

TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI
PHÂN HIỆU TẠI TP. HỒ CHÍ MINH
BỘ MÔN CÔNG NGHỆ THÔNG TIN



BÁO CÁO ĐỒ ÁN TỐT NGHIỆP
ĐỀ TÀI: XÂY DỰNG HỆ THỐNG TƯ VẤN MÔN HỌC TỰ
CHỌN CHO SINH VIÊN

Giảng viên hướng dẫn: ThS. PHẠM THỊ MIÊN

Sinh viên thực hiện: TRẦN THANH SANG

Lớp : CÔNG NGHỆ THÔNG TIN

Khoá : 57

Tp. Hồ Chí Minh, năm 2021

TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI
PHÂN HIỆU TẠI TP. HỒ CHÍ MINH
BỘ MÔN CÔNG NGHỆ THÔNG TIN



BÁO CÁO ĐỒ ÁN TỐT NGHIỆP
ĐỀ TÀI: XÂY DỰNG HỆ THỐNG TƯ VẤN MÔN HỌC TỰ
CHỌN CHO SINH VIÊN

Giảng viên hướng dẫn: ThS. PHẠM THỊ MIÊN

Sinh viên thực hiện: TRẦN THANH SANG

Lớp : CÔNG NGHỆ THÔNG TIN

Khoá : 57

Tp. Hồ Chí Minh, năm 2021

NHIỆM VỤ THIẾT KẾ TỐT NGHIỆP
BỘ MÔN: CÔNG NGHỆ THÔNG TIN

-----***-----

Mã sinh viên:5751071033

Họ tên SV: Trần Thanh Sang

Khóa: 57

Lớp: CQ.57.CNTT

1. Tên đề tài: XÂY DỰNG HỆ THỐNG TƯ VẤN MÔN HỌC TỰ CHỌN CHO SINH VIÊN.

2. Mục đích, yêu cầu:

-Xây dựng hệ thống tư vấn môn học tự chọn cho sinh viên, dựa trên kết quả của các môn học ở kì trước.

3. Nội dung và phạm vi đề tài:

-Xây dựng hệ thống tư vấn môn học tự chọn cho sinh viên ngành Công nghệ thông tin.

4. Công nghệ, công cụ và ngôn ngữ lập trình:

- Ngôn ngữ lập trình C#, Visual Studio, Mô hình mạng Nơ-ron (Neural Network).

5. Các kết quả chính dự kiến sẽ đạt được và ứng dụng:

-Ứng dụng mạng Nơ-ron (Neural Network) để dự đoán số điểm có thể đạt được của các môn học tự chọn của sinh viên.

6. Giáo viên và cán bộ hướng dẫn:

Họ tên: Phạm Thị Miên

Đơn vị công tác: Trường Đại học Giao thông Vận tải Phân hiệu tại Thành phố Hồ Chí Minh.

Điện thoại: 0961170638

Email: ptmien@utc2.edu.vn

Ngày tháng 03 năm 2021
Trưởng BM Công nghệ Thông tin

Đã giao nhiệm vụ TKTN
Giáo viên hướng dẫn

Trần Phong Nhã

Phạm Thị Miên

Đã nhận nhiệm vụ TKTN

Sinh viên: Trần Thanh Sang

Điện thoại: 0935597203

Ký tên:

Email: sangtranwp@gmail.com

LỜI CẢM ƠN

Lời đầu tiên, em xin phép gửi lời cảm ơn chân thành đến Ban giám hiệu Trường Đại học Giao thông Vận tải Phân hiệu tại Thành phố Hồ Chí Minh cùng toàn thể các thầy cô, đặc biệt là các thầy cô trong bộ môn Công nghệ thông tin, đã giúp đỡ và tạo điều kiện tốt nhất cho em trong suốt quá trình học tập tại trường cũng như trong quá trình thực hiện đồ án tốt nghiệp. Đặc biệt, em xin bày tỏ lòng biết ơn sâu sắc đến cô Phạm Thị Miên, là giảng viên hướng dẫn, nhận xét, đánh giá và giúp đỡ em rất nhiều trong việc thực hiện đồ án tốt nghiệp.

Tuy nhiên vì thời gian có hạn cũng như vốn kiến thức của em còn hạn hên nên trong quá trình thực hiện và viết bài báo cáo này, khó có thể tránh khỏi những thiếu sót. Do đó, mong các thầy cô có thể bỏ qua và góp ý thêm để em có thể hoàn thiện tốt hơn đề tài của mình.

Một lần nữa, em xin chân thành cảm ơn các thầy cô!

Tp. Hồ Chí Minh, ngày ... tháng ... năm

Sinh viên thực hiện

Trần Thanh Sang

[illegible]

Giáo viên hướng dẫn

ThS. Phạm Thị Miên

MỤC LỤC

MỞ ĐẦU.....	1
Đặt vấn đề	1
Mục tiêu của đề tài	1
Nội dung nghiên cứu	2
Ý nghĩa của đề tài.....	2
Cấu trúc báo cáo đồ án tốt nghiệp.....	2
CHƯƠNG 1. CƠ SỞ LÝ THUYẾT.....	3
1.1 Khai phá dữ liệu.....	3
1.1.1 Khai phá dữ liệu là gì.....	3
1.1.2 Các ứng dụng của khai phá dữ liệu.....	3
1.1.3 Quy trình phát hiện tri thức.....	4
1.1.4 Nhiệm vụ chính trong khai phá dữ liệu.....	5
1.1.5 Các phương pháp khai phá dữ liệu.....	6
1.1.6 Lợi thế của khai phá dữ liệu so với phương pháp cơ bản.....	7
1.1.7 Lựa chọn phương pháp.....	8
1.1.8 Những thách thức.....	8
1.2 Mạng Nơ-ron nhân tạo (Artificial Neural Network).....	10
1.2.1 Tìm hiểu về Nơ-ron thần kinh.....	10
1.2.2 Giới thiệu về mạng Nơ-ron nhân tạo.....	11
1.2.3 Các đặc trưng cơ bản của mạng nơ-ron.....	11
1.2.4 Các thành phần cơ bản của mạng nơ-ron nhân tạo.....	12
1.2.5 Các hình trạng của mạng.....	16
1.2.6 Các hình thức huấn luyện mạng nơ-ron nhân tạo.....	17
1.2.7 Mạng truyền thẳng và thuật toán lan truyền ngược.....	17
1.3 Tổng quan về ngôn ngữ lập trình C#.....	20

1.3.1 Đặc trưng của C#.....	20
1.3.2 Tính ứng dụng.....	20
CHƯƠNG 2. ỨNG DỤNG MÔ HÌNH MẠNG NƠ RON VÀO DỰ BÁO ĐIỂM	
MÔN TỰ CHỌN CHO SINH VIÊN.....	21
2.1 Đặt vấn đề.....	21
2.2 Áp dụng mạng Nơ-ron vào tư vấn môn học.....	23
2.3 Nghiên cứu thực tiễn	23
2.4 Công cụ lập trình	25
2.5 Cấu trúc tệp dữ liệu.....	26
2.6 Yêu cầu hệ thống.....	26
CHƯƠNG 3. CHẠY CHƯƠNG TRÌNH VÀ NGHIỆM THU KẾT QUẢ.....	27
3.1 Giao diện đăng nhập chương trình.....	27
3.2 Thao tác với chương trình	27
3.2.1 Tải dữ liệu vào chương trình.....	27
3.2.2 Thông số mạng.....	29
3.2.3 Tiến hành huấn luyện.....	29
3.2.4 Kiểm thử dữ liệu.....	30
3.2.5 Mối liên hệ.....	31
3.2.6 Dự báo điểm cho sinh viên.....	32
3.2.7 Dự báo điểm (dùng cho Giảng viên - Cố vấn học tập).....	32
3.3 Hướng dẫn cài đặt	34
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....	38
Kết quả đạt được	38
Hạn chế của đề tài.....	38
Hướng phát triển	38
TÀI LIỆU THAM KHẢO.....	39

DANH MỤC CHỮ VIẾT TẮT

STT	Mô tả	Ý nghĩa	Ghi chú
1	CSDL	Cơ sở dữ liệu	
2	CVHT	Cổ vấn học tập	
3	KDD	Knowledge Discovery in Database	Phát hiện tri thức trong cơ sở dữ liệu
4	KPDL	Khai phá dữ liệu	
5	LTM	Lập trình mạng	
6	MLP	Multilayer Perceptron	Mô hình mạng truyền thẳng nhiều tầng.
7	SOM	Self-Organizing Map	Mạng tự tổ chức
8	SVM	Support Vector Machine	

DANH MỤC HÌNH ẢNH

Hình 1.1 Quy trình phát hiện tri thức	4
Hình 1.2 Cấu trúc nơ-ron thần kinh.....	110
Hình 1.3 Đơn vị xử lý.....	12
Hình 1.4 Hàm đồng nhất.....	14
Hình 1.5 Hàm bước nhị phân.....	14
Hình 1.6 Hàm Sigmoid.....	15
Hình 1.7 Hàm Sigmoid lưỡng cực.....	15
Hình 1.8 Mạng truyền thẳng.....	16
Hình 1.9 Mạng hồi quy	16
Hình 1.10 Mạng truyền thẳng nhiều lớp	18
Hình 2.1 Các học phần ở học kì thứ 5	21
Hình 2.2 Các học phần tự chọn ở kì 6	22
Hình 2.3 Mô hình tư vấn môn học tự chọn	23
Hình 2.4 Bảng điểm ở học kì thứ 5	24
Hình 2.5 Học phần tự chọn	24
Hình 2.6 Mô hình áp dụng vào thực tiễn	25
Hình 3.1 Giao diện đăng nhập	27
Hình 3.2 Tải vào dữ liệu điểm LTM	28
Hình 3.3 Cài đặt thông số mạng	29
Hình 3.4 Tiến hành huấn luyện	30
Hình 3.5 Kiểm thử dữ liệu	30
Hình 3.6 Trọng số liên kết	31
Hình 3.7 Dự báo điểm cho sinh viên	32

Hình 3.8 Dự báo dùng cho giảng viên	33
Hình 3.9 Xuất báo cáo ra file Excel	33
Hình 3.10 Truy cập vào folder FinalProjet.....	34
Hình 3.11 Truy cập folder Setup1	35
Hình 3.12 Truy cập vào folder Debug	35
Hình 3.13 Dữ liệu điểm	36
Hình 3.14 Cài đặt thành công	36
Hình 3.15 Khởi chạy chương trình.....	37

DANH MỤC BẢNG BIỂU

Bảng 3.1.....	28
Bảng 3.2.....	28
Bảng 3.3.....	31

MỞ ĐẦU

Đặt vấn đề

Trong những năm gần đây, khi công nghệ thông tin đang ngày càng phát triển mạnh mẽ thì vấn đề khai phá dữ liệu đã và đang trở thành một trong những hướng nghiên cứu chính trong lĩnh vực khoa học máy tính và công nghệ tri thức. Khai phá dữ liệu được ứng dụng thành công vào rất nhiều các lĩnh vực khác nhau như thương mại, tài chính, thị trường chứng khoán, y học, thiên văn học, sinh học, giáo dục và viễn thông..v.v.

Hình thức đào tạo tín chỉ là một phương pháp đào tạo tiên tiến trên thế giới và đang được áp dụng rộng rãi tại một số trường đại học của Việt Nam hiện nay. Phương pháp đào tạo theo hệ thống tín chỉ ra đời vào năm 1873 tại Đại học Harvard, Hoa Kỳ và sau đó được lan tỏa tới các nước khác, trước hết là các nước tây Âu từ những năm 1966 và hiện nay đã phổ biến trên toàn thế giới. Ưu điểm của hình thức đào tạo này là tính liên thông giữa các hệ đào tạo, giữa các trường để tạo điều kiện cho việc hội nhập với giáo dục thế giới. Tuy nhiên hình thức đào tạo này vẫn còn một số điểm hạn chế như kiến thức bị phân mảnh và khó tạo nên sự gắn kết trong sinh viên.

Hình thức này sẽ bao gồm: các học phần bắt buộc và các học phần tự chọn. Đối với các học phần bắt buộc, sinh viên sẽ thực hiện đăng ký theo chương trình đào tạo của nhà trường, mà không có sự lựa chọn nào khác. Bên cạnh đó là những học phần tự chọn, những môn học mà tùy vào sở thích hoặc định hướng của mỗi sinh viên, các bạn sẽ có lựa chọn cho riêng mình.

Vậy vấn đề ở đây là liệu có một tiêu chí hay phương pháp nào, mà sinh viên có thể dựa vào đó để đưa ra lựa chọn cho các môn học tự chọn một cách hợp lý và hiệu quả hơn. Đó cũng là lý do mà thúc đẩy em tiến hành nghiên cứu đề tài này.

Mục tiêu của đề tài

Nắm được kiến thức cơ bản về khai phá dữ liệu.

Tìm hiểu về mạng Nơ-ron (Neural Network).

Áp dụng mô hình mạng Nơ-ron (Neural Network) vào tư vấn môn học tự chọn cho sinh viên.

Hỗ trợ cho sinh viên và CVHT có thể tư vấn cho sinh viên trong việc lựa chọn môn học tự chọn, từ đó nâng cao hiệu quả học tập và giảng dạy.

Nội dung nghiên cứu

Ứng dụng khai phá dữ liệu và mô hình mạng nơ-ron nhân tạo để giải quyết bài toán tư vấn môn học tự chọn cho sinh viên.

Ý nghĩa của đề tài

Xuất phát từ mong muốn tạo ra một công cụ hỗ trợ cho việc tư vấn môn học tự chọn cho sinh viên, từ đó giúp cho việc xây dựng lộ trình học tập phù hợp cho sinh viên cũng như hỗ trợ công tác giảng dạy của giảng viên.

Cấu trúc báo cáo đồ án tốt nghiệp

Mở đầu

Chương 1: Trình bày cơ sở lý thuyết của khai phá dữ liệu và mạng nơ-ron nhân tạo.

Chương 2: Ứng dụng mạng nơ-ron vào dự báo điểm môn học tự chọn cho sinh viên.

Chương 3: Chạy chương trình và nghiệm thu kết quả.

Kết luận và hướng phát triển.

CHƯƠNG 1. CƠ SỞ LÝ THUYẾT

1.1 Khai phá dữ liệu

1.1.1 Khai phá dữ liệu là gì

Khai phá dữ liệu (Data Mining) được định nghĩa như là một quá trình chất lọc hay khai phá tri thức từ một lượng lớn dữ liệu. Một ví dụ hay được sử dụng là việc khai thác vàng từ đá và cát, Data Mining được ví như công việc "Đãi cát tìm vàng" trong một tập hợp lớn các dữ liệu cho trước. Thuật ngữ Data Mining ám chỉ việc tìm kiếm một tập hợp nhỏ có giá trị từ một số lượng lớn các dữ liệu thô. Có nhiều thuật ngữ hiện được dùng cũng có nghĩa tương tự với từ Data Mining như Knowledge Mining (khai phá tri thức), Knowledge Extraction (chất lọc tri thức), Data/Pattern Analysis (phân tích dữ liệu/mẫu), Data Archaeology (khảo cổ dữ liệu), Data Dredging (nạo vét dữ liệu),...

Định nghĩa: Khai phá dữ liệu là một tập hợp các kỹ thuật được sử dụng để tự động khai thác và tìm ra các mối quan hệ lẫn nhau của dữ liệu trong một tập hợp dữ liệu không lồ và phức tạp, đồng thời cũng tìm ra các mẫu tiềm ẩn trong tập dữ liệu đó.[5]

1.1.2 Các ứng dụng của khai phá dữ liệu

Trong lĩnh vực tài chính: ứng dụng của Data Mining trong lĩnh vực này được dùng để tăng độ trung thành của khách hàng bằng cách thu thập và phân tích dữ liệu hành vi của khách hàng.

Trong lĩnh vực chăm sóc sức khỏe: việc tìm ra mối quan hệ giữa các loại bệnh và hiệu quả của phương pháp điều trị sẽ giúp thay đổi các loại thuốc mới hoặc đảm bảo rằng bệnh nhân được chăm sóc phù hợp, kịp thời.

Trong lĩnh vực viễn thông: ứng dụng của data mining giúp ngành viễn thông đạt được lợi thế cạnh tranh và giảm chi phí khách hàng bằng cách hiểu các đặc điểm nhân khẩu học và dự đoán hành vi của khách hàng.

Trong lĩnh vực Marketing và Sales: khai phá dữ liệu cho phép các doanh nghiệp hiểu các điều ẩn đằng sau dữ liệu giao dịch mua bán của khách hàng.

Trong lĩnh vực thương mại điện tử: nhiều công ty đang áp dụng ứng dụng của Data Mining để bán hàng qua nhiều nước thông qua các trang web của họ, tiêu biểu như Amazon.

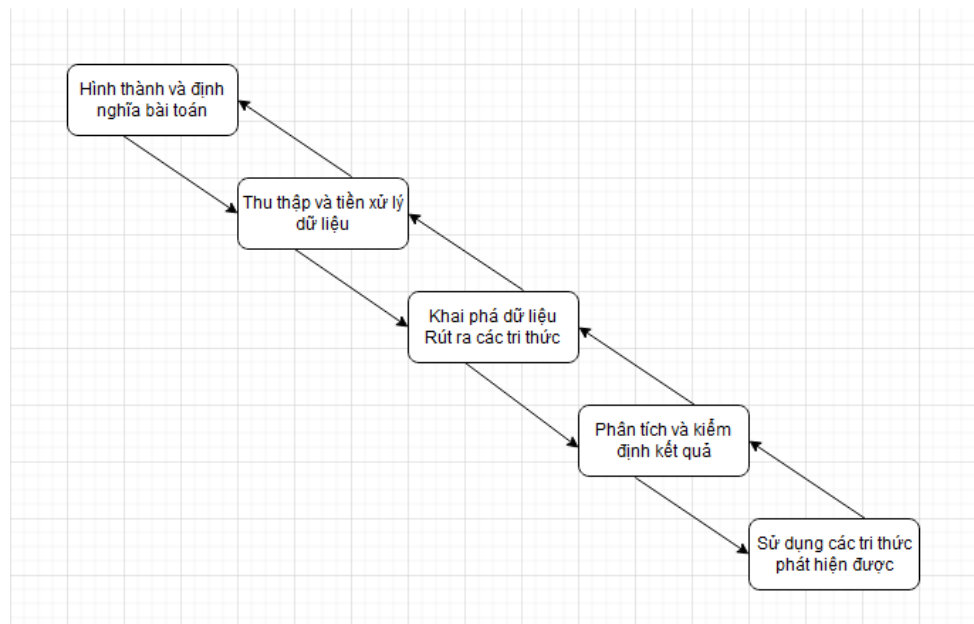
Trong lĩnh vực giáo dục: sử dụng khai thác dữ liệu để đưa ra quyết định chính xác và cũng như để dự đoán kết quả của học sinh. Từ đó cải tiến các phương pháp dạy học hiệu quả hơn.

Trong lĩnh vực kỹ thuật sản xuất: thường được sử dụng để khám phá các mẫu (patterns) trong quy trình sản xuất phức tạp.[11]

1.1.3 Quy trình phát hiện tri thức

Năm 1989, Fayyad, Piatetsky-Shapiro và Smyth đã dùng khái niệm Phát hiện tri thức trong cơ sở dữ liệu (Knowledge Discovery in Database – KDD) để chỉ toàn bộ quá trình phát hiện các tri thức có ích từ các tập dữ liệu lớn. Trong đó, KPDL là một bước đặc biệt trong toàn bộ tiến trình, sử dụng các giải thuật đặc biệt để chiết xuất ra các mẫu (pattern) (hay các mô hình) từ dữ liệu.[2]

Quy trình phát hiện tri thức thường tuân theo các bước sau:



Hình 1.1 Quy trình phát hiện tri thức

Bước thứ nhất: Hình thành, xác định và định nghĩa bài toán.

Là tìm hiểu lĩnh vực ứng dụng từ đó hình thành bài toán, xác định các nhiệm vụ cần phải hoàn thành.

Bước thứ hai: Thu thập và tiền xử lý dữ liệu.

Là thu thập và xử lý thô, còn được gọi là tiền xử lý dữ liệu nhằm loại bỏ nhiễu (làm sạch dữ liệu), xử lý việc thiếu dữ liệu (làm giàu dữ liệu), biến đổi dữ liệu và rút gọn dữ liệu nếu cần thiết. Sau bước này, dữ liệu sẽ nhất quán, đầy đủ, được rút gọn và rời rạc hoá.

Bước thứ ba: Khai phá dữ liệu, rút ra các tri thức.

Là khai phá dữ liệu, hay nói cách khác là trích ra các mẫu hoặc các mô hình ẩn dưới các dữ liệu. Tùy theo bài toán xác định được mà ta lựa chọn các phương pháp khai phá dữ liệu cho phù hợp.

Bước thứ 4: Phân tích và kiểm định kết quả.

Là hiểu tri thức đã tìm được, đặc biệt là làm sáng tỏ các mô tả và dự đoán.

Bước thứ 5: Sử dụng các tri thức phát hiện được. Là hiểu tri thức đã tìm được, đặc biệt là làm sáng tỏ các mô tả và dự đoán.

Các bước trên có thể lặp đi lặp lại một số lần, kết quả thu được có thể được lấy trung bình trên tất cả các lần thực hiện. Do các kết quả có thể là các dự đoán hoặc các mô tả nên chúng có thể được đưa vào các hệ thống hỗ trợ ra quyết định nhằm tự động hoá quá trình này.

Tóm lại: KDD là một quá trình kết xuất ra tri thức từ kho dữ liệu mà trong đó khai phá dữ liệu là công đoạn quan trọng nhất.

1.1.4 Nhiệm vụ chính trong khai phá dữ liệu

Quá trình khai phá dữ liệu là quá trình phát hiện ra mẫu thông tin.

Trong đó, giải thuật khai phá tìm kiếm các mẫu đáng quan tâm theo dạng xác định như các luật, phân lớp, hồi quy, cây quyết định, ...

Phân lớp (phân loại - classification)

Là việc xác định một hàm ánh xạ từ một mẫu dữ liệu vào một trong số các lớp đã được biết trước đó. Mục tiêu của thuật toán phân lớp là tìm ra mối quan hệ nào đó giữa thuộc tính dự báo và thuộc tính phân lớp.

Ví dụ: Một mục biểu diễn thông tin về nhân viên có các thuộc tính dự báo là: họ tên, tuổi, giới tính, trình độ học vấn, ... và thuộc tính phân loại là trình độ lãnh đạo của nhân viên.

Hồi quy (regression)

Là việc học một hàm ánh xạ từ một mẫu dữ liệu thành một biến dự đoán có giá trị thực. Nhiệm vụ của hồi quy tương tự như phân lớp, điểm khác nhau chính là ở chỗ thuộc tính để dự báo là liên tục chứ không phải rời rạc.

Ứng dụng của hồi quy là rất nhiều, ví dụ: ước lượng xác suất người bệnh có thể chết bằng cách kiểm tra các triệu chứng; dự báo nhu cầu của người dùng đối với một sản phẩm, ...

Phân nhóm (clustering)

Là việc mô tả chung để tìm ra các tập hay các nhóm, loại mô tả dữ liệu. Các nhóm có thể tách nhau hoặc phân cấp hay gộp lên nhau. Các ứng dụng khai phá dữ liệu có nhiệm vụ phân nhóm như phát hiện tập các khách hàng có phản ứng giống nhau trong CSDL tiếp thị; xác định các quang phổ từ các phương pháp đo tia hồng ngoại, ...

Tổng hợp (summarization)

Là công việc liên quan đến các phương pháp tìm kiếm một mô tả tập con dữ liệu, thường áp dụng trong việc phân tích dữ liệu có tính thăm dò và báo cáo tự động. Nhiệm vụ chính là sản sinh ra các mô tả đặc trưng cho một lớp.

Mô hình hoá sự phụ thuộc (dependency modeling)

Là việc tìm kiếm một mô hình mô tả sự phụ thuộc giữa các biến, thuộc tính theo hai mức: Mức “*cấu trúc của mô hình*” mô tả (thường dưới dạng đồ thị) và mức “*định lượng mô hình*” mô tả mức độ phụ thuộc. Những phụ thuộc này thường được biểu thị dưới dạng theo luật “*nếu - thì*” (nếu tiền đề là đúng thì kết luận đúng).

Phát hiện sự biến đổi và độ lệch (change and deviation detection)

Tập trung vào khám phá hầu hết sự thay đổi có nghĩa dưới dạng độ đo đã biết trước hoặc giá trị chuẩn, phát hiện độ lệch đáng kể giữa nội dung của tập con dữ liệu thực và nội dung mong đợi. Hai mô hình độ lệch hay dùng là “*lệch theo thời gian*” hay “*lệch theo nhóm*”.

Vì những nhiệm vụ này yêu cầu số lượng và các dạng thông tin rất khác nhau nên chúng thường ảnh hưởng đến việc thiết kế và chọn phương pháp khai phá dữ liệu khác nhau..

1.1.5 Các phương pháp khai phá dữ liệu

Có thể kể ra đây một vài phương pháp như: sử dụng công cụ truy vấn, xây dựng cây quyết định, dựa theo khoảng cách (K-láng giềng gần), giá trị trung bình, phát hiện luật kết hợp, ...

Các thành phần của giải thuật khai phá dữ liệu

Giải thuật khai phá dữ liệu bao gồm 3 thành phần chính như sau: biểu diễn mô hình, kiểm định mô hình và phương pháp tìm kiếm.

-Biểu diễn mô hình: mô hình được biểu diễn theo một ngôn ngữ L nào đó để miêu tả các mẫu có thể khai thác được.

Mô hình ban đầu được xác định bằng cách kết hợp biến đầu ra (phụ thuộc) với các biến độc lập mà biến đầu ra phụ thuộc vào. Sau đó phải tìm những tham số mà bài toán cần tập trung giải quyết.

-Kiểm định mô hình (model evaluation): Là việc đánh giá, ước lượng các mô hình chi tiết, chuẩn trong quá trình xử lý và phát hiện tri thức với sự ước lượng có dự báo chính xác hay không và có thoả mãn cơ sở logic hay không

-Phương pháp tìm kiếm: bao gồm hai thành phần: “tìm kiếm tham số” và “tìm kiếm mô hình”.

Phương pháp suy diễn / quy nạp

Có hai kỹ thuật chính để thực hiện việc này là suy diễn và quy nạp.

-Phương pháp suy diễn: nhằm rút ra thông tin là kết quả logic của các thông tin trong cơ sở dữ liệu. Phương pháp suy diễn dựa trên các sự kiện chính xác để suy ra các tri thức mới từ các thông tin cũ.

-Phương pháp quy nạp: phương pháp quy nạp suy ra các thông tin được sinh ra từ cơ sở dữ liệu. Có nghĩa là nó tự tìm kiếm, tạo mẫu và sinh ra tri thức chứ không phải bắt đầu với các tri thức đã biết trước

Phương pháp ứng dụng K-láng giềng gần

Kỹ thuật K-láng giềng gần là một phương pháp tìm kiếm đơn giản. Tuy nhiên, nó có một số mặt hạn chế giới hạn phạm vi ứng dụng của nó. Đó là thuật toán này có độ phức tạp tính toán là lũy thừa bậc 2 theo số bản ghi của tập dữ liệu.

Phương pháp sử dụng cây quyết định và luật

Với kỹ thuật phân lớp dựa trên cây quyết định, kết quả của quá trình xây dựng mô hình sẽ cho ra một cây quyết định. Cây này được sử dụng trong quá trình phân lớp các đối tượng dữ liệu chưa biết hoặc đánh giá độ chính xác của mô hình.

Phương pháp phát hiện luật kết hợp

Phương pháp này nhằm phát hiện ra các luật kết hợp giữa các thành phần dữ liệu trong cơ sở dữ liệu. Mẫu đầu ra của giải thuật khai phá dữ liệu là tập luật kết hợp tìm được.

1.1.6 Lợi thế của khai phá dữ liệu so với phương pháp cơ bản

Học máy (Machine Learning)

Mặc dù người ta đã cố gắng cải tiến các phương pháp học máy để có thể phù hợp với mục đích khai phá dữ liệu nhưng sự khác biệt giữa cách thiết kế, các đặc điểm của cơ sở dữ liệu đã làm cho phương pháp học máy trở nên không phù hợp với mục đích này.

Phương pháp hệ chuyên gia

Phương pháp này khác với khai phá dữ liệu ở chỗ các ví dụ của chuyên gia thường ở mức chất lượng cao hơn rất nhiều so với các dữ liệu trong cơ sở dữ liệu, và chúng thường chỉ bao được các trường hợp quan trọng.

Phát kiến khoa học

Khai phá dữ liệu rất khác với phát kiến khoa học ở chỗ khai phá trong CSDL ít có chủ tâm và có điều kiện hơn. Hơn nữa, các nhà khoa học có thể tạo lại các thí nghiệm và có thể tìm ra rằng các thiết kế ban đầu không thích hợp. Trong khi đó, các nhà quản lý cơ sở dữ liệu hầu như không thể xa xỉ đi thiết kế lại các trường dữ liệu và thu thập lại dữ liệu.

Phương pháp thống kê

Sự khác nhau cơ bản giữa khai phá dữ liệu và thống kê là ở chỗ khai phá dữ liệu là một phương tiện được dùng bởi người sử dụng đầu cuối chứ không phải là các nhà thống kê.

1.1.7 Lựa chọn phương pháp

Các giải thuật khai phá dữ liệu tự động vẫn mới chỉ ở giai đoạn phát triển ban đầu. Người ta vẫn chưa đưa ra được một tiêu chuẩn nào trong việc quyết định sử dụng phương pháp nào và trong trường hợp nào thì có hiệu quả.

Mỗi phương pháp đều có điểm mạnh và yếu của nó, nhưng hầu hết các điểm yếu đều có thể khắc phục được.

1.1.8 Những thách thức

Ở đây, ta đưa ra một số khó khăn trong việc nghiên cứu và ứng dụng kỹ thuật khai phá dữ liệu.

1.1.8.1 Các vấn đề về cơ sở dữ liệu

-Dữ liệu lớn: cho đến nay, các cơ sở dữ liệu với hàng trăm trường và bảng, hàng triệu bản ghi và với kích thước đến gigabytes đã là chuyện bình thường. Các phương pháp giải quyết hiện nay là đưa ra một ngưỡng cho cơ sở dữ liệu, lâu mẫu, các phương pháp xấp xỉ, xử lý song song (Agrawal et al, Holsheimer et al).

-Kích thước lớn: không chỉ có số lượng bản ghi lớn mà số các trường trong cơ sở dữ liệu cũng nhiều. Vì vậy mà kích thước của bài toán trở nên lớn hơn. Biện pháp khắc phục là làm giảm kích thước tác động của bài toán và sử dụng các tri thức biết trước để xác định các biến không phù hợp.

-Dữ liệu động: đặc điểm cơ bản của hầu hết các cơ sở dữ liệu là nội dung của chúng thay đổi liên tục. Dữ liệu có thể thay đổi theo thời gian và việc khai phá dữ liệu cũng bị ảnh hưởng bởi thời điểm quan sát dữ liệu. Vấn đề này được giải quyết bằng các giải pháp tăng trưởng để nâng cấp các mẫu và coi những thay đổi như là cơ hội để khai thác bằng cách sử dụng nó để tìm kiếm các mẫu bị thay đổi.

-Các trường không phù hợp: một đặc điểm quan trọng khác là tính không thích hợp của dữ liệu, nghĩa là mục dữ liệu trở thành không thích hợp với trọng tâm hiện tại của việc khai thác.

-Các giá trị bị thiếu: sự có mặt hay vắng mặt của giá trị các thuộc tính dữ liệu phù hợp có thể ảnh hưởng đến việc khai phá dữ liệu.

-Các trường bị thiếu: một quan sát không đầy đủ cơ sở dữ liệu có thể làm cho các dữ liệu có giá trị bị xem như có lỗi. Việc quan sát cơ sở dữ liệu phải phát hiện được toàn bộ các thuộc tính có thể dùng để giải thuật khai phá dữ liệu có thể áp dụng nhằm giải quyết bài toán.

-Độ nhiễu và không chắc chắn: đối với các thuộc tính đã thích hợp, độ nghiêm trọng của lỗi phụ thuộc vào kiểu dữ liệu của các giá trị cho phép. Các giá trị của các thuộc tính khác nhau có thể là các số thực, số nguyên, chuỗi và có thể thuộc vào tập các giá trị định danh. Các giá trị định danh này có thể sắp xếp theo thứ tự từng phần hoặc đầy đủ, thậm chí có thể có cấu trúc ngữ nghĩa.

-Mối quan hệ phức tạp giữa các trường: các thuộc tính hoặc các giá trị có cấu trúc phân cấp, các mối quan hệ giữa các thuộc tính và các phương tiện phức tạp để diễn tả tri thức về nội dung của cơ sở dữ liệu yêu cầu các giải thuật phải có khả năng sử dụng một cách hiệu quả các thông tin này

1.1.8.2 Một số vấn đề khác

-“Quá phù hợp” (Overfitting): khi một giải thuật tìm kiếm các tham số tốt nhất cho đó sử dụng một tập dữ liệu hữu hạn, nó có thể sẽ bị tình trạng “quá độ” dữ liệu (nghĩa là tìm kiếm quá mức cần thiết gây ra hiện tượng chỉ phù hợp với các dữ liệu đó mà không có khả năng đáp ứng cho các dữ liệu lạ), làm cho mô hình hoạt động rất kém đối với các dữ liệu thử.

-Đánh giá tầm quan trọng thống kê: ta có thể sử dụng phương pháp điều chỉnh thống kê trong kiểm tra như một hàm tìm kiếm, ví dụ như điều chỉnh Bonferroni đối với các kiểm tra độc lập.

-Khả năng biểu đạt của mẫu: trong rất nhiều ứng dụng, điều quan trọng là những điều khai thác được phải càng dễ hiểu với con người càng tốt. Vì vậy, các giải pháp thường bao gồm việc diễn tả dưới dạng đồ họa, xây dựng cấu trúc luật với các đồ thị có hướng (Gaines), biểu diễn bằng ngôn ngữ tự nhiên (Matheus et al.) và các kỹ thuật khác nhằm biểu diễn tri thức và dữ liệu.

-Sự tương tác với người sử dụng và các tri thức sẵn có: rất nhiều công cụ và phương pháp khai phá dữ liệu không thực sự tương tác với người dùng và không dễ dàng kết hợp cùng với các tri thức đã biết trước đó. [1]

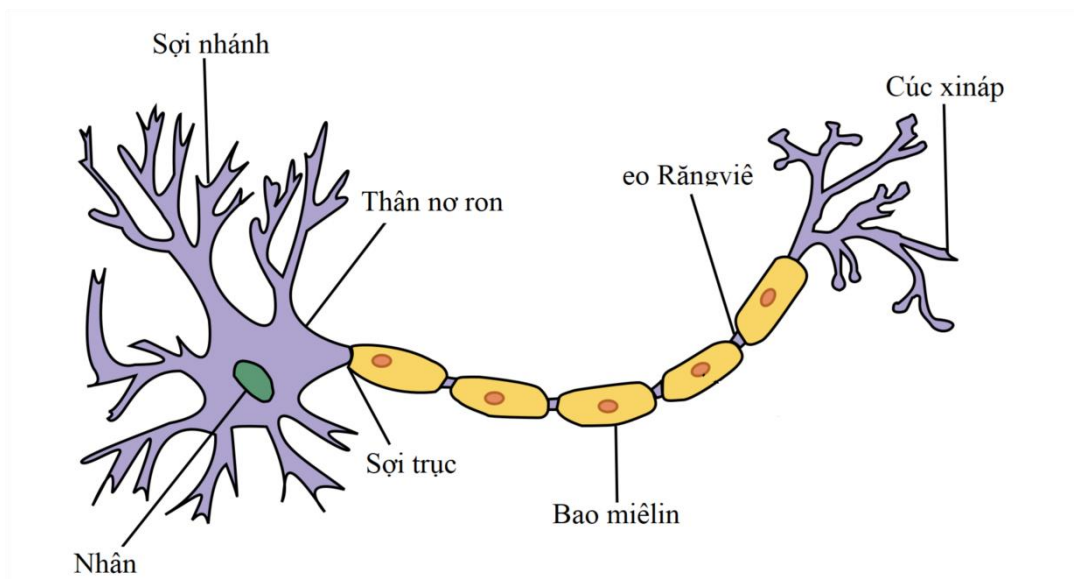
1.2 Mạng Nơ-ron nhân tạo (Artificial Neural Network)

1.2.1 Tìm hiểu về Nơ-ron thần kinh

Neuron là tế bào thần kinh có khả năng cảm ứng, phát sinh xung thần kinh và dẫn truyền xung điện này, là đơn vị cơ bản cấu tạo hệ thần kinh của hầu hết các loài động vật và là thành phần quan trọng bậc nhất của não. Thân và sợi nhánh của các neuron tạo thành chất xám. Sợi trục (nếu đi chung với nhau thành bó gọi là dây thần kinh) cấu tạo chất trắng trong não.

1.2.1.1 Cấu tạo

Mỗi Neuron gồm: thân chứa nhân (chịu trách nhiệm xử lý thông tin); các sợi nhánh (tiếp nhận thông tin); sợi trục (truyền thông tin xử lý đến nơ-ron khác), bao gồm các xynap, bao quanh sợi trục có các bao mielin, khoảng cách giữa các bao mielin là các eo răng-vi-ê.



Hình 1.2 Cấu trúc nơ-ron thần kinh

“Nguồn: Bài 43 Giới thiệu chung hệ thần kinh, trường THPT Nguyễn Huệ”

1.2.1.2 Chức năng

Cảm ứng và dẫn truyền xung thần kinh dưới dạng các tín hiệu hóa học trong các hoạt động điều khiển, điều hoà và phối hợp mọi hoạt động của các cơ quan và hệ cơ quan trong cơ thể nhằm tạo sự thích nghi với các thay đổi từ môi trường bên trong và bên ngoài.

1.2.1.3 Phân loại

- Theo hướng dẫn truyền xung thần kinh, bao gồm: neuron hướng tâm, neuron trung gian và neuron li tâm.

- Theo chức năng, bao gồm: neuron cảm giác, neuron chuyên tiếp, neuron vận động.[9]

1.2.2 Giới thiệu về mạng Nơ-ron nhân tạo

1.2.2.1 Khái niệm

Mạng nơ-ron nhân tạo (ANN) là một mô hình toán học được xây dựng dựa trên cơ sở các mạng nơ-ron sinh học gồm một số lượng lớn các phần tử (gọi là nơ-ron) kết nối với nhau thông qua các liên kết (gọi là trọng số liên kết) làm việc như một thể thống nhất để giải quyết các vấn đề cụ thể như nhận dạng mẫu, phân loại dữ liệu, v.v... thông qua một quá trình học từ tập các mẫu huấn luyện.

1.2.2.2 Lịch sử phát triển

- Nghiên cứu từ năm 1943: mô hình nơ-ron McCulloch & Pitts.
- Perceptron (Rosenblatt, 1958): mô phỏng thị giác người.
- Giữa thập niên 1980: Mạng nơ-ron Giải thuật lan truyền ngược: Rumelhart, Williams, Hinton.
- Mạng SOM: (Kohonen, 1984).
- Mạng nơ-ron tích chập: Hinton, LeCun, Bengio.
- Mạng SVM: (Vapnik et al., 1995).
- Deep learning: (Hinton et al., 2006).

1.2.3 Các đặc trưng cơ bản của mạng nơ-ron

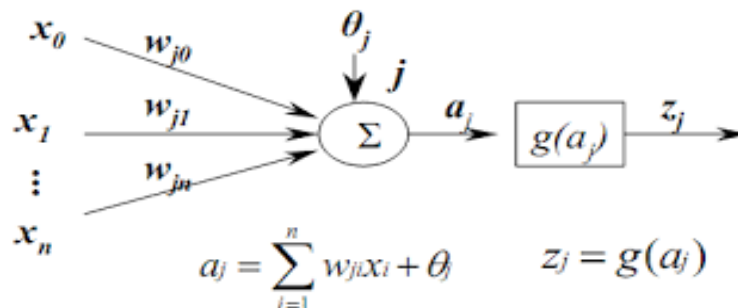
- Gồm một tập các đơn vị xử lý (các nơ-ron nhân tạo).
- Trạng thái kích hoạt hay đầu ra của đơn vị xử lý.

- Liên kết giữa các đơn vị. Xét tổng quát, mỗi liên kết được định nghĩa bởi một trọng số W_{jk} cho ta biết hiệu ứng mà tín hiệu của đơn vị j có trên đơn vị k .
- Một luật lan truyền quyết định cách tính tín hiệu ra của từng đơn vị từ đầu vào của nó.
- Một hàm kích hoạt, hay hàm chuyển (activation function, transfer function), xác định mức độ kích hoạt khác dựa trên mức độ kích hoạt hiện tại.
- Một đơn vị điều chỉnh (độ lệch).
- Phương pháp thu thập thông tin (luật học - learning rule), môi trường hệ thống có thể hoạt động.

1.2.4 Các thành phần cơ bản của mạng nơ-ron nhân tạo

1.2.4.1 Đơn vị xử lý

Còn được gọi là một nơron hay một nút (node), thực hiện một công việc rất đơn giản: nó nhận tín hiệu vào từ các đơn vị phía trước hay một nguồn bên ngoài và sử dụng chúng để tính tín hiệu ra sẽ được lan truyền sang các đơn vị khác.



Hình 1.3 Đơn vị xử lý

“Nguồn: Lý thuyết mạng Nơron – Đại học Duy Tân”

Trong đó:

- x_i : các đầu vào
- w_{ji} : các trọng số tương ứng với các đầu vào
- θ_j : độ lệch (bias)
- a_j : đầu vào mạng (net-input)
- z_j : đầu ra của nơron
- $g(x)$: hàm chuyển (hàm kích hoạt).

Trong một mạng nơron có ba kiểu đơn vị:

- Các đơn vị đầu vào (Input units): nhận tín hiệu từ bên ngoài.
- Các đơn vị đầu ra (Output units): gửi dữ liệu ra bên ngoài.
- Các đơn vị ẩn (Hidden units), tín hiệu vào (input) và ra (output) của nó nằm

trong mạng. Mỗi đơn vị j có thể có một hoặc nhiều đầu vào nhưng chỉ có một đầu ra z_j . Một đầu vào tới một đơn vị có thể là dữ liệu từ bên ngoài mạng, hoặc đầu ra của một đơn vị khác, hoặc là đầu ra của chính nó.

1.2.4.2 Hàm kết hợp

Mỗi một đơn vị (sigma units) trong một mạng kết hợp các giá trị đưa vào nó thông qua các liên kết với các đơn vị khác, sinh ra một giá trị gọi là net input. Hàm thực hiện nhiệm vụ này gọi là hàm kết hợp (combination function), được định nghĩa bởi một luật lan truyền cụ thể.

$$a_j = \sum_{i=1}^n w_{ji}x_i + \theta_j$$

Nếu $w_{ji} > 0$, nơron ở trong “trạng thái kích thích”; còn nếu $w_{ji} < 0$, nơron ở “trạng thái kiềm chế”. Trong một vài trường hợp người ta cũng có thể sử dụng các luật lan truyền phức tạp hơn.

Rất nhiều hàm kết hợp sử dụng một "độ lệch" hay "ngưỡng" để tính net input tới đơn vị. Đối với một đơn vị đầu ra tuyến tính, thông thường, θ_j được chọn là hằng số và trong bài toán xấp xỉ đa thức $\theta_j = 1$.

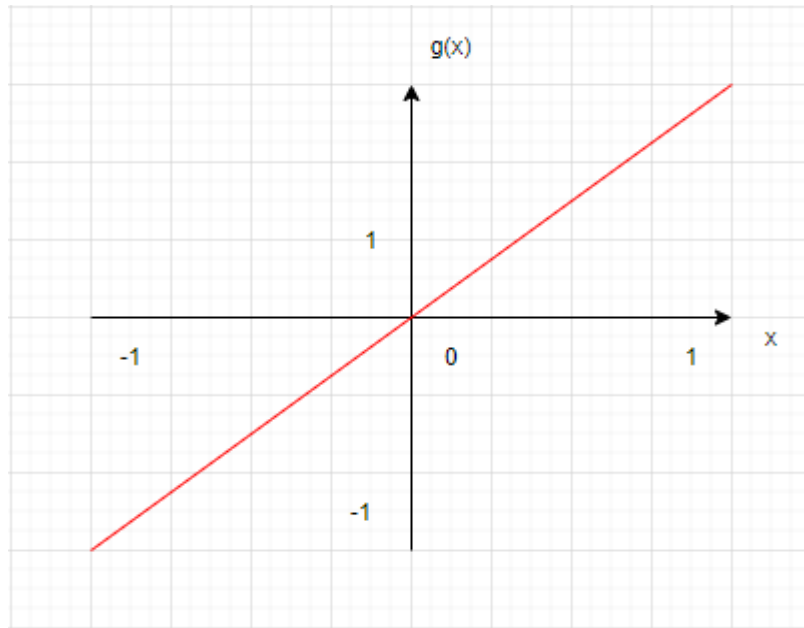
1.2.4.3 Hàm kích hoạt

Phần lớn các đơn vị trong mạng nơron chuyển net input bằng cách sử dụng một hàm vô hướng (scalar-to-scalar function) gọi là hàm kích hoạt, kết quả của hàm này là một giá trị gọi là mức độ kích hoạt của đơn vị (unit's activation). Loại trừ khả năng đơn vị đó thuộc lớp ra, giá trị kích hoạt được đưa vào một hay nhiều đơn vị khác. Các hàm kích hoạt thường bị ép vào một khoảng giá trị xác định. Các hàm kích hoạt hay được sử dụng là:

Hàm đồng nhất

Nếu coi các đầu vào là một đơn vị thì chúng sẽ sử dụng hàm này. Đôi khi một hằng số được nhân với net-input để tạo ra một hàm đồng nhất.

$$g(x)=x$$

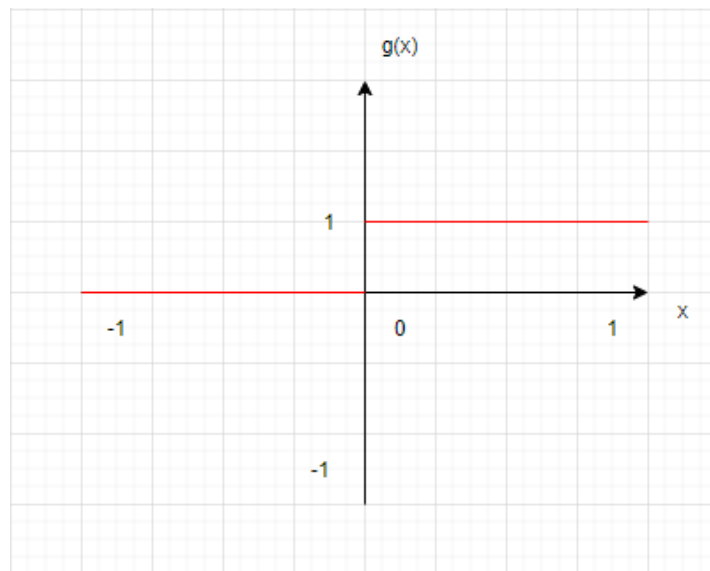


Hình 1.4 Hàm đồng nhất

Hàm bước nhị phân

Dạng hàm này được sử dụng trong các mạng chỉ có một lớp, đầu ra của hàm này được giới hạn vào một trong hai giá trị:

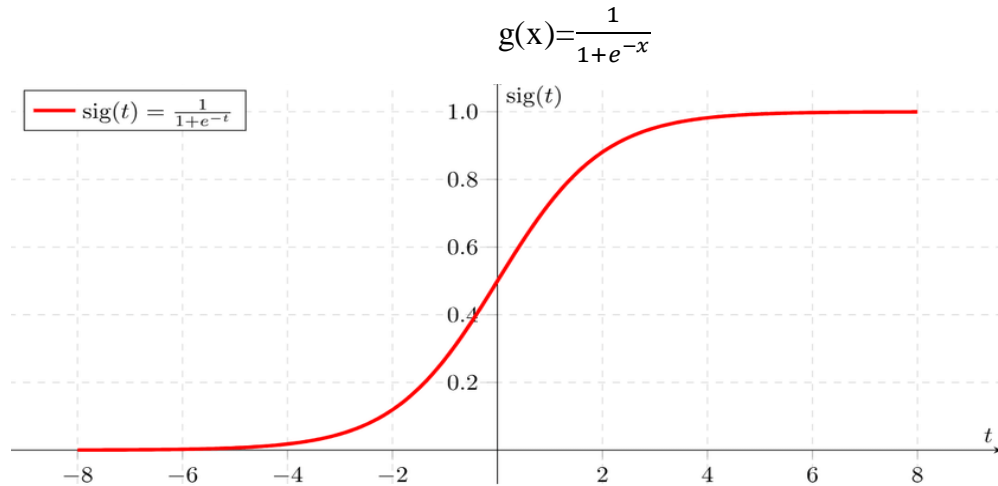
$$g(x) = \begin{cases} 1, & x \geq \theta \\ 0, & x < \theta \end{cases}$$



Hình 1.5 Hàm bước nhị phân

Hàm sigmoid

Hàm này đặc biệt thuận lợi khi sử dụng cho các mạng được huấn luyện (trained) bởi thuật toán “Lan truyền ngược” (back-propagation), bởi vì nó dễ lấy đạo hàm, do đó có thể giảm đáng kể tính toán trong quá trình huấn luyện, đầu ra của hàm rơi vào khoảng $[0,1]$.

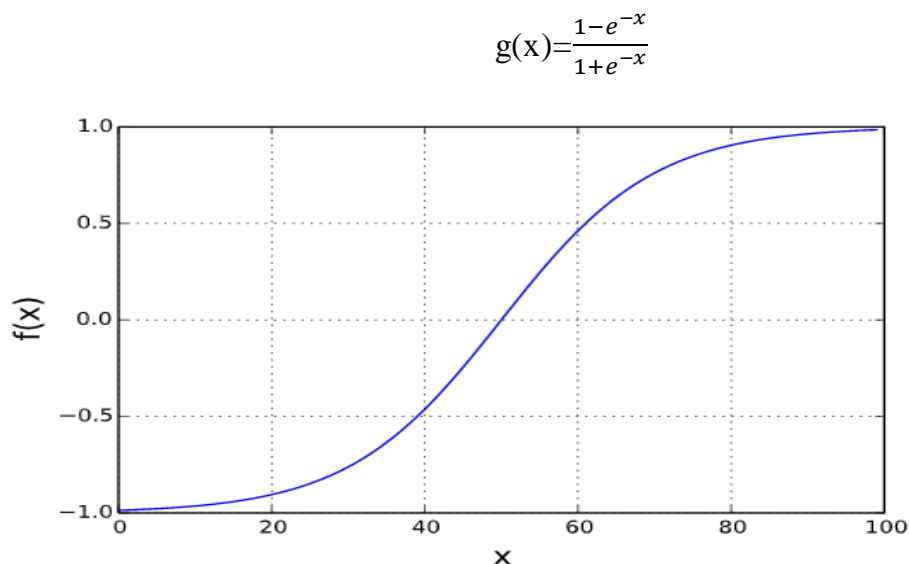


Hình 1.6 Hàm Sigmoid

“Nguồn: Plot of the sigmoid function-Researchget”

Hàm sigmoid lưỡng cực

Hàm này có các thuộc tính tương tự hàm sigmoid. Nó làm việc tốt đối với các ứng dụng có đầu ra yêu cầu trong khoảng $[-1,1]$.



Hình 1.7 Hàm Sigmoid lưỡng cực

“Nguồn: Bipolar Sigmoid- An overview of activation functions used in neural networks”

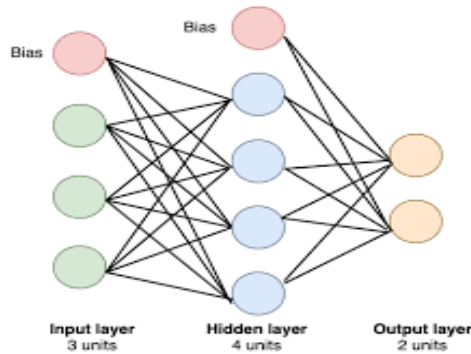
Các hàm chuyển của các đơn vị ẩn (hidden units) là cần thiết để biểu diễn sự phi tuyến vào trong mạng. Đối với luật học lan truyền ngược, hàm phải khả vi và sẽ có ích nếu như hàm được gán trong một khoảng nào đó. Do vậy, hàm sigmoid là lựa chọn thông dụng nhất.

1.2.5 Các hình trạng của mạng

Hình trạng của mạng được định nghĩa bởi: số lớp (layers), số đơn vị trên mỗi lớp, và sự liên kết giữa các lớp như thế nào. Dựa trên cách thức liên kết các đơn vị, có 2 loại:

Mạng truyền thẳng

Dòng dữ liệu từ đơn vị đầu vào đến đơn vị đầu ra chỉ được truyền thẳng. Việc xử lý dữ liệu có thể mở rộng ra nhiều lớp, nhưng không có các liên kết phản hồi.

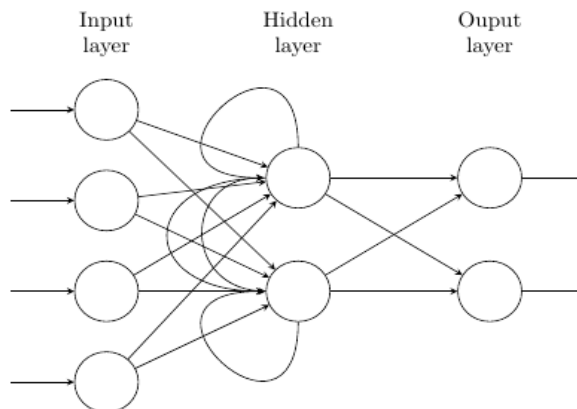


Hình 1.8 Mạng truyền thẳng

“Nguồn: ProqramerSought.com”

Mạng hồi quy

Có chứa các liên kết ngược. Khác với mạng truyền thẳng, các thuộc tính động của mạng mới quan trọng. [5]



Hình 1.9 Mạng hồi quy

“Nguồn: Theappsolution”

1.2.6 Các hình thức huấn luyện mạng nơ-ron nhân tạo

Theo hình thức học, các thuật toán phổ biến bao gồm: Học có giám sát, Học không giám sát, Học bán giám sát và Học tăng cường.

Học có giám sát

Là thuật toán dự đoán đầu ra của một dữ liệu mới dựa trên các cặp (input, outcome) đã biết từ trước. Bao gồm 2 loại chính: Phân loại (Classification) và Hồi quy (Regression).

Học không giám sát

Trong thuật toán này, chúng ta không biết được outcome hay nhãn mà chỉ có dữ liệu đầu vào, bao gồm 2 loại chính: Clustering và Association.

Học bán giám sát

Các bài toán khi chúng ta có một lượng lớn dữ liệu XX nhưng chỉ một phần trong chúng được gán nhãn.

Học tăng cường

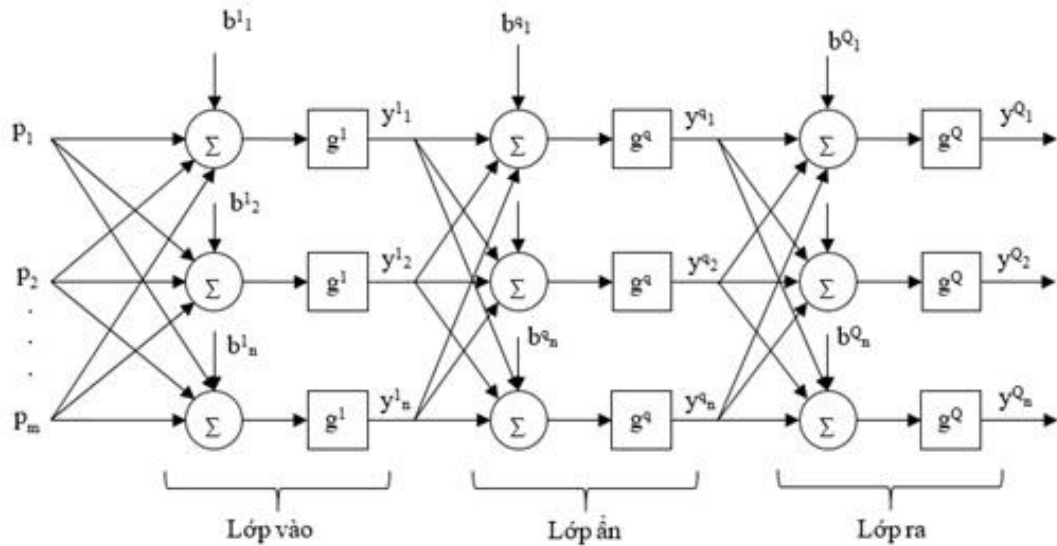
Là các bài toán giúp cho một hệ thống tự động xác định hành vi dựa trên hoàn cảnh để đạt được lợi ích cao nhất (maximizing the performance). [8]

1.2.7 Mạng truyền thẳng và thuật toán lan truyền ngược

Mạng truyền thẳng

Mạng truyền thẳng trình bày trong phần này là mạng truyền thẳng có nhiều lớp (MLP - MultiLayer Perceptron). Đây là một trong những mạng truyền thẳng điển hình, thường được sử dụng trong các hệ thống nhận dạng.

Một mạng truyền thẳng nhiều lớp bao gồm: một lớp vào, một lớp ra và một hoặc nhiều lớp ẩn. Các nơron đầu vào nhận tín hiệu từ bên ngoài, các nơron ở lớp ẩn và lớp ra thực hiện các tính toán, kết quả được định dạng bởi hàm đầu ra (hàm chuyển).



Hình 1.10 Mạng truyền thẳng nhiều lớp

“Nguồn: Bộ tài nguyên môi trường”

Phương trình mô tả mạng:

$$y_i^q = g_i^q(u_i^q) = g_i^q\left(\sum_{j=1}^{n^{q-1}} w_{ij}^q y_j^{q-1} + b_i^q\right)$$

Trong đó:

y_i^q : đầu ra của nơron thứ i ở lớp thứ q, ($i=1...n^q$), ($q=1,...Q$)

n^q : số nơron ở lớp thứ q.

Q: Số lớp nơron của mạng.

b_i^q : ngưỡng của nơron thứ i ở lớp thứ q.

w_{ij}^q : trọng số liên kết giữa đầu ra thứ j của lớp thứ q-1 đến nơron thứ i của

lớp thứ q, $j=1...n^{q-1}$

Đầu vào mạng $y^0 = p$.

Đầu ra mạng $y = y^q$

Hàm năng lượng của mạng có thể tính theo biểu thức:

$$E(w) = \frac{1}{2} \sum_{j=1}^{n^Q} (d_i - y_i^Q)^2$$

Trong đó: d_i là đầu ra mong muốn của nơron thứ i ở lớp ra.

Hầu hết các mạng nơron thực tế chỉ có 2 hoặc 3 lớp, rất hiếm mạng có từ 4 lớp trở lên. Các đặc điểm của tín hiệu đầu ra sẽ quyết định hàm truyền của mạng ở lớp ra.[7]

Thuật toán lan truyền ngược

Các mạng nơ-ron truyền thẳng nhiều lớp được luyện bằng phương pháp học có kiểm soát.

Thuật toán lan truyền ngược là dạng tổng quát của thuật toán trung bình bình phương tối thiểu (Least Means Square-LMS).

Thuật toán lan truyền ngược sẽ bao gồm 2 giai đoạn:

Giai đoạn lan truyền thẳng

Bước 1: Vector thông số đầu vào được nhập vào các nơ-ron ở lớp nhập.

$$a^{(0)} = x$$

Bước 2: Tại nơ-ron lớp ẩn thứ j , giá trị tín hiệu nhận từ lớp nhập sẽ được tính bằng tổng trọng số hóa của tất cả các dữ liệu được nhập bằng tổng của tích các đầu vào và trọng số liên kết của nó.

$$z_{in j} = b_{oj} + \sum_{i=1}^n x_i v_{ij}$$

Bước 3: Sau đó, hàm kích hoạt sẽ được sử dụng để chuyển giá trị nhận được thành giá trị đầu ra.

$$z_j = f(z_{in j})$$

Tiếp theo, giá trị đầu ra tại nơ-ron lớp ẩn j tiếp tục được truyền đến nơ-ron lớp xuất k giống với phương thức từ lớp nhập đến lớp ẩn.

$$y_{in k} = w_{ok} + \sum_{j=1}^n z_j w_{jk}$$

Sau đó, ta lại sử dụng hàm truyền để tính giá trị đầu ra của nơ-ron tại lớp xuất.

$$y_k = f(y_{in k})$$

Giai đoạn lan truyền ngược

Bước 4: Với mỗi bộ số liệu bao gồm đầu vào và giá trị thực tế, ta tính được sai số đầu ra tương ứng.

$$J = t_k - y_k$$

Bước 5: Tính đạo hàm của hàm này theo trọng số giữa lớp ẩn-lớp xuất và lớp nhập-lớp ẩn.

$$\Delta w_{jk} = \frac{\partial J}{\partial w_{jk}}$$

$$\Delta v_{ij} = \frac{\partial J}{\partial v_{ij}}$$

Bước 6: Cập nhật lại giá trị trọng số liên kết giữa lớp ẩn-lớp nhập và lớp ẩn-lớp xuất (α là hệ số học).[3]

$$w_{jk}(\text{new}) = w_{jk}(\text{old}) + \alpha \cdot \Delta w_{jk}$$

$$v_{ij}(\text{new}) = v_{ij}(\text{old}) + \alpha \cdot \Delta v_{ij}$$

1.3 Tổng quan về ngôn ngữ lập trình C#

C# (hay C sharp) là một ngôn ngữ lập trình đơn giản, được phát triển bởi đội ngũ kỹ sư của Microsoft vào năm 2000, trong đó người dẫn đầu là Anders Hejlsberg và Scott Wiltamuth.

C# là ngôn ngữ lập trình hiện đại, hướng đối tượng và nó được xây dựng trên nền tảng của hai ngôn ngữ mạnh nhất là C++ và Java.

1.3.1 Đặc trưng của C#

Là ngôn ngữ cơ bản và hiện đại.

Là ngôn ngữ lập trình hướng đối tượng.

Là ngôn ngữ ít từ khóa.

1.3.2 Tính ứng dụng

Xây dựng ứng dụng Desktop.

Công cụ phát triển game đa nền tảng Unity.

Ứng dụng Web và Mobile. [10]

CHƯƠNG 2. ỨNG DỤNG MÔ HÌNH MẠNG NƠI RON VÀO DỰ BÁO ĐIỂM MÔN TỰ CHỌN CHO SINH VIÊN

2.1 Đặt vấn đề

-Ngày nay chương trình đào tạo tại các trường đại học, cao đẳng đang được tổ chức theo hình thức tín chỉ, nhằm tạo điều kiện phù hợp cho lực học của sinh viên.

-Hình thức này sẽ bao gồm: các “học phần bắt buộc” và các “học phần tự chọn”. Đối với các học phần bắt buộc, sinh viên sẽ thực hiện đăng ký theo chương trình đào tạo của nhà trường, mà không có sự lựa chọn nào khác. Bên cạnh đó là những học phần tự chọn, những môn học mà tùy vào sở thích hoặc định hướng của mỗi sinh viên, các bạn sẽ có lựa chọn cho riêng mình.

-Như đã nói ở trên, vì đối với các học phần bắt buộc thì mỗi sinh viên phải hoàn thành theo chương trình khung của nhà trường nên đề tài này tập trung chủ yếu vào các học phần tự chọn, cụ thể là tìm ra mối quan hệ giữa điểm các môn của học kì trước có ảnh hưởng như thế nào đến điểm số của các học phần tự chọn ở học kỳ kế tiếp. Từ đó, có thể đưa ra dự báo về số điểm mà sinh viên có thể đạt được ở các môn học đó.

-Sau đây, chúng ta cùng xem qua về khung chương trình của ngành Công nghệ thông tin của trường Đại học Giao thông Vận tải Phân hiệu tại Thành phố Hồ Chí Minh.

♥ Học kỳ 1 năm học 2018_2019

1	GDTCC	Chứng chỉ Giáo dục thể chất		Xem	Giáo dục đại cương		1	0	0	0	1
2	CPM06.3	Phân tích thiết kế hệ thống	7	Xem			3	30	15	15	1
3	CPM211.3	Lập trình trực quan	8	Xem			3	30	15	0	1
4	GDT05.1	Giáo dục thể chất F5	8	Xem	Kỳ 5		1	0	30	0	1
5	KHM07.3	An toàn và bảo mật thông tin	9.30	Xem			3	30	30	0	1
6	MHT06.3	Mạng máy tính	7.60	Xem			3	30	15	0	1
7	ANHB1.4	Tiếng Anh B1		Xem	Tiếng anh B1		4	45	45	0	1
8	KHM08.3	Phân tích thiết kế thuật toán	8.70	Xem	Tự chọn 1	3	3	30	15	15	1
9	KHM202.3	Lý thuyết thông tin và Entropy		Xem			3	30	15	15	1
10	MHT09.2	Thiết kế cơ sở dữ liệu		Xem	Tự chọn 2	2	2	15	15	15	1
11	MHT10.2	Hệ điều hành Unix	8.80	Xem			2	15	0	30	1

Hình 2.1 Các học phần ở học kì thứ 5

4	KHM09.3	Đồ họa máy tính	8.50	Xem	Tự chọn 1	3	3	30	15	15	2
5	MHT22.3	Lập trình mạng		Xem			3	30	15	15	2
6	CPM08.3	Công nghệ phần mềm	6.50	Xem	Tự chọn 2	3	3	30	30	0	2
7	CPM212.3	Lập trình sử dụng API		Xem			3	30	15	15	2
8	MHT21.3	Thực hành mạng		Xem	Tự chọn 3	3	3	15	15	45	2
9	MHT234.3	Lập trình thiết bị di động	10	Xem			3	30	15	15	2
10	KHM11.3	Nguyên lý các ngôn ngữ lập trình	8	Xem	Tự chọn 4	3	3	30	15	15	2
11	KHM13.3	Chương trình dịch		Xem			3	30	15	15	2

Hình 2.2 Các học phần tự chọn ở kì 6

(Lưu ý: Ở đây chúng ta xem xét các học phần ở học kỳ 5 và các môn học tự chọn ở học kỳ 6)

-Vậy sinh viên đang gặp phải những vấn đề gì với việc đăng ký môn học tự chọn như hiện nay. Thứ nhất, chọn môn học nào để phù hợp với năng lực của bản thân ; Thứ hai, chọn môn học nào phù hợp với định hướng nghề nghiệp trong tương lai ; Thứ ba, chọn môn học nào để đạt được kết quả tốt nhất với lực học hiện tại.

-Từ đó, em thấy được sự cần thiết phải có một phương pháp, để phần nào giúp cho việc thực hiện đăng ký môn học tự chọn được khoa học hơn, cũng như giúp cho công tác giảng dạy và học tập trở nên hiệu quả hơn.

-Những lợi ích có thể mang lại của việc tư vấn môn học tự chọn cho sinh viên:

+) Đối với sinh viên: cung cấp cho sinh viên thêm một phương pháp, tiêu chí để chọn môn học tự chọn phù hợp.

+) Đối với giảng viên:

Cố vấn học tập: hỗ trợ sinh viên tốt hơn trong việc lựa chọn môn học, từ đó hoàn thành tốt vai trò của một cố vấn học tập.

Đối với giảng viên bộ môn: có thể tiếp cận được nhóm sinh viên có năng lực phù hợp với môn học, từ đó nâng cao chất lượng giảng dạy.

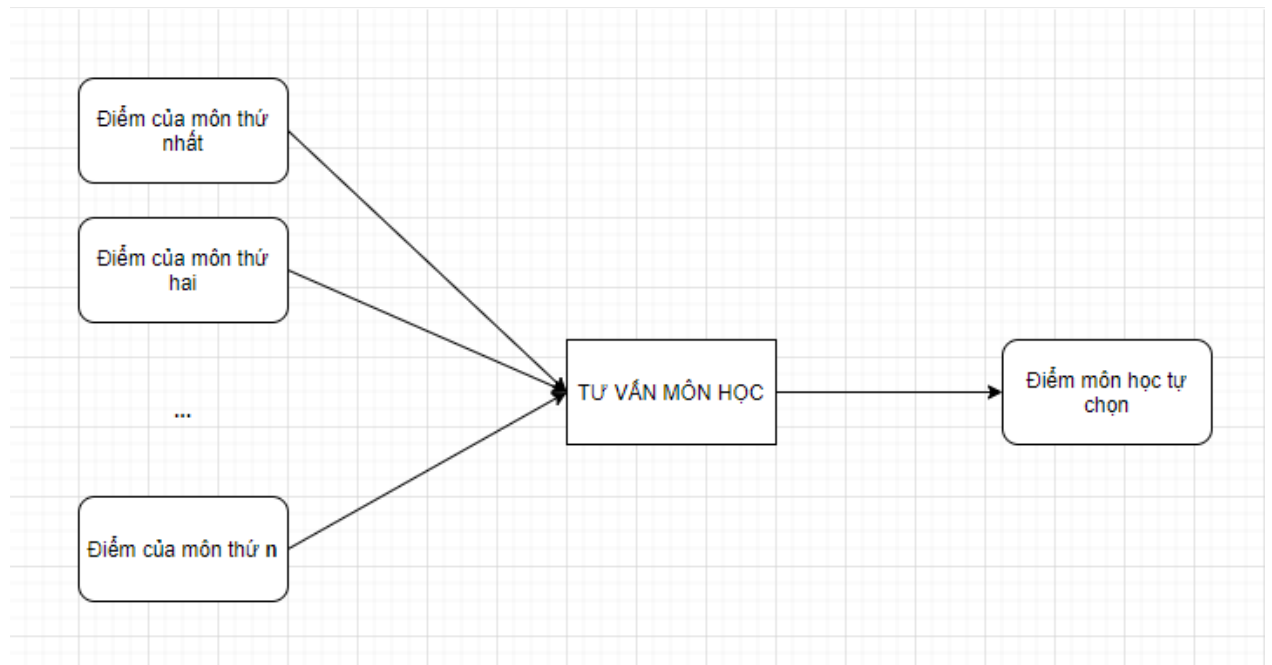
+) Đối với nhà trường: cải thiện chất lượng giáo dục-đào tạo của trường, định hướng con đường sự nghiệp tốt hơn cho sinh viên. Từ đó, nâng cao uy tín, vị thế của trường trong ngành giáo dục nói chung và từng chuyên ngành đào tạo nói riêng, dẫn đến thu hút các bạn học sinh đến học tập tại trường.

Qua đó cho thấy, xây dựng một giải pháp hỗ trợ tư vấn môn học tự chọn cho sinh viên là một việc hết sức cần thiết và hứa hẹn sẽ đem lại hiệu quả đáng mong đợi. Hiện nay, có nhiều cách thức có thể được áp dụng để giải quyết vấn đề này, nhưng xét thấy tính khả dụng của mạng nơ-ron nên ở đây, em áp dụng để đưa ra một giải pháp cho bài toán này.

2.2 Áp dụng mạng Nơ-ron vào tư vấn môn học

- Qua quá trình tìm hiểu, em thấy các học phần của kì trước hầu hết có ảnh hưởng đến các học phần ở kì tiếp theo. Ở đây, mô hình dự báo dùng để tư vấn môn học tự chọn cho sinh viên sẽ gồm các đầu vào là điểm của các môn học của kì trước, và đầu ra sẽ là điểm số của môn học tự chọn ở kì tiếp theo “có thể đạt được” của sinh viên. Dưới đây là mô hình mô tả luồng thực hiện:

-Điểm của mỗi môn khi đưa vào chương trình sẽ được gán với một giá trị gọi là trọng số, hệ thống tiến hành tính toán và điều chỉnh trọng số sao cho phù hợp. Sau khi đã xác định được các trọng số này, nếu chúng ta truyền vào dữ liệu điểm các môn học kì trước, chúng ta sẽ nhận được số điểm của môn tự chọn mà ta có thể đạt được.



Hình 2.3 Mô hình tư vấn môn học tự chọn

2.3 Nghiên cứu thực tiễn

- Trong phạm vi đề tài này, chúng ta sẽ xem xét chương trình đào tạo của ngành Công nghệ thông tin thuộc trường Đại học Giao thông Vận tải Phân hiệu tại Thành phố Hồ Chí Minh. Như đã nói ở phần trên, các học phần ở đây cũng được chia thành 2 nhóm: “các học phần bắt buộc” và “các học phần tự chọn”. Sinh viên buộc phải đăng ký các học phần bắt buộc và có thể tùy chọn học phần tự chọn.

- Kết thúc mỗi học kì, sinh viên sẽ được biết điểm các môn học của mình:

<MSSV>	<exchange>	<PTTKHT>	<LTTQ>	<ATBMTT>	<MMT>	<PTTKTT>	<HDH>
2001110002	T	6.3	8.5	7.2	5.1	6.1	6
2001110003	T	5.2	8.5	6.7	5.9	6	6.8
2001110008	T	5.1	8.5	6.1	4.6	5.7	6.2
2001110012	T	5.1	8	6.4	5.2	5.9	5.1
2001110013	T	5.7	8.5	6.3	5.9	6.1	6.9
2001110014	T	6.2	9	6.4	5.9	6.4	7.4
2001110016	T	6.1	8	5.8	4.2	5.4	6.1
2001110018	T	6.6	9	7.1	5.9	5.8	5.9
2001110020	T	5.5	7	5.6	5	6.4	5.6
2001110025	T	5.3	9	7.1	3.9	5.7	6.5
2001110026	T	3.8	9.3	6.2	3.6	6.4	6.6
2001110028	T	7.4	9.5	6	6	6.5	7.2
2001110033	T	8.4	10	7.4	5.8	6.4	7.4
2001110036	T	7.5	9.5	7.9	5.5	6.2	6.1
2001110037	T	3.8	8.5	5.7	4.7	5.4	5.6
2001110040	T	6.3	8	5.9	5.4	6.5	6.3
2001110042	T	5.8	8	5.2	5.2	6.6	6.6
2001110043	T	6.6	9.5	7.5	6.2	6.7	6.9
2001110045	T	6.6	9	7.2	5.4	6.4	6.9
2001110046	T	6.9	9	7.9	5.8	7.4	7.6

Hình 2.4 Bảng điểm ở học kì thứ 5

Ở đây chúng ta xem xét bảng điểm các môn học ở học kì thứ 5, bao gồm các môn:

- +) Phân tích thiết kế hệ thống (PTTKHT).
- +) Lập trình trực quan (LTTQ).
- +) An toàn và bảo mật thông tin (ATBMTT).
- +) Mạng máy tính (MMT).
- +) Phân tích thiết kế thuật toán (PKTKTT) .
- +) Hệ điều hành Unix (HDH).

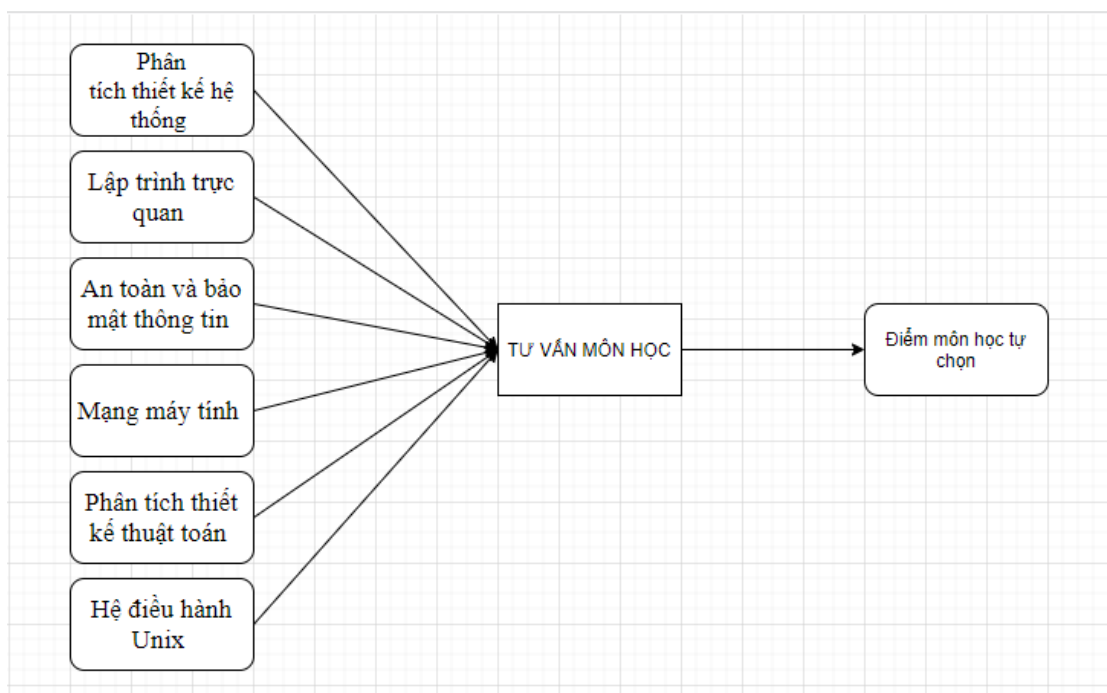
-Các môn tự chọn ở học kì tiếp theo:

4	KHM09.3	Đồ họa máy tính	8.50	Xem	Tự chọn 1	3	3	30	15	15	2
5	MHT22.3	Lập trình mạng		Xem			3	30	15	15	2
6	CPM08.3	Công nghệ phần mềm	6.50	Xem	Tự chọn 2	3	3	30	30	0	2
7	CPM212.3	Lập trình sử dụng API		Xem			3	30	15	15	2
8	MHT21.3	Thực hành mạng		Xem	Tự chọn 3	3	3	15	15	45	2
9	MHT234.3	Lập trình thiết bị di động	10	Xem			3	30	15	15	2
10	KHM11.3	Nguyên lý các ngôn ngữ lập trình	8	Xem	Tự chọn 4	3	3	30	15	15	2
11	KHM13.3	Chương trình dịch		Xem			3	30	15	15	2

Hình 2.5 Học phần tự chọn

Như đã thấy, ở đây chúng ta có nhiều học phần tự chọn, nhưng đâu là cách lựa chọn các học phần này. Thực tế ở trường hiện nay, các học phần tự chọn được sinh viên đăng ký ngẫu nhiên, đăng ký theo sự lựa chọn của các khóa trước, hay đăng ký theo lựa chọn của nhóm bạn chơi thân, thậm chí đăng ký theo học phần có nhiều sinh viên tham gia để đủ số lượng mở lớp học phần. Qua đó, em nhận ra thực trạng đáng quan ngại của tình trạng đăng ký môn học hiện nay. Do đó, ta cần phải tiến hành nghiên cứu tìm một giải pháp phù hợp, hiệu quả cho vấn đề này.

-Áp dụng mô hình tư vấn môn học tự chọn vào đây, em thu được mô hình như sau:



Hình 2.6 Mô hình áp dụng vào thực tiễn

2.4 Công cụ lập trình

- Chương trình được lập trình trên ứng dụng Visual Studio 2019 của Microsoft. Visual Studio bao gồm một trình soạn thảo mã hỗ trợ Intellisense cũng như cải tiến mã nguồn. Trình gỡ lỗi tích hợp hoạt động cả về trình gỡ lỗi mức độ mã nguồn và gỡ lỗi mức độ máy. Công cụ tích hợp khác bao gồm một mẫu thiết kế các hình thức xây dựng giao diện ứng dụng, thiết kế web, thiết kế lớp và thiết kế giản đồ cơ sở dữ liệu. Nó chấp nhận các plug-in nâng cao các chức năng ở hầu hết các cấp bao gồm thêm hỗ trợ cho các hệ thống quản lý phiên bản (như Subversion) và bổ sung thêm bộ công cụ mới như biên tập

và thiết kế trực quan cho các miền ngôn ngữ cụ thể hoặc bộ công cụ dành cho các khía cạnh khác trong quy trình phát triển phần mềm.

- Chương trình được viết bằng ngôn ngữ lập trình C#, chạy trên máy tính ASUS X555UJ Intel Core I5 6200U RAM 8GB, ROM 120GB, hệ điều hành Windows phiên bản Windows 10 Pro.

2.5 Cấu trúc tập dữ liệu

- Trong đề tài này, em sử dụng 2 tập dữ liệu môn “Lập trình mạng”, 1 tập hơn 400 bản ghi cho training và 1 tập dữ liệu gần 200 bản ghi để predict .

-Ở tập dữ liệu training, bao gồm điểm của các môn học kì trước và điểm của môn học tùy chọn tương ứng.

-Ở tập dữ liệu predict, chỉ bao gồm điểm của các môn học kì trước mà thôi.

(Lưu ý: tên các học phần ở đây theo như chương trình khung của trường. Còn dữ liệu điểm, do gặp một số khó khăn nhất định, nên em tạm lấy dữ liệu điểm ngẫu nhiên. Tuy nhiên, việc này không ảnh hưởng đến sự đúng đắn của thuật toán và chương trình.)

2.6 Yêu cầu hệ thống

-Yêu cầu chức năng: dự báo điểm của môn học tự chọn tương ứng, bao gồm chức năng “dự báo cho sinh viên”, “dự báo cho giảng viên” và “xuất báo cáo”.

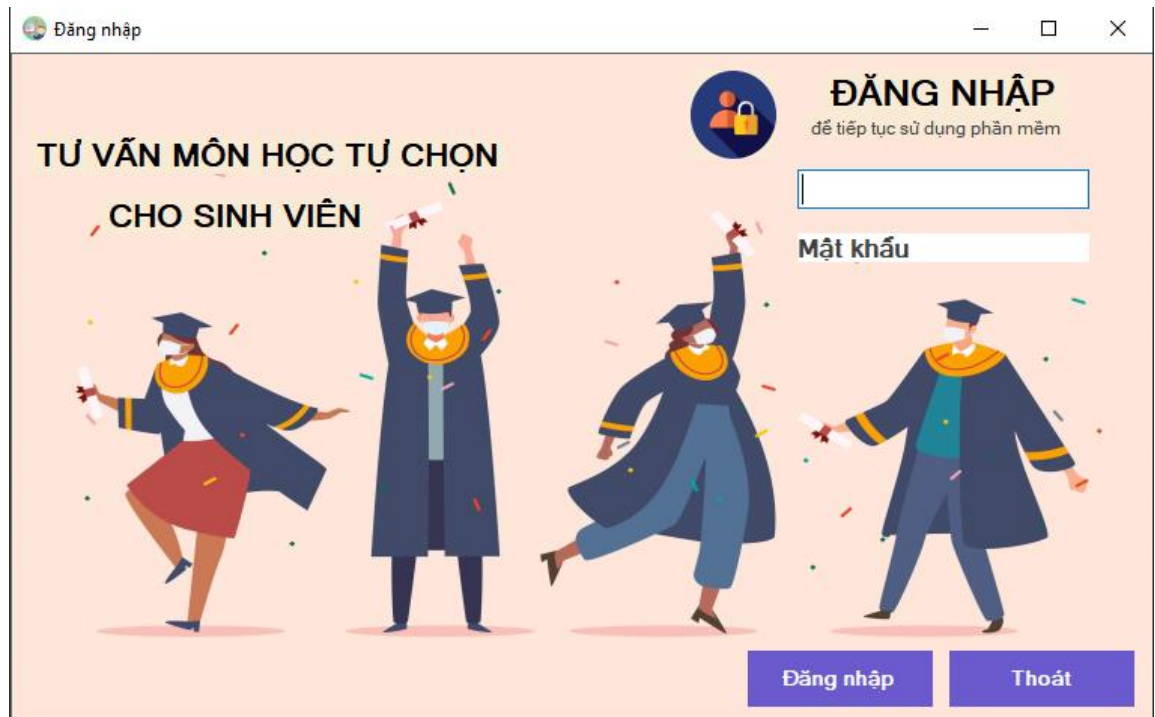
-Yêu cầu phi chức năng:

+) Sai số ở mức chấp nhận được.

+) Chương trình chạy ổn định, không chiếm quá nhiều thời gian.

CHƯƠNG 3. CHẠY CHƯƠNG TRÌNH VÀ NGHIỆM THU KẾT QUẢ

3.1 Giao diện đăng nhập chương trình



Hình 3.1 Giao diện đăng nhập

Tiến hành nhập “Tên tài khoản” và “Mật khẩu” để sử dụng chương trình.

3.2 Thao tác với chương trình

3.2.1 Tải dữ liệu vào chương trình

Sau khi khởi chạy chương trình, tiến hành chọn tệp dữ liệu (tệp dữ liệu training) ứng với môn học tự chọn mà ta muốn dự báo điểm và tải dữ liệu vào (như hình trên).

Sau đó, ta tiến hành chọn một vài bản ghi (dòng dữ liệu) để thêm vào “phần dữ liệu kiểm tra” và “phần dữ liệu huấn luyện”. Thông thường, “phần dữ liệu huấn luyện” và “phần dữ liệu kiểm tra” sẽ được chia theo tỉ lệ (80%:20%) hoặc (90%:10%) kích thước (số bản ghi hay số dòng dữ liệu) của tệp dữ liệu.

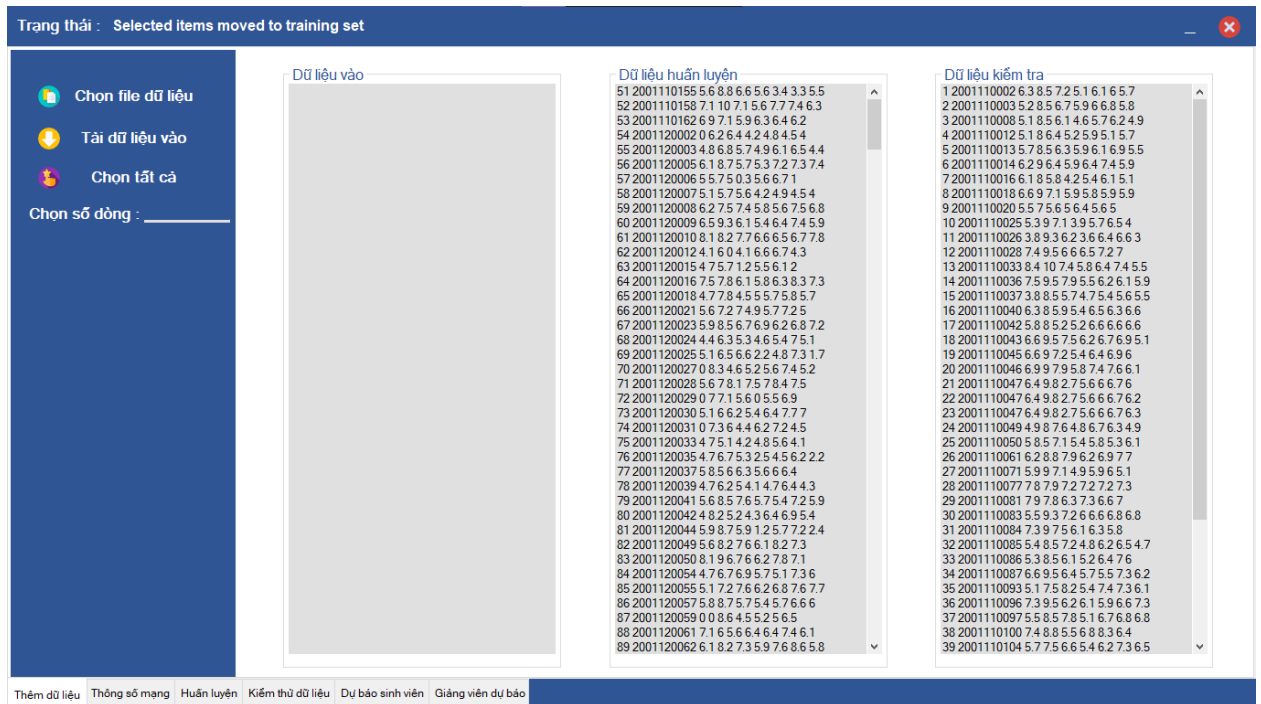
	Số lần lặp	Sai số huấn luyện	Sai số kiểm thử
Tỉ lệ: 90/10	10000	0.0851	0.1096
	20000	0.0820	0.1120
	40000	0.0784	0.1183

Bảng 3.1

	Số lần lặp	Sai số huấn luyện	Sai số kiểm thử
Tỉ lệ: 80/20	10000	0.0730	0.1274
	20000	0.0743	0.1400
	40000	0.0666	0.1436

Bảng 3.2

Ở đây, nhìn vào 2 bảng trên, em nhận thấy: sai số huấn luyện giảm dần khi tăng số lần lặp và kích thước tập kiểm thử. Ngược lại, sai số kiểm thử lại tăng lên. Vì vậy, để hạn chế sai số trong kiểm thử, đồng thời không khiến sai số huấn luyện quá lớn, em sẽ chọn tỉ lệ (90%:10%), với số lần lặp là 10000 (lần).

**Hình 3.2 Tải vào dữ liệu điểm LTM**

Ở đây, em tiến hành trên tập dữ liệu của môn “Lập trình mạng”.

3.2.2 Thông số mạng

Ở đây em tiến hành chọn thông số mạng, bao gồm:

- Lookback: số lớp đầu vào (Input Layer).
- Node ẩn: số nơ-ron ở lớp ẩn.
- Dự đoán thuộc tính: môn học tự chọn mà chúng ta muốn tiến hành dự báo điểm.
- Sigmoid Alpha: giá trị mặc định là 2.
- Tốc độ học (hằng số học): có vai trò điều tiết mức độ thay đổi của các trọng số trong mỗi lần cập nhật, mặc định là 0.001.
- Momentum (hằng số quán tính): giúp cho giải thuật không bị dừng ở tối ưu cục bộ, mặc định là 0.



Hình 3.3 Cài đặt thông số mạng

3.2.3 Tiến hành huấn luyện

- Ta chọn số lần lặp. Mặc định ở đây là 100000 (lần). Đương nhiên, ta có thể cập nhật lại số lần lặp mà ta mong muốn để dừng hoặc dừng khi cần thiết.

Trạng thái : Lần lặp thứ: 10000 Sai số: 0.0857092176847547

Số lần lặp: 10000

Cập nhật số lần lặp

Bắt đầu huấn luyện

Dừng huấn luyện

Phục hồi huấn luyện

Lưu thông số mạng

Lưu Vectors

Dữ liệu huấn luyện

51 56 88 66 56 34 33
52 71 10 71 56 77 74
53 69 71 59 63 64
54 62 64 42 48 45
55 48 68 57 49 61 65
56 61 87 57 53 72 73
57 55 75 35 66 67
58 51 57 56 42 49 45
59 62 75 74 58 56 75
60 65 93 61 54 64 74
61 81 82 77 66 66 56 7
62 41 6 41 66 66 7
63 47 57 12 55 61
64 75 78 61 58 63 83
65 47 78 45 55 75 8
66 56 72 74 95 77 2
67 59 85 67 69 62 68
68 44 63 53 46 54 7
69 51 65 66 22 48 73
70 83 46 52 56 74
71 56 78 17 57 84
72 77 15 6 55
73 51 66 25 46 47 7
74 73 64 46 27 2
75 47 51 42 48 56
76 47 67 53 25 45 62
77 58 56 63 56 6
78 47 62 54 14 76 4
79 56 85 76 57 54 72
80 48 25 43 46 46 9
81 59 87 59 12 57 72
82 56 82 76 61 82
83 81 96 76 62 78
84 47 67 69 57 51 73
85 51 72 76 62 68 76
86 58 87 57 54 57 66
87 86 45 52 5
88 71 65 66 46 47 4
89 61 82 73 59 76 86

Kết quả thực tế

51 55
52 63
53 62
54 4
55 44
56 74
57 1
58 4
59 68
60 59
61 78
62 43
63 2
64 73
65 57
66 5
67 72
68 51
69 17
70 52
71 75
72 69
73 7
74 45
75 41
76 22
77 64
78 43
79 59
80 54
81 24
82 73
83 71
84 6
85 77
86 6
87 65
88 61
89 58

Dự báo của

54212828665012
6263514318212
6095091122495
3769581296220
4106176659893
7123523825195
0948700127132
3814805196908
6668114525309
5854969852549
7573218906931
4267695249748
1950272731710
7051193039873
5674347152266
486093050318
7150856669441
5089224974404
1466874930187
5057089858435
7418670428603
6783133756010
6780840237916
4383857218010
3935354297292
1959529144177
6278869320903
4227423163667
5791686106444
5293152828721
2367966497180
7141804101363
6987677840834
5942541131392
7493320673404
5987029654861
6467151257268
6056307515784
5615951903012

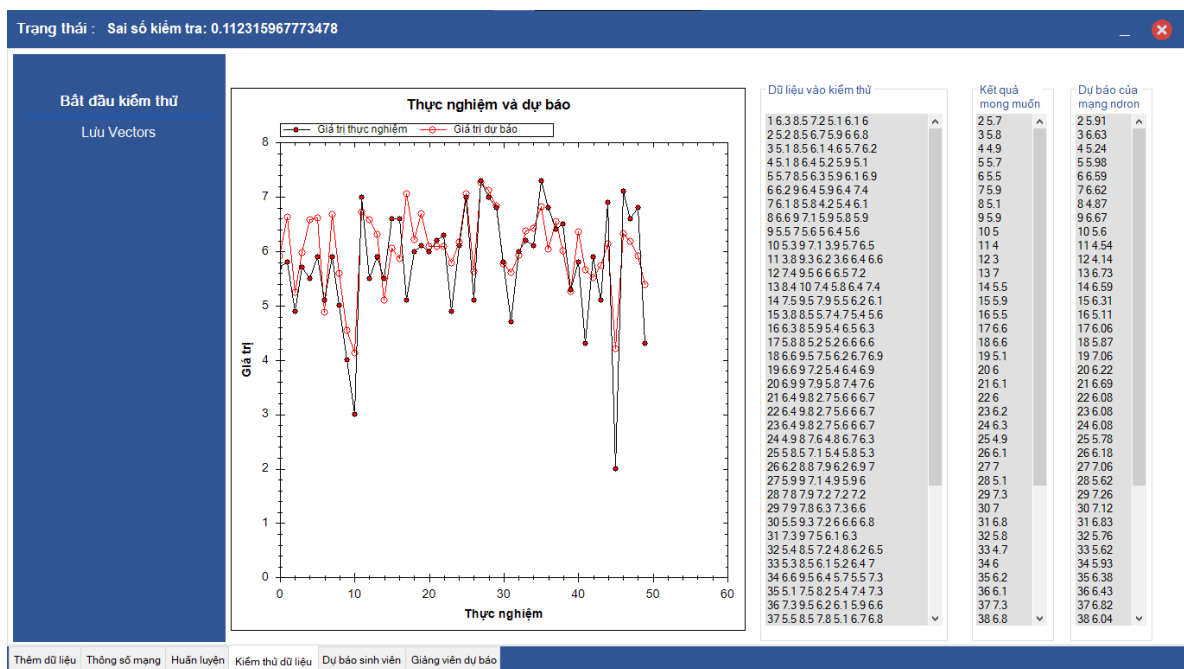
Thêm dữ liệu Thông số mạng Huấn luyện Kiểm thử dữ liệu Dự báo sinh viên Giảng viên dự báo

Hình 3.4 Tiến hành huấn luyện

Ở đây, em cập nhật lại số lần lặp là 10000 (lần), ta thu được kết quả như hình. Ở bước này, em chọn lưu thông số mạng (không bắt buộc).

3.2.4 Kiểm thử dữ liệu

- Để kiểm thử dữ liệu, ta chuyển sang chức năng kiểm thử và tiến hành kiểm thử dữ liệu.



Hình 3.5 Kiểm thử dữ liệu

Như hình, ta thấy được sự chênh lệch giữa giá trị thực nghiệm và giá trị dự báo. Trong đó, đường màu đen trên biểu đồ tương ứng với giá trị thực nghiệm. Còn đường màu đỏ ứng với giá trị dự báo.

3.2.5 Mỗi liên hệ:

Ở bước tiến hành huấn luyện, em đã chọn chức năng lưu thông số mạng. Lúc này, em sẽ có được một file như hình:

Ở đây, em thấy rằng: khi trọng số của của môn thứ 4 tăng (hay giảm) thì trọng số của môn thứ 7 (môn tự chọn) cũng tăng (tăng hay giảm theo). Đối chiếu với tập dữ liệu, em nhận thấy đó là môn Mạng máy tính và Lập trình mạng.

Như vậy, có thể thấy rằng, thông qua huấn luyện, mạng nơ-ron đã tìm ra được mối quan hệ hay sự ảnh hưởng của điểm số môn Mạng máy tính đến điểm số môn Lập trình mạng.

3.2.6 Dự báo điểm cho sinh viên

- Ở chức năng “Dự báo sinh viên”, sinh viên tiến hành nhập điểm tương ứng của các môn học kì trước và chương trình sẽ tiến hành dự báo số điểm có thể đạt được của môn học tự chọn tương ứng.

Nhập dữ liệu	
Điểm môn 1	1
Điểm môn 2	2
Điểm môn 3	3.5
Điểm môn 4	9
Điểm môn 5	4
Điểm môn 6	4
Dự báo	

Kết quả dự báo	
8.69	A

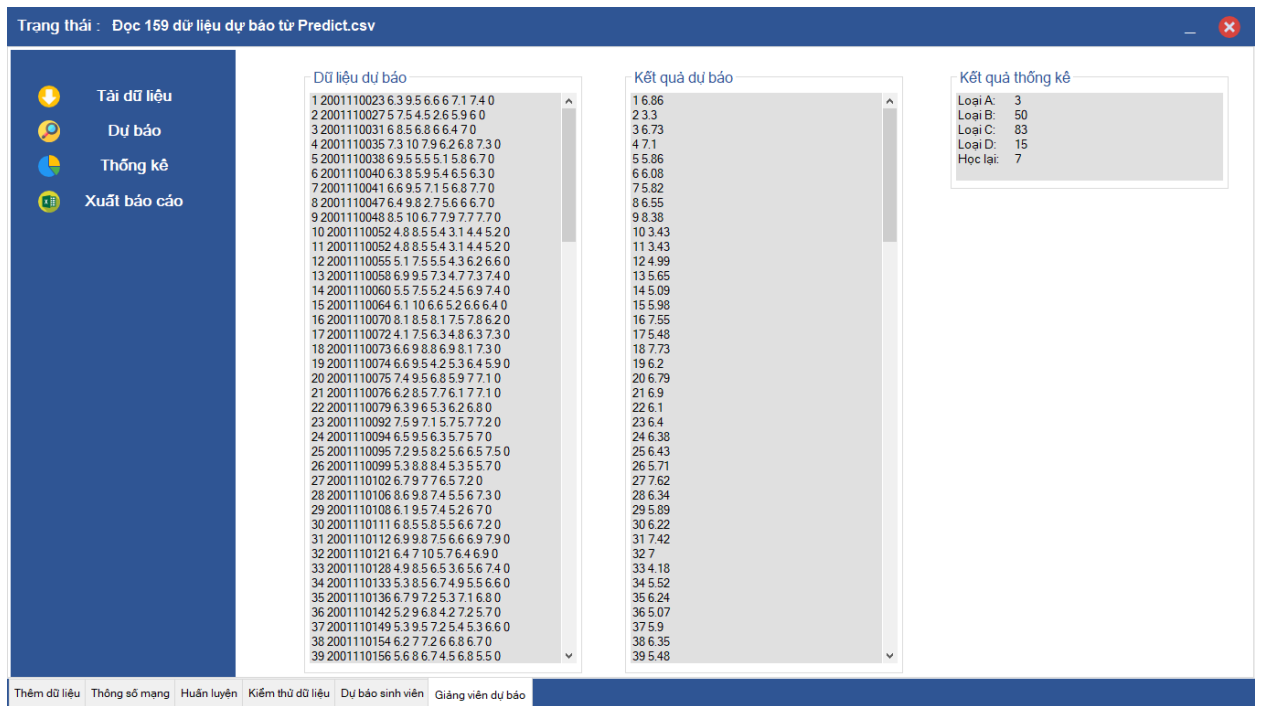
Thêm dữ liệu | Thông số mạng | Huấn luyện | Kiểm thử dữ liệu | **Dự báo sinh viên** | Giảng viên dự báo

Hình 3.7 Dự báo điểm cho sinh viên

Ở đây ta có thể thấy, chương trình dự báo điểm của môn tự chọn (Lập trình mạng) của sinh viên này là: 8.69 và đạt xếp loại A. Và ta cũng có thể thấy được phần nào sự ảnh hưởng của môn “Mạng máy tính” đến điểm số của môn “Lập trình mạng”.

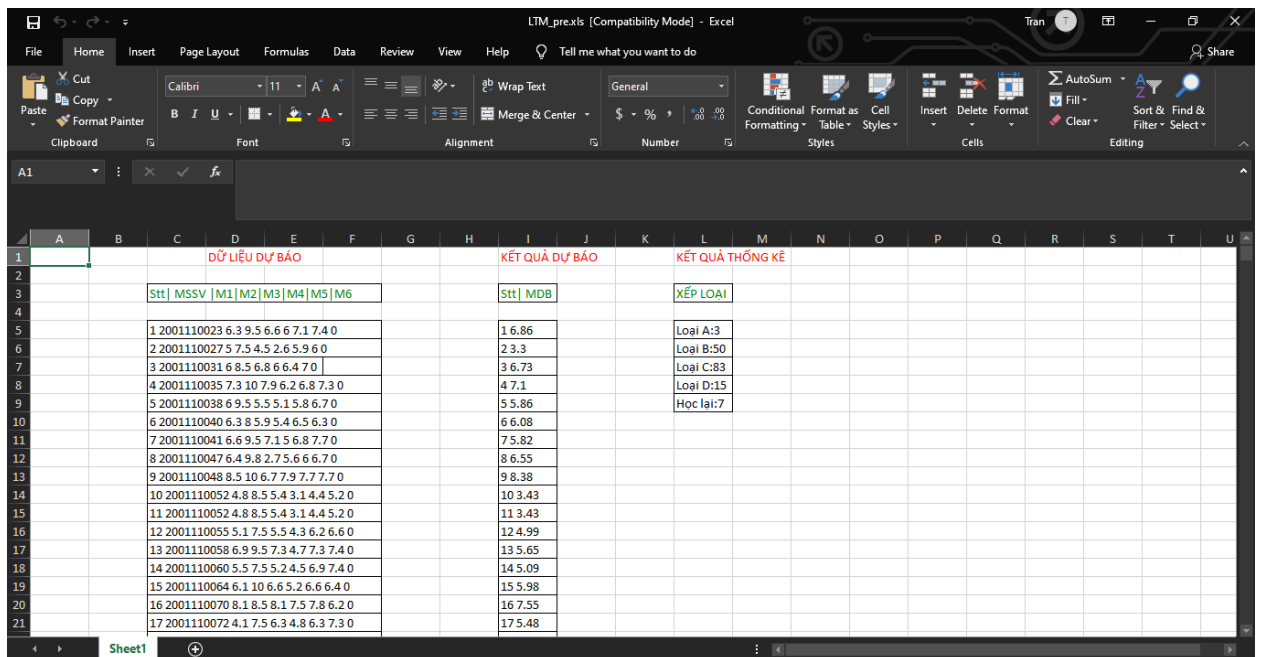
3.2.7 Dự báo điểm (dùng cho Giảng viên - Cố vấn học tập)

- Ở chức năng “Dự báo giảng viên”, ta tiến hành tải vào dữ liệu predict (dự đoán). Sau đó, ta tiến hành dự báo và thống kê điểm của các sinh viên.



Hình 3.8 Dự báo dùng cho giảng viên

Ngoài ra, với chức năng “Xuất báo cáo” giảng viên cũng có thể xuất kết quả trên ra Excel để tiện cho việc tư vấn môn học.



Hình 3.9 Xuất báo cáo ra file Excel

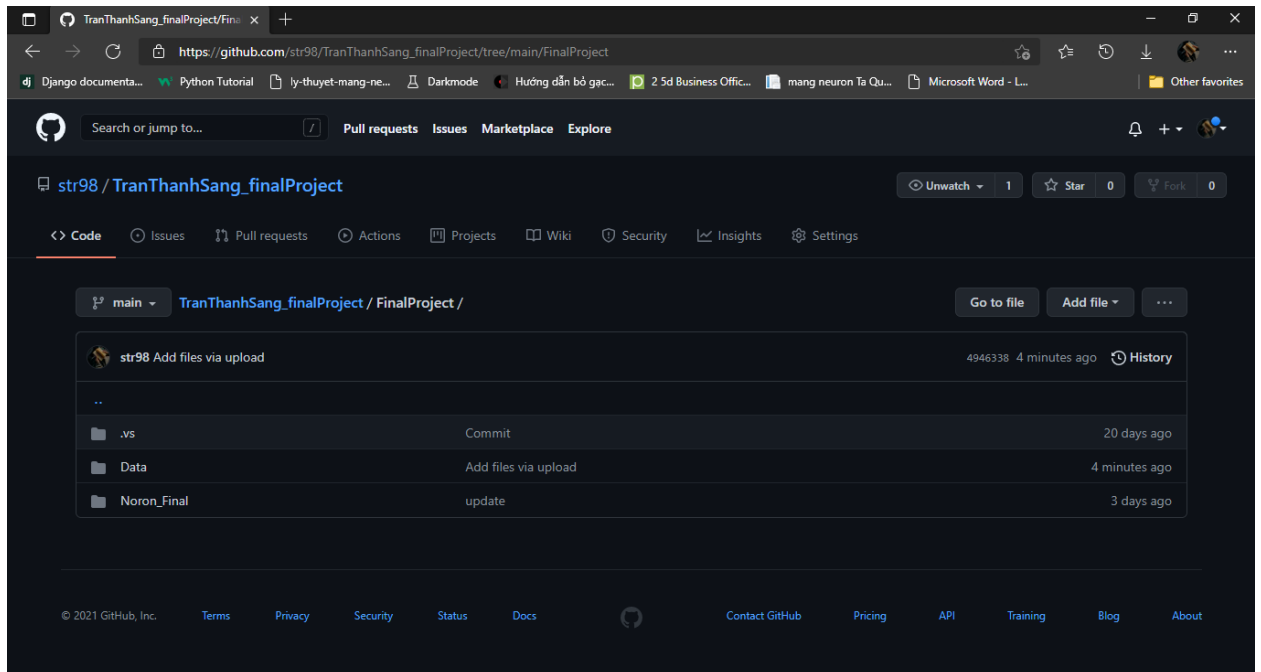
Từ kết quả thu được ở trên, giảng viên có thể tiến hành tư vấn cho sinh viên môn học có thể lựa chọn.

Qua đó, có thể thấy, chương trình đã chạy thành công. Và đây có thể là một giải pháp để giải quyết vấn đề dự báo điểm môn học tự chọn, từ đó đưa ra một tiêu chí để sinh viên có thể lựa chọn môn học cho phù hợp và hiệu quả hơn. Đồng thời, giúp giảng viên và nhà trường có thể nâng cao chất lượng dạy và học.

3.3 Hướng dẫn cài đặt

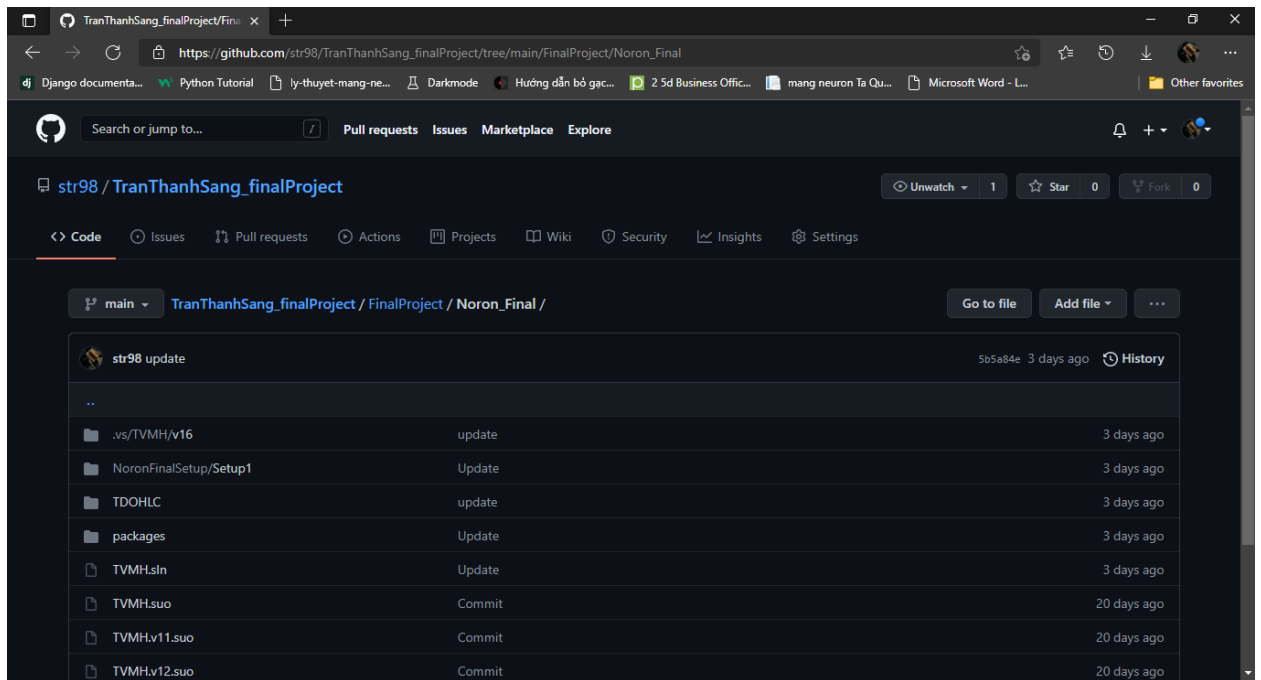
Bước 1: truy cập đường dẫn https://github.com/str98/TranThanhSang_finalProject.

Bước 2: vào thư mục FinalProject.



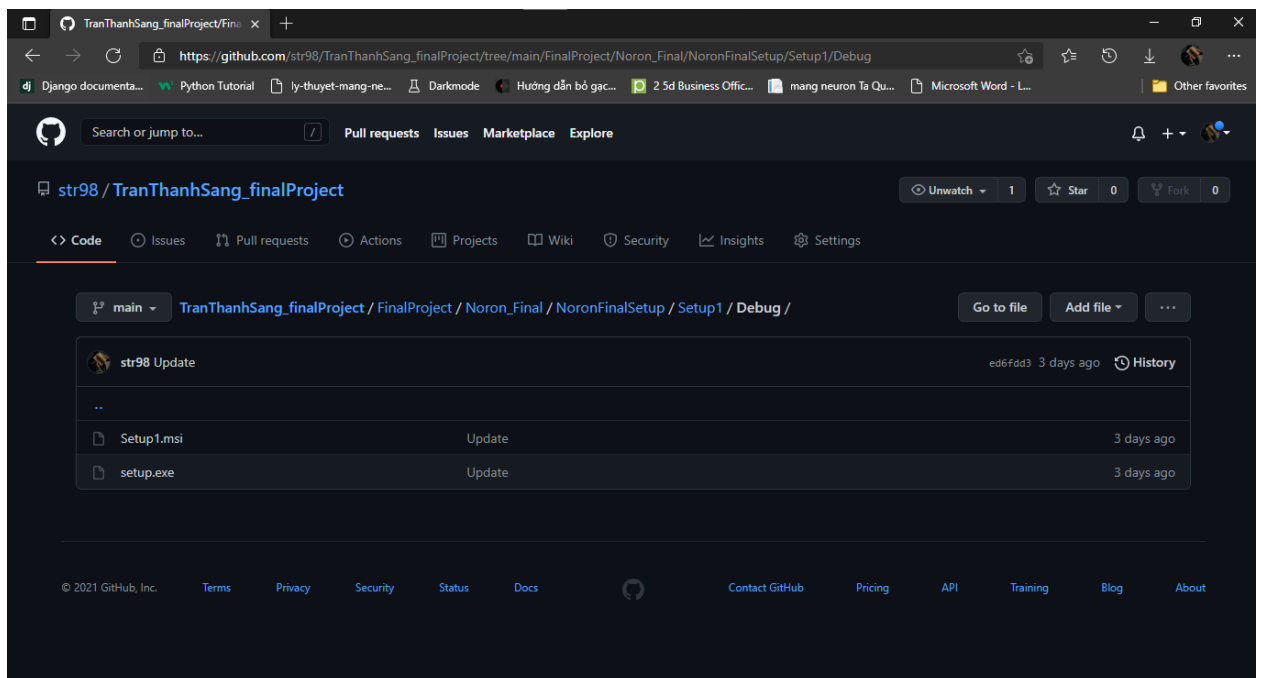
Hình 3.10 Truy cập vào folder FinalProject

Bước 3: vào thư mục Noron_Final, sau đó vào thư mục NoronFinalSetup/Setup1:



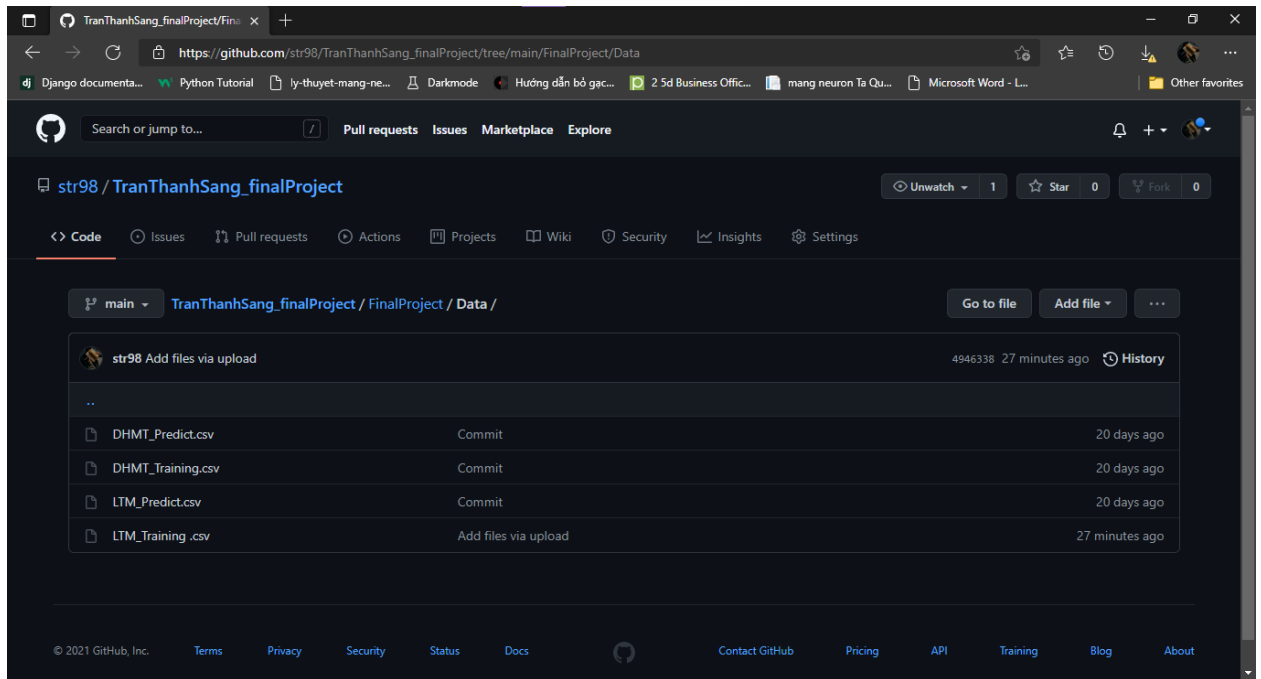
Hình 3.11 Truy cập folder Setup1

Bước 4: Vào thư mục Debug và tiến hành tải file cài đặt về.



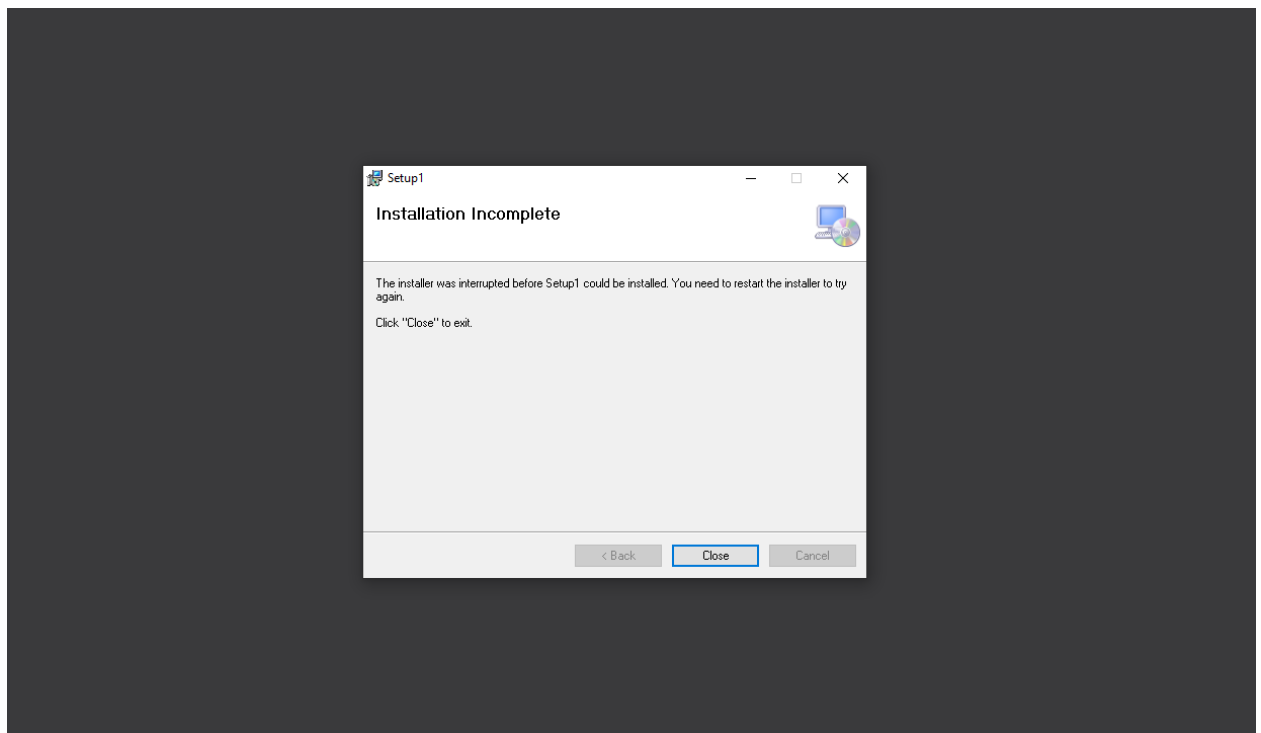
Hình 3.12 Truy cập vào folder Debug

Dữ liệu để chạy ta có thể tải về ở thư mục Data.



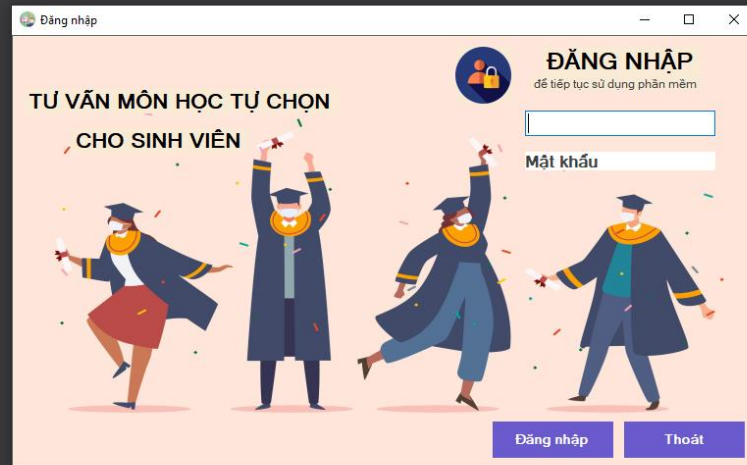
Hình 3.13 Dữ liệu điể

Quá trình cài đặt sẽ nhanh chóng hoàn thành.



Hình 3.14 Cài đặt thành công

Và bây giờ, ta đã có thể bắt đầu sử dụng chương trình.



Hình 3.15 Khởi chạy chương trình

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Kết quả đạt được

Khai phá dữ liệu là một lĩnh vực quan trọng và ngày càng được ứng dụng nhiều hơn trong cuộc sống hiện nay. Và trong đề tài này em cũng đã áp dụng kỹ thuật khai phá dữ liệu, thông qua việc xây dựng mạng nơ-ron nhân tạo để giải quyết vấn đề chọn môn tự chọn cho sinh viên các khối ngành đại học, cao đẳng mà thực nghiệm tại ngành Công nghệ thông tin ở trường Đại học Giao Thông Vận tải Phân hiệu tại Thành phố Hồ Chí Minh.

Đề tài đã tiến hành khai thác dữ liệu điểm của sinh viên bao gồm điểm của các môn học ở kỳ trước để tiến hành dự báo điểm có thể đạt được của môn tự chọn ở kì kế tiếp, từ đó tiến hành chọn môn học phù hợp, dựa trên mô hình mạng nơ-ron nhân tạo.

Hạn chế của đề tài

Nhìn chung, đề tài đã hoàn thành cơ bản được các mục tiêu đề ra. Dù vẫn còn tồn tại một số nhược điểm như: dữ liệu đầu vào chưa nhiều, vẫn mất một khoảng thời gian tương đối cho việc huấn luyện chương trình, giao diện chưa thân thiện với người dùng... do còn mang nhiều yếu tố của một bài nghiên cứu và khó tiếp cận với đại đa số sinh viên và giảng viên. Tuy nhiên, qua việc tiến hành chạy thử nghiệm và kết quả thu được, có thể thấy đây là một phương pháp khả quan trong việc hỗ trợ tư vấn môn học cho sinh viên và giảng viên.

Hướng phát triển

Với những kết quả thu được ở trên, em mong rằng đề tài sẽ được tiếp tục nghiên cứu và phát triển, không chỉ với ngành Công nghệ thông tin mà còn mở rộng sang các ngành học khác như Kinh tế, Giáo dục,... Bên cạnh đó, cũng cần cải thiện hay nghiên cứu các thuật toán và mô hình khác giúp cho việc khai phá dữ liệu được hiệu quả, mất ít thời gian và chi phí hơn. Đồng thời có thể phù hợp với những bài toán lớn hơn, ứng dụng trong các lĩnh vực khác của đời sống, chứ không riêng trong ngành giáo dục. Từ đó, đem lại những công cụ hỗ trợ đắc lực trong công việc, và phần nào nâng cao chất lượng cuộc sống của chính chúng ta.

TÀI LIỆU THAM KHẢO

Tiếng Việt

Giáo trình, tạp chí khoa học:

[1]. Bài giảng Khai phá dữ liệu, Đại học Hàng Hải Việt Nam, 2011.

[2] Tổng quan về phát hiện tri thức và khai phá dữ liệu. (Tạp chí Khoa học & Công nghệ. Số 5. 2011)

<https://www.hau.edu.vn/media/26/ufpdf26980.pdf>

[3] Ứng dụng mạng nơ-ron nhân tạo trong dự báo rỗng. Tác giả: Tạ Quốc Dũng - Lê Thế Hà - Phạm Duy Khang (Trích từ Tạp chí Dầu Khí số 7 năm 2019).

<http://www.pvn.vn/DataStore/Documents/2019/TapchiDK/so%207.%202019/mang%20neuron%20Ta%20Quoc%20Dung.pdf>

Liên kết:

[4]. http://uet.vnu.edu.vn/~thuyhq/Student_Thesis/K44_Do_Thi_Dieu_Ngoc_Thesis.pdf

[5] <http://kdienu.duytan.edu.vn/media/50176/ly-thuyet-mang-neural.pdf>

[6]. Giới thiệu chung hệ thần kinh.

<https://thcsnguyenhueq4.hcm.edu.vn/chuyen-muc-tin/sinh-8-tuan-24-tiet-45-bai-43-gioi-thieu-chung-he-than-kinh-cm86216-422830.aspx>

[7]. Giới thiệu mạng nơ-ron truyền thẳng nhiều lớp và giải thuật lan truyền ngược.

(Trích từ Trung tâm Quy hoạch và Điều tra Tài nguyên Nước Quốc gia, thuộc Bộ Tài nguyên và Môi trường.)

http://nawapi.gov.vn/index.php?option=com_content&view=article&id=3239%3Agii-thieu-mng-nron-truyn-thng-nhiu-lp-va-gii-thut-lan-truyn-ngc&catid=70%3Anhim-v-chuyen-mon-ang-thc-hin&Itemid=135&lang=vi

[8]. Machine Learning cơ bản.

<https://machinelearningcoban.com/2016/12/27/categories/>

[9]. Neuron.

<https://vi.wikipedia.org/wiki/Neuron>

[10]. Tổng quan về C#.

<https://eduking.edu.vn/c-la-gi-tong-quan-ve-c/>

[11]. Ứng dụng của Data Mining trong các lĩnh vực.

<https://insight.isb.edu.vn/ung-dung-cua-data-mining-trong-cac-linh-vuc/>

Tiếng Anh

[13]. Bipolar Sigmoid Function.

<https://adl1995.github.io/an-overview-of-activation-functions-used-in-neural-networks.html>

[14]. Default Sigmoid Alpha.

<http://www.aforgenet.com/framework/docs/html/70c19d16-2497-1382-511c-d540971c4624.html>

[15]. Feed Forward Network.

<https://programmersought.com/article/15035806233/>

[16]. Recurrent Network

<https://theappsolutions.com/blog/development/recurrent-neural-networks/>

[17]. Sigmoid Function.

https://www.researchgate.net/figure/Plot-of-the-sigmoid-function_fig7_325226633

[18]. Understand the impact of learning rate on neural network performance.

<https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/>

[19]. What is momentum in neural network.

<https://datascience.stackexchange.com/questions/84167/what-is-momentum-in-neural-network>