

KHAI PHÁ KẾT QUẢ TÍCH LŨY CHỨNG CHỈ MOS EXCEL CỦA SINH VIÊN TRƯỜNG ĐẠI HỌC KINH TẾ, ĐẠI HỌC HUẾ

Mai Thu Giang*

Ngày nhận bài: 20/01/2023

Ngày nhận bản sửa: 10/03/2023

Ngày duyệt đăng: 28/03/2023

Tóm tắt. Nghiên cứu trình bày kết quả khai phá dữ liệu kết quả tích lũy chứng chỉ Microsoft Office Specialist Excel (MOS Excel) của sinh viên Trường Đại học Kinh tế, Đại học Huế từ tháng 9/2017 đến nay. Đây là nghiên cứu đầu tiên về vấn đề này tại Việt Nam nói chung và Đại học kinh tế nói riêng. Kết quả nghiên cứu không chỉ có ý nghĩa thực tiễn đối người dạy, người học và Trường Đại học Kinh tế trong xu thế nâng cao chất lượng theo hướng chuẩn quốc tế, mà còn là bài học kinh nghiệm cho các cơ sở giáo dục đang bắt đầu tiếp cận chứng chỉ MOS. Để đạt được mục tiêu của nghiên cứu, chúng tôi sử dụng bộ phần mềm mã nguồn mở WEKA (Waikato Environment for Knowledge Analysis) với việc áp dụng kỹ thuật trích chọn thuộc tính để tìm ra được 04 thuộc tính có ảnh hưởng đến kết quả tích lũy cuối cùng và xây dựng mô hình dự báo phân lớp dựa trên giải thuật cây quyết định với kết quả phân lớp tốt nhất thuộc về mô hình giải thuật J48 với tỷ lệ phân lớp chính xác đạt trên 77%.

Từ khóa: Cây quyết định; MOS Excel; Phân lớp; Trích chọn thuộc tính.

1. Đặt vấn đề

Khai phá dữ liệu là trích xuất và khai thác những thông tin hữu ích, tiềm ẩn của dữ liệu. Công việc này giải quyết các vấn đề bằng cách phân tích lượng dữ liệu lớn hiện có để khám phá ra các xu hướng và các quy tắc có ý nghĩa (Brijesh, 2011). Triển khai công cụ khai phá dữ liệu là một cách để phân tích và quản lý khối lượng lớn dữ liệu sao cho có thể khám phá được các mô hình hữu ích cho giải quyết vấn đề và hỗ trợ ra quyết định (Yehuala, 2015). Baradwaj & Pal (2012) cho rằng kết quả học tập nói chung hay kết quả tích lũy chứng chỉ kỹ năng của sinh viên chịu ảnh hưởng của nhiều yếu tố như các đặc điểm riêng của từng cá nhân, đặc điểm kinh tế xã hội và các yếu tố liên quan đến môi trường sống. Biết rõ những yếu tố này và ảnh hưởng của chúng đến quá trình tích lũy kiến thức, kỹ năng của sinh viên có thể giúp cho không chỉ sinh viên mà cả các

* Trường Đại học Kinh tế, Đại học Huế, email: mtgiang@hce.edu.vn

nhà quản lý giáo dục triển khai công tác đào tạo một cách hiệu quả và đáp ứng tốt yêu cầu xã hội.

Thực tế cho thấy, nhu cầu tích lũy chứng chỉ kỹ năng MOS Excel của sinh viên Đại học Kinh tế, Đại học Huế ngày càng tăng bởi vì nhiều lý do. Thứ nhất, sử dụng thành thạo tin học văn phòng ngày càng quan trọng khi các kỹ năng đó góp phần giúp người dùng làm việc và học tập hiệu quả hơn (Certiport, 2022). Thứ hai, Formby & cộng sự (2017) tin rằng việc thành thạo kỹ năng tin học văn phòng sẽ đem đến nhiều cơ hội việc làm cho sinh viên sau khi tốt nghiệp, giúp họ trở thành ứng viên tiềm năng trong mắt nhà tuyển dụng, nhanh chóng tìm được công việc tốt trong tương lai sau khi ra trường. Việc sử dụng thành thạo tin học văn phòng sẽ giúp người học và người làm việc tối ưu được quá trình làm việc của họ. Thứ ba, rõ ràng không thể phủ nhận rằng chương trình học phần Tin học Ứng dụng hiện nay hay học phần Tin học Đại cương trước đây mà sinh viên trường Đại học Kinh tế, Đại học Huế đã và đang được học cung cấp cho sinh viên Đại học Kinh tế, Đại học Huế kiến thức cơ bản và toàn diện về máy tính điện tử, hệ thống thông tin và các thành phần của nó cùng với các kiến thức và kỹ năng từ cơ bản đến nâng cao đối với các phần mềm ứng dụng hỗ trợ các công việc hằng ngày như Microsoft Word, Microsoft Excel, Microsoft Powerpoint. Tuy nhiên, một số lượng lớn sinh viên Đại học Kinh tế nhận thức chứng chỉ là một phương tiện để làm minh chứng về mức độ thành thạo với các kỹ năng khả thi và cần thiết khi tham gia vào thị trường lao động như khẳng định trong kết quả nghiên cứu của Michael & Christopher (2005).

Theo báo cáo tổng kết của IIG Việt Nam, tại Việt Nam đã có nhiều cơ sở giáo dục đại học áp dụng chuẩn đầu ra chứng chỉ MOS cho sinh viên một số ngành hoặc cho toàn cơ sở giáo dục. Nhận thức được lợi ích của loại chứng chỉ này đem lại và nắm bắt được xu hướng và sự cần thiết của chứng chỉ MOS trong việc tích lũy kỹ năng đáp ứng đi làm, từ tháng 9/2017 trường Đại học Kinh tế là trường duy nhất trong Đại học Huế ký kết hợp tác với IIG Việt Nam trong việc tổ chức thi chứng chỉ MOS. Đến nay, số lượt sinh viên tự nguyện tham gia tích lũy chứng chỉ MOS lên đến xấp xỉ 5000 lượt sinh viên thông qua số tài khoản ôn tập Gmetrix giữa Trường và IIG Việt Nam. Trong số đó, số bài thi đăng ký trực tiếp thi tại địa điểm trường Đại học Kinh tế, Đại học Huế là trên 2985 bài thi.

Trên thế giới đã có một số nghiên cứu với mục đích chỉ ra tầm quan trọng của MOS Excel đối với sinh viên kinh tế (Nesrin Bakir, 2019). Bên cạnh đó, tác giả Ngo-Ye và cộng sự (2018) đã nghiên cứu hướng dẫn triển khai hiệu quả chương trình học lồng

ghép giảng dạy và tổ chức thi chứng chỉ MOS. Tuy nhiên, cho đến nay chưa có một nghiên cứu nào được thực hiện để khai phá kết quả học – thi chứng chỉ MOS Excel thực tế từ một tổ chức giáo dục đã triển khai để từ đó xác định được những đặc điểm nào hay những thuộc tính nào ảnh hưởng đến kết quả tích lũy cuối cùng của sinh viên Đạt hay Không đạt chứng chỉ; và mô hình dự báo nào có thể đưa ra được kết quả dự báo đó đạt tỷ lệ chính xác cao nhất. Để giải quyết những khoảng trống này, chúng tôi sẽ khai phá cơ sở dữ liệu đã được lưu trữ và thu thập trong quá trình tổ chức và hướng dẫn cho sinh viên của trường Đại học Kinh tế, Đại học Huế trong quá trình tích lũy chứng chỉ của họ.

2. Tổng quan nghiên cứu và phương pháp nghiên cứu

2.1 Tổng quan nghiên cứu

2.1.1 Nghiên cứu liên quan đến khai phá dữ liệu giáo dục

Romero & Sebastian (2007) đã cho rằng khai phá dữ liệu giáo dục là một lĩnh vực nghiên cứu tiềm năng và đầy hứa hẹn. Trong một nghiên cứu về sử dụng khai phá dữ liệu, Bayer & cộng sự (2012) đã kết hợp công cụ phân tích mạng xã hội với kỹ thuật khai phá dữ liệu bao gồm Cây quyết định và Naïve Bayes để dự báo kết quả học của sinh viên. Kết quả nghiên cứu cho thấy mô hình xây dựng trên giải thuật Cây quyết định (J48) mang lại tỷ lệ phân lớp chính xác hơn Naïve Bayes. Ngoài ra, Kapoor & Reena (2015) từng nêu trong công bố trong nghiên cứu của họ rằng, giải thuật cây quyết định J48 là một trong những giải thuật tốt nhất trong việc xây dựng mô hình dự báo phân lớp. Thêm vào đó, để đánh giá tỷ lệ phân lớp một cách chính xác, Kohavi cho thấy đánh giá chéo 10 lần là phương pháp tốt nhất khi xây dựng mô hình phân lớp mặc dù cần phải thực hiện các thao tác tính toán nhiều hơn (Kohavi, 1995).

Trong khai phá dữ liệu, trích chọn thuộc tính là một vấn đề nghiên cứu được nhiều tác giả quan tâm. Trong đó, giải thuật và đánh giá thuộc tính BestFirst–CfsSubsetEval của WEKA được nhiều tác giả lựa chọn. Ví dụ, Lei & Pingfan (1988) đã chứng minh rằng phương pháp lựa chọn thuộc tính theo BestFirst là một giải thuật lựa chọn thuộc tính tối ưu, cho ra tập thuộc tính ít hơn nhiều so với các phương pháp tìm kiếm khác. Thêm vào đó, Aggarwal & Amrita (2013) đã sử dụng CfsSubsetEval làm công cụ đánh giá một tập hợp con của các thuộc tính bằng cách xem xét khả năng riêng của từng thuộc tính cùng với mức độ dư thừa của chúng và đưa ra tập thuộc tính của mô hình với độ phân lớp chính xác tới 99,95%.

2.1.2 Nghiên cứu liên quan đến tầm quan trọng của việc thành thạo kỹ năng Microsoft Excel và tích lũy chứng chỉ tin học đối với sinh viên kinh tế

Nghiên cứu của Formby & cộng sự (2017) đã chỉ ra rằng hơn 80% sinh viên kinh tế khẳng định mục tiêu của họ là kiếm được một công việc được trả lương cao và nhiều doanh nghiệp đang yêu cầu nâng cao kỹ năng Microsoft Excel. Nghiên cứu đã chỉ rõ đối với sinh viên chuyên ngành kinh tế, kiến thức về bảng tính là bắt buộc để có khả năng thành công trong thị trường việc làm. Sử dụng thành thạo với Microsoft Excel ngày càng trở nên một điều kiện tiên quyết cho sinh viên khi tiếp cận thị trường việc làm. Bên cạnh đó, nghiên cứu của Gomillion chỉ ra rằng sinh viên chuyên ngành kinh tế được hưởng lợi từ việc tích lũy chứng chỉ kỹ năng công nghệ thông tin (Gomillion, 2017). Tuy nhiên, trong các nghiên cứu trên không đề cập đến việc sinh viên có thể dùng phương tiện nào để chứng minh cho nhà tuyển dụng về kỹ năng của mình. Nghiên cứu của Randall & Zirkle (2005) đã giải quyết điều đó. Nhóm nghiên cứu kết luận rằng, chứng chỉ là một phương tiện để cung cấp cho sinh viên để làm minh chứng về mức độ thành thạo với các kỹ năng khả thi và cần thiết khi tham gia vào thị trường lao động và chứng chỉ có ảnh hưởng đáng kể đến khả năng được tuyển dụng của nhân viên (Certiport, 2022).

Chứng chỉ MOS (Microsoft Office Specialist) là bài thi đánh giá khả năng tin học văn phòng uy tín nhất do Microsoft cấp và được Certiport triển khai. Chứng chỉ được dùng để đánh giá kỹ năng nghề nghiệp được công nhận trên toàn cầu, chứng minh được năng lực sử dụng các ứng dụng tin học văn phòng – công cụ hữu hiệu nâng cao hiệu quả và năng suất công việc. Chứng chỉ MOS đang được áp dụng trên 150 quốc gia và vùng lãnh thổ.

Nghiên cứu liên quan đến chứng chỉ MOS Excel: trên thế giới có nghiên cứu của tác giả Ngo-Ye & cộng sự (2018) đã chỉ ra những kinh nghiệm mà tổ chức giáo dục nơi tác giả làm việc có thể giúp sinh viên chuẩn bị cho dự thi chứng chỉ MOS Excel như thế nào. Ngoài ra, trong nghiên cứu về bài học kinh nghiệm của kết hợp chương trình giảng dạy với chương trình giảng dạy chứng chỉ MOS Excel của trường Đại học West Texas A&M University dành cho các sinh viên ngành kinh tế (Nesrin Bakir, 2019), trường đã đưa chương trình MOS vào giảng dạy cho sinh viên. Nghiên cứu này tập trung vào phân tích mối liên hệ và cách thức triển khai, kết hợp với chương trình hiện tại để giúp sinh viên đạt kết quả tích lũy chứng chỉ được cao nhất. Tuy nhiên, các nghiên cứu đã được thực hiện chưa chỉ ra được yếu tố hay thuộc tính nào trong cơ sở dữ liệu lưu trữ về người học có ảnh hưởng đến kết quả tích lũy chứng chỉ của sinh viên.

2.2 Phương pháp nghiên cứu

2.2.1 Thu thập số liệu và chuẩn hóa

Dữ liệu được thu thập từ dữ liệu hồ sơ lưu trữ quá trình học – thi MOS của sinh viên Đại học Kinh tế, Đại học Huế từ năm 2017 tháng 6/2022. Trong đó, để phục vụ cho mục đích nghiên cứu, dữ liệu về các bản ghi của sinh viên tham gia với mục tiêu tích lũy MOS Excel được lọc riêng. Việc gán nhãn “Đạt” hay “Không đạt” được thực hiện bằng cách sử dụng hàm Vlookup dựa vào khóa là số Chứng minh nhân dân hoặc Căn cước công dân (thông tin này được lưu trữ từ giai đoạn sinh viên đăng ký mua tài khoản học Gmetrix từ IIG Việt Nam) và kết quả IIG báo cáo sau khi thi. Nhãn Đạt được gán cho sinh viên thi “Pass” chứng chỉ MOS Excel với điểm số $\geq 700/1000$. Bản ghi được gán nhãn lớp “Không đạt” bao gồm bản ghi của các thí sinh có kết quả dự thi được IIG báo về là “Fail” hoặc mua tài khoản Gmetrix và tham gia học nhưng không tham gia dự thi.

Sau đó, thực hiện lệnh lọc ra kết quả 2905 bản ghi mang nhãn “Đạt” và được lưu trong file Dat.xlsx là bản ghi lưu trữ dữ liệu của những thí sinh tham gia học thi và thi đạt chứng chỉ MOS Excel của sinh viên Trường Đại học Kinh tế, Đại học Huế. Đồng thời, thực hiện lệnh lọc tiếp theo trên cơ sở dữ liệu gốc ra tập dữ liệu lưu bản ghi mang nhãn “Không đạt” được và lưu trong file Khongdat.xlsx gồm 1800 bản ghi.

Tiếp theo, với mục tiêu cân đối dữ liệu, nghiên cứu sử dụng hàm ngẫu nhiên trong Excel để lấy ngẫu nhiên 1800 bản ghi trong tập Dat.xlsx và kết hợp với cơ sở dữ liệu gồm 1800 bản ghi mang nhãn lớp Không đạt để được cơ sở dữ liệu sử dụng dùng để khai phá trên đó, đặt tên tập dữ liệu đó là Dulieu.xlsx. Cuối cùng, tập mang phần mở rộng .xlsx được chuyển lưu dưới dạng .csv file để thuận lợi cho quá trình áp dụng các kỹ thuật khai phá dữ liệu sử dụng WEKA.

Trong đó, dữ liệu thu được quản lý bao gồm các bảng dữ liệu liên quan đến 03 nhóm thông tin: Thông tin người học, Thông tin học tập và kết quả tích lũy. Thứ nhất, thông tin hồ sơ người học bao gồm: họ tên, số căn cước/chứng minh nhân dân, giới tính, thời điểm tham gia học là sinh viên năm thứ mấy? chuyên ngành, điểm môn tin học ứng dụng, điểm tự đánh giá kỹ năng sử dụng excel theo thang điểm 10. Tuy nhiên, trong cơ sở dữ liệu dùng để thực hiện khai phá, dữ liệu về họ tên, số căn cước công dân/chứng minh nhân dân được loại bỏ bởi vì dữ liệu là không mang tính đại diện cho nhóm. Thứ hai, thông tin lưu trữ về học tập bao gồm: thời gian tham gia học trực tiếp, số lần hoàn thành sản phẩm, số lần nộp bài sau buổi học, thời gian luyện tập gmetrix bài

1, thời gian luyện tập gmetrix bài 2, thời gian luyện tập gmetrix bài 3, thời gian hoàn thành kết quả tốt nhất bài 1, thời gian hoàn thành kết quả tốt nhất bài 2, thời gian hoàn thành kết quả tốt nhất bài 3, Kết quả tối đa đạt được bài 1, Kết quả tối đa đạt được bài 2, Kết quả tối đa đạt được bài 3, tổng số lần tham gia thi thử, trung bình thời gian hoàn thành, trung bình điểm kết quả thi thử đã đạt được. Cuối cùng, thông tin về kết quả tích lũy sẽ được quản lý bởi trường Ketqua. Giá trị của trường là “Đạt” hoặc “Không đạt”.

2.2.2 Xử lý số liệu

Ứng dụng phần mềm mở WEKA để tiến hành khai phá dữ liệu đã thu thập ở bước trên. Giải thuật BestFirst-CfsSubsetEval được sử dụng để trích chọn thuộc tính. Kỹ thuật phân lớp theo giải thuật cây quyết định được áp dụng tiếp sau đó để đánh giá tập dữ liệu thuộc tính thu được đối với kết quả dự báo phân lớp. Giải thuật cây quyết định được sử dụng để xây dựng mô hình dự báo kết quả tích lũy để đánh giá kết quả tập thuộc tính sau trích chọn.

Các giải thuật phân lớp cây quyết định được xây dựng trong có tên là Weka Decision Tree Tool bao gồm giải thuật J48, Decision Stump, HoeffdingTree, LMT, RandomForest, RandomTree và REPTree được Rashid Naseem & cộng sự (2020) phân tích chi tiết trong nghiên cứu của họ. Do đó, nghiên cứu này với mục đích áp dụng dữ liệu nghiên cứu đã thu thập được vào bộ công cụ có sẵn đó để so sánh kết quả phân lớp giữa các phương pháp nhau trong cùng một ứng dụng khai phá dữ liệu.

Cuối cùng, so sánh tỷ lệ dự báo phân lớp chính xác của các mô hình để từ đó lựa chọn tập thuộc tính tương ứng với mô hình phân lớp cho ra kết quả dự báo phân lớp với tỷ lệ chính xác cao nhất, dựa trên phương pháp đánh giá chéo 10 lần (Kohavi, 1995).

3. Kết quả và thảo luận

3.1 Kết quả mô tả về cơ sở dữ liệu lưu trữ

3.1.1 Kết quả thu thập theo giới tính

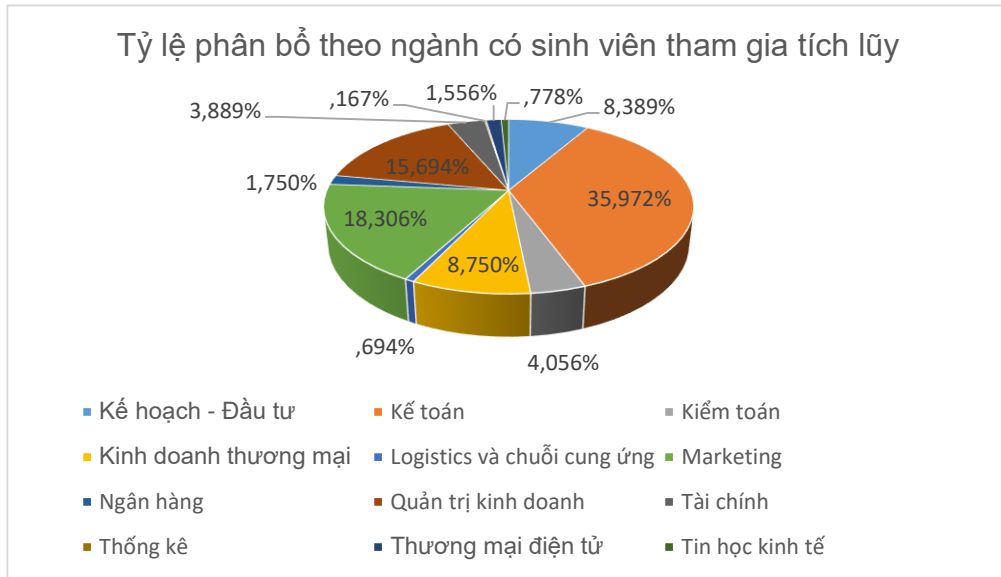
Bảng 1: Kết quả thu thập theo giới tính

Giới tính	Số lượng tích lũy
Nam	1089
Nữ	2511

Nguồn: Tổng hợp từ cơ sở dữ liệu nghiên cứu

Kết quả thể hiện trong Bảng 1 cho thấy, số lượng sinh viên Nữ tham gia tích lũy cao hơn rất nhiều so với số lượng sinh viên Nam. Cụ thể, với 2511 sinh viên nữ tham gia tích lũy trong khi đó nam giới là 1089.

3.1.2 Kết quả về phân bố theo ngành

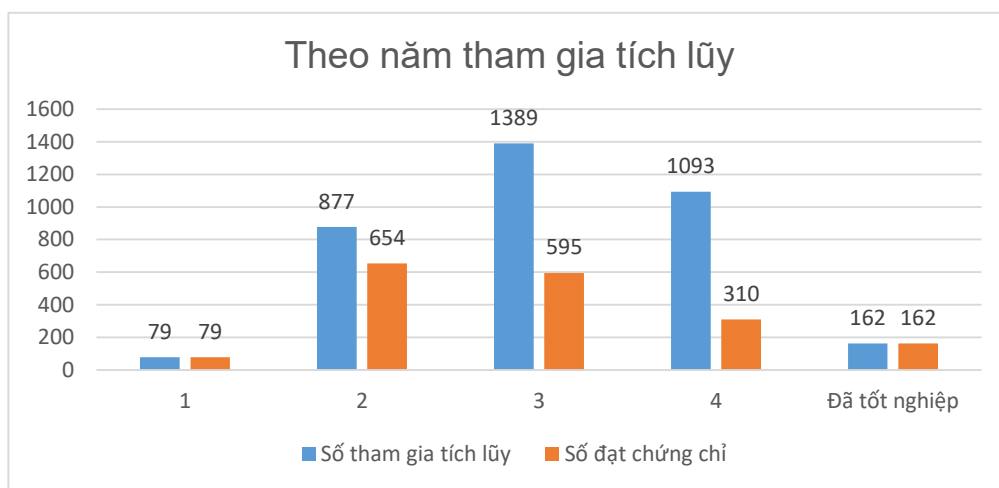


Hình 1: Tỷ lệ phân bố theo ngành có sinh viên tham gia tích lũy

Nguồn: Tổng hợp từ cơ sở dữ liệu nghiên cứu

Kết quả trình bày trong Hình 1 cho thấy, tỷ lệ sinh viên ngành Kế toán, Marketing, và Quản trị kinh doanh chiếm tỷ lệ lớn trong tổng số sinh viên trường Đại học Kinh tế, Đại học Huế đã tích lũy chứng chỉ, với tỷ lệ chiếm 35.97%, 18.31% và 15.69% theo thứ tự. Tiếp theo đó là sinh viên ngành Kinh doanh thương mại và Kế hoạch đầu tư với tỷ lệ tương ứng chiếm 8.75 và 8.39% trong tổng số sinh viên đã tích lũy. Sinh viên ngành Tin học Kinh tế, Logistic và chuỗi cung ứng và Thống kê kinh doanh có số sinh viên tích lũy chứng chỉ ít nhất với 0.17%, 0.69% và 0.78% theo thứ tự.

3.1.3 Kết quả thống kê về thời điểm tham gia tích lũy chứng chỉ tương ứng với số lượng tham gia và số lượng đạt chứng chỉ:

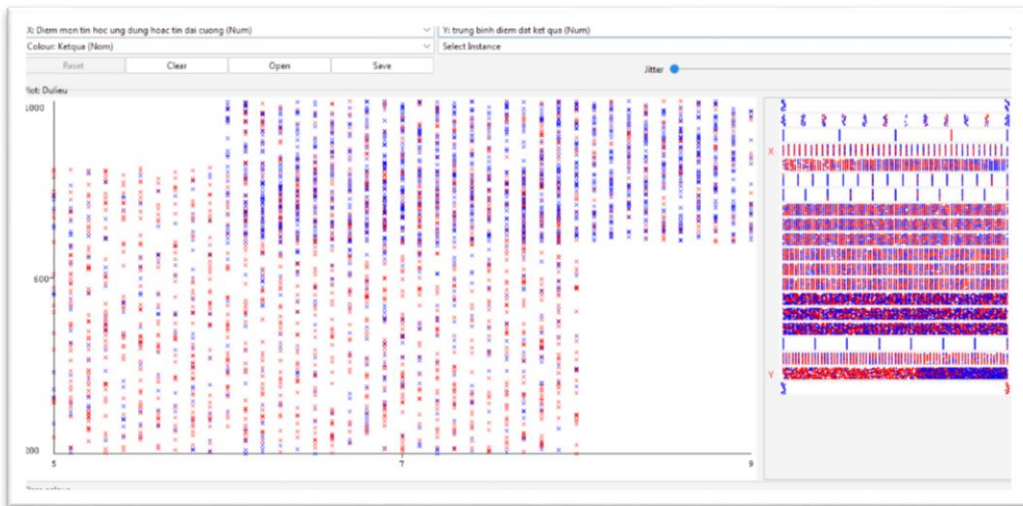


Hình 2: Thời điểm tham gia tích lũy, số lượng tham gia và số lượng đạt chứng chỉ

Nguồn: Tổng hợp từ cơ sở dữ liệu nghiên cứu

Tại Hình 2 thể thấy rằng nhu cầu tích lũy khi sinh viên học năm 3,4 cao hơn nhiều so với sinh viên năm 1,2 và sinh viên đã tốt nghiệp. Tuy nhiên, tỷ lệ sinh viên năm 2 tích lũy chứng chỉ thành công cao hơn nhiều so với nhóm sinh viên năm 3,4. Đặc biệt, mặc dù số sinh viên năm 1 và số sinh viên đã tốt nghiệp tham gia khóa học với số lượng thấp hơn nhiều các nhóm còn lại, chỉ với 78 và 162 sinh viên, nhưng hai nhóm sinh này đều có tỷ lệ đều đạt kết đạt 100%. Điều này có thể là vì khi sinh viên năm 1 vừa được học chương trình Tin học ứng dụng hoặc tin học đại cương (trước đây) và tham gia thi MOS Excel thì nền tảng đã được trang bị trong chương trình học đã giúp ích rất nhiều khi sinh viên tích lũy chứng chỉ MOS. Trong trường hợp đối với nhóm sinh viên đã tốt nghiệp và sinh viên có nhu cầu tích lũy chứng chỉ, tỷ lệ sinh viên tham gia thi và đạt kết quả 100% có thể là vì sinh viên đã đi làm luôn ý thức việc học và tích lũy chứng chỉ là cần thiết nên đầu tư thời gian và nỗ lực để đạt kết quả tích lũy như mong đợi.

3.1.4 Kết quả thể hiện mối liên hệ giữa điểm môn tin học ứng dụng/tin học đại cương với điểm trung bình làm các bài thi thử MOS



Hình 3: Mối liên hệ giữa điểm môn Tin học ứng dụng/Tin học đại cương đến điểm làm đề thi thử MOS

Nguồn: Kết quả phân tích, tổng hợp từ WEKA trên cơ sở dữ liệu nghiên cứu

Kết quả Hình 3 thể hiện trực hoành của biểu đồ là biểu diễn dữ liệu điểm trung bình môn tin học đại cương hoặc tin học ứng dụng của sinh viên tham gia tích lũy thường tập trung trong khoảng điểm từ 5 đến 9 theo hệ số 10. Trong khi đó, trục tung của biểu đồ biểu diễn điểm số đạt được trung bình khi tham gia thi thử với miền giá trị từ 100 đến 1000. Kết quả mối liên hệ giữa điểm trung bình môn Tin học đại cương hoặc tin học ứng dụng và kết quả trung bình đạt được khi thi thử cho thấy sinh viên có điểm trung bình môn học Tin học đại cương hoặc tin học ứng dụng từ 6.0 trở lên thì tỷ lệ thi có điểm trên 700 chiếm đa số, tương ứng với kết quả đạt (màu xanh dương) phủ dày đặc hơn so với với kết quả điểm trung bình tin học đại cương và tin học ứng dụng dưới 6.0.

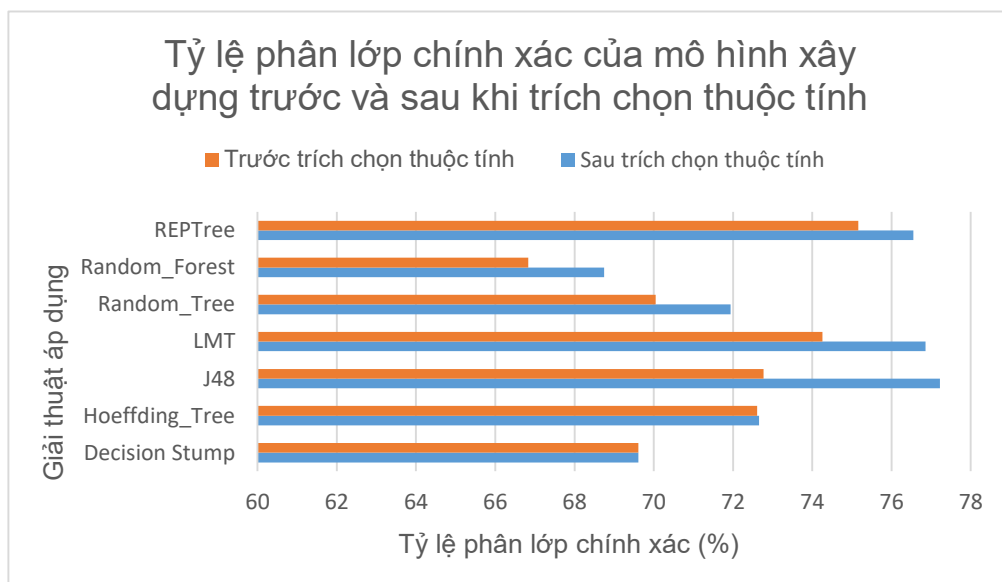
3.2 Kết quả trích chọn thuộc tính và mô hình dự báo phân lớp dựa trên giải thuật cây quyết định

3.2.1 Kết quả trích chọn thuộc tính

Giải thuật BestFirst-CfsSubsetEval được sử dụng để trích chọn thuộc tính. Kết quả cho thấy 04 thuộc tính được giữ lại bao gồm: Thời điểm điểm tham gia tích lũy (học năm thứ mấy), thời gian học trực tiếp, tổng số lần tham gia làm bài thi thử, và trung bình kết quả thi thử đạt được.

3.2.2 Kết quả mô hình dự báo phân lớp dựa trên giải thuật cây quyết định

Kết quả tỷ lệ phân lớp chính xác áp dụng giải thuật cây quyết định cung cấp bởi WEKA cho tập dữ liệu trước và sau khi trích chọn thuộc tính như sau:



Hình 5: So sánh tỷ lệ phân lớp chính xác giữa các mô hình phân lớp

Nguồn: Kết quả tỷ lệ phân lớp chính xác dựa trên các mô hình được chạy bởi WEKA trên cơ sở dữ liệu nghiên cứu

Giải thuật Cây quyết định được sử dụng để xây dựng mô hình dự báo phân lớp Kết quả. Các giải thuật này bao gồm J48, Decision Stump, HoeffdingTree, LMT, RandomForest, RandomTree và REPTree. Mô hình dự báo được xây dựng đồng thời trên các tập dữ liệu trước và sau khi trích chọn thuộc tính để so sánh và kiểm chứng ý nghĩa về kết quả của trích chọn thuộc tính. Đánh giá phân lớp chính xác của mô hình đối với cả tập dữ liệu trước và sau trích chọn đều dựa trên phương pháp đánh giá chéo 10 lần (Kohavi, 1995).

Hình 5 cho kết quả về tỷ lệ phân lớp chính xác của các mô hình được áp dụng dựa trên tập dữ liệu trước và sau trích chọn thuộc tính. Kết quả trước trích chọn là thanh màu vàng; kết quả sau trích chọn thể hiện bởi thanh màu xanh.

Kết quả cho thấy, kết quả mô hình phân lớp sau khi trích chọn thuộc tính luôn cho kết quả phân lớp chính xác bằng hoặc cao hơn tập giải thuật. Ví dụ, đối với giải thuật REPTree, tỷ lệ mô hình phân lớp trên tập dữ liệu sau khi trích chọn thuộc tính cho độ chính xác cao hơn so với tập dữ liệu gốc là xấp xỉ 2%. Tương tự, giải thuật Random-

Tree, Random_Forest, LMT, J48 mô hình được xây dựng dựa dựa trên tập thuộc tính trích chọn cao hơn tập thuộc tính ban đầu từ 2.5-4%. Chỉ giải thuật Decision Stump, Hoeffding_Tree không thể hiện sự khác nhau nhiều về tỷ lệ phân lớp của mô hình xây dựng trước và sau trích chọn.

Trong tất các các giải thuật được áp dụng, mô hình phân lớp dựa trên giải thuật J48 cho ra kết quả phân lớp chính xác sau trích chọn cao nhất với trên 77%. Tiếp theo là mô hình được xây dựng dựa trên giải thuật LMT và Rep_Tree với xấp xỉ 76.5%.

4. Kết luận và đề xuất

Thứ nhất, kết quả nghiên cứu cho thấy sự khác biệt lớn trong tỷ lệ tích lũy thành công chứng chỉ MOS đối với nhóm sinh viên nam và nhóm sinh viên nữ. Nghiên cứu đề tìm ra lý do của sự khác biệt này sẽ có vai trò quan trọng trong việc nâng cao kỹ năng đồng bộ cho cả sinh viên nam và nữ khi tốt nghiệp.

Thứ hai, số lượng sinh viên chuyên ngành Kế toán, Quản trị kinh doanh và Marketing là 03 chuyên ngành có tỷ lệ sinh viên tham gia tích lũy nhiều nhất trong toàn trường trong giai đoạn từ khi trường Đại học Kinh tế, Đại học Huế phối hợp tổ chức thi MOS cho sinh viên từ tháng 9/2017 tới nay. Thiết nghĩ, khi Trường Đại học Kinh tế, Đại học Huế đang từng bước nâng cao chuẩn đầu ra Tin học cho sinh viên và tìm kiếm chuyên ngành sinh viên có nhu cầu tích lũy chứng chỉ MOS để từng bước áp chuẩn cho một số ngành thì đây sẽ là cơ sở để các nhà quản lý có thể tham khảo khi đưa ra quyết định.

Thứ ba, kết quả điểm tin học ứng dụng hay tin học đại cương của sinh viên có ảnh hưởng lớn đến kết quả tích lũy chứng chỉ cuối cùng (Đạt chứng chỉ) của sinh viên. Như vậy, việc tạo nền tảng tốt cho sinh viên ngay từ khi tham gia học môn học cơ bản là cơ sở tốt để sinh viên có thể tích lũy chứng chỉ tin học quốc tế MOS Excel thành công. Nghiên cứu này cũng đưa ra hướng cần suy nghĩ đó là việc kết hợp giữa chương trình học phần Tin học ứng dụng giảng dạy hiện tại với chương trình tích lũy chứng chỉ MOS để nâng cao kỹ năng cho toàn bộ sinh viên chứ không chỉ có nhóm sinh viên tích lũy theo nhu cầu.

Thứ tư, trong 19 thuộc tính được lưu trữ trong cơ sở dữ liệu của 3600 bản ghi sinh viên từng tham gia tích lũy chứng chỉ MOS tại Trường Đại học Kinh tế, Đại học Huế, sau khi được áp dụng giải thuật trích chọn thuộc tính, chỉ 04 thuộc tính được giữ lại bao gồm Thời điểm điểm tham gia tích lũy (học năm thứ mấy), thời gian học trực tiếp, tổng

số lần tham gia làm bài thi thử, và trung bình kết quả thi thử đạt được. Ngược lại, các thuộc tính khác như kết quả luyện tập tài khoản Gmetrix về thời gian và kết quả không được giữ lại sau quá trình trích chọn.

Thứ năm, mặc dù mô hình phân lớp Cây quyết định áp dụng trên tập thuộc tính đã trích chọn là cao hơn so với tập thuộc tính lưu trữ từ đầu. Tuy nhiên, tỷ lệ dự báo phân lớp chính xác mới đang đạt trên 77%. Trong đó, mô hình phân lớp cây quyết định dựa trên giải thuật J48 đạt tỷ lệ phân lớp cao nhất với 77.2%.

Cuối cùng, trong khi chứng chỉ Tin học quốc tế MOS chưa được coi là chuẩn đầu ra về tin học của sinh viên trường Đại học Kinh tế, Đại học Huế. Nhưng với nguyện vọng tích lũy chứng chỉ và kỹ năng đi làm, Trường Đại học Kinh tế đã tạo điều kiện cho sinh viên được thi và tích lũy bằng cách phối hợp với IIG Việt Nam trong việc tổ chức thi từ năm 2017 đến nay. Kết quả nghiên cứu là tài liệu hữu ích cho Trường Đại học Kinh tế, Đại học Huế trong việc nhìn nhận kết quả và thực trạng của việc tích lũy chứng chỉ tin học quốc tế nói chung và chứng chỉ MOS Excel nói riêng của sinh viên trường Đại học Kinh tế, Đại học Huế để từ đó có giải pháp cho việc thúc đẩy, triển khai nâng cao chất lượng đầu ra của sinh viên với mục tiêu đáp ứng yêu cầu nguồn nhân lực xã hội yêu cầu. Ngoài ra, đối với các cơ sở giáo dục đang có nguyện vọng triển khai và hỗ trợ sinh viên tiếp cận với chứng chỉ quốc tế MOS có thể tham khảo kết quả nghiên cứu làm bài học kinh nghiệm trong việc triển khai cho cơ sở đào tạo của mình để kết quả tích lũy của sinh viên đạt cao nhất có thể.

TÀI LIỆU THAM KHẢO

- Aggarwal, M., & Amrita. (2013). Performance Analysis Of Different Feature Selection Methods In Intrusion Detection. *Educational Data Mining*.
- Baradwaj, B., & Pal, S. (2012). Mining educational data to analyze students' performance. *IJACSA* 2:, 4(4), 63-69.
- Bayer, J., Bydzovska, H., & G'eryk, J. (2012). Predicting drop-out from social behaviour of students. *International Data Mining*.
- Brijesh, B. (2011, 6). Mining Educational Data to Analyze Students' Performance. *International Journal of Advanced Computer Science and Applications*, 5(7), 65-75.
- Certiport. (2022). *Microsoft Office success stories*. (Microsoft) Retrieved 5 6, 2022, from <https://certiport.pearsonvue.com/Certifications/Microsoft/MOS/Success-stories>
- Dekker, G., & Pechenizkiy, M. (2009). Predicting students drop out: A case study. *International Conference on Educational Data Mining*, (pp. 41-50).

-
- Formby, Medlin, Ellington. (2017). Microsoft Excel®: Is It An Important Job Skill for College Graduates? *Information Systems Education Journal*, 15(3), 55-63.
- Gomillion, D. L. (2017). The Role of Industry Certifications in an AACSB-Accredited Institution. *Information Systems Education Journal (ISEDJ)*, 15(1), 68-79.
- Kapoor, P., & Reena, R. (2015). Efficient Decision Tree Algorithm Using J48 and Reduced Error Pruning. *International Journal of Engineering Research and General Science*, 5(7), 67-90.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy. *International Joint Conference on Artificial Intelligence*. Stanford University.
- Lei, X., & Pingfan, Y. (1988). Best first strategy for feature selection. *Pattern Recognition*, (pp. 1040-1043).
- Merceron, A., & Ycef, K. (2005). Educational Data mining: A case study. *International Conference on Artificial*. The Netherlands.
- Michael H. Randall, C. J. (2005). Information Technology Student-Based Certification in Formal Education Settings: Who Benefits and What is Needed. *Journal of Information Technology Education*, 4, 287-306.
- Nesrin Bakir, K. D. (2019). Certifying Business Students in Microsoft Office Specialist Certification Excel Core Exam: Lessons Learned. *Information Systems Education Journal*, 17(6), 4-11.
- Ngo-Ye, Jae, Gittens. (2018). Helping Students Prepare for Microsoft Office Specialist Certification Exams. *Information Systems*, 9(4), 122-131.
- Nithya, P., Umamaheswari, B., & Umadevi, A. (2016). A survey on educational data mining in field of education. *Journal Computer Science Software Development*, 7(8), 1-6.
- Rashid Naseem, Bilal Khan, Arshad Ahmad, Ahmad Almogren, Saima Jabeen, Bashir Hayat, Muhammad Arif Shah. (2020). Investigating Tree Family Machine Learning Techniques for a Predictive System to Unveil Software Defects. *Hindawi Complexity*, 2020. doi:<https://doi.org/10.1155/2020/6688075>
- Romero, C., & Sebastian, V. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 3(5), 135-146.
- Sharma, P. (2014). Comparative Analysis of Various Decision Tree Classification Algorithms using WEKA. *International Journal on Recent and Innovation Trends in Computing and Communication* , 3(2), 684 – 690.

Yehuala, M. (2015). Application Of Data Mining Techniques For Student Success And Failure Prediction. *International Journal Of Scientific & Technology Research*, 4(4), 342-250.

MINING THE RESULTS OF ACCUMULATING INTERNATIONAL CERTIFICATES MOS EXCEL OF STUDENTS AT THE UNIVERSITY OF ECONOMICS, HUE UNIVERSITY

Mai Thu Giang

Abstract. The study focuses on presenting the results of mining the data set which have collected and stored since September 2017 about the results getting MOS Excel certificates of students at the University of Economics, Hue University. The research results have high practical significance for students, lecturers with relevant expertise, and training managers of the University of Economics in the trend of improving investment quality, approaching international standards in the trend of globalization. In addition, the study provide a lesson for educational institutions that are starting to approach and pay attention to this issue. To achieve the goal of the study, we use the open source data mining software suite WEKA (Waikato Environment for Knowledge Analysis) with the application of an attribute extraction algorithm using the CfsSubsetEval function to find out. the important attributes stored in the accumulation process that have an influence on predicting the cumulative results of students' MOS Excel. At the same time, we apply the classification technique based on the decision tree algorithms to both data sets before and after extraction to evaluate the obtained results. The results show that the attributes time to study, time to learn directly, the total number of times to participate in the mock test, and the average result of the mock test are four important attributes that are retained after extraction. Finally, the classification model based on the J48 decision tree algorithm is the best model proposed to predict cumulative results with an accuracy rate of above 77%.

Keywords: Decision Tree; MOS Excel; Classification; Feature selection.