

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

ĐINH CHUNG DŨNG

**NGHIÊN CỨU VÀ ÁP DỤNG KỸ THUẬT
KHAI PHÁ DỮ LIỆU TRÊN BỘ DỮ LIỆU SINH VIÊN ĐẠI HỌC
PHỤC VỤ CÔNG TÁC CỐ VẤN HỌC TẬP**

Ngành: Công nghệ thông tin.

Chuyên ngành: Truyền dữ liệu và Mạng máy tính

Mã số:

**TÓM TẮT LUẬN VĂN THẠC SĨ
CÔNG NGHỆ THÔNG TIN**

Người hướng dẫn khoa học: TS. NGUYỄN TRUNG TUẤN

HÀ NỘI, 2017

MỞ ĐẦU

1. Lý do lựa chọn đề tài

Khai phá dữ liệu và phát hiện tri thức đang là lĩnh vực được các nhà khoa học quan tâm nghiên cứu trong nhiều năm gần đây. Ứng dụng khai phá dữ liệu được thực hiện trong nhiều lĩnh vực khác nhau như giáo dục, y tế, tài chính, ngân hàng, kinh doanh... Đặc biệt, trong thời gian gần đây, khai phá dữ liệu và phát hiện tri thức trong lĩnh vực giáo dục đang được quan tâm nghiên cứu. Đối với bậc giáo dục Đại học hiện nay, sinh viên đang học tập tại các trường Đại học theo hình thức đào tạo tín chỉ. Đối với hình thức đào tạo này yêu cầu sinh viên phải có sự chủ động cao, có nhiều sự lựa chọn mềm dẻo các môn học trong chuyên ngành đào tạo. Sinh viên sẽ phải tự mình phân bổ các môn học cho từng kỳ sao cho đủ số tín chỉ theo quy chế đào tạo, sinh viên có thể học nhanh để ra trường sớm hoặc đúng hạn với số điểm cao. Trên thực tế đã có rất nhiều trường hợp thời gian học đã hết nhưng các em vẫn chưa hoàn thành đủ tín chỉ, còn nợ môn chuyên ngành. Các sinh viên chưa quen và gặp rất nhiều khó khăn trong định hướng học tập, làm ảnh hưởng đến quá trình học tập của mình cũng như ảnh hưởng đến kết quả đào tạo của nhà trường. Chính vì vậy công tác cố vấn học tập cho sinh viên đã được đặt ra là một công việc quan trọng trong hình thức đào tạo theo tín chỉ. Đây cũng là bài toán được đặt ra cho lĩnh vực khai phá dữ liệu khi có số liệu lớn về sinh viên và quá trình học tập của sinh viên trong nhà trường nhằm trợ giúp cho cố vấn học tập đạt được hiệu quả cao hơn.

Hiện nay tôi đang công tác tại Trường Đại học Kinh tế quốc dân, trước những thực trạng đang tồn tại ở nơi làm việc cùng với lĩnh vực tôi đang theo học, được sự đồng ý của TS. Nguyễn Trung Tuấn tôi chọn đề tài luận văn: ***“Nghiên cứu và áp dụng kỹ thuật khai phá dữ liệu trên bộ dữ liệu sinh viên đại học phục vụ công tác cố vấn học tập”***, luận văn góp phần vào việc giải quyết các vấn đề hết sức cấp bách và cần thiết trong thực tế.

2. Mục tiêu của đề tài

Đề tài luận văn tập trung thực hiện các mục tiêu sau:

- Hiểu được các kỹ thuật khai phá dữ liệu và phát hiện tri thức cơ bản. Tập trung vào hai kỹ thuật là khai phá luật kết hợp và cây quyết định.

- Hiểu các quy chế, quy định, thông tư hướng dẫn về triển khai thực hiện đào tạo đại học theo hệ thống tín chỉ, các văn bản liên quan đến việc quy định cố vấn học tập, hệ thống dữ liệu đào tạo chính quy tại trường Kinh tế Quốc dân và những vấn đề đặt ra đối với việc cố vấn cho sinh viên trong quá trình lựa chọn môn học, phân lớp dự báo khả năng sinh viên rơi vào tình trạng ra trường đúng hạn hay không.

- Phát hiện một số luật, cây phân loại dự báo khả năng ra trường đúng hạn phục vụ cho việc ra quyết định trong cố vấn học tập thông qua bộ dữ liệu thực tế của trường Đại học kinh tế quốc dân bằng việc áp dụng kỹ thuật khai phá luật kết hợp và cây quyết định với sự trợ giúp của công cụ đã có.

3. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu:

- Các vấn đề về phát hiện tri thức và khai phá dữ liệu.
- Các kỹ thuật khai phá dữ liệu.
- Đặc trưng của bài toán cố vấn học tập trong đào tạo tín chỉ tại Trường Đại học kinh tế quốc dân.
- Đặc trưng của bộ dữ liệu thử nghiệm.
- Một số công cụ hỗ trợ khai phá dữ liệu và phát hiện tri thức.

Phạm vi nghiên cứu:

- Nghiên cứu về kỹ thuật khai phá dữ liệu dựa trên luật kết hợp và cây quyết định.
- Bài toán cố vấn học tập cho sinh viên chính quy đào tạo theo hình thức tín chỉ tại trường Đại học kinh tế quốc dân.
- Dữ liệu thử nghiệm được thực hiện trên 01 khóa sinh viên đã ra trường.

- Công cụ hỗ trợ khai phá là Business Intelligence Development Studio (BIDS) trong bộ SQL Server 2008 R2.

4. Phương pháp nghiên cứu

Đề tài luận văn được tác giả thực hiện nghiên cứu sử dụng phương pháp nghiên cứu hỗn hợp giữa nghiên cứu lý thuyết bằng cách tổng hợp tài liệu và nghiên cứu thực nghiệm.

- *Nghiên cứu lý thuyết*: Tổng hợp các vấn đề lý thuyết liên quan từ các nguồn tài liệu như Giáo trình, Bài giảng, Internet, chuyên đề, luận văn, luận án và tìm hiểu thực tiễn đào tạo theo học chế tín chỉ tại trường Đại học kinh tế quốc dân để phát biểu bài toán cổ vấn học tập cần giải quyết.

- *Nghiên cứu thực nghiệm*:

+ Nghiên cứu cài đặt, thực hành công cụ hỗ trợ khai phá dữ liệu BIDS

+ Sử dụng bộ dữ liệu thực tế được thu thập tại trường Đại học kinh tế quốc dân để giải bài toán cổ vấn học tập bằng các phương pháp lý thuyết đã nghiên cứu.

+ Dựa vào bài toán cổ vấn học tập đã phát biểu, chọn phương pháp giải, biến đổi dữ liệu thô đã thu thập thành bộ dữ liệu con phù hợp với kỹ thuật lựa chọn và bài toán đã phát biểu. Sau khi có dữ liệu tiến hành các bước khai phá dữ liệu trên công cụ BIDS.

+ Mô tả các kết quả thu được có ý nghĩa thực tiễn phù hợp với bài toán đã phát biểu.

5. Ý nghĩa thực tiễn và khoa học của đề tài

Hiện nay Bộ giáo dục và đào tạo đã chính thức ban hành quy chế đào tạo theo tín chỉ cho các Trường Đại Học. Hầu hết tất cả các trường Đại Học trên cả nước đã áp dụng hình thức đào tạo này, bên cạnh đó đã có nhiều hội thảo và bài báo nói về việc học theo hình thức tín chỉ và công tác được quan tâm là cổ vấn học tập trong đào tạo tín chỉ cho sinh viên. Trong nhiều năm qua Trường Đại

Học Kinh tế Quốc dân đã có chủ trương ứng dụng công nghệ thông tin vào các công tác quản lý đào tạo. Hệ thống dữ liệu đào tạo đã góp phần quản lý tốt cho công tác đào tạo đại học chính quy. Bộ phận cố vấn học tập cũng có những quy định để góp phần giúp sinh viên có một tiến trình học hiệu quả. Qua đề tài việc ứng dụng vào dữ liệu thực tế để đưa ra các tri thức cần thiết để hỗ trợ ra quyết định cố vấn là điều hết sức cấp bách và cần thiết cho bộ phận cố vấn học tập. Góp phần vào đảm bảo nâng cao chất lượng đào tạo chung của nhà trường. Khai phá dữ liệu cũng đang là lĩnh vực mà các nhà khoa học quan tâm trong nhiều năm gần đây và cũng đã được nghiên cứu áp dụng vào thực tiễn mang lại nhiều lợi ích.

Nội dung của đề tài mang ý nghĩa khoa học khi phát biểu, phân tích các đặc trưng của bài toán cố vấn học tập và bộ dữ liệu tại trường Đại học kinh tế quốc dân, quy trình áp dụng các kỹ thuật khai phá dữ liệu và phát hiện tri thức trên bộ dữ liệu này. Đề tài luận văn cũng mang ý nghĩa thực tiễn khi tác giả thực nghiệm các phương pháp trên với các dữ liệu thực tế, mô tả các kết quả đạt được có ý nghĩa với bài toán cố vấn học tập, từ đó đưa ra những đề xuất, kiến nghị. Góp phần có thêm nhiều cơ sở thông tin giúp ích cho bộ phận cố vấn học tập ra quyết định cố vấn trong quá trình sinh viên học tập, nâng cao chất lượng đào tạo của nhà trường.

6. Bố cục luận văn

Luận văn được trình bày trong 3 chương chính ngoài phần Mở đầu, Kết luận, Tài liệu tham khảo, Danh mục các hình và Phụ lục. Cụ thể như sau:

Chương 1. Tổng quan về phát hiện tri thức và khai phá dữ liệu

Trong chương này sẽ trình bày những vấn đề cơ bản về phát hiện tri thức và khai phá dữ liệu, bao gồm những nội dung cơ bản: tổng quan về khai phá dữ liệu và phát hiện tri thức; ứng dụng của khai phá dữ liệu và phát hiện tri thức; các phương pháp và kỹ thuật khai phá dữ liệu và phát hiện tri thức.

Chương 2. Bài toán cố vấn học tập và đặc điểm bộ dữ liệu sinh viên tại trường Đại học Kinh tế Quốc dân

Nội dung của chương này sẽ trình bày những vấn đề về cố vấn học tập trong đào tạo đại học chính quy theo hình thức tín chỉ, những vấn đề gặp phải

trong quá trình cố vấn học tập. Từ đó phân tích và hình thành bài toán cần giải quyết trong công tác cố vấn học tập tại trường Đại học Kinh tế Quốc dân. Giới thiệu và mô tả đặc điểm của bộ dữ liệu sinh viên chính quy đã thu thập được trường Đại học Kinh tế quốc dân để phục vụ cho quá trình thực nghiệm; mô tả về các bộ dữ liệu con được trích rút từ bộ dữ liệu lớn phục vụ cho các mục đích phân tích khác nhau theo yêu cầu của bài toán cố vấn học tập.

Chương 3. Ứng dụng thử nghiệm khai phá dữ liệu sinh viên phục vụ cố vấn học tập tại trường Đại học Kinh tế Quốc dân

Chương này sẽ giới thiệu về một số công cụ khai phá dữ liệu và phát hiện tri thức thông dụng và công cụ BIDS của Microsoft SQL Server 2008. Quy trình thực nghiệm khai phá và phát hiện tri thức với bài toán cố vấn học tập. Trình bày và đánh giá các kết quả khai phá dữ liệu trên 02 bài toán cố vấn học tập: Tư vấn lựa chọn môn học theo tổ hợp lựa chọn từng ngành, chuyên ngành; Phân lớp dự báo sinh viên có khả năng ra trường đúng hạn hay không đúng hạn.

Ngoài ra, phần Mở đầu của luận văn sẽ giới thiệu chung về những nội dung và phương pháp thực hiện nghiên cứu đề tài luận văn. Phần Kết luận của luận văn sẽ trình bày về tóm tắt về những kết quả đã đạt được, những hạn chế và hướng nghiên cứu tiếp theo của đề tài luận văn.

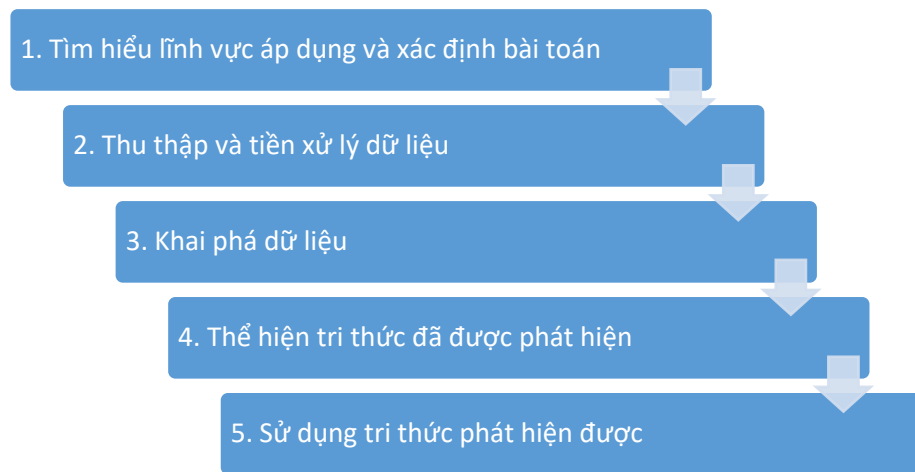
CHƯƠNG 1. TỔNG QUAN VỀ KHAI PHÁ DỮ LIỆU VÀ PHÁT HIỆN TRI THỨC.

1.1. Giới thiệu chương

1.2. Tổng quan về phát hiện tri thức và khai phá dữ liệu

Khai phá dữ liệu (Data mining) là một khâu trong quá trình khám phá tri thức mà trong đó ta có thể áp dụng những thuật toán khai phá dữ liệu với những giới hạn có thể chấp nhận được về độ phức tạp tính toán để tìm ra những mẫu hoặc mô hình trong dữ liệu [4].

1.3. Quá trình phát hiện tri thức và khai phá dữ liệu



1.4. Các phương pháp khai phá dữ liệu

- **Phân lớp:** là việc xác định một hàm ánh xạ các mục dữ liệu vào một trong nhiều lớp đã được xác định trước.
- **Hồi quy:** là việc xác định một hàm ánh xạ một mục dữ liệu đến một giá trị dữ liệu thực của biến dự báo.
- **Phân cụm:** là công việc mang tính mô tả thông thường, nó sẽ xác định tập hữu hạn các nhóm hoặc các cụm để mô tả dữ liệu.
- **Tổng quát hoá:** bao gồm các phương pháp để tìm kiếm một mô tả ngắn gọn và tổng quát cho một tập con dữ liệu.
- **Mô hình hoá sự phụ thuộc:** bao gồm việc tìm một mô hình mô tả những sự phụ thuộc cơ bản giữa các biến.
- **Phát hiện thay đổi và chênh lệch:** tập trung vào việc phát hiện những thay đổi đáng chú ý trên dữ liệu từ những giá trị được đo trước đó.
- **Biểu diễn mô hình:** là phương pháp để mô tả những mẫu hoặc mô hình có thể được phát hiện. Nếu biểu diễn này bị hạn chế và có nhiều ràng buộc thì khi đó không thể tìm được mô hình đúng đắn cho dữ liệu.
- **Đánh giá mô hình:** dự tính khả năng đáp ứng của một mẫu hoặc mô hình và các tham số của nó với các tiêu chí kết quả của tiến trình KDD.

- **Phương pháp tìm kiếm:** có hai bài toán là tìm kiếm tham số và tìm kiếm mô hình.

1.5. Các vấn đề cần nghiên cứu của phát hiện tri thức và khai phá dữ liệu

- Phương pháp luận khai phá dữ liệu
- Tương tác với người dùng

1.6. Các lĩnh vực ứng dụng của phát hiện tri thức và khai phá dữ liệu

Giáo dục, y tế, thương mại, tài chính, chứng khoán, văn bản, Web, sinh học, thể thao, thiên văn học, an ninh chống khủng bố.....

1.7. Kỹ thuật khai phá luật kết hợp

1.7.1. Lý thuyết về luật kết hợp

1.7.2. Định nghĩa luật kết hợp

Quy trình khai phá luật kết hợp được thực hiện lần lượt theo hai bài toán sau:

- Bài toán 1: Tìm tất cả các tập mục mà có độ hỗ trợ lớn hơn độ hỗ trợ tối thiểu do người dùng xác định. Các tập mục thoả mãn độ hỗ trợ tối thiểu được gọi là các tập mục phổ biến (theo ngưỡng minsupp).
- Bài toán 2: Dùng các tập mục phổ biến để sinh ra các luật mong muốn. Ý tưởng chung là nếu gọi ABCD và AB là các tập mục phổ biến, thì chúng ta có thể xác định luật nếu $AB \Rightarrow CD$ giữ lại với tỷ lệ độ tin cậy:

$$conf = \frac{sup(ABCD)}{sup(AB)}$$

nếu $conf \geq minconf$ thì luật được giữ lại (luật này sẽ thoả mãn độ hỗ trợ tối thiểu vì ABCD là phổ biến)

1.7.3. Một số hướng tiếp cận trong khai phá luật kết hợp

- Luật kết hợp nhị phân
- Luật kết hợp có thuộc tính số và thuộc tính hạng mục
- Luật kết hợp tiếp cận theo hướng tập thô
- Luật kết hợp nhiều mức
- Luật kết hợp mờ

- Luật kết hợp với thuộc tính được đánh trọng số
- Khai thác luật kết hợp song song

1.8. Kỹ thuật khai phá cây quyết định

Một cây quyết định là một cấu trúc hình cây, trong đó:

- Mỗi đỉnh trong (đỉnh có thể khai triển được) biểu thị cho một phép thử đối với một thuộc tính.
- Mỗi nhánh biểu thị cho một kết quả của phép thử.
- Các đỉnh lá (các đỉnh không khai triển được) biểu thị các lớp hoặc các phân bổ lớp.
- Đỉnh trên cùng trong một cây được gọi là gốc.

Việc sinh cây quyết định bao gồm hai giai đoạn:

+ Giai đoạn 1: Xây dựng cây

- Tại thời điểm khởi đầu, tất cả các cây (case) dữ liệu học đều nằm tại gốc.
- Các cây dữ liệu được phân chia đệ quy trên cơ sở các thuộc tính được chọn.

+ Giai đoạn 2: Rút gọn cây

- Phát hiện và bỏ đi các nhánh chứa các điểm dị thường và nhiễu trong dữ liệu.

1.9. Tổng kết chương 1

Chương 1 đã tóm tắt được cơ sở lý thuyết liên quan đến phát hiện tri thức và khai phá dữ liệu, nêu những ý tưởng chính trong hai kỹ thuật khai phá luật kết hợp và cây quyết định.

CHƯƠNG 2. BÀI TOÁN CỔ VẤN HỌC TẬP VÀ ĐẶC TRUNG BỘ DỮ LIỆU SINH VIÊN TẠI TRƯỜNG ĐẠI HỌC KINH TẾ QUỐC DÂN

2.1. Những vấn đề về cổ vấn học tập theo hình thức đào tạo tín chỉ tại trường Đại học Kinh tế Quốc dân.

2.1.1. Tổ chức hệ thống cổ vấn học tập

2.1.2. Chức năng của cổ vấn học tập

2.1.3. Nhiệm vụ của cổ vấn học tập

2.1.3.1. Nhiệm vụ chung của CVHT chuyên trách và CVHT kiêm nhiệm

2.1.3.2. Nhiệm vụ cụ thể

2.2. Bài toán cổ vấn học tập tại trường Đại học kinh tế quốc dân

2.2.1. Vấn đề thực tế xung quanh bài toán

2.2.2. Phát biểu bài toán

Bài toán 1: Cổ vấn cho sinh viên đăng ký các học phần tự chọn theo các tổ hợp trên định hướng chuyên ngành. Vào đầu mỗi kỳ học khi phòng đào tạo thông báo mở các lớp học phần, sinh viên thường phải tự sắp xếp thời khóa biểu của mình và chủ động đăng ký môn học. Họ thường gặp khó khăn trong việc lựa chọn, băn khoăn không biết nên học môn nào trong một tổ hợp, và đa số phải tham khảo ý kiến của cán bộ cổ vấn học tập để xin tư vấn, định hướng lựa chọn các học phần tự chọn trong kỳ học đó sao cho phù hợp với năng lực sở thích và quy chế đào tạo.

Bài toán 2: Phân lớp, dự báo cho sinh viên có khả năng ra trường đúng thời hạn hay không đúng hạn. Theo thống kê của phòng đào tạo, hàng năm có từ 10 đến 15 phần trăm sinh viên ra trường muộn. Để giải quyết vấn đề này thì vai trò của cán bộ cổ vấn học tập là phải đưa ra quyết định cảnh báo học tập kịp thời. Sau khi kết thúc mỗi kỳ học, cán bộ cổ vấn học tập thường phải theo dõi kết quả học tập của sinh viên, tổng hợp kết quả từng kỳ học. Nếu phát hiện những sinh viên chưa đủ số tín chỉ và xếp loại học lực yếu thì phải thông báo cho sinh viên biết sớm, giúp sinh viên nhanh chóng điều chỉnh kế hoạch và thái độ học tập, bổ sung đủ tín chỉ, cải thiện điểm thì mới hoàn thành tốt nghiệp đúng thời hạn theo quy chế đào tạo của nhà trường.

2.2.3. Mục tiêu và ý nghĩa của bài toán

Mục tiêu, ý nghĩa bài toán 1: Làm thế nào có thêm nhiều cơ sở thông tin giúp cho cán bộ cố vấn học tập dựa vào đó để làm phương tiện cố vấn, giải quyết những vấn đề thực tế của sinh viên. Bằng phương pháp khai phá dữ liệu dựa trên luật kết hợp, tìm ra mối quan hệ kết hợp giữa các môn học (môn học nào hay được sinh viên kết hợp đăng ký cùng nhau), kết quả sinh ra được một tập luật kết hợp giữa các môn học, luật này mạnh và có ích với khả năng xảy ra cao. Ý nghĩa từ bảng tập luật đó giúp cán bộ cố vấn trả lời hai câu hỏi của sinh viên.

- Nếu đăng ký học phần A ở tổ hợp này, và học phần C ở tổ hợp kia, thì thường hay đăng ký học phần nào ở tổ hợp khác, theo từng chuyên ngành khác nhau.
- Trong các tổ hợp học phần lựa chọn, học phần lựa chọn nào hay được chọn đăng ký cùng với nhau.

Từ đó cán bộ cố vấn học tập sẽ nắm được xu hướng lựa chọn học phần, phân tích xem nhu cầu ra sao, cố vấn cho phòng đào tạo điều chỉnh tăng, giảm, thay đổi số lượng lớp học phần cho phù hợp phân bổ chương trình môn học hợp lý cho giáo viên và sinh viên.

Mục tiêu, ý nghĩa bài toán 2: Từ kết quả phân lớp xác định được sinh viên nào đang bị rơi vào tình trạng cảnh báo ra trường không đúng hạn, đúng hạn. Nếu cán bộ cố vấn học tập có những cảnh báo nhanh chóng và kịp thời thì hàng năm tỉ lệ ra trường muộn sẽ giảm đi rất nhiều. Bằng phương pháp khai phá dữ liệu, phân lớp sinh viên dựa trên cây quyết định. Giúp cán bộ cố vấn học tập trong quá trình theo dõi kết quả học tập của sinh viên chính xác hơn, cảnh báo, dự báo tiến trình học tập cho sinh viên sau mỗi kỳ học, góp phần tăng tỷ lệ ra trường đúng hạn của nhà Trường đạt tối đa lên mục tiêu 100% sinh viên ra trường đúng hạn.

Sau khi khai phá dữ liệu bằng kỹ thuật phân lớp dựa vào cây quyết định. Ý nghĩa kết quả sau khi thực hiện phân lớp sinh viên là dựa vào số tín chỉ đã tích lũy và điểm chung bình chung tích lũy của các kỳ học sẽ giúp cán bộ cố vấn học tập có khả năng ra quyết định cảnh báo, dự báo sinh viên A có khả năng rơi vào

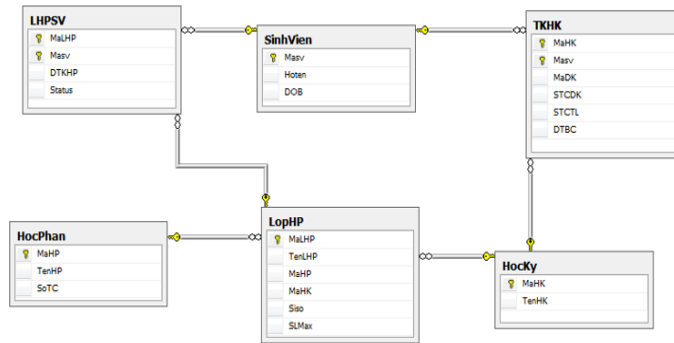
trường hợp ra trường đúng hạn hay không đúng hạn, nếu không đúng hạn thì sinh viên đó sớm đăng ký học phần bổ sung cho kịp ra trường.

2.2.4. Sơ đồ phác thảo thực hiện 2 bài toán

2.3. Đặc trưng dữ liệu sinh viên trường Đại học kinh tế quốc dân

2.3.1. Hệ thống quản lý đào tạo, quản lý sinh viên trong quá trình học tập

2.3.2. Mô tả một phần cơ sở dữ liệu quản lý sinh viên dựa trên những thông tin đã thu thập.



2.4. Tổng kết chương 2

Chương 2 đã tóm tắt các vấn đề liên quan đến cố vấn học tập và đào tạo theo học chế tín chỉ tại trường Đại học kinh tế quốc dân, phát biểu 2 bài toán cố vấn học tập cụ thể, xác định mục tiêu và ý nghĩa thực tiễn của 2 bài toán. Phác thảo sơ đồ thực hiện giải bài toán.

CHƯƠNG 3. ỨNG DỤNG THỬ NGHIỆM GIẢI BÀI TOÁN CỐ VẤN HỌC TẬP TẠI TRƯỜNG ĐẠI HỌC KINH TẾ QUỐC DÂN

3.1. Giới thiệu một số công cụ khai phá dữ liệu và phát hiện tri thức

3.1.1. Weka

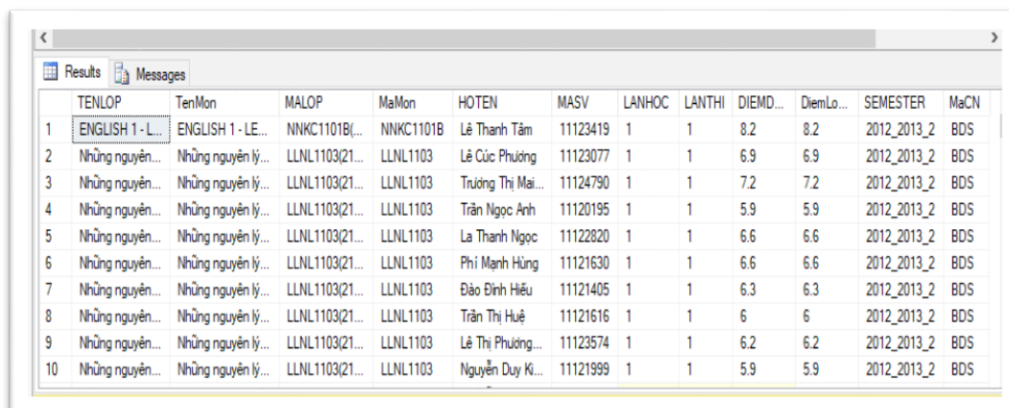
3.1.2. Ngôn ngữ R

3.1.3. SQL Datamining

3.2. Quy trình thực hiện khai phá dữ liệu sinh viên và phát hiện tri thức với bài toán cố vấn học tập tại Trường Đại học Kinh tế Quốc dân.

3.3. Khai phá dữ liệu bằng luật kết hợp giải bài toán 1

3.3.1. Từ dữ liệu thô thu thập được



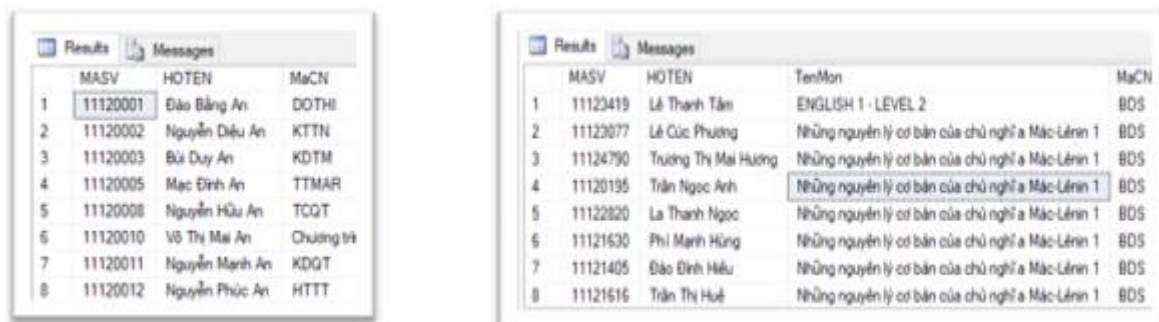
	TENLOP	TenMon	MALOP	MaMon	HOTEN	MASV	LANHOC	LANTHI	DIEMD...	DiemLo...	SEMESTER	MaCN
1	ENGLISH 1 - L...	ENGLISH 1 - LE...	NNKC1101B(...	NNKC1101B	Lê Thanh Tâm	11123419	1	1	8.2	8.2	2012_2013_2	BDS
2	Những nguyên...	Những nguyên lý...	LLNL1103(21...	LLNL1103	Lê Cúc Phương	11123077	1	1	6.9	6.9	2012_2013_2	BDS
3	Những nguyên...	Những nguyên lý...	LLNL1103(21...	LLNL1103	Trương Thị Mai...	11124790	1	1	7.2	7.2	2012_2013_2	BDS
4	Những nguyên...	Những nguyên lý...	LLNL1103(21...	LLNL1103	Trần Ngọc Anh	11120195	1	1	5.9	5.9	2012_2013_2	BDS
5	Những nguyên...	Những nguyên lý...	LLNL1103(21...	LLNL1103	La Thanh Ngọc	11122820	1	1	6.6	6.6	2012_2013_2	BDS
6	Những nguyên...	Những nguyên lý...	LLNL1103(21...	LLNL1103	Phí Mạnh Hùng	11121630	1	1	6.6	6.6	2012_2013_2	BDS
7	Những nguyên...	Những nguyên lý...	LLNL1103(21...	LLNL1103	Đào Đình Hiếu	11121405	1	1	6.3	6.3	2012_2013_2	BDS
8	Những nguyên...	Những nguyên lý...	LLNL1103(21...	LLNL1103	Trần Thị Huệ	11121616	1	1	6	6	2012_2013_2	BDS
9	Những nguyên...	Những nguyên lý...	LLNL1103(21...	LLNL1103	Lê Thị Phượng...	11123574	1	1	6.2	6.2	2012_2013_2	BDS
10	Những nguyên...	Những nguyên lý...	LLNL1103(21...	LLNL1103	Nguyễn Duy Ki...	11121999	1	1	5.9	5.9	2012_2013_2	BDS

Hình 3.1 Dữ liệu thu thập

- Dữ liệu dưới dạng file excel của khóa học đã ra trường, có 4000 sinh viên, 12 thuộc tính, hàng chục nghìn bản ghi. Trên mỗi file có 5 sheet, mỗi sheet có hàng nghìn bản ghi là thể hiện của 1 năm học, mỗi năm có 2 kỳ học. (có thể coi đây như là một bộ các giao dịch đăng ký môn học của sinh viên).
- Toàn bộ dữ liệu thể hiện được quá trình đăng ký tất cả các môn học phần của sinh viên trong tất cả các chuyên ngành trong khóa học đó, tách theo từng kỳ học.

3.3.2. Tiến hành biến đổi dữ liệu theo bài toán 1

Bộ dữ liệu con thu được dùng trong mô hình khai phá gồm có 2 view như sau: (DanhSachSV, và SV_DangKy_MonHoc)



	MASV	HOTEN	MaCN
1	11120001	Đào Bằng An	DOTH
2	11120002	Nguyễn Diệu An	KTTN
3	11120003	Bùi Duy An	KDTM
4	11120005	Mạc Đình An	TTMAR
5	11120008	Nguyễn Hữu An	TCQT
6	11120010	Võ Thị Mai An	Chương tr
7	11120011	Nguyễn Mạnh An	KDQT
8	11120012	Nguyễn Phúc An	HTTT

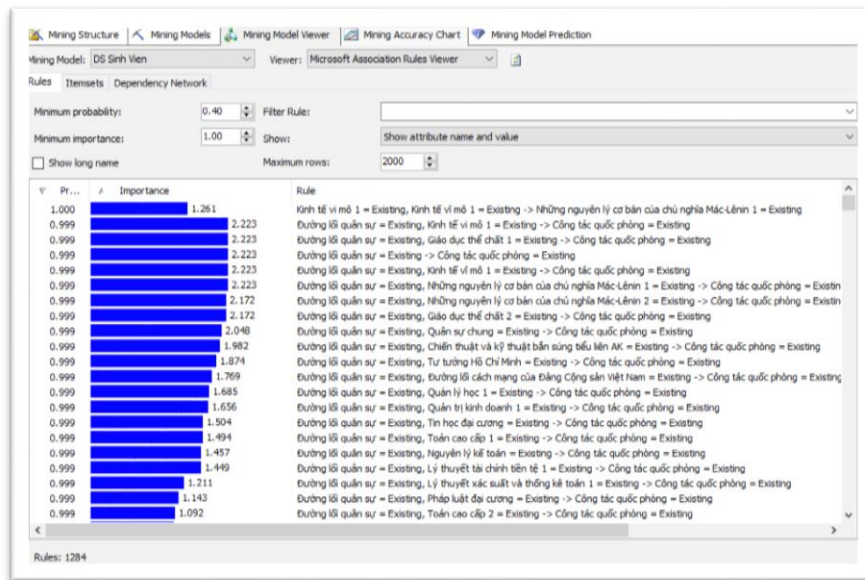
	MASV	HOTEN	TenMon	MaCN
1	11123419	Lê Thanh Tâm	ENGLISH 1 - LEVEL 2	BDS
2	11123077	Lê Cúc Phương	Những nguyên lý cơ bản của chủ nghĩa Mác-Lênin 1	BDS
3	11124790	Trương Thị Mai Hương	Những nguyên lý cơ bản của chủ nghĩa Mác-Lênin 1	BDS
4	11120195	Trần Ngọc Anh	Những nguyên lý cơ bản của chủ nghĩa Mác-Lênin 1	BDS
5	11122820	La Thanh Ngọc	Những nguyên lý cơ bản của chủ nghĩa Mác-Lênin 1	BDS
6	11121630	Phí Mạnh Hùng	Những nguyên lý cơ bản của chủ nghĩa Mác-Lênin 1	BDS
7	11121405	Đào Đình Hiếu	Những nguyên lý cơ bản của chủ nghĩa Mác-Lênin 1	BDS
8	11121616	Trần Thị Huệ	Những nguyên lý cơ bản của chủ nghĩa Mác-Lênin 1	BDS

Hình 3.2 Dữ liệu cho khai phá luật kết hợp

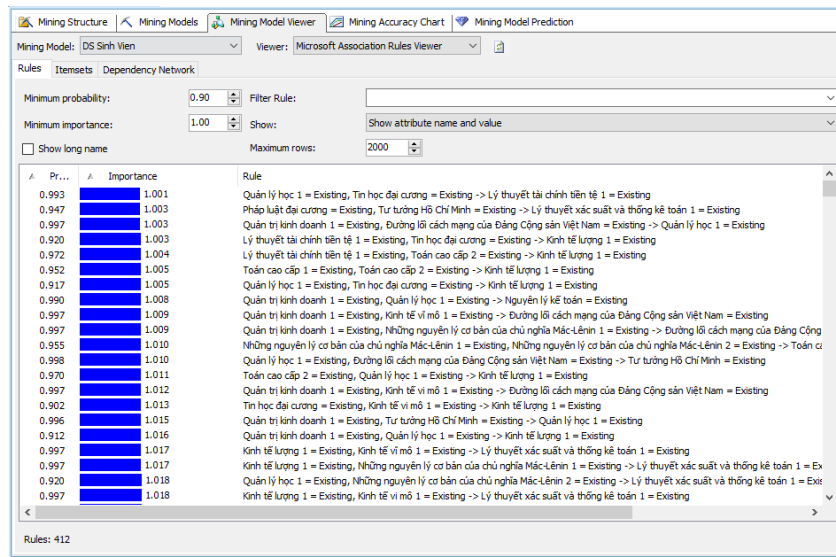
3.3.3. Thực hiện thử nghiệm trên công cụ BIDS

Kết quả khai phá dữ liệu giải bài toán 1 sử dụng kỹ thuật khai phá luật kết hợp trong 3 lần thử nghiệm chính như sau:

Lần 1: Với tất cả dữ liệu đăng ký môn học phần (*cả tự chọn và bắt buộc*) của khóa học với tất cả các chuyên ngành và đưa 2 view vào hệ quản trị cơ sở dữ liệu SQL. Kết quả hai lần chạy với tham số khác nhau trong lần 1.



Hình 3.3 L1.1: minsupport=0.4 và minprobability = 0.4



Hình 3.4 L1.2: minsupport=0.4 và minprobability = 0.9

Nhận xét: Công cụ đã cho ra các luật như mong muốn, nhưng vì trên các luật không có thông tin chuyên ngành (vì dữ liệu gồm tất cả các môn của tất cả các chuyên ngành), nên muốn tư vấn cho từng chuyên ngành lại phải tìm xem môn đó thuộc chuyên ngành nào việc tư vấn cho từng chuyên ngành là khó khăn. Có quá nhiều luật và luật lại kết hợp cả học phần tự chọn và học phần bắt buộc nên lần 1 chạy là không khả thi, tiến hành thử nghiệm lần 2.

Lần 2: Vẫn dữ liệu như lần 1 và có thay đổi sau:

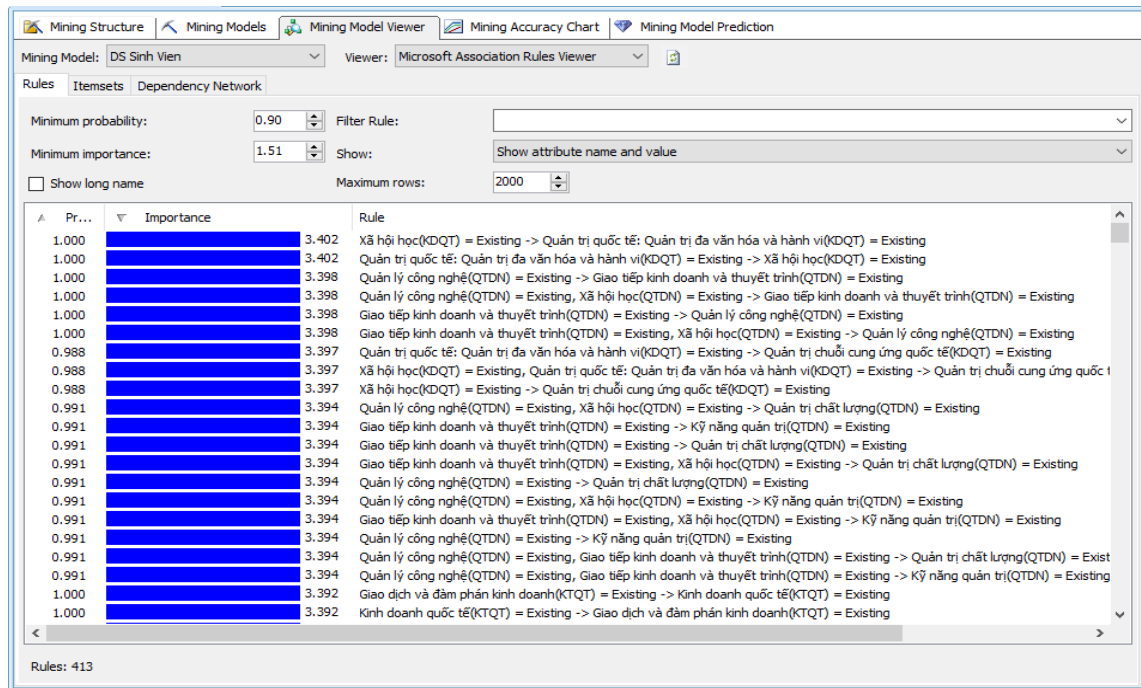
- Loại bỏ các học phần bắt buộc ra khỏi dữ liệu (còn lại các học phần tự chọn).
- Thêm mã chuyên ngành vào sau các môn học phần tự chọn (ví dụ; xã hội học(KDQT), quản lý công nghệ(QTDN)).

Kết quả: Chạy với minsupp= 0.03, minprobability= 0.54, thu được 663 luật

Sau đó thay đổi: minsupp= 0.03, minprobability= 0.9, thu được 413 luật

Pr...	Importance	Rule
1.000	3.402	Xã hội học(KDQT) = Existing -> Quản lý quốc tế: Quản trị đa văn hóa và hành vi(KDQT) = Existing
1.000	3.402	Quản lý quốc tế: Quản trị đa văn hóa và hành vi(KDQT) = Existing -> Xã hội học(KDQT) = Existing
1.000	3.398	Quản lý công nghệ(QTDN) = Existing -> Giao tiếp kinh doanh và thuyết trình(QTDN) = Existing
1.000	3.398	Giao tiếp kinh doanh và thuyết trình(QTDN) = Existing -> Quản lý công nghệ(QTDN) = Existing
1.000	3.398	Quản lý công nghệ(QTDN) = Existing, Xã hội học(QTDN) = Existing -> Giao tiếp kinh doanh và thuyết trình(QTDN) = Existing
1.000	3.398	Giao tiếp kinh doanh và thuyết trình(QTDN) = Existing, Xã hội học(QTDN) = Existing -> Quản lý công nghệ(QTDN) = Existing
0.988	3.397	Xã hội học(KDQT) = Existing, Quản lý quốc tế: Quản trị đa văn hóa và hành vi(KDQT) = Existing -> Quản trị chuỗi cung ứng quốc tế(KDQT) = Existing
0.988	3.397	Quản lý quốc tế: Quản trị đa văn hóa và hành vi(KDQT) = Existing -> Quản trị chuỗi cung ứng quốc tế(KDQT) = Existing
0.988	3.397	Xã hội học(KDQT) = Existing -> Quản trị chuỗi cung ứng quốc tế(KDQT) = Existing
0.991	3.394	Quản lý công nghệ(QTDN) = Existing, Xã hội học(QTDN) = Existing -> Quản trị chất lượng(QTDN) = Existing
0.991	3.394	Giao tiếp kinh doanh và thuyết trình(QTDN) = Existing -> Kỹ năng quản trị(QTDN) = Existing
0.991	3.394	Giao tiếp kinh doanh và thuyết trình(QTDN) = Existing -> Quản trị chất lượng(QTDN) = Existing
0.991	3.394	Quản lý công nghệ(QTDN) = Existing, Giao tiếp kinh doanh và thuyết trình(QTDN) = Existing -> Kỹ năng quản trị(QTDN) = Existing
0.991	3.394	Quản lý công nghệ(QTDN) = Existing -> Kỹ năng quản trị(QTDN) = Existing
0.991	3.394	Quản lý công nghệ(QTDN) = Existing, Giao tiếp kinh doanh và thuyết trình(QTDN) = Existing -> Quản trị chất lượng(QTDN) = Existing
0.991	3.394	Giao tiếp kinh doanh và thuyết trình(QTDN) = Existing, Xã hội học(QTDN) = Existing -> Quản trị chất lượng(QTDN) = Existing
0.991	3.394	Giao tiếp kinh doanh và thuyết trình(QTDN) = Existing, Xã hội học(QTDN) = Existing -> Kỹ năng quản trị(QTDN) = Existing
0.991	3.394	Quản lý công nghệ(QTDN) = Existing -> Quản trị chất lượng(QTDN) = Existing
0.991	3.394	Quản lý công nghệ(QTDN) = Existing, Xã hội học(QTDN) = Existing -> Kỹ năng quản trị(QTDN) = Existing
1.000	3.392	Giao dịch và đàm phán kinh doanh(KTQT) = Existing -> Kinh doanh quốc tế(KTQT) = Existing
1.000	3.392	Kinh doanh quốc tế(KTQT) = Existing -> Giao dịch và đàm phán kinh doanh(KTQT) = Existing

Hình 3.5 L2.1: minsupp= 0.03, minprobability= 0.54



Hình 3.6 L2.2: $\text{minsupp}=0.03$, $\text{minprobability}=0.9$

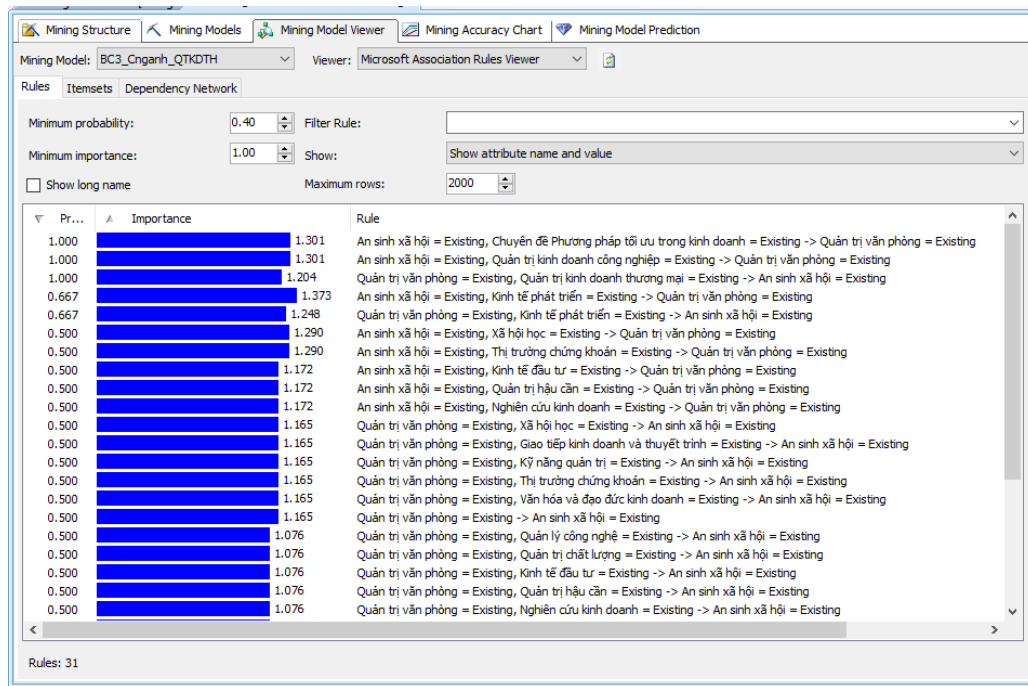
Nhận xét: Tất cả luật thu được ở lần chạy 2 đều như mong muốn, tăng $\text{minprobability}=0.9$, cũng thu được 413 luật cho nhiều chuyên ngành với xác suất cao, nhưng không đủ cho tất cả các chuyên ngành, hơn nữa muốn tư vấn theo chuyên ngành thì phải dùng công cụ lọc (*Filter Rule*) theo mã chuyên ngành, không có ý nghĩa với bài toán, Lần chạy 2 không khả thi, tiến hành thử nghiệm lần 3.

Lần 3: Vẫn là dữ liệu đã được loại bỏ học phần bắt buộc như lần chạy thứ hai và có một vài thay đổi như sau:

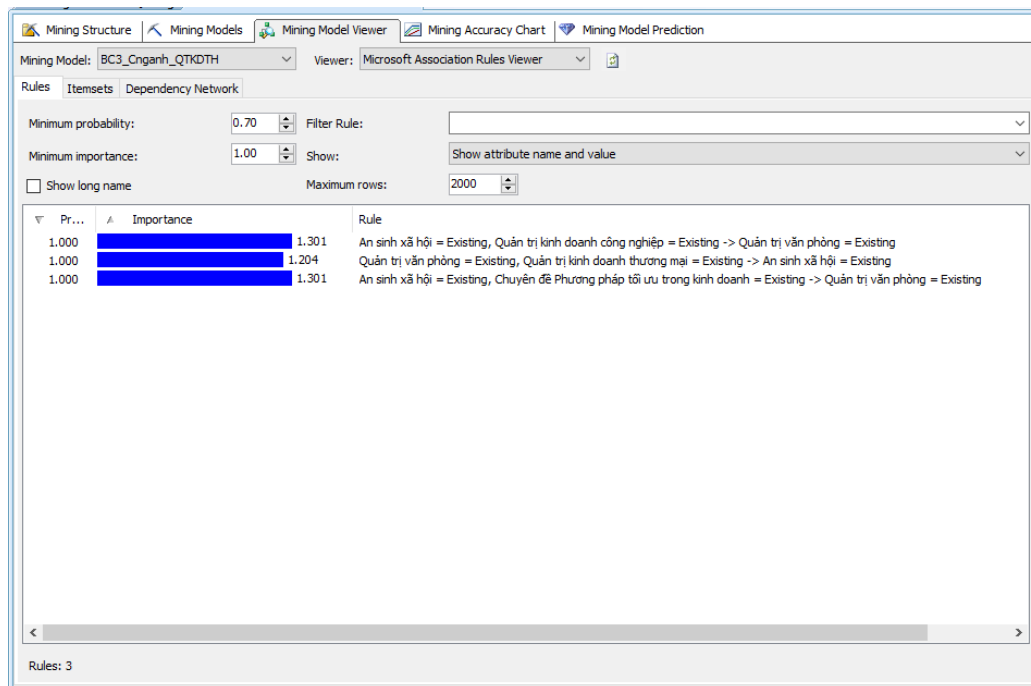
Dữ liệu mới lúc này được tách ra mỗi chuyên ngành một bộ 2 view thể hiện sinh viên, môn học của chuyên ngành đó và quá trình đăng ký các học phần tự chọn. Tiến hành chạy thử với chuyên ngành Quản trị kinh doanh tổng hợp.

Kết quả: chạy với $\text{minsupport}=0.01$, $\text{minprobability}=0.4$, thu được 31

Sau đó tăng $\text{minprobability}=0.7$, giữ nguyên $\text{minsupport}=0.01$, thu được 3 luật với xác suất xảy ra là 100%.



Hình 3.7 L3.1: minsupport=0.01, minprobability= 0.4



Hình 3.8 L3.2: minsupport=0.01, minprobability= 0.7

Nhận xét: Dễ dàng nhận thấy kết quả các luật thu được trong lần 3 này là phù hợp với ý nghĩa bài toán đã phát biểu nhất, có giá trị đề tư vấn cho sinh viên đăng ký các học phần tự chọn của chuyên ngành quản trị kinh doanh tổng hợp. Có ý nghĩa rất phù hợp với yêu cầu bài toán 1, làm cơ sở thông tin cho cố vấn học tập tư vấn đăng ký môn học cho sinh viên.

Ví dụ Luật : An sinh xã hội, Quản trị kinh doanh công nghiệp → Quản trị kinh doanh văn phòng (xảy ra với xác suất 100%)

Phát biểu luật tư vấn: Sinh viên khi đăng ký môn *An sinh xã hội* trong tổ hợp 1 kiến thức lựa chọn của ngành, và đăng ký môn *Quản trị kinh doanh công nghiệp* trong tổ hợp 5 kiến thức lựa chọn của ngành, thì thường sẽ đăng ký môn *Quản trị kinh doanh văn phòng* trong tổ hợp 6 kiến thức lựa chọn của ngành với xác suất là 100%.

- Do đó với các lần thử nghiệm tiếp theo chỉ chạy với bộ dữ liệu như lần thử nghiệm thứ 3 riêng cho các chuyên ngành và ngành khác nhau. Luận văn đã chọn những ngành, chuyên ngành có số lượng sinh viên lớn nhất để chạy thử nghiệm, kết quả thu được rất hữu ích cho cán bộ cố vấn tư vấn đăng ký môn học cho sinh viên. Phần kết quả cho các chuyên ngành khác và phát biểu luật tư vấn được trình bày trong phần phụ lục.

Kết luận thực nghiệm giải bài toán 1: Thử nghiệm được tiến hành nhiều lần với nhiều chuyên ngành khác nhau và kết quả thu được có ý nghĩa với bài toán 1 giúp cán bộ cố vấn học tập có thêm cơ sở tư vấn lựa chọn môn học cho sinh viên (*bảng kết quả tập luật lưu trong phần phụ lục*).

3.4. Khai phá dữ liệu bằng cây quyết định giải bài toán 2

3.4.1. Từ dữ liệu thô thu thập được

Dựa vào yêu cầu bài toán 2 để phân loại dự báo sinh viên có ra trường đúng hạn hay không và dựa vào quy chế đào tạo về số lượng tín chỉ, xếp loại học lực mỗi kỳ, mô hình cây quyết định được xây dựng để giải bài toán 2 sử dụng thông tin đầu vào liên quan đến thuộc tính số tín chỉ, xếp loại học lực từ điểm trung bình chung.

Lớp chuyên ngành KTQT 54A, KY 2010_2011_1																						
TT	Mã SV	Họ tên	GT	0	1	2	3	4	5	6	7	8	9	10	11	12	13	TC K5	TC TLK5	TBC K5	TBC TLK5	
1	11120001	Nguyễn Ngọc Anh	Nữ										7.3	5				9	85	2.66	1.98	
2	11120002	Nguyễn Tuấn Anh	Nam		5								3	5				6	84	0.75	2.17	
3	11120003	Trần Văn Dũng	Nam											5			6.8	5	84	1.9	2.07	
4	11120004	Đặng Ngọc Dương	Nam													5		6	83	1.83	2.18	
5	11120005	Nguyễn Tài Hạnh	Nữ													0	0		73	0	2.24	
6	11120006	Lê Quang Hiệp	Nam																65	0	2.12	
7	11120007	Trịnh Tiến Hiếu	Nam			6.3	5.3						6.2	5				15	77	2.14	2.38	
8	11120008	Lê Ngọc Hùng	Nam						8					5				5	89	1.64	2.01	
9	11120009	Bùi Văn Huy	Nam							7.2						6.9		5	80	2.8	2.16	
10	11120010	Nguyễn Thanh Hưng	Nam	0			0									0			85	0	1.98	
11	11120011	Nguyễn Cảnh Linh	Nam										4.8			7.6		4	84	2	2.16	
12	11120012	Nguyễn Hữu Nguyên	Nam							7.4	6.2							7	92	2.43	2.58	
13	11120013	Lưu Vũ Phi	Nam			0											4.1	2	40	0.4	2.11	
14	11120014	Phạm Văn Sự	Nam													7.3	0	2	65	1.5	1.98	
15	11120015	Nguyễn Thương	Nam													8		2	77	3.5	2.07	

Hình 3.10 Bảng điểm từng chuyên ngành theo kỳ sau khi biến đổi

3.4.2. Tiến hành biến đổi dữ liệu theo bài toán 2

Dữ liệu được biến đổi sang dạng bảng điểm từng lớp chuyên ngành theo kỳ (các môn học được quay lên các cột, mỗi sinh viên cùng với điểm là một bản ghi),

- Làm tương tự với các kỳ và chuyên ngành khác, cuối cùng gộp tất cả các chuyên ngành theo kỳ, dữ liệu tổng kết tương ứng với từng kỳ thu được như sau:

	A	B	C	D	AA	AB	AC	AD
1	QTKDTH 54A, KỲ 2010_2011_1							
2								
3	TT	Mã SV	Họ tên	GT	TC K5	TC TLK5	TBC K5	TBC TLK5
4	1	11120001	Nguyễn Ngọc Anh	Nữ	9	85	2.66	1.98
5	2	11120002	Nguyễn Tuấn Anh	Nam	6	84	0.75	2.17
6	3	11120003	Trần Văn Dũng	Nam	5	90	1.9	2.07
7	4	11120004	Đặng Ngọc Dương	Nam	6	83	1.83	2.18
8	5	11120005	Nguyễn Tài Hạnh	Nữ		73	0	2.24
9	6	11120006	Lê Quang Hiệp	Nam		65	0	2.12
10	7	11120007	Trịnh Tiến Hiếu	Nam	18	77	2.14	2.38
11	8	11120008	Lê Ngọc Hùng	Nam	5	89	1.64	2.01
12	9	11120009	Bùi Văn Huy	Nam	5	80	2.8	2.16
13	10	11120010	Nguyễn Thanh Hưng	Nam		85	0	1.98
14	11	11120011	Nguyễn Cảnh Linh	Nam	4	84	2	2.16
15	12	11120012	Nguyễn Hữu Nguyên	Nam	7	92	2.43	2.58
16	13	11120013	Lưu Vũ Phi	Nam	2	90	0.4	2.11
17	14	11120014	Phạm Văn Sự	Nam	2	65	1.5	1.98
18	15	11120015	Nguyễn Thương	Nam	2	77	3.5	2.07
19	16	11120016	Lê Bá Toán	Nam		89	0	2.04
20	17	11120017	Đặng Anh Tuấn	Nam	2	80	1.5	2.00

Hình 3.11 Bảng điểm tổng kết của một kỳ, tất cả chuyên ngành (ví dụ là kỳ 5)

- Theo bài toán phân lớp dự đoán sinh viên có ra trường đúng hạn hay không, thực tế năm thứ 3 trở đi sinh viên thường hay đi làm thêm và có nhiều nguyên nhân dẫn đến lười học. Cán bộ cố vấn thường xuyên phải theo dõi 3 kỳ cuối. Do đó 3 mô hình khai phá được đề xuất cho 3 kỳ cuối là kỳ 5, kỳ 6, kỳ 7 được xây

dựng dựa trên cây quyết định. Dự báo kịp thời vào 3 kỳ cuối có ý nghĩa với bài toán 2 đã phát biểu ở trên.

- Bảng điểm của 3 kỳ 5, 6, 7 (*dạng số*) sau khi thu được sẽ được biết đổi về dạng rời rạc (*các giá trị rời rạc*), 4 thuộc tính đầu vào (*input*) được thay bằng giá trị rời rạc như sau:

+ TCKy5, TCKy6, TCKy7 mà lớn hơn 15 tín chỉ là giá trị “đủ”, nhỏ hơn 15 tín chỉ là giá trị “không đủ”

+ TCTichLuyK5 mà lớn hơn 75 tín chỉ thì nhận giá trị “đủ”, nhỏ hơn 75 tín chỉ nhận giá trị “không đủ”

+ TCTichLuyK6 mà lớn hơn 90 tín chỉ thì nhận giá trị “đủ”, nhỏ hơn 90 tín chỉ nhận giá trị “không đủ”

+ TCTichLuyK7 mà lớn hơn 105 tín chỉ thì nhận giá trị “đủ”, nhỏ hơn 112 tín chỉ là “không đủ”

+ HlucKy5,6,7 và HlucDenKy5,6,7: nằm trong các khoảng sau: $3.6 < \text{xuất sắc} < 4$ | $3.2 < \text{Giỏi} < 3.6$ | $2.5 < \text{Khá} < 3,5$ | $2 < \text{tb} < 2.5$ | < 2 là Yếu.

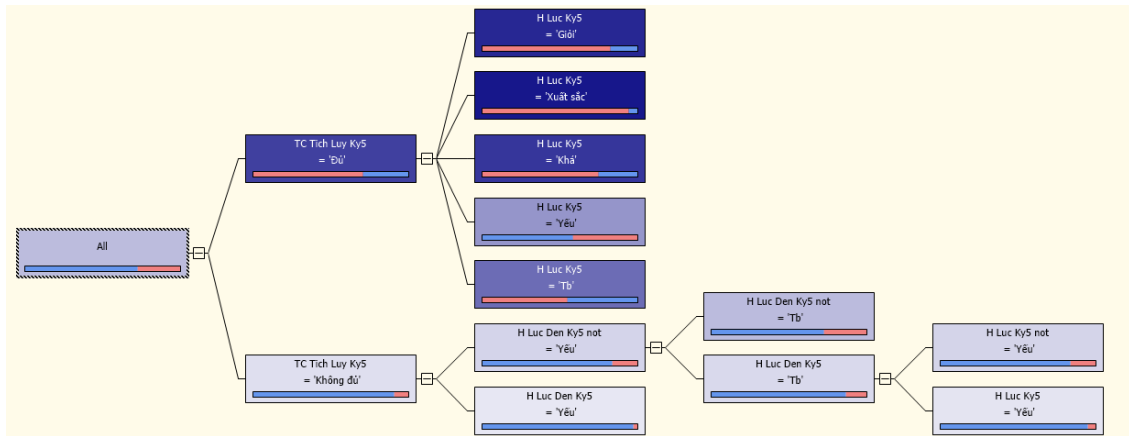
- Thêm cột thuộc tính dự báo được lấy từ dữ liệu là cột: ” tình trạng sinh viên” với 2 giá trị phân lớp (*đúng hạn, không đúng hạn*), dữ liệu sau khi biến đổi cuối cùng để đưa vào công cụ khai phá có dạng sau (*bốn cột thuộc tính đầu vào, một cột dự báo*):

	A	B	C	D	E	F
1	MaSV	TCKy1	HLucKy1	TCTichLuyKy1	HLucDenKy1	Dự Báo
2	11120001	Không đủ	Yếu	Không đủ	Yếu	Không đúng hạn
3	11120002	Không đủ	Khá	Không đủ	Khá	Không đúng hạn
4	11120003	Đủ	Tb	Đủ	Tb	Không đúng hạn
5	11120004	Không đủ	Yếu	Không đủ	Yếu	Không đúng hạn
6	11120005	Không đủ	Khá	Không đủ	Khá	Không đúng hạn
7	11120006	Không đủ	Tb	Đủ	Khá	Đúng hạn
8	11120007	Không đủ	Tb	Không đủ	Tb	Không đúng hạn
9	11120008	Không đủ	Yếu	Không đủ	Yếu	Không đúng hạn
10	11120009	Không đủ	Tb	Không đủ	Tb	Đúng hạn
11	11120010	Không đủ	Yếu	Không đủ	Yếu	Không đúng hạn
12	11120011	Không đủ	Khá	Không đủ	Khá	Không đúng hạn
13	11120012	Không đủ	Tb	Không đủ	Tb	Đúng hạn
14	11120013	Không đủ	Yếu	Không đủ	Yếu	Không đúng hạn
15	11120014	Không đủ	Tb	Không đủ	Tb	Không đúng hạn

Hình 3.12 Bảng dữ liệu đưa vào khai phá

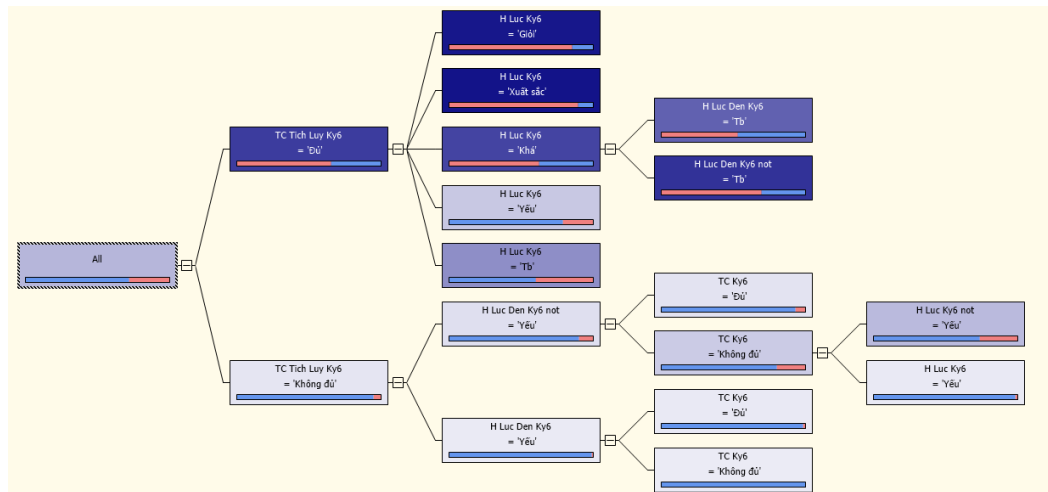
3.4.3. Thực hiện thử nghiệm trên công cụ BIDS

Lần 1: Mô hình phân lớp dự báo cho kỳ 5. Trên cây quyết định phân lớp thu được, phần xanh đậm là lớp dự đoán cho khả năng sinh viên ra trường đúng hạn.



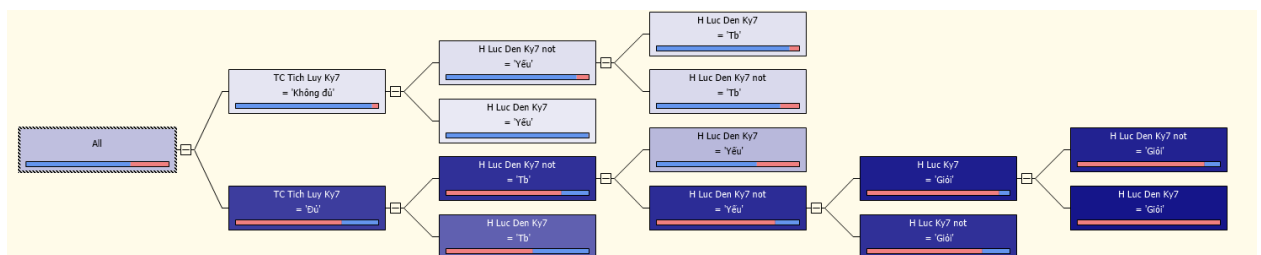
Hình 3.13 Cây quyết định phân lớp kỳ 5

Lần 2: Mô hình phân lớp dự báo cho kỳ 6



Hình 3.14 Cây quyết định phân lớp kỳ 6

Lần 3: Mô hình phân lớp dự báo cho kỳ 7



Hình 3.15 Cây quyết định phân lớp kỳ 7

Kết luận: Kết quả thu được sau khi thử nghiệm là 3 cây quyết định dùng để phân lớp dự đoán. Dựa vào 3 cây quyết định của kỳ 5, kỳ 6, kỳ 7, cán bộ cố vấn học tập sẽ phân loại được nhóm ra trường đúng hạn hay không đúng hạn. Có ý nghĩa làm cơ sở để cảnh báo kịp thời cho sinh viên, phù hợp với bài toán 2 đã phát biểu.

Từ cây quyết định có thể hiểu thành các luật như sau:

Ví dụ: Nếu TC Tích Luy Ky7 = 'Đủ' and H Luc Den Ky7 not = 'Tb' and H Luc Den Ky7 not = 'Yếu' and H Luc Den Ky7 not = 'Khá' thì ra trường Đúng hạn.

Kết quả bảng luật cụ thể có ý nghĩa góp phần giúp cán bộ cố vấn học tập ra quyết định cảnh báo học tập. Được lưu trong phần phụ lục.

3.5. Một số đề xuất, kiến nghị

- Qua thực nghiệm và thu được kết quả ý nghĩa với bài toán cố vấn học tập trong thực tiễn, có một số đề xuất kiến nghị như sau:
- Quy chế và chương trình đào tạo tại trường đại học kinh tế quốc dân nên được thống nhất giữa các khóa để dễ dàng theo dõi thống kê, quản lý dữ liệu, phục vụ cho việc phân tích khai phá tốt hơn, nhằm tìm ra các mẫu có ích với thực tiễn.
- Từ kết quả đạt được qua thực nghiệm trong luận văn, tác giả đề xuất Trường đại học kinh tế quốc dân nên xây dựng một hệ thống cố vấn học tập cho sinh viên chính quy. Giúp bộ phận cố vấn có thêm công cụ trực quan tư vấn cho sinh viên kế hoạch học tập sao cho phù hợp nhất.
- Hệ thống hoàn chỉnh gồm có cơ sở dữ liệu lớn và các bài toán cố vấn nhằm giải quyết tất cả những vấn đề học tập trong 4 năm học của sinh viên.
- Triển khai được hệ thống cũng là góp phần nâng cao việc ứng dụng công nghệ thông tin vào công tác quản lý chung của nhà trường. Nâng cao chất lượng phục vụ, đào tạo và học tập của sinh viên.

KẾT LUẬN

Sau một thời gian nghiên cứu và thực hiện đề tài, luận văn đã đạt được mục tiêu đã đề ra, thu được những kết quả ý nghĩa với thực tiễn.

Đã tóm tắt được lý thuyết liên quan đến phát hiện tri thức và khai phá dữ liệu, đặt biệt nhấn mạnh vào hai phương pháp khai phá dữ liệu cơ bản là luật kết hợp và cây quyết định.

Đã hiểu được quy định chung trong đào tạo theo học chế tín chỉ, những vấn đề còn tồn tại trong công tác cố vấn học tập, thu thập và tìm hiểu về dữ liệu quản lý đào tạo sinh viên đại học.

Đã đề xuất được bài toán mà mục tiêu là trợ giúp cho các hoạt động cố vấn học tập. Có thêm cơ sở thông tin cho cán bộ cố vấn học tập hoàn thành nhiệm vụ.

Sau khi áp dụng thử nghiệm trên công cụ BIDS để khai thác dữ liệu giải bài toán dựa vào kỹ thuật thuật cây quyết định và luật kết hợp đã thu được các kết quả có ý nghĩa với mục tiêu bài toán đã phát biểu.

Hạn chế:

Luận văn mới sử dụng dữ liệu của một khóa (do các khóa có các chương trình môn học và quy chế khác nhau)

Hướng phát triển:

- Nghiên cứu thêm các kỹ thuật khai phá dữ liệu và các công cụ khác nữa
- Phân tích sâu hơn về các phương pháp KPDL để lựa chọn phương pháp tối ưu nhất cho các bài toán cố vấn học tập.
- Thu thập và xử lý thêm dữ liệu của các khóa khác để tăng độ chính xác.
- Phát biểu thêm các bài toán cố vấn học tập khác nữa, nhằm có thêm nhiều cơ sở giúp ích cho hoạt động cố vấn học tập thêm ý nghĩa.
- Xây dựng một hệ thống gồm nhiều bài toán cố vấn học tập, hỗ trợ cho Trường và đội ngũ cố vấn, giúp ích cho nâng cao chất lượng đào tạo.