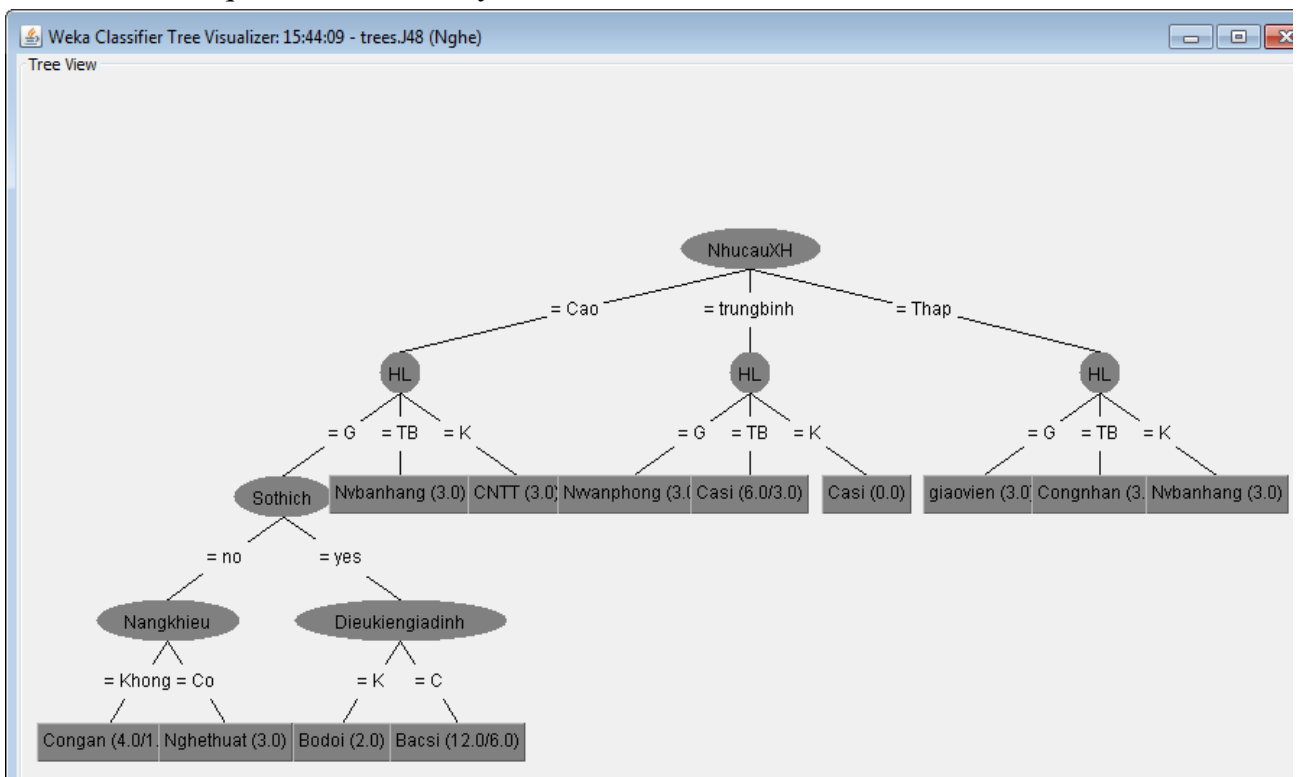
=== Summary ===

<i>Correctly Classified Instances</i>	28
62.2222 %	
<i>Incorrectly Classified Instances</i>	17
37.7778 %	
<i>Kappa statistic</i>	0.575
<i>Mean absolute error</i>	0.0684
<i>Root mean squared error</i>	0.199
<i>Relative absolute error</i>	37.9693 %
<i>Root relative squared error</i>	65.9877 %
<i>Total Number of Instances</i>	45

Tóm tắt kết quả xác nhận phân lớp:

- Trường hợp phân lớp chính xác: 28 chiếm 62.22%
- Trường hợp không chính xác: 17 chiếm 37.78%
- Các thống kê lỗi.

Kết quả hiển thị của cây như sau:



Hình 3.10: Cây quyết định

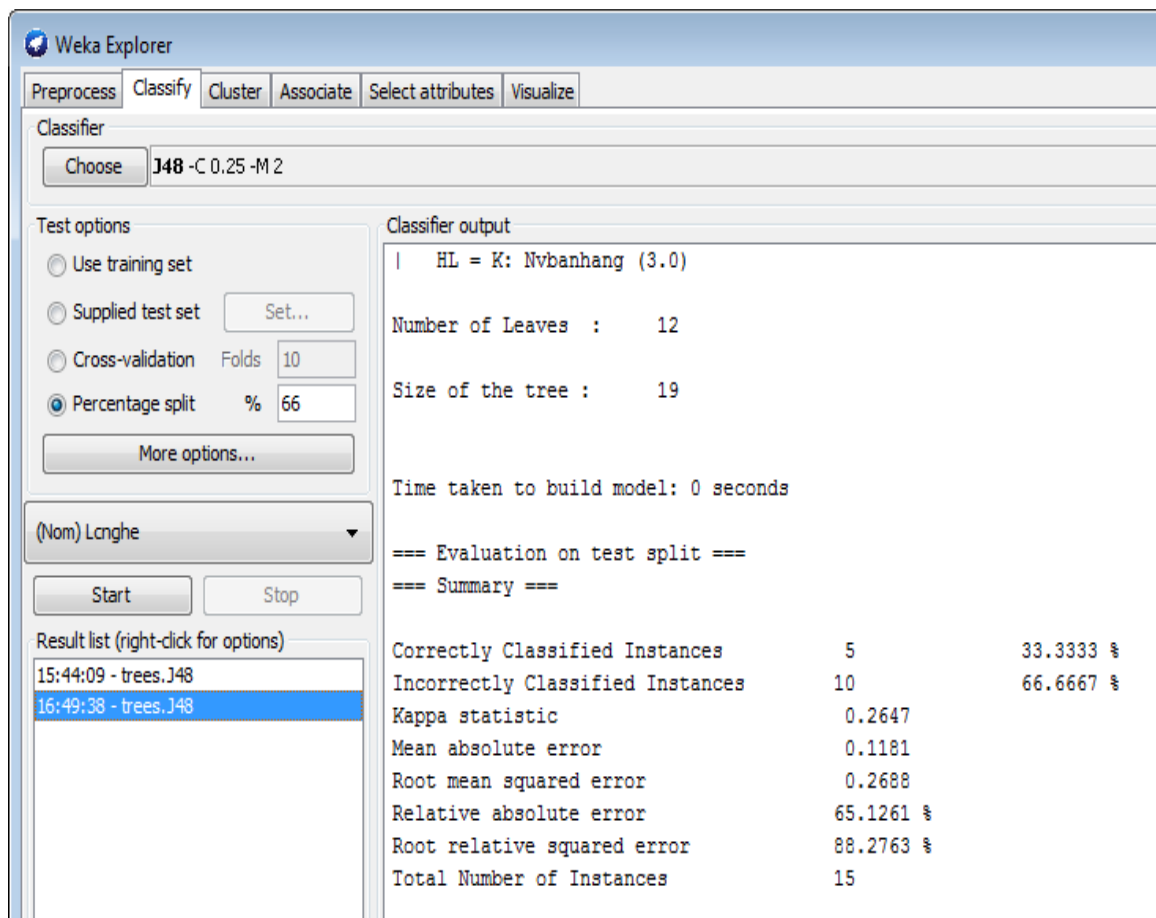
Tạo luật từ cây quyết định:

1. Rule 1: If NhucauXH = “Cao” and HL = G and Sothich =no and Nangkhieu= “Khong” then Lcnghe = “Congan”
2. Rule 2: If NhucauXH = “Cao” and HL = G and Sothich =no and nangkhieu= “Co” then Lcnghe = “Nghethuat”
3. Rule 3: If NhucauXH = “Cao” and HL = G and Sothich =yes and Dieukiengiadinhh =K then Lcnghe = “Bodoi”
4. Rule 4: If NhucauXH = “Cao” and HL = G and Sothich =yes and Dieukiengiadinhh =C then Lcnghe = “Bacsi”
5. Rule 5: If NhucauXH = “Cao” and HL = TB then Lcnghe = “Nvbanhang”
6. Rule 6: If NhucauXH = “Cao” and HL = K then Lcnghe = “CNTT”
7. Rule 7: If NhucauXH = “trungbinh” and HL = G then Lcnghe = “Nvvanphong”
8. Rule 8: If NhucauXH = “trungbinh” and HL = TB then Lcnghe = “Casi”
9. Rule 9: If NhucauXH = “trungbinh” and HL = K then Lcnghe = “Casi”
10. Rule 10: If NhucauXH = “Thap” and HL = G then Lcnghe = “Giaovien”
11. Rule 11: If NhucauXH = “Thap” and HL = TB then Lcnghe = “Congnhan”
12. Rule 12: If NhucauXH = “Thap” and HL = K then Lcnghe = “Nvbanhang”

3.2.5.2. Phương pháp tách phần trăm (Percentage split)

Cho biết tỉ lệ phân chia là bao nhiêu % thì đạt hiệu quả phân lớp cao nhất.

Lần thứ nhất: với tỉ lệ phân chia là 66% thì ta có kết quả như sau:

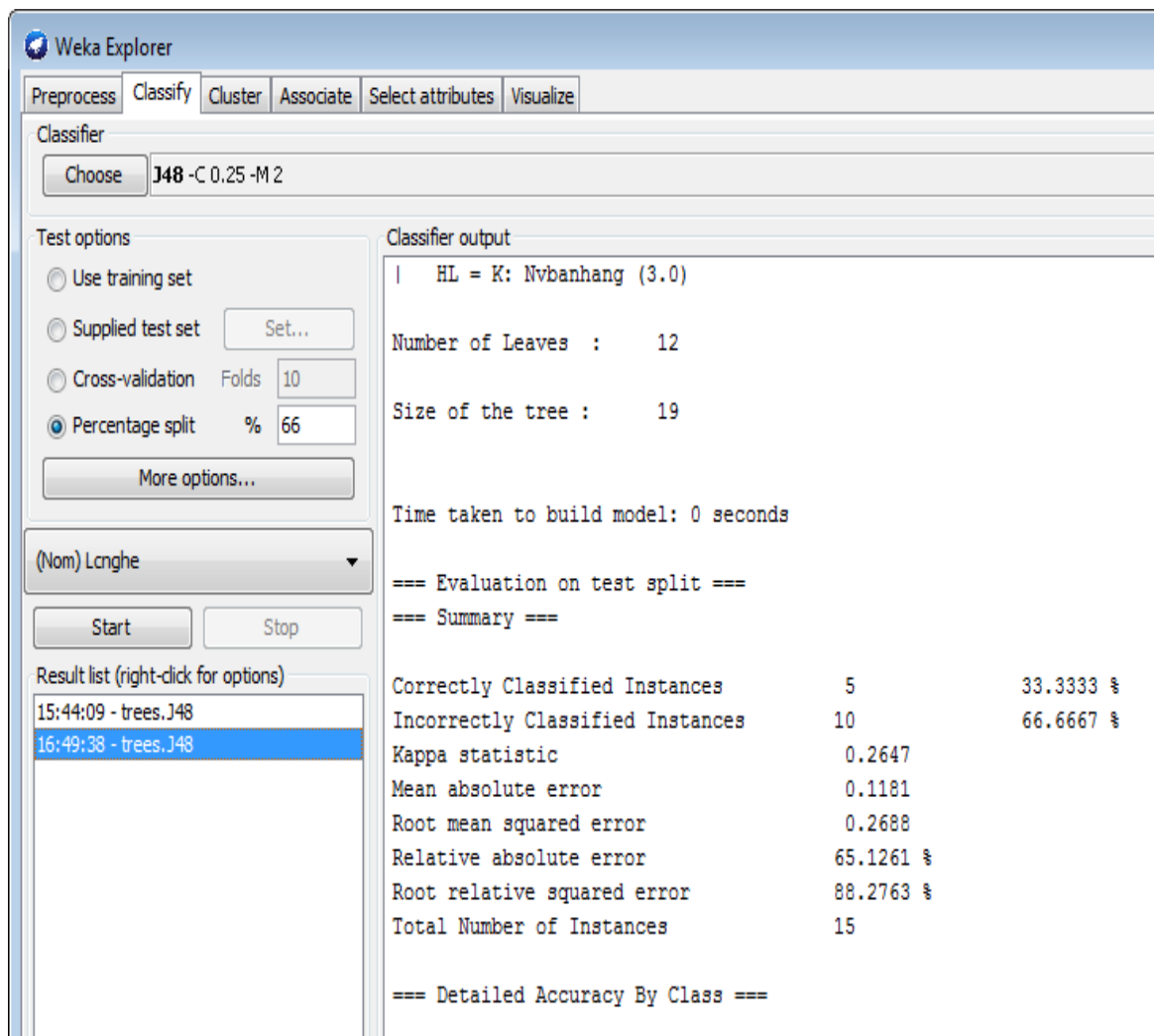


Hình 3.11. Màn hình Weka

=== Evaluation on test split ===			
=== Summary ===			
Correctly Classified Instances	5	33.33	%
Incorrectly Classified Instances	10	66.67	%

Hình 3.12. Đánh giá kết quả

Lần thứ hai: với tỉ lệ phân chia là 60% thì ta có kết quả như sau:

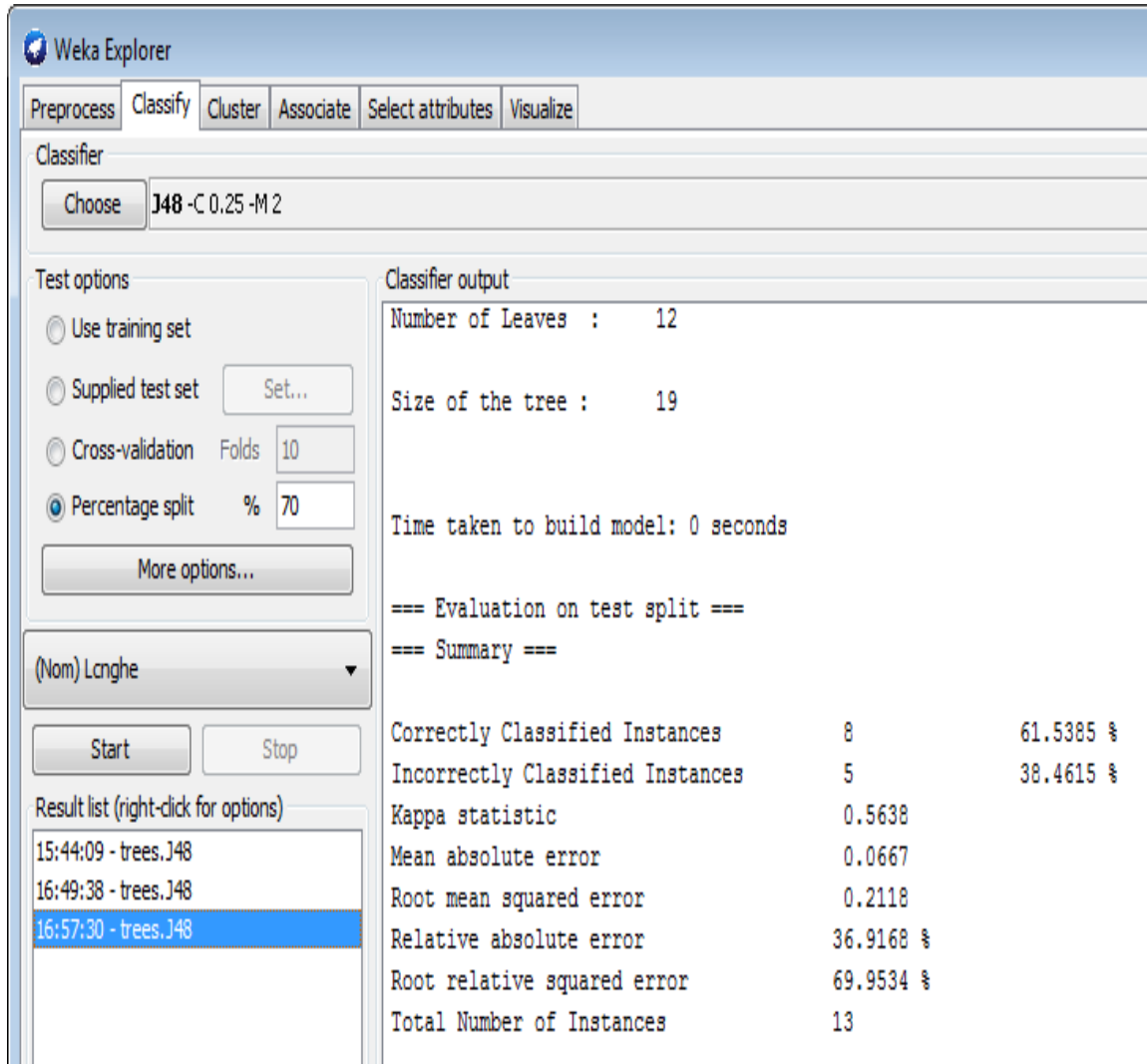


Hình 3.13. Màn hình Weka lần hai

=== Evaluation on test split ===			
=== Summary ===			
Correctly Classified Instances	8	44.44	%
Incorrectly Classified Instances	10	55.56	%

Hình 3.14. Đánh giá đối với lần hai

Lần thứ ba: với tỉ lệ phân chia là 70% thì ta có kết quả như sau:

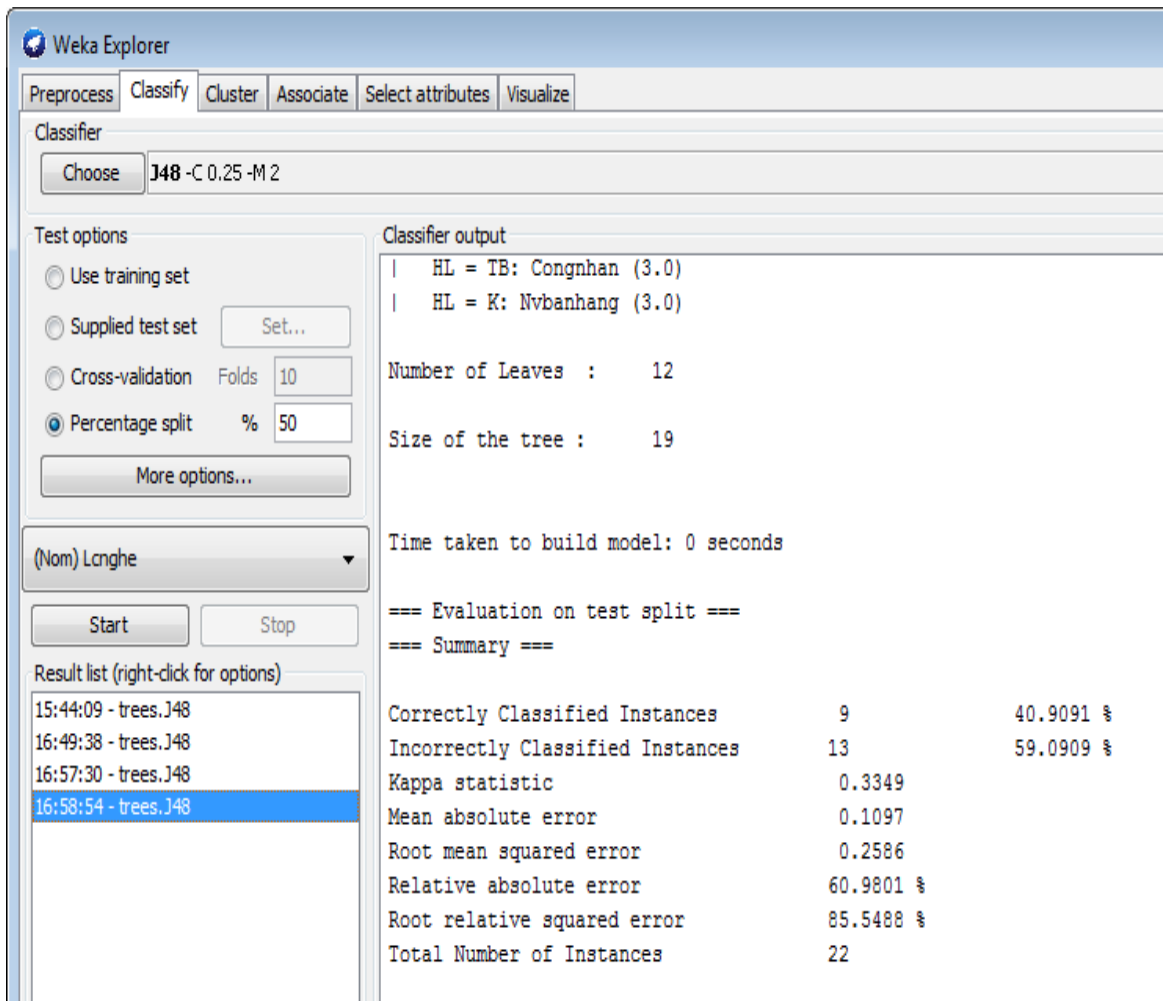


Hình 3.15. Màn hình đối với lần ba

=== Evaluation on test split ===			
=== Summary ===			
Correctly Classified Instances	8	61.54	%
Incorrectly Classified Instances	5	38.46	%

Hình 3.16. Đánh giá đối với lần ba

Lần thứ tư: với tỉ lệ phân chia là 50% thì ta có kết quả như sau:



Hình 3.17. Màn hình đối với lần thứ tư

=== Evaluation on test split ===			
=== Summary ===			
Correctly Classified Instances	9	40.9	%
Incorrectly Classified Instances	13	59.1	%

Hình 3.18. Đánh giá đối với lần thứ tư

Lần thứ năm: với tỉ lệ phân chia là 40% thì ta có kết quả như sau:

The screenshot shows the Weka Explorer interface with the Classifier tab selected. The classifier chosen is J48 -C 0.25 -M 2. The test options are set to Percentage split at 40%. The classifier output is displayed in the right pane, showing the following summary:

=== Evaluation on test split ===		
=== Summary ===		
Correctly Classified Instances	9	33.3333 %
Incorrectly Classified Instances	18	66.6667 %
Kappa statistic	0.2659	
Mean absolute error	0.1284	
Root mean squared error	0.3093	
Relative absolute error	70.7483 %	
Root relative squared error	101.0363 %	
Total Number of Instances	27	

Below the summary is a detailed accuracy table by class:

=== Detailed Accuracy By Class ===							
	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0	0	0	0	0	0.5	Congan
	1	0.542	0.188	1	0.316	0.653	Bo doi
	0.333	0	1	0.333	0.5	0.903	Bacsi
	1	0.115	0.25	1	0.4	0.942	Casi
	0	0	0	0	0	0.5	Nghe thuat
	0	0	0	0	0	0.5	giao vien
	1	0	1	1	1	1	Congnhan
	0	0	0	0	0	0.9	CNTT
	1	0.08	0.5	1	0.667	0.98	Nvbanhang
	0	0	0	0	0	0.98	Nvvanphong
Weighted Avg.	0.333	0.07	0.252	0.333	0.229	0.716	

The status bar at the bottom shows 'OK'.

Hình 3.18. Màn hình đôi với lần thứ năm

=== Evaluation on test split ===			
=== Summary ===			
Correctly Classified Instances	9	33.3	%
Incorrectly Classified Instances	18	66.7	%

Hình 3.19. Đánh giá đôi với lần thứ năm

3.2.5.3. Đánh giá kết quả:

Sau khi thực hiện chạy 5 lần J48 để xây dựng cây với các tham số đầu vào khác nhau ta thấy ứng với lần chạy thứ 3 với các tham số lựa chọn là: bộ dữ liệu dùng khởi tạo cây 70%, bộ dữ liệu thử là 30% đạt được tỉ lệ phân lớp chính xác là 61, 54%. Số lượng mẫu thử không quá nhỏ; cho nên người ta chọn cây quyết định sinh ra tại lần chạy thứ 3 để sử dụng cho các mẫu thử bất kỳ sau này.

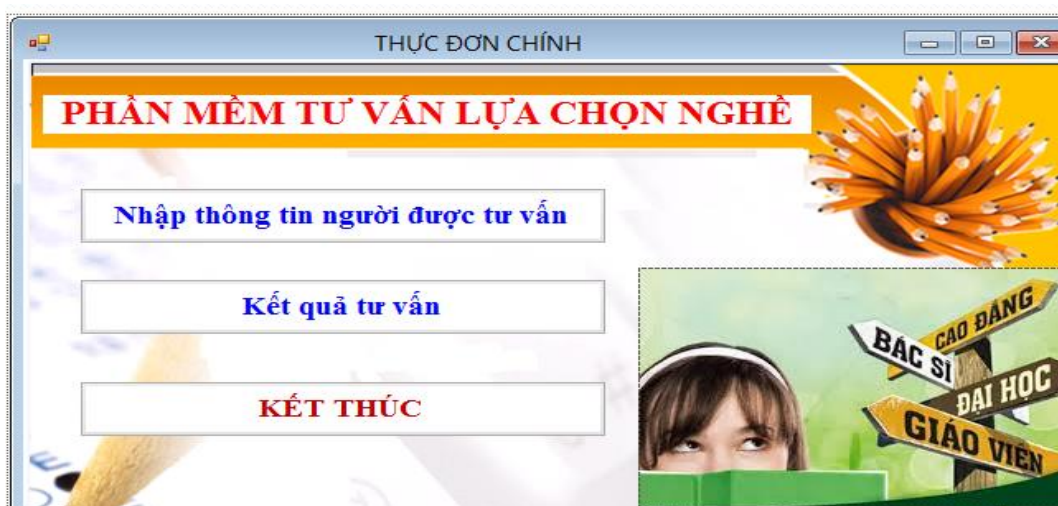
3.3. Đánh giá kết quả sử dụng cây quyết định

3.3.1. Giải pháp xây dựng hệ thống

Giải pháp đề được đưa ra để xây dựng hệ thống theo các bước tạo luật từ cây quyết định bằng công cụ phổ biến Weka đưa về kết quả và độ chính xác cao; tiến hành xây dựng chương trình các khuôn dạng Windows bằng ngôn ngữ lập trình C#, việc thể hiện thuật toán bằng ngôn ngữ lập trình sẽ được rút ngắn, đề tài sẽ đi sâu vào việc xây dựng mô hình khai phá dữ liệu.

3.3.2. Giao diện phần mềm

Giao diện chính của chương trình:



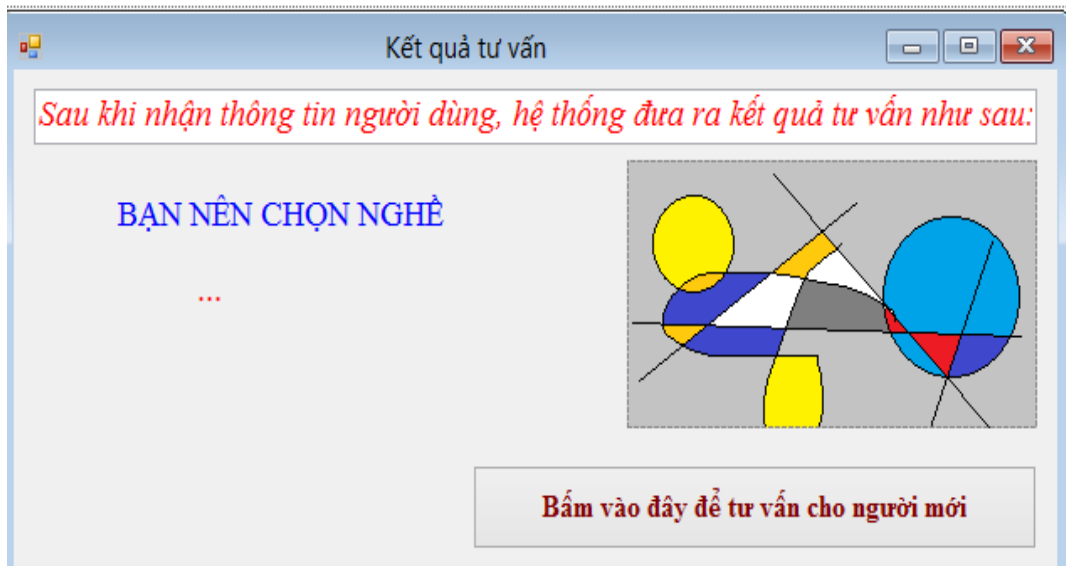
Hình 3.20. Giao diện chương trình hỗ trợ ra quyết định trong tư vấn nghề nghiệp

Giao diện lựa chọn các yêu cầu ra quyết định:

HÃY NHẬP THÔNG TIN ĐỂ ĐƯỢC TƯ VẤN	
Học lực	Em có thích đi làm ngay không?
<input type="text"/>	<input type="text"/>
Theo em, thị trường lao động ntn?	Điều kiện gia đình em có khá không?
<input type="text"/>	<input type="text"/>
Bạn có khả năng đi làm không?	Bấm vào đây để được tư vấn
<input type="text"/>	

Hình 3.12. Giao diện nhập các thông tin để được tư vấn

Giao diện lựa chọn các yêu cầu ra quyết định:



Hình 3.13. Giao diện ra quyết định tư vấn

3.4. Kết luận chương

Chương 3 đã trình bày việc ứng dụng phần mềm Weka vào việc xây dựng cây quyết định từ đó tạo được tập luật và thử nghiệm công tác ra quyết định, dựa trên cơ sở dữ liệu về tuyển sinh đã được phân tích, thiết kế, xây dựng.

Luận văn đã tiến hành phân tích, nghiên cứu, tìm hiểu được các bước, quy trình của công việc tư vấn nghề trong thực tế và ứng dụng của máy tính vào việc tư vấn;

Nắm được các phương pháp và mô hình khai phá dữ liệu, áp dụng để giải quyết yêu cầu bài toán đặt ra. Cụ thể là nghiên cứu và vận dụng thuật toán j48 xây dựng Cây quyết định để khai phá dữ liệu công việc tư vấn lựa chọn nghề nghiệp.

Ứng dụng minh họa thực hiện thành công, đáp ứng được yêu cầu đặt ra. Thực hiện truy vấn và trả về kết quả phù hợp với yêu cầu của chương trình đã trình bày ở trên. Giao diện và cách tổ chức thuật toán rất thân thiện, giúp người dùng dễ dàng sử dụng.

Về mặt thực tiễn, luận văn đã nêu được giải pháp kỹ thuật để vận dụng và xây dựng hệ thống trợ giúp ra quyết định trong công tác tư vấn nghề nghiệp cho học sinh trung học phổ thông.

Việc kết hợp lý thuyết về mô hình khai phá dữ liệu và thuật toán xây dựng Cây quyết định là rất cần thiết, nó giúp giảm thiểu đáng kể thời gian trong việc tìm kiếm, xác định thông tin, giúp cho các em học sinh đưa ra các quyết định đúng khi tốt nghiệp THPT.

KẾT LUẬN

1. Những kết quả đạt được của luận văn

Hệ trợ giúp ra quyết định tuy không phải là một đề tài mới mẻ, với rất nhiều đề tài nghiên cứu và các nỗ lực áp dụng thực tế nó đã dần dà trở thành một bộ phận quan trọng của các hệ thống thông tin hiện đại. Tuy nhiên các áp dụng của nó vẫn ở mức sơ khai và chưa có một chuẩn thống nhất. Trong khuôn khổ luận văn em đã tìm hiểu một cách tổng quan Hệ trợ giúp ra quyết định.

Sau thời gian nghiên cứu và hoàn thiện luận văn, tôi đã thực hiện được các mục tiêu đề ra trong đề cương: (i) tìm hiểu về hệ thống thông tin và hệ trợ giúp quyết định; (ii) hiện trạng của công tác định hướng nghề và nhu cầu về bài toán ra quyết định; (iii) thử nghiệm về bài toán định hướng nghề cho học sinh trung học phổ thông.

Các kết quả nghiên cứu trên hi vọng là tài liệu bổ ích, giúp học sinh ngay khi ngồi trên ghế nhà trường cũng đã có định hướng đúng, khoa học về lựa chọn nghề nghiệp cho tương lai.

Các kết quả đạt được trong luận văn

Về mặt khoa học, luận văn đã (i) phân tích, nghiên cứu, tìm hiểu được các bước, quy trình của công việc tư vấn nghề trong thực tế; (ii) nắm được các phương pháp và mô hình khai phá dữ liệu; (iii) thử nghiệm bài toán đặt ra.

Về mặt thực tiễn, luận văn đã (i) nêu giải pháp kỹ thuật để vận dụng và xây dựng hệ thống trợ giúp ra quyết định trong công tác tư vấn nghề; (ii) kết hợp lý thuyết về mô hình khai phá dữ liệu và thuật toán xây dựng cây quyết định.

Tuy nhiên, hạn chế của các kết quả là : (i) đặc thù của việc tư vấn nghề còn mang nhiều tính chất định tính, có nhiều yếu tố nhạy cảm, vì vậy hệ thống được xây dựng chỉ mang tính hỗ trợ là chính; (ii) mặt khác, công tác tư vấn còn nhiều quy định mà dữ liệu training không thể thể hiện hết được, dẫn đến kết quả chưa theo mong muốn. Hiện tại, hệ thống chỉ chạy thử nghiệm với phạm vi hẹp.

2. Hướng phát triển

Phần mềm sử dụng trong luận văn cần được tiếp tục thử nghiệm. học viên xác định được rằng cần hoàn thiện ứng dụng để nó có thể áp dụng một cách thực tế: (i) về dữ liệu, cần sử dụng các dữ liệu sát với điều kiện thực tế; (ii) về phần mềm, giao diện thân thiện hơn và gần gũi hơn....; (iii) các kết quả tính toán cần đáp ứng được các yêu cầu và độ chính xác cao hơn.

Học viên công tác tại cơ sở giáo dục, luôn nhận thức được vai trò của ứng dụng công cụ công nghệ thông tin và truyền thông để tin học hóa công tác quản lí và hướng nghiệp cho học sinh.

Học viên sẽ cố gắng tìm hiểu các đặc trưng của nghề và các yếu tố đánh giá chính xác về đối tượng được tư vấn nghề, từ đó hoàn thiện hệ thống để hệ thống mang tính ứng dụng và độ chính xác cao hơn.

TÀI LIỆU THAM KHẢO

Tài liệu tiếng Việt

- [1]. Hoàng Xuân Huân, Hệ thống trợ giúp quyết định, bài giảng trường Đại học Công nghệ, Đại học Quốc gia Hà Nội, 2009
- [2]. Vũ Xuân Nam, Nguyễn Văn Huân, Hệ thống thông tin quản lý, bài giảng trường Đại học Thái Nguyên, 2008
- [3]. Đỗ Trung Tuấn, Hệ trợ giúp quyết định, nxb. Đại học Quốc gia Hà Nội, 2016

Tài liệu tiếng Anh

- [4]. Delic, K.A., Douillet, L. and Dayal, U., "Towards an architecture for real-time decision support systems: challenges and solutions, 2001
- [5]. E. Turban, Decision support and expert systems, Prentice Hall, 1995
- [6]. I.M. Makarov, T.M. Vinogradskaya, Rubchinsky, V.B. Sokolov, The Theory of choice and decision making, Mir Publishers Moskow, 1987
- [7]. M.W.Davis, Applied Decision Support, Prentice Hall, 1988
- [8]. Power, D. J. Web-based and model-driven decision support systems: concepts and issues. in proceedings of the Americas Conference on Information Systems, Long Beach, California, 2000
- [9]. Sauter, V. L. Decision support systems: an applied managerial approach. New York, John Wiley, 1997
- [10]. Sprague R, Decision Support Systems, Ed. Prentice Hall, 1998
- [11]. Sprague, R. H. and H. J. Watson. Decision support systems: putting theory into practice. Englewood Clifts, N.J., Prentice Hall, 1993
- [12]. W. H. Inmon, Big Data - getting it right: A checklist to evaluate your environment, <Http://DSS.resources>, 2014

Website:

- [13]. Ho Tu Bao, Chapter 3: Data with Decision Tree [http:// en.wikipedia.org /wiki/ Decision_support_system](http://en.wikipedia.org/wiki/Decision_support_system), 2014
- [14]. [http:// baodautu.vn/nhan-luc-cong-nghe-thong-tin-mong-so-luong-yeu-chat-luong-d61495.html](http://baodautu.vn/nhan-luc-cong-nghe-thong-tin-mong-so-luong-yeu-chat-luong-d61495.html)<http://en.wikipedia.org/wiki/DSS>, 2014
- [15]. <http://tcdn.gov.vn/vi/hoi-dap-trao-doi/trao-doi/Pages/Tieu-chi-nao-cho-truong-nghe>