**DANH SÁCH THAM GIA ĐỀ TÀI**

**THÀNH VIÊN THAM GIA ĐỀ TÀI:**

|  |  |  |
| --- | --- | --- |
| **Họ và Tên** | **Đơn vị công tác** | **Nhiệm vụ** |
| Nguyễn Thành Trung | Trung tâm Đào tạo E-Learning | Chủ nghiệm đề tài |
| Trần Lan Thu | Trung tâm Đào tạo E-Learning | Nghiên cứu viên |
| Nguyễn Danh Tú | Khoa Toán tin – ĐHBKHN | Nghiên cứu viên |
| Bùi Minh Tâm | Trung tâm Đào tạo E-Leanring | Thư ký |

**ĐƠN VỊ PHỐI HỢP CHÍNH:**

|  |  |
| --- | --- |
| **TÊN ĐƠN VỊ** | **LĨNH VỰC PHỐI HỢP** |
| Trung tâm Phát triển Đào tạo | Tư vấn hành vi người học |
| Công ty CP ĐT & PT ĐT Edutop 64 | Tập hợp dữ liệu từ phần mềm |

# Danh mục hình vẽ và bảng biểu

[**Hình 1.** Thống kê ảnh hưởng của nghỉ học tới trường học, học viên và xã hội. 6](#_Toc311440178)

[**Hình 2.** Thống kê các nguyên dẫn dẫn tới NGHỈ HỌC của học viên 8](#_Toc311440179)

[**Hình 3.** Khái niệm khai phá dữ liệu 10](#_Toc311440180)

[**Hình 4.** Thống kê các nguồn dữ liệu 14](#_Toc311440181)

[**Hình 5.** Giao diện chính phần mềm 16](#_Toc311440182)

[**Hình 6.** Chức năng phân tích tập dữ liệu 17](#_Toc311440183)

[**Hình 7.** Chức năng tiền xử lý dữ liệu 18](#_Toc311440184)

[**Hình 8.** Tùy chỉnh tham số mạng 19](#_Toc311440185)

[**Hình 9.** Đồ thị hàm lỗi của mạng trên các tập dữ liệu 20](#_Toc311440186)

[**Hình 10.** Thiết lập kiến trúc mạng 20](#_Toc311440187)

[**Hình 11.** Phần trăm đúng CCR trên các tập dữ liệu tham gia luyện mạng 21](#_Toc311440188)

[**Hình 12.** Phần trăn đúng dựa trên dữ liệu xác minh 22](#_Toc311440189)

[**Hình 13.** Phần trăm đúng trên tập dữ liệu Test 22](#_Toc311440190)

[**Hình 14.** Kết quả truy vấn 1 23](#_Toc311440191)

[**Hình 15.** Kết quả truy vấn 2 24](#_Toc311440192)

# DANH MỤC CÁC TỪ VIẾT TẮT

|  |  |  |
| --- | --- | --- |
| **STT** | **Nội dung** | **Viết tắt** |
|  | Knowledge discovery in databases | KDD |
|  | School Managent System | CRM |
|  | Customer Relationalship System | CRM |
|  | Cố vấn học tập | CVHT |
|  | Tư vấn tuyển sinh | TVTS |
| 6 | Chương trình đào tạo hệ từ xa theo phương thức E-Learning do Trung tâm Đào tạo E-Learning quản lý | HOU E-Learning |
| 7 | Công nghệ thông tin và truyền thông | ICT |

MỞ ĐẦU

Cùng với sự phát triển mạnh mẽ về quy mô đào tạo E-Learning, Trung tâm Đào tạo E-Learning cũng như các đơn vị đào tạo từ xa và đào tạo trực tuyến khác trên thế giới đối mặt với tình trạng bỏ học của một bộ phận không nhỏ học viên. Vấn đề này ảnh hưởng trực tiếp đến quy mô phát triển, chất lượng và hiệu quả kinh tế của chương trình đào tạo. Tìm hiểu và dự báo được khả năng nghỉ học của các nhóm học viên, phát hiện được những thành tố liên quan trực tiếp đến quá trình nghỉ học sẽ giúp các nhà quản lý đưa ra các phương pháp phù hợp để giảm bớt tình trạng này mang lại hiệu quả cho Viện, cho Trung tâm trong cả ngắn hạn và dài hạn.

Mặc dù trên thế giới đã có những nghiên cứu về vấn đề học viên nghỉ học (drop-out) trong môi trường người học là người lớn, trong giáo dục mở và từ xa. Chúng ta phải thừa nhận rằng điều kiện xã hội, điều kiện kinh tế, động cơ học tập của người học tại các nước khác nhau có sự khác biệt lớn. Đề tài **Khai phá dữ liệu: dự báo tỉ lệ nghỉ học của các nhóm học viên hệ từ xa theo phương thức E-Learning** có mục đích xây dựng những nền tảng ban đầu trong nghiên cứu vấn đề học viên nghỉ học trong môi trường E-Learning đặc thù của HOU.

# Tiếp cận vấn đề học viên nghỉ học trong các cơ sở đào tạo

Bảng sau tổng kết một số tác động chính lên trường học, học viên, và xã hội về mặt ngắn hạn và dài hạn:

|  |  |  |
| --- | --- | --- |
| **Đối tượng** | **Ngắn hạn** | **Dài hạn** |
| Trường học | Doanh thu giảm | Thương hiệu bị đánh giá thấp |
| Chi phí tăng | Số học viên tuyển sinh bị giảm các năm học sau |
| Học viên | Chi phí học tập học viên tăng mà không đạt được mục đích | Không có việc làm |
| Tâm lý bị tiêu cực, chán nản | Thu nhập bị sụt giảm |
| Không có bằng cấp và kỹ năng nghề nghiệp | Khả năng phạm tội tăng |
| Xã hội | Nạn thất nghiệp gia tăng | Chi phí công đầu tư cho giáo dục, phúc lợi xã hội tăng |
| Chi phí đầu tư giáo dục tăng | Tỷ lệ phạm tội gia tăng |

##### Thống kê ảnh hưởng của nghỉ học tới trường học, học viên và xã hội.

**Tổng hợp các nguyên nhân nghỉ học:**

Chúng tôi đã tổng hợp một số nguyên nhân dẫn đến tình trạng nghỉ học của học viên:

| **Nguyên nhân** | **Trước quá trình học** | **Trong quá trình học** |
| --- | --- | --- |
| Trường học | Tư vấn chưa đúng (khác so với quá trình học)  **Vergidis & Panagiotakopoulos, 2002**  Đề cương chương trình không đầy đủ, thiếu chi tiết  Giảng viên không đủ tiêu chuẩn/thiếu kinh nghiệm | Môi trường học tập và phương pháp giảng dạy không thu hút HV  **Keller,1987**  Mức độ hài lòng thấp của học viên v̀ môi trường học trực tuyến (Vấn đề về cách thức quản lý của CVHT vấn đề về công nghệ,…)  **Gortan & ereb, 2007; Keller; Levy, 2007; Liaw**  Học liệu không: đầy đủ, trực quan và đa tương tác  Thiếu giám sát, quản lý, thiếu sự hộ trợ sinh viên |
| Học viên | Nền tảng kiến thức của HV chưa đáp ứng  Không đủ tài chính để học tập  Kỹ năng công nghệ của học viên không đủ đáp ứng  **Parker; Boyd, 2004**  Mục tiêu giáo dục của học viên sai khác so với mục tiêu giáo dục của chương trình đào tạo  **Parker, 1995**  Kĩ năng mềm của học viên (ghi nhớ, quản lý thời gian, đọc, viết,..) KHÔNG đủ đáp ứng  **Nash,2005; Liaw, 2008** | Học viên không đủ niềm tin, kiên trì để học trong môi trường từ xa **Chacon-Duque, 1987; Liaw, 2008**  Học viên không sắp xếp được thời gian để theo học  Học viên Không đủ điều kiện tốt nghiệp  Mục tiêu giáo dục của học viên thay đổi trong quá trình học  Học viên không thu được kiến thức và kĩ năng mà họ đặt ra  **Running head: DROPOUTS IN E-LEARNING COURSES**  HV không đánh giá đúng thời gian và sự nỗ lực cần thiết cho khóa học  **Nash, 2005; Seifert et al., 2008; Vergidis & Panagiotakopoulos** |
| Môi trường xã hội | Áp lực gia đình, áp lực công việc làm thêm  Xu hướng xã hội không phù hợp | Công nghệ giảng dạy không phù hợp, lỗi thời  Sự cố cuộc sống không mong đợi xảy ra (tai nạn, gia đình, tệ nạn, ...)  Xu hướng ngành nghề xã hội bị thay đổi (tác động giảm) |

##### Thống kê các nguyên dẫn dẫn tới NGHỈ HỌC của học viên

## Các hướng giải quyết cho vấn đề nghỉ ở HOU E-learning

Để giải quyết một phần vấn đề này Trung tâm đào tạo từ xa Elearning đã đưa một số định hướng như sau.

Đối với học viên sắp hoặc đã nghỉ, các biện pháp là:

* Tái tư vấn thông tin
* Bảo lưu kết quả
* Hỗ trợ chuẩn hóa hồ̀ sơ
* Chuyển đổi ngành học cho học viên
* Thay đổi chương trình đào tạo
* Hỗ trợ tư vấn tài chính

Để tránh thụ động trong giải quyết vấn đề nghỉ học này, chúng tôi đã đưa ra một số phương pháp ngăn ngừa.

Đối với các học viên đang theo học, chúng tôi:

* Dự báo tiềm năng NGHỈ HỌC dựa trên thông tin trong quá trình học với các thông tin về tình trạng đóng học phí của học viên (thời hạn đóng học phí, số tiền đóng, …), thông tin quá trình học tập, thông tin tương tác học viên với môi trường học (tình trạng tham gia lớp học qua môi trường mạng, đăng nhập, trao đổi thảo luận trên diễn dàn, H24/72, 3D, làm bài luyện tập, bài tập về nhà, bài tập kỹ năng, nhóm,…), hồ sơ học viên.
* Tổ chức các hoạt động tăng cường tương tác 2 chiều
* Hỗ trợ các hoạt động tăng cường kỹ năng (kỹ năng cá nhân, kỹ năng giao tiếp, kỹ năng làm việc nhóm, kỹ năng học tập và làm việc qua môi trường mạng, kỹ năng phỏng vấn, nhập môn Elearning, ….)

Song song với đó, chúng tôi chủ động giải quyết vấn đề nghỉ học của học viên ngay tại thời điểm trước khi vào học (trong quá trình tư vấn). Một số biện pháp chủ yếu được đưa ra là:

* Tư vấn chính xác
* Chọn lọc đúng đối tượng
* Kiểm tra tư vấn chéo
* **Dự báo tiềm năng NGHỈ HỌC dựa trên hồ̀ sơ thông tin học viên**

Vấn đề dự báo khả năng của học viên hiện tại vẫn đang dựa trên kinh nghiệm của các cố vấn học tập và chưa có một quy trình cụ thể. Chính vì vậy, chúng tôi đã thực hiện nghiên cứu đề tài với nội dung: **KHAI PHÁ DỮ LIỆU: DỰ BÁO TỶ LỆ NGHỈ HỌC CỦA CÁC NHÓM HỌC VIÊN HỆ TỪ XA THEO PHƯƠNG THỨC E-LEARNING**. Kết quả của đề tài nghiên cứu sẽ là một số mô hình cho phép dự báo được khả năng của học viên dựa trên thông tin hồ sơ của học viên.

# Tóm tắt cơ sở lý thuyết và những kết quả nghiên cứu

## Khái niệm về khai phá dữ liệu

Khai phá dữ liệu (Data mining): là việc trích rút ra được các mẫu hoặc tri thức quan trọng từ một lượng dữ liệu lớn.

Một số tên gọi khác: Khám phá tri thức trong các cơ sở dữ liệu (knowledge discovery (mining) in database-KDD), trích rút tri thức (Knowledge extraction), phân tích mẫu, phân tích dữ liệu (Data/pattern analysis) , …

Khai phá dữ liệu khác với: tìm kiếm thông tin (information retrieval), xử lý câu truy vấn (SQL) đối với các cơ sở dữ liệu

CÁC MẪU

KHAI PHÁ DỮ LIỆU

DỮ LIỆU

##### Khái niệm khai phá dữ liệu

## Kết quả nghiên cứu từ những trường đại học khác

Dưới đây là một số nghiên cứu về vấn đề NGHỈ HỌC của những trường đại học khác trên thế giới chúng tôi đã tổng hợp:

| **Bài báo** | **Tác giả/ Trường** | **Năm** | **Tóm tắt kết quả** |
| --- | --- | --- | --- |
| Flight risk: identifying potential drop outs | Prakash V. Arumugam  School of Business and Admisistration  Wawasan Open University | 2010 | Mô hình dự báo khả năng bỏ học của học viên sau đại học. Đề xuât nghiên cứu làm thế nào để cải thiện tỉ lệ duy trì ở các trường ĐH |
| Improving graduation rates of open and distance learners via online community | Richard Ng  Director, Perak Regional Learning Centre  Open University Malaysia | 2010 | Nghiên cứu tập trung vào đội ngũ trực tuyến giúp học viên cam kết ở lại các chương trình của họ. |
| A review of online course drop out research and suggestions for future research | Youngju Lee, Jaeho Choi, & Patrick McGuire  University of Virginia | 2010 | Xác định được 3 nhóm nhân tố ảnh hưởng tới quyết định bỏ học của học viên.Phân tích mức độ ảnh hưởng của từng nhân tố. Phân tích các chiến lược khắc phục drop-out. |
| An artificial neural network for predicting student graduation outcomes | Stamos T. Karamouzis and Andreas Vrettos | 2008 | Sử dụng mạng Perceptron 3 lớp dự báo khả năng tốt nghiệp. Luyện với 1407 hồ sơ của các học viên trường Waubonsee College; Tập dữ liệu chia làm 2 tập. Kết quả: Trên tập luyện với số mẫu 1,100 là 77% và tập dữ liệu kiểm tra với số mẫu 307 là 68%. |
| Student dropout analysis with application of data mining methods | M. Jadrić, Ž. Garača, M. Ćukušić | 2010 | Trình ứng dụng phân tích sinh viên NGHỈ HỌC sử dụng phương pháp phân lớp (hồi qui, mạng nơ ron, cây quyết định). Luyện với 40% tập dữ liệu 2000 mẫu trên 715 sinh viên của Faculty of Economics Information System. |
| Identifying factors that predict student success in a community college online distance learning course | Johnelle Bryson Welsh, A. G. S., B. S., M. S.  / University Of North Texas | 2007 | Xác định các nhân tố để dự báo học viên tốt nghiệp trong chương trình từ xa trực tuyến. Phân tích trên mẫu dữ liệu gồm 926 học viên. |
| Enrollment prediction models using data mining | Ashutosh Nandeshwar; Subodh Chaudhari | 2009 | Mô hình dự báo tuyển sinh dùng khai phá dữ liệu bằng WEKA phân lớp sử dụng cây quyết định. Kết quả dự báo chính xác trong khoảng 83%-84%. Dữ liệu được lấy tại West Virginia University |
| Data mining algorithms to classify students | Cristóbal Romero, Sebastián Ventura, Pedro G. Espejo and César Hervás  / Computer Science Department, Córdoba University, Spain | 2008 | So sánh câc phương pháp khai phá dữ liệu để phân lớp sinh viên dựa vào 438 dữ liệu sinh viên trong 7 khóa học từ hệ thống Moodle của trường Cordoba University .Phát triển và tích hợp công cụ Moodle Mining Tool hỗ trợ các giảng viên trực tuyến. |

## Phân tích vấn đề nghỉ học ở HOU – Elearning

Tập dữ liệu học viên được phân thành bốn lớp: Đang học, Đã tốt nghiệp, Đã NGHỈ HỌC, Đang tư vấn. Nên bản chất bài toán của chúng tôi gặp phải tại HOU-Elearning là bài toán phân lớp. Trong “phân lớp” có những kĩ thuật như:

* Mạng Noron
* Mạng Bayes
* Cây quyết định
* Mô hình Markov ẩn
* Luật cơ sở
* Trí nhớ dựa trên các nguyên nhân
* Giải thuật di truyền

Do mạng noron có những ưu điểm như:

* Cho phép xấp xỉ những ánh xạ phi tuyến tùy ý
* Là hệ thống xử lý song song làm tăng tốc độ tính toán cho phép đáp ứng khả năng tính toán thời gian thực và chính xác
* Là hệ học và thích nghi, khi mạng được huấn luyện từ các dữ liệu quá khứ, đồng thời có khả năng khái quát hóa khi dữ liệu vào bị thiếu hoặc không đầy đủ
* Phù hợp với các hệ thống nhận dạng, chuẩn đoán kỹ thuật...

Nên chúng tôi chọn kĩ thuật mạng nơron để giải quyết bài toán “"**DỰ BÁO TỶ LỆ NGHỈ HỌC CỦA CÁC NHÓM HỌC VIÊN HỆTỪ XA THEO PHƯƠNG THỨC E-LEARNING**.

# Quá trình và kết quả nghiên cứu và phát triển

## Thu thập dữ liệu

Chúng tôi đã thu thập được các dữ liệu cần thiết cho các mô hình dự báo như: Dữ liệu về hồ sơ học viên gồm có 7476 học viên và 30 cột thuộc tính (Tuổi, giới tính, bằng tốt nghiệp cấp 3, dân tộc thiểu số YN, tôn giáo,…) từ ngày bắt đầu tuyển sinh đến ngày 5/10/2011.

Dưới đây là thống kê về các nguồn dữ liệu đã lấy được:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Loại dữ liệu** | **Nguồn** | **SL trường thông tin** | **SL bản ghi thu thập** |
| 1 | Hồ sơ học viên | SCM | 30 | 7,476 |
| 2 | Kết quả thi PTTH | CRM | 4 | 7,476 |
| 3 | Ngành học, đối tượng học | CVHT | 3 | 20 |
| 4 | Thông tin quá trình tư vấn | CRM | 5 | 712,564 |
| 5 | Bằng cấp, chứng chỉ | SCM | 10 | 1,024 |
| 6 | Hồ sơ học tập học viên | SCM | 12 | 642,715 |
| 7 | Lộ trình thu học phí | CVHT | 7 | 160 |
| 8 | Thông tin đóng học phí | SCM | 8 | 21,542 |
| 9 | Qui trình tuyển sinh | TVTS | 4 | 20 |
| 10 | Thông tin quá trình học | LMS | 3 | 614,523 |

##### Thống kê các nguồn dữ liệu

## Các mô hình dự báo

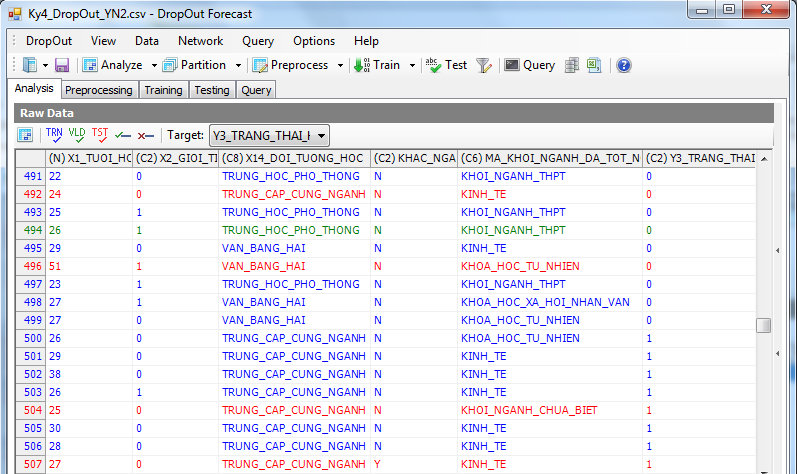
Thông qua quá trình nghiên cứu trên yêu cầu của Trung tâm và các loại dữ liệu thu thập được, chúng tôi đã đề xuất một số mô hình dự báo để tiến hành nghiên cứu và triển khai như sau:

|  |  |  |
| --- | --- | --- |
| **Mô hình** | **Đầu vào** | **Đầu ra** |
| Dự báo tình trạng nghỉ học của học viên | Hồ sơ của những học viên đã tốt nghiệp và những học viên đã NGHỈ HỌC | **Trạng thái học viên** 0 – NGHỈ HỌC 1 – Tốt nghiệp |
| Dự báo tình trạng học tập của học viên qua 1 kỳ xác định (Ký hiệu: kỳ x =2,3,4) | Hồ sơ của những học viên đã học qua kỳ x và những học viên đã NGHỈ HỌC trước kỳ x | **Trạng thái học viên** 0 – NGHỈ HỌC  1 – Học vượt qua kỳ x |
| Dự báo tiềm năng nghỉ học của học viên trong quá trình học | Kỳ x;  Hồ sơ học viên;  Qui mô lớp; Điểm tích lũy trung bình; Số lần đăng nhập diễn đàn; Tỉ lệ NGHỈ HỌC kỳ trước | **Trạng thái học viên** Màu xanh (Học thêm được ít nhất 2 kỳ nữa)  Màu vàng (Kỳ tới sẽ NGHỈ HỌC) Màu đỏ (Kỳ này NGHỈ HỌC) |

## Sản phẩm phần mềm và các bước xử lý

### Sản phẩm phần mềm

Trong quá trình thực hiện đề tài chúng tôi đã xây dựng thành công sản phẩm phần mềm mô phỏng kỹ thuật mô hình mạng Neuro cài đặt luật học thích nghi. Mục đích của sản phẩm phần mềm là chạy kiểm thử tập dữ liệu thu thập được, kiểm chứng độ chính xác kết quả các mô hình thực hiện trong đề tài. Hỗ trợ truy vấn kết quả của các mô hình với tập dữ liệu mới tương ứng.

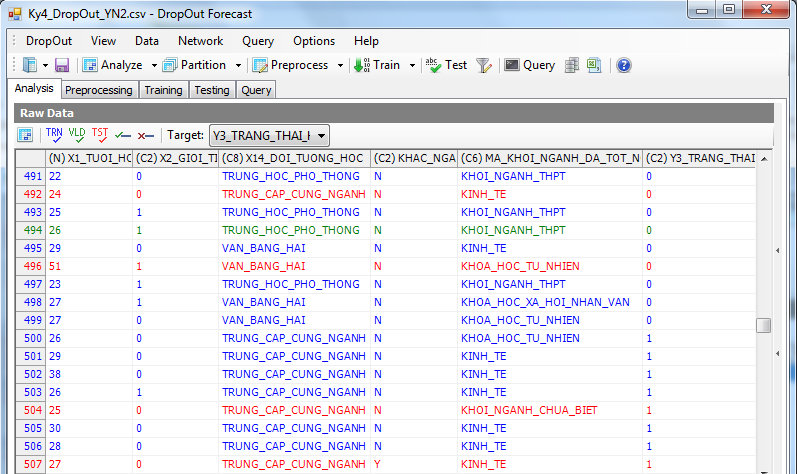


##### Giao diện chính phần mềm

Nhằm mục đích giải quyết vấn đề thế giới thực, chương trình chỉ bao gồm một số tính năng cơ bản và khá dễ dàng sử dụng. Sản phẩm phần mềm dự báo được chúng tôi xây dựng có các chức năng sau( Hình ảnh minh họa là mô hình “Kỳ 4 NGHỈ HỌC Có/Không”:

#### Chức năng: phân tích tập dữ liệu

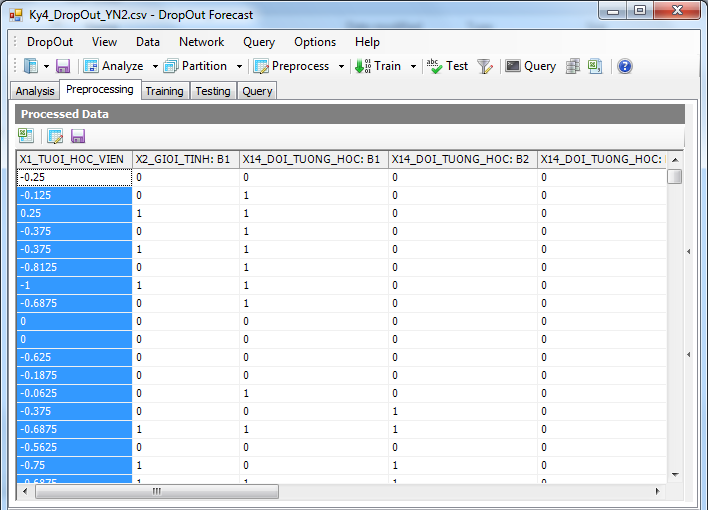
Dữ liệu thô được đưa vào chương trình tại tab Analysis, các bản ghi nhiễu, thiếu hoặc dữ liệu dị thường sẽ được tự động lọc tại đây trước khi chuyển sang bước tiền xử lý dữ liệu. Tại tab này các cột (thuộc tính) dữ liệu được tự động nhận dạng kiểu dữ liệu (có thể tùy chỉnh bằng tay). Đồng thời tập dữ liệu được phân chia làm ba phần dữ liệu tương ứng với tập dữ liệu luyện mạng (blue), tập dữ liệu đánh giá mạng (green), tập dữ liệu kiểm tra mạng (red) và mặc định được chia tỉ lệ 0.68:0.16:0.16.



##### Chức năng phân tích tập dữ liệu

#### Chức năng: Tiền xử lý dữ liệu

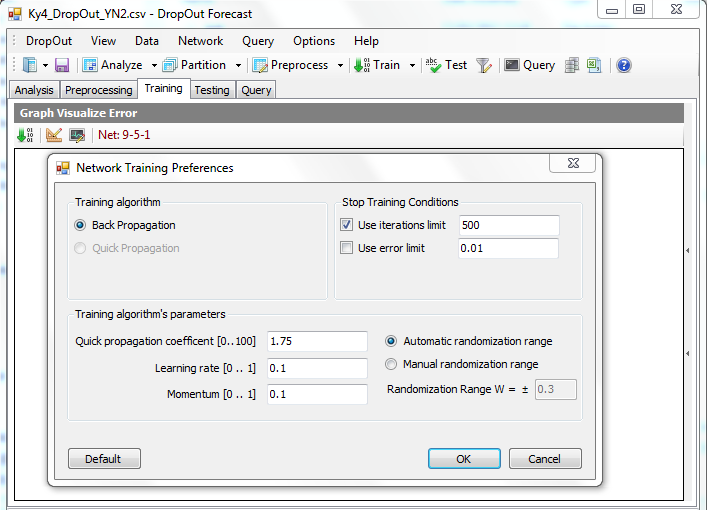
Tại Analysis tab click Preprocess button dữ liệu được tự động xử lý sau quá trình phân tích và phân chia tập dữ liệu tương ứng các kiểu dữ liệu đưa tại bước trước. Các dữ liệu được chuyển đổi từ kiểu dữ liệu phổ thông được trình bày ở phần trước sang kiểu dữ liệu phù hợp với mạng như định lại tỉ lệ kiểu dữ liệu có thứ tự, mã hóa dữ liệu kiểu văn bản, kiểu dữ liệu ngày tháng, …



##### Chức năng tiền xử lý dữ liệu

#### Chức năng: Luyện mạng

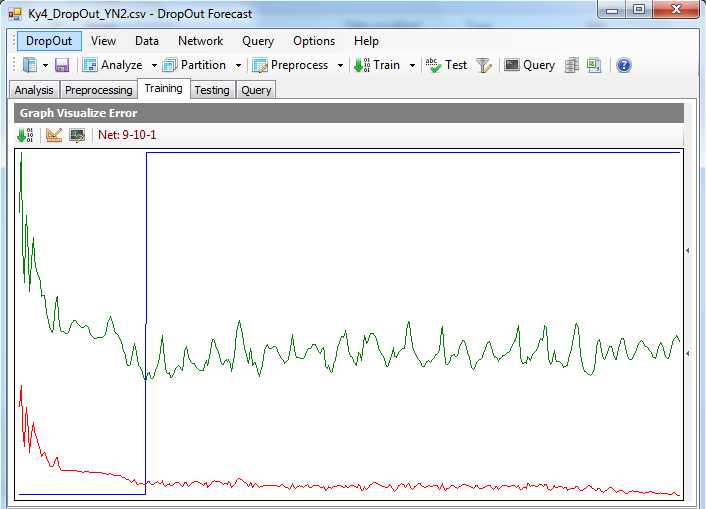
Với bộ dữ liệu được phần tích và xử lý. Tại tab Training chúng ta thấy hệ thống hỗ trợ tự động thiết kế cấu trúc mạng mặc định, trước khi luyện mạng các tham số mạng và kiến trúc mạng có thể được thay đổi bởi người sử dụng để mạng có thể hoạt động được tốt nhất.



##### Tùy chỉnh tham số mạng

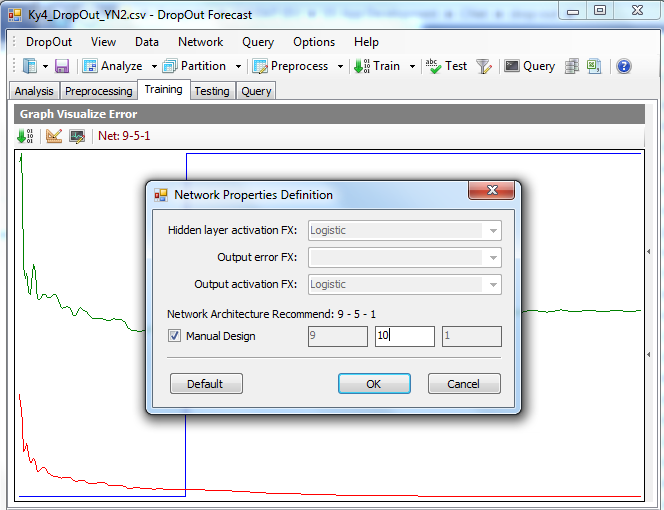
Sau khi hoàn tất bước trên mạng được tự động luyện với bộ dữ liệu vào và các tham số tùy chọn người dùng. Kết quả luyện mạng được trực quan bằng đồ thị hàm lỗi của mạng trên các tập dữ liệu trực tiếp tham gia trong quá trình luyện mạng tương ứng (training set, validation set).

Nhận thấy rằng, đồ thị hàm lỗi của mạng trên tập dữ liệu luyện mạng (training set) là thấp hơn rất nhiều so với đồ thị hàm lỗi của mạng trên tập dữ liệu kiểm tra mạng (validation set), lý do là mạng có năng lực để làm cho các bộ trọng của mạng khớp nhất với những gì mà nó được học. Đó là lý do tại sao ta đưa thêm vào mạng bộ dữ liệu kiểm tra mạng. Vậy tại thời điểm mà lỗi trên tập dữ liệu kiểm tra mạng không giảm nữa bắt đầu tăng ta có được mạng với bộ trọng tốt nhất cho dự báo với độ chính xác tốt nhất.



##### Đồ thị hàm lỗi của mạng trên các tập dữ liệu

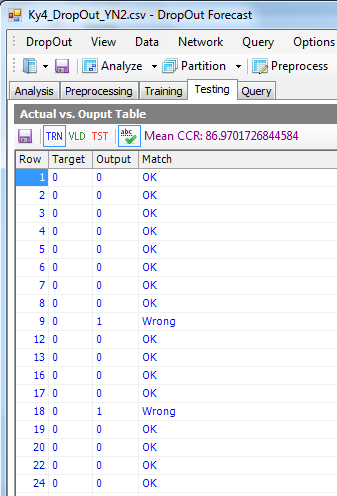
#### Chức năng: Thiết lập kiến trúc mạng



##### Thiết lập kiến trúc mạng

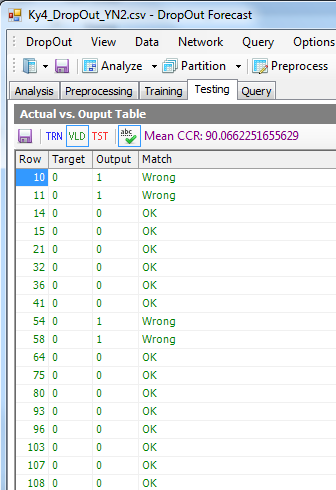
#### Chức năng: Kiểm tra kết quả luyện mạng

Kết quả luyện mạng ta có thể thấy được tại tab Test. Rõ ràng với mạng được chọn kết quả dự đoạn trên các tập dữ liệu là khác nhau. Và kết phần trăm đúng CCR trên tập dữ liệu Test được sử dụng để dự báo trên dữ liệu học viên mới



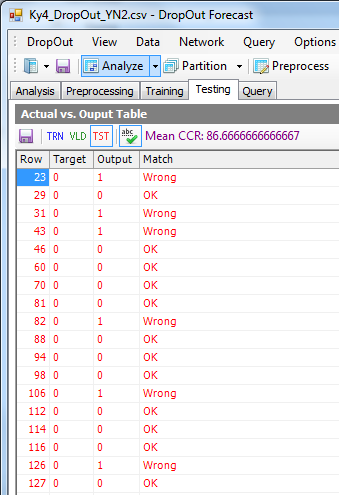
##### Phần trăm đúng CCR trên các tập dữ liệu tham gia luyện mạng

#### Chức năng: kiểm tra tính phù hợp của mạng



##### Phần trăn đúng dựa trên dữ liệu xác minh

#### Chức năng: *Kiểm tra*



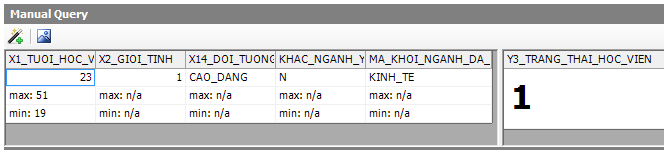
##### Phần trăm đúng trên tập dữ liệu Test

#### Chức năng: Truy vấn kết quả

Hỏi với với hồ sơ của một học viên thì trạng thái của học viên này tại kỳ 4 là như thế nào?

Ví dụ1: ta có hồ sơ của một học viên như sau.

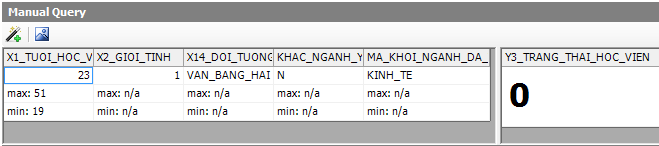
* Tuổi: 23
* Giới tính: 1 (nam)
* Đối tượng học: Cao đẳng
* Khác ngành: Yes
* Mã khối ngành đã tốt nghiệp: Kinh tế



##### Kết quả truy vấn 1

Ví dụ2: Ta có hồ sơ của một học viên như sau.

* Tuổi: 23
* Giới tính: 1 (nam)
* Đối tượng học: Văn bằng hai
* Khác ngành: N
* Mã khối ngành đã tốt nghiệp: Kinh tế



##### Kết quả truy vấn 2

## Báo cáo phân tích dữ liệu nghỉ học

Căn cứ trên 10 loại dữ liệu thập được gồm 87 trường thông tin và khoảng 2 triệu bản ghi dữ liệu, thông qua quá trình tiền xử lý (trích lọc thông tin, loại nhiễu,…) và đưa vào mô hình Mạng Noron để luyện, chúng tôi thu được kết quả như sau:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Dữ liệu mẫu** | **Tỷ lệ mẫu** | **NGHỈ HỌC**  **YN** | **NGHỈ HỌC SAU KỲ** | | |
| **Kỳ 2** | **Kỳ 3** | **Kỳ 4** |
| **Tập dữ liệu thu thập** | **100%** | **78** | **1364** | **1351** | **938** |
| **Tập dữ liệu luyện** | **68%** | **53** | **927** | **918** | **637** |
| **Tập dữ liệu kiểm tra phù hợp** | **16%** | **12** | **218** | **216** | **150** |
| **Tập dữ liệu kiểm tra** | **16%** | **12** | **218** | **216** | **150** |
| **Kết quả dự báo đúng** |  | **91.67%** | **68.81%** | **77.68%** | **86.00%** |

KẾT LUẬN VÀ KIẾN NGHỊ

Giảm thiểu tỉ lệ học viên bỏ học là thách thức lớn trong môi trường đào tạo E-Learning. Mặc dù đã có nhiều nghiên cứu và đánh giá tại các cơ sở giáo dục ứng dụng E-Learning trên thế giới, HOU E-Learning vẫn phải tìm ra những hướng đi mới của mình để có thể áp dụng hiệu quả trong môi trường Việt Nam nhằm giảm thiểu tỉ lệ bỏ học tại HOU E-Learning.

**Kết quả đạt được**

Phương pháp dự báo bằng khai phá dữ liệu bằng phần mềm cho phép tự động hóa quá trình phát hiện học viên có xác xuất bỏ học cao trong tương lai. Kết quả của phương pháp này cho phép Trung tâm đào tạo E-Learning triển khai các biện pháp chủ động như tư vấn sớm cho học viên về lựa chọn tiến độ học tập cũng như định hướng cuộc sống phù hợp để có thể kết thúc được khóa học.

Phương pháp này cũng dự đoán được các tiêu chí có độ nhạy lớn nhất với xác xuất nghỉ học. Đây là một trong những phát kiến quan trọng của đề tài. Tuy nhiên mối liên hệ trực tiếp giữa các tiêu chí này với các nguyên nhân bỏ học chính vẫn còn đòi hỏi các nghiên cứu tiếp theo.

Về mặt mô hình dự báo đã xác định được bước đầu mô hình dựa trên các yếu tố địa lý, nhân khẩu học và thông tin từ hồ sơ nhập học.

Về sản phẩm phần mềm nhóm nghiên cứu đã có được phần mềm trên nền tảng .NET cho phép tự động hóa phân tích dữ liệu được tập hợp từ phần mềm quản lý học viên SCM của Trung tâm.

**Hướng nghiên cứu**

Tiếp tục mở rộng mô hình phân tích là điểm quan trọng cần quan tâm trong thời gian tiếp theo. Đối tượng chính tiếp theo của mô hình sẽ là các kết quả học tập của học viên và hoạt động của học viên trong khóa học.

Để có thể thực hiện các phân tích một các thường xuyên hơn, phần mềm dự báo khả năng nghỉ học cần được tích hợp chặt chẽ vào các hệ thống thu thập dữ liệu như CRM, SCM, LMS,

Để có thể xác định kết quả nghiên cứu rõ ràng hơn các biện pháp kiểm định cần được xác định bằng các phương pháp khác như phỏng vấn, phỏng vấn nhóm nhỏ.

Vì là chương trình mới, phương pháp đào tạo của Trung tâm Đào tạo E-Learning có sự thay đổi theo từng thời gian, cùng thời gian đó uy tín của chương trình E-learning cũng có thay đổi (ví dụ có học viên tốt nghiệp), điều này dẫn đến mô hình và phương pháp phân tích cần có sự điều chỉnh phù hợp. Nhóm đề tài dự kiến sẽ phân tích và đưa ra mô hình mới theo từng 6 tháng.

**Kiến nghị**

Để xác định được tính chính xác và có thể điều chỉnh, mở rộng mô hình phân tích một cách hiệu quả, Trung tâm đào tạo E-Learning cần có kế hoạch ứng dụng kết quả nghiên cứu vào từng nhóm học viên. Thực hiện so sánh, đánh giá hiệu quả và đưa ra những góp ý với nhóm nghiên cứu.

Để mở rộng và tăng cường tính chính xác của mô hình phân tích, chúng ta cũng cần có sự đơn giản và đồng bộ hóa các dữ liệu đầu vào. Trên quan điểm quản trị đây chính là đòi hỏi về tính đồng đều trong chất lượng tư vấn và chất lượng dịch vụ học viên tại các Trung tâm địa phương.

Trong khuôn khổ vấn đề nghỉ học của học viên có những nguyên nhân mới mà có thể chỉ gặp phải tại Việt Nam và một số nước có trình độ phát triển tương đương như hiểu biết của người học về tin học, khả năng kết nối. Những điều kiện cơ bản này cần phải được cải thiện thông qua quá trình xây dựng chương trình đào tạo, chuẩn đầu ra và hỗ trợ cộng đồng …Những vấn đề này cần được các cấp quản lý quan tâm và giải quyết đồng bộ trong chiến lược phát triển kinh tế xã hội như các đề án xây dựng xã hội học tập, phát triển văn hóa nông thôn, phổ cập internet ở vùng sâu vùng xa…

TÀI LIỆU THAM KHẢO

**Martinez, M.**  (2003). High attrition rates in e-Learning: Challenges, predictors and solutions. The e-Learning Developers Journal, 17(11).

Ormond Simson, International Forum on Student Retention in Higher Education, 2009, UKOU Press.

**Parker, A.**  (1999, December). A study of variables that predict dropout from distance education. International Journal of Educational Technology.