**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**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: NGUYÊN CỨU CÁC THUẬT TOÁN TRONG XỬ LÝ NGÔN NGỮ TỰ NHIÊN VÀ XÂY ĐỰNG WEBSITE HỖ TRỢ SINH VIÊN UTC2**

Giảng viên hướng dẫn: TRẦN PHONG NHÃ

Sinh viên thực hiện: NGUYỄN MINH MẪN

Lớp: CÔNG NGHỆ THÔNG TIN

Khoá:57

TP. Hồ Chí Minh, tháng 05 năm 2020

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**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: NGUYÊN CỨU CÁC THUẬT TOÁN TRONG XỬ LÝ NGÔN NGỮ TỰ NHIÊN VÀ XÂY ĐỰNG WEBSITE HỖ TRỢ SINH VIÊN UTC2**

Giảng viên hướng dẫn: TRẦN PHONG NHÃ

Sinh viên thực hiện: NGUYỄN MINH MẪN

Lớp: CÔNG NGHỆ THÔNG TIN

Khoá:57

TP. Hồ Chí Minh, tháng 05 năm 2020

TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI **CỘNG HÒA XÃ HỘI CHỦ NGHIÃ VIỆT NAM**

**PHÂN HIỆU TẠI THÀNH PHỐ HỒ CHÍ MINH** Độc lập – Tự do – Hạnh phúc

# NHIỆM VỤ THIẾT KẾ TỐT NGHIỆP

BỘ MÔN: **CÔNG NGHỆ THÔNG TIN**

-------\*\*\*-------

**Mã sinh viên:** 5751071024 **Họ tên SV:** Nguyễn Minh Mẫn

**Khóa:** 57 **Lớp:** Công Nghệ Thông Tin

1. **Tên đề tài.**

***“NGUYÊN CỨU CÁC THUẬT TOÁN TRONG XỬ LÝ NGÔN NGỮ TỰ NHIÊN VÀ XÂY ĐỰNG WEBSITE HỖ TRỢ SINH VIÊN UTC2”.***

1. **Mục đích, yêu cầu.**
   1. **Mục đích.**

* Xây dựng trang website hỗ trợ sinh viên UTC2.

- Xây dựng hệ thống diễn đàn sinh viên tự động trả lời (đề xuất các câu trả lời tương tự trên hệ thống).

- Xây dựng hệ thống dự đoán điểm và tùy chỉnh dự đoán để đặt ra mục tiêu học tập.

- Xây dựng hệ thống hỗ trợ trực tuyến ( live chat ) để sinh viên được tư vấn trực truyến từ các phòng ban trên hệ thống.

- Xây dựng hệ thống giúp giảng viên có thể giải đáp thắc mắc của sinh viên một cách nhanh chóng.

- Xây dựng hệ thống quản lý trên website hỗ trợ sinh viên UTC2.

* 1. **Yêu cầu.**
* Tìm hiểu về Machine Learning.
* Nghiên cứu về xử lý ngôn ngữ tự nhiên.
* Nghiên cứu thuật toán TF IDF trong xử lý ngôn ngữ tự nhiên.
* Nghiên cứu về thuật toán Linear Regression để dự đoán điểm.
* Tìm hiểu về Deep Learning.
* Nguyên cứu Neural Networks trong xử lý ngôn ngữ tự nhiên.
* Thu thập dữ liệu câu hỏi, câu trả lời trên diễn đàn nghe nói.
* Thu thập điểm của sinh viên UTC2.

1. **Nội dung và phạm vi đề tài.**
   1. **Nội dung đề tài.**

* Giới thiệu tổng quan về trí tuệ nhân tạo và Machine Learning, Deep Learning
* Giới thiệu về thuật toán TF IDF trong xử lý ngôn ngữ tự nhiên.
* Nghiên cứu, phân tích, đánh giá thuật toán.
* Sản phảm hoàn chỉnh website hỗ trợ sinh viên.
* Ứng dụng thuật toán vào bài toán thực tế, cụ thể là đề xuất câu hỏi tương tự trên diễn đàn.
  1. **Phạm vi đề tài.**

- Nghiên cứu thuật toán TF IDF vào phân loại câu hỏi thuộc phòng ban nào.

* Nghiên cứu thuật toán TF IDF trả lời câu hỏi tự động.
* Nguyên cứu Neural Networks để tạo ra 2 mô hình trên.
* Nguyên cứu thuật toán Linear Regression trong dự đoán điểm

1. **Công nghệ, công cụ và ngôn ngữ lập trình.**
   1. **Công nghệ:** Python, JavaScript, Odoo.
   2. **Công cụ:** Một số thư viện mã nguồn mở của Python: Scikit-learn, pandas, numpy, matplotlib, keras
   3. **Ngôn ngữ lập trình:** Python, JavaScript, XML
2. **Các kết quả chính dự kiến sẽ đạt được và ứng dụng**

* Hoàn chỉnh cuốn báo cáo đề tài.
* Khái quát được tổng quan về Machine Learning.
* Khái quát được tổng quan về Deep Learning và Neural Networks.
* Nắm được thuật toán TF IDF, Linear Regression và có thể áp dụng được thuật toán cho bất kỳ bài toán nào liên quan.
* Nắm được các bước trong xử lý ngôn ngữ tự nhiên.
* Nắm được các ưu, nhược điểm của thuật toán, phương pháp tối ưu cho thuật toán.
* Sử dụng Odoo để xây dựng được website và quản lý các chức năng của nó.

1. **Giáo viên và cán bộ hướng dẫn**

Họ tên: TRẦN PHONG NHÃ

Đơn vị công tác: Bộ môn Công Nghệ Thông Tin – Trường Đại học Giao thông Vận tải phân hiệu tại TP HCM

Điện thoại: 0906761014 Email: tpnha@utc2.edu.vn

|  |  |
| --- | --- |
| **Ngày 20 tháng 10 năm 2018**  **BM Công Nghệ Thông Tin** | **Đã giao nhiệm vụ TKTN**  **Giáo viên hướng dẫn** |
|  | **Trần Phong Nhã** |

Đã nhận nhiệm vụ TKTN

Sinh viên: Nguyễn Minh Mẫn Ký tên:

Điện thoại: 0349183111

Email: [minhmanit98@gmail.com](mailto:minhmanit98@gmail.com)

**LỜI CẢM ƠN**

Lời nói đầu tiên, em xin kính gửi lời cảm ơn chân thành nhất tới Quý thầy cô trong Bộ môn Công Nghệ Thông Tin, cũng như Ban Giám Hiệu Trường Đại học Giao thông Vận tải phân hiệu tại Thành phố Hồ Chí Minh, đã cho phép em thực hiện đề tài tốt nghiệp: ***“ NGUYÊN CỨU CÁC THUẬT TOÁN TRONG XỬ LÝ NGÔN NGỮ TỰ NHIÊN VÀ XÂY ĐỰNG WEBSITE HỖ TRỢ SINH VIÊN UTC2”***.

Để hoàn thành nhiệm vụ được giao này, ngoài sự nỗ lực học hỏi không ngừng của bản thân còn có sự hướng dẫn tận tình của thầy giáo **Trần Phong Nhã**, người đã hướng dẫn cho em những hướng đi, truyền đạt cho em những kiến thức, kỹ năng để em có thể hoàn thành đề tài tốt nghiệp này.

Mặc dù đã cố gắng hết sức để hoàn thành đề tài, nhưng chắc chắn rằng sẽ khó tránh khỏi những thiếu sót. Em rất mong nhận được những sự đánh giá, góp ý của Quý thầy cô để em có thể rút ra cho mình những bài học, kinh nghiệm quý báu.

Sau cùng, em cũng không biết nói gì hơn ngoài kính chúc Quý thầy cô trong Bộ môn Công Nghệ Thông Tin và đặc biệt là thầy giáo **Trần Phong Nhã** thật dồi dào sức khỏe và ngày càng gặt hái được nhiều thành công hơn nữa trong cuộc sống cũng như trong sự nghiệp giảng dạy của mình.

Em xin chân thành cảm ơn !

***TP. Hồ Chí Minh, ngày 15 tháng 01 năm 2019***

**Sinh viên thực hiện**

**Nguyễn Minh Mẫn**

**NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN**

|  |
| --- |
| ***TP. Hồ Chí Minh, ngày 15 tháng 01 năm 2019***  **Giáo viên hướng dẫn**  **Trần Phong Nhã** |

# 

# MỤC LỤC

[DANH MỤC HÌNH ẢNH, BẢNG BIỂU iii](#_Toc45199707)

[THÔNG TIN KẾT QUẢ NGHIÊN CỨU CỦA ĐỀ TÀI iv](#_Toc45199708)

[CHƯƠNG 1: TỔNG QUAN 1](#_Toc45199709)

[1.1. Đặt vấn đề 1](#_Toc45199710)

[1.2. Tình hình nghiên cứu 2](#_Toc45199711)

[1.3 Quá trình thực hiện nghiên cứu 2](#_Toc45199712)

[CHƯƠNG 2: CƠ SỞ LÝ THUYẾT 5](#_Toc45199713)

[2.1 Công nghệ sử dụng 5](#_Toc45199714)

[2.1.1 Odoo 5](#_Toc45199715)

[2.1.1.1 Khái niệm [1] 5](#_Toc45199716)

[2.1.1.2 Ưu điểm của Odoo 5](#_Toc45199717)

[2.1.2 Cơ sở dữ dữ liệu postgressql 5](#_Toc45199718)

[2.2 Ngôn ngữ lập trình 6](#_Toc45199719)

[2.2.1 Python [5] 6](#_Toc45199720)

[2.2.2 Javascript [7] [8] 6](#_Toc45199721)

[2.2.3 XML [9] 6](#_Toc45199722)

[2.3. Thuật toán áp dụng 6](#_Toc45199723)

[2.3.1 Bài toán phân loại câu hỏi và đề xuất câu hỏi tương tự. 6](#_Toc45199724)

[2.3.1.1 Đặt vấn đề 6](#_Toc45199725)

[2.3.1.2 Giải pháp 7](#_Toc45199726)

[2.3.1.3 Quá trình xử lý khi đưa câu hỏi lên hệ thống. 9](#_Toc45199727)

[2.3.1.4 Giai đoạn phân tag cho câu hỏi 10](#_Toc45199728)

[2.3.2 Bài toán dự đoán điểm. 10](#_Toc45199729)

[2.3.2.1 Đặt vấn đề. 10](#_Toc45199730)

[2.3.2.2 Giải pháp 11](#_Toc45199731)

[CHƯƠNG 3: PHÂN TÍCH VÀ THIẾT KẾ HỆ THỐNG 1](#_Toc45199732)

[3.1 Biểu đồ usecase của hệ thống 1](#_Toc45199734)

[3.2 Biểu đồ hoạt động 2](#_Toc45199737)

[3.2.1 Tạo tài khoản 2](#_Toc45199738)

[3.2.2 Đăng nhập 3](#_Toc45199739)

[3.2.3 Xem thông tin trên web 3](#_Toc45199740)

[3.2.4 Thao tác với bài viết 4](#_Toc45199741)

[3.2.5 Tìm kiếm bài viết câu hỏi 5](#_Toc45199742)

[3.2.6 Dự đoán điểm 6](#_Toc45199743)

[3.2.7 Hỗ trợ chat trực tuyến 7](#_Toc45199744)

[3.2.8 Quản lý người dùng 8](#_Toc45199745)

[3.2.9 Quản lý hỗ trợ trực tuyến 8](#_Toc45199746)

[3.2.10 Quản lý diễn đàn 9](#_Toc45199747)

[3.5.11 Báo cáo thống kê. 11](#_Toc45199748)

[CHƯƠNG 4: XÂY DỰNG CHƯƠNG TRÌNH 12](#_Toc45199749)

[4.1 Tạo tài khoản và đăng nhập 12](#_Toc45199750)

[4.2 Giao diện trang chủ website 13](#_Toc45199751)

[4.3 Tương tác bài viết 13](#_Toc45199752)

[4.4 Chat trực tuyến. 15](#_Toc45199753)

[4.4 Dự đoán điểm 16](#_Toc45199754)

[4.5 Nhóm chức năng quản lý 17](#_Toc45199755)

[4.5.1 Quản lý người dùng 18](#_Toc45199756)

[4.5.2 Quản lý chat trực tuyến. 19](#_Toc45199757)

[4.5.3 Quản lý dự đoán điểm 19](#_Toc45199758)

[CHƯƠNG 5: KẾT LUẬN 21](#_Toc45199759)

[5.1. Kết quả đạt được 21](#_Toc45199760)

[5.2. Nhược điểm 21](#_Toc45199761)

[5.3. Hướng phát triển 21](#_Toc45199762)

[TÀI LIỆU THAM KHẢO 21](#_Toc45199763)

**DANH MỤC HÌNH ẢNH, BẢNG BIỂU**

[Bảng 1.1 Các phòng ban/trung tâm 3](#_Toc45199764)

[Hình 2.1: Quy trình hóa phân loại văn bản. 7](#_Toc45199765)

[Hình 2.2: Kết quả phân loại thực tế. 8](#_Toc45199766)

[Hình 2.3: Quá trình tạo mô hình véc- tơ và ma trận câu hỏi. 8](#_Toc45199767)

[Hình 2.4: Qúa trình chọn câu hỏi-câu trả lời tương tự trên diễn đàn 9](#_Toc45199768)

[Hình 2.5: Quá trình phân tag cho câu hỏi 10](#_Toc45199769)

[Hình 2.6: Công thức dự đoán điểm 11](#_Toc45199770)

[Hình 3.1: Sơ đồ usecase tổng quan của hệ thống. 2](#_Toc45199771)

[Hình 3.2: Sơ đồ hoạt động tạo tài khoản. 2](#_Toc45199772)

[Hình 3.3: Hình Sơ đồ hoạt động đăng nhập 3](#_Toc45199773)

[Hình 3.4: Sơ đồ hoạt động xem thông tin trên trang web 3](#_Toc45199774)

[Hình 3.5: Sơ đồ hoạt động thao tác với bài viết 4](#_Toc45199775)

[Hình 3.6: Sơ đồ tìm kiếm bài viết câu hỏi. 5](#_Toc45199776)

[Hình 3.7: Sơ đồ hoạt động dự đoán điểm. 6](#_Toc45199777)

[Hình 3.8 Sơ đồ hoạt động chat trực tuyến 7](#_Toc45199778)

[Hình 3.9: Sơ đồ hoạt động quản lý người dùng. 8](#_Toc45199779)

[Hình 3.10: Sơ đồ hoạt động hỗ trợ trực tuyến 8](#_Toc45199780)

[Hình 3.11: Sơ đồ hoạt động quản lý diễn đàn. 9](#_Toc45199781)

[Hình 3.12: Sơ đồ hoạt động quản lý dự đoán điểm. 10](#_Toc45199782)

[Hình 3.13: Sơ đồ hoạt động quản lý báo cáo thống kê 11](#_Toc45199783)

[Hình 4.1: Giao diện tạo tài khoản đăng nhập 12](#_Toc45199784)

[Hình 4.2: Giao diện đăng nhập 12](#_Toc45199785)

[Hình 4.3: Giao diện trang chủ của website 13](#_Toc45199786)

[Hình 4.4: Giao diện đặt câu hỏi trên diễn đàn. 13](#_Toc45199787)

[Hình 4.5: Giao diện trả lời câu hỏi 14](#_Toc45199788)

[Hình 4.6: Các chức năng thao tác với câu hỏi 14](#_Toc45199789)

[Hình 4.7: Giao diện chức năng tìm kiếm câu hỏi. 15](#_Toc45199790)

[Hình 4.8: Giao diện chức năng chat trực tuyến 15](#_Toc45199791)

[Hình 4.9: Giao diện người dùng sử dụng chức năng chat trực tuyến. 16](#_Toc45199792)

[Hình 4.10: Giao diện chức năng dự đoán điểm 16](#_Toc45199793)

[Hình 4.11: Giao diện bảng điểm sau khi người dùng chọn chức năng dự đoán điểm. 17](#_Toc45199794)

[Hình 4.12: Giao diện nhóm chức năng quản lý 17](#_Toc45199795)

[Hình 4.13: Giao diện quản lý người dùng 18](#_Toc45199796)

[Hình 4.14: Giao diện hiện thị người dùng truy cập trang web 18](#_Toc45199797)

[Hình 4.15: Giao diện quản lý chat trực **tuyến**. 19](#_Toc45199798)

[Hình 4.16: Giao diện theo dõi lịch sử chat. 19](#_Toc45199799)

[Hình 4.17: Giao diện quản lý điểm 20](#_Toc45199800)

# TỔNG QUAN

## 1. Đặt vấn đề

Trên thế giới cũng như ở Việt Nam, công nghệ thông tin đã trở thành một ngành công nghệ mũi nhọn. Bất kỳ một ngành nghề nào, lĩnh vực nào trong xã hội cũng cần đến sự góp sức của công nghệ thông tin để giải quyết một số vấn đề nhất định như xây dựng hệ thống quản lý danh sách nhân viên, tiền lương của một công ty, quản lý tiền điện, tiền nước hay hệ thống quản lý sách trong một thư viện, ….

Ở nước ta hiện nay, việc ứng dụng công nghệ thông tin tại các cơ quan, trường học, xí nghiệp, tổ chức đang rất phổ biến và dần trở nên cần thiết. Bởi ngành nghề nào cũng đòi hỏi con người phải xử lý khối lượng công việc khổng lồ, phức tạp với dữ liệu lớn, những kiến thức và đào tạo chuyên sâu. Cụ thể như trong quá trình học tập tại trường nhiều bạn sinh viên, đặc biệt là các bạn sinh viên năm nhất mới bước chân vào giảng đường chắc chắn sẽ có nhiều thắc mắc về quá trình học tập cũng như các quy định rèn luyện thi đua, nhưng đa phần các bạn thường không biết nên hỏi ai và không thể xác định các thông tin chính xác về các vấn đề của mình.

Nhằm khắc phục được vấn đề này, nhiều trường đại học đã thực hiện một số biện pháp để giải quyết như: xây dựng các fanpage, các kênh diễn đàn để sinh viên có thể đặt các câu hỏi cũng như chia sẻ các khó khăn của mình.Tuy nhiên, cách làm này vẫn tồn tại nhiều phát sinh, các câu hỏi của các bạn sinh viên không thể ngay lập tức đến với những người có thể giải đáp mà phải trải qua quy trình chung gian, là admin của các fanpage lọc câu hỏi và tiếp tục gửi đi tới các bộ phận liên quan có thể giải quyết. Điều này dẫn đến việc các bạn sinh viên phải chờ đợi lâu hơn, quy trình giúp đỡ các bạn sinh viên trở nên rườm rà và còn thiếu tính linh động tương tác.

Xuất phát từ tình hình thực tế đó, nhóm đã nghĩa ra một ý tưởng đó là xây dựng một hệ thống cố vấn học tập online cho Phân hiệu trường đại học Giao thông Vận Tải tại tp. Hồ Chí Minh để công tác hỗ trợ sinh viên diễn ra tiện lợi linh hoạt hơn.

Trường Đại học Giao thông vận tải Phân hiệu tại Tp. Hồ Chí Minh, hiện nay có gần 7000 học viên và sinh viên các hệ, mỗi sinh viên đều được nhà trường cung cấp cho một tài khoản gmail cá nhân. Trường hiện tại có 11 phòng ban trung tâm với các vai trò khác nhau. Mỗi phòng ban phụ trách các vấn đề riêng tuy nhiên tất cả đều có một mục đích chung đó là làm cho trường đại học GTVT phát triển tốt hơn, các bạn sinh viên có môi trường học tập, rèn luyện hoàn chỉnh hơn.

Các ban ngành, phong ban trung tâm, giáo viên cố vấn và các giảng viên luôn tích cực cố gắng giảng dạy, hỗ trợ và giúp đỡ các bạn sinh viên. Tuy nhiên, không thể nào như cấp 3 giáo viên luôn quan sát nhắc nhở học sinh. Môi trường đại học các bạn sinh viên cần có tính chủ động và tự giác, khi gặp các vấn đề cần chủ động hỏi để được giúp đỡ, hơn nữa sự thiếu hụt về nguồn nhân lực của nhà trường, công tác hỗ trợ sinh viên còn nhiều khó khăn

Nhóm đề xuất nghiên cứu xây dựng một hệ thống cố vấn học tập online cho Phân hiệu trường đại học Giao thông Vận tải tại tp.Hồ Chí Minh dựa trên phần mềm odoo -giải pháp quản lí doanh nghiệp dẽ dàng để tạo ra một ứng dụng dễ sử dụng, chi phí thấp,có nhiều chức năng hữu dụng và dễ dàng cập nhập phục vụ với nhà trường và nhu cầu sử dụng của các bạn sinh viên.

## Tình hình nghiên cứu

Hiện nay tại Phân hiệu Trường Đại học Giao Thông Vận Tải tại TP.HCM chưa có ứng dụng phục vụ hỗ trợ cố vấn học tập online trên các trang web để tạo môi trường hỏi đáp, giúp đỡ sinh viên.

Trên thị trường phần mềm Odoo có vai trò hỗ trợ đắc lực của các phần mềm quản lý doanh nghiệp với nhiều chức năng khác nhau như: bán hàng, chăm sóc khách hàng, quản trị dự án, quản lý tài chính và nguồn nhân lực, quản trị sản xuất…phần mềm này được lập trình sẵn 18 ngôn ngữ giúp cho công ty ở nhiều nước có thể dễ dàng tùy chỉnh sử dụng. Bên cạch đó còn mang đến nhiều lợi ích cho doanh nhiệp như: dễ dàng sử dụng, tiết kiệm chi phí, dễ dàng cài đặt và có nhiều chức năng hữu dụng, công nghệ được cập nhật liên tục.

## 3. Quá trình thực hiện nghiên cứu

Để hiểu rõ hơn về công tác giải đáp hỗ trợ sinh viên. Nhóm đã lựa chọn PangPage Diễn đàn nghe sinh viên nói, và tìm hiểu vai trò của các phòng ban/trung tâm của trường đại học Giao Thông vận tải phân hiệu tại thành phố Hồ Chí Minh để khảo sát, tìm hiểu và phát triển ứng dụng

Hiện nay tại Phân hiệu Trường Đại học GTVT tại TP. HCM có các ban ngành phòng bạn sau:

Bảng 1: Các phòng ban/trung tâm

|  |  |
| --- | --- |
| **STT** | **Tên các ban ngành/phòng ban** |
| **1** | Phòng tổ chức hành chính |
| **2** | Phòng đào tạo |
| **3** | Phòng công tác sinh viên |
| **4** | Phòng khảo thí và đảm bảo chất lượng đào tạo |
| **5** | Phòng tài chính kế toán |
| **6** | Phòng thiết bị quản trị |
| **7** | Phòng khoa học công nghệ và đối ngoại |
| **8** | Ban quản lý ký túc xá |
| **9** | Ban thanh tra |
| **10** | Trung tâm thông tin thư viện |
| **11** | Trung tâm đào tạo thực hành |

Việc trả lời các thắc mắc cho các bạn sinh viên cần được diễn ra chính xác và nhanh chóng, tiết kiệm được các quy trình nhân lực trong việc xử lí. Đối tượng sử dụng là các bạn sinh viên, giáo viên; các chức năng chính cần có của hệ thống gồm có:

* **Đối với sinh viên:**
* Đăng nhập và đăng xuất
* Đăng bài trong diễn đàn
* Trả lời câu hỏi trong diễn đàn
* Thao tác với bài viết trên diễn đàn (yêu thích, cắm cờ, báo cáo quản trị, ...)
* Tạo trang cá nhân
* Cập nhật thông tin cá nhân
* Tích điểm nâng cấp tài khoản (cấp bậc tài khoản gồm Đồng Bạc Vàng)
* Quản lý trang cá nhân
* Quản lý bài đăng
* Nhận tin tức, thông báo từ nhà trường
* **Đối với giảng viên:**
* Giải đáp thắc mắc
* Trả lời chat online
* Cập nhật thông tin cá nhân
* Duyệt bài đăng
* Thao tác với bài đăng (yêu thích, xóa bài đăng vi phạm, xóa bình luận vi phạm, …)
* **Đối với người quản lý:**
* Quản lý giảng viên
* Quản lý sinh viên
* Quản lý diễn đàn
* Quản lý trang cá nhân của thành viên
* Quản lý các tin tức, thông báo
* Thống kê, báo cáo

# CHƯƠNG 1: CƠ SỞ LÝ THUYẾT

* 1. **Machine Learning**

**2.2 Odoo**

* + - 1. Khái niệm
* ***Odoo*** là một phần mềm quản trị doanh nghiệp mã nguồn mở sử dụng ngôn ngữ lập trình Python 7 (còn có thêm Javascript và XML) [1]. Bao gồm các module bán hàng, chăm sóc khách hàng, quản trị dự án, quản trị kho, quản trị sản xuất, quản lý tài chính và quản trị nguồn nhân lực, …
* ***Odoo/OpenERP*** được tích hợp công nghệ điện toán đám mây, cực kỳ phù hợp với các doanh nghiệp vừa và nhỏ trong mọi ngành nghề, lĩnh vực [2].

#### 2.1.1.2 Ưu điểm của Odoo

* Các ưu điểm của Odoo [3]:
* Là một mã nguồn mở nên nhiều công ty tin học nhỏ có thể tham gia cung cấp triển khai và phát triển bổ sung các module phụ trợ.
* Odoo dễ cài, vận hành thử trên nhiều nền tảng OS
* Công nghệ được cập nhật liên tục
* Kết nối thông minh
* Dễ dàng tùy chỉnh
* Không cần trả phí bản quyền
* Các tiện ích của Odoo:
* Thương mại điện tử.
* Bán hàng.
* Bất động sản.
* Phân tích dữ liệu
* Nắm bắt được những ưu điểm trên, nhóm chúng mình đã sử dụng phần mềm Odoo để xây dựng nên trang web….
  + 1. Cơ sở dữ dữ liệu postgressql
* ***PostgreSQL*** là một hệ thống quản trị cơ sở dữ liệu quan hệ-đối tượng *(object-relational database management system)*, hệ thống cơ sở dữ liệu mã nguồn mở tiên tiến nhất hiện nay [4].
* ***PostgreSQL*** được thiết kế để chạy trên các nền tảng tương tự UNIX. Tuy nhiên, sau đó được điều chỉnh linh động để có thể chạy được trên nhiều nền tảng khác nhau như Mac OS X, Solaris và Windows [4].
* ***PostgreSQL*** là một phần mềm mã nguồn mở miễn phí. Vì vậy, bạn sẽ được tự do sử dụng, sửa đổi và phân phối PostgreSQL dưới mọi hình thức.
  1. **Ngôn ngữ lập trình**
     1. Python

Python là một ngôn ngữ lập trình bậc cao cho các mục đích lập trình đa năng. Python là ngôn ngữ có hình thức rất sáng sủa, cấu trúc rõ ràng [5].

Ban đầu, Python được phát triển để chạy trên nền Unix. Nhưng sau đó, Python dần mở rộng sang nhiều hệ điều hành từ MS-DOS đến Mac OS, OS/2, Windows, Linux và các hệ điều hành khác thuộc họ Unix [6].

* + 1. Javascript
* JavaScript là ngôn ngữ kịch bản cho phép tạo ra trang web động - cập nhật nội dung theo ngữ cảnh, điều khiển đa phương tiện, hoạt cảnh các hình ảnh [7].
* JavaScript là ngôn ngữ lập trình phổ biến nhất trên thế giới trong suốt 20 năm qua [8].
  + 1. XML
* XML là ngôn ngữ đánh dấu với mục đích chung do W3C đề nghị, để tạo ra các ngôn ngữ đánh dấu khác. Đây là một tập con đơn giản của SGML (một hệ thống tổ chức và gắn thẻ tài liệu.), có khả năng mô tả nhiều loại dữ liệu khác nhau [9].
* Mục đích chính của XML là đơn giản hóa việc chia sẻ dữ liệu giữa các hệ thống khác nhau, đặc biệt là các hệ thống được kết nối với Internet.

## 2.3. Thuật toán áp dụng

### **2.3.1 Bài toán phân loại câu hỏi và đề xuất câu hỏi tương tự.**

#### 2.3.1.1 Đặt vấn đề

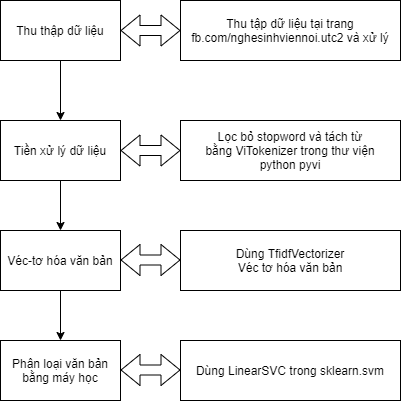
Câu hỏi và câu trả lời trong hệ thống hỏi đáp sinh viên mang những đặc điểm riêng, đó là ở dưới dạng văn bản tự do, không theo một loại câu hỏi nhất định nào, cũng không theo một chủ đề nhất định nào cả. Do đó, một phần hết sức quan trọng trong hệ thống này là phân tích câu hỏi như thế nào lấy được thông tin nhiều nhất khi mà câu hỏi như thế nào để lấy được thông tin nhiều nhất mà câu hỏi không hề có một cấu trúc nhất định nào cả. Hầu hết các hệ thống hỏi-đáp truyền thống đều chỉ trả lời cho các câu hỏi thuộc về một loại câu hỏi nào đó. Do đó, phương pháp mà chúng tôi chọn thử nghiệm cho hệ thống diễn đàn sinh viên là phương pháp dựa trên từ khóa, trích từ khóa và đánh trọng số cho các từ khóa trong văn bản đề tìm kiếm câu hỏi. Ngoài ra, nhằm cải thiện hiệu quả hệ thống, giảm không gian tìm kiếm, trước khi tìm kiếm, các cặp hỏi-đáp được phân thành các cụm gồm các câu hỏi tương tự nhau.

Chúng tôi tiến hành thử nghiệm các phương pháp đề xuất, cải thiện hiệu quả hệ thống trên mỗi bước phù hợp với dữ liệu của diễn đàn sinh viên.

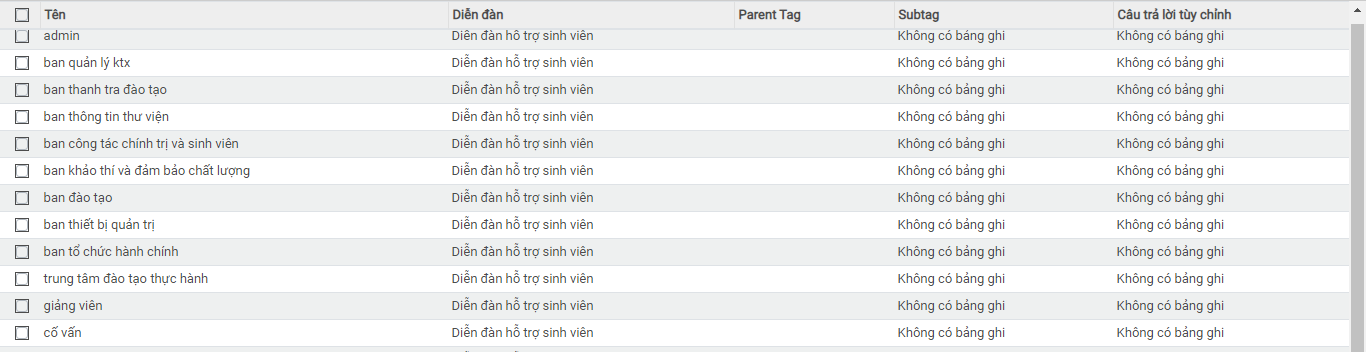
#### 2.3.1.2 Giải pháp

Kiến thúc hệ thống diễn đàn sinh viên gồm 2 phần chính yếu nhất là phân loại tag (phân loại câu hỏi), tìm kiếm bài viết/ câu hỏi tương tự trên diễn đàn để người dùng tham gia sau khi đặt câu hỏi.

Qúa trình mô hình hóa phân loại văn bản được chia làm 4 giai đoạn:

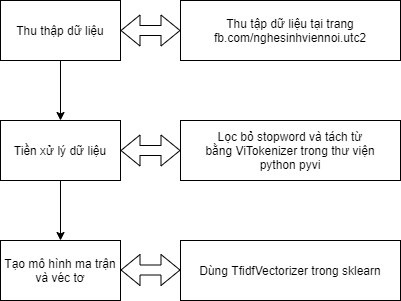


##### *Hình 2.1: Quy trình hóa phân loại văn bản.*



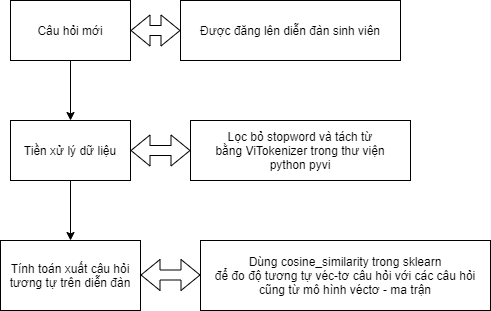
##### *Hình 2.2: Kết quả phân loại thực tế.*

Qúa trình tạo mô hình véc-tơ và ma trận câu hỏi:



##### *Hình 2.3: Quá trình tạo mô hình véc- tơ và ma trận câu hỏi.*

#### 2.3.1.3 Quá trình xử lý khi đưa câu hỏi lên hệ thống.



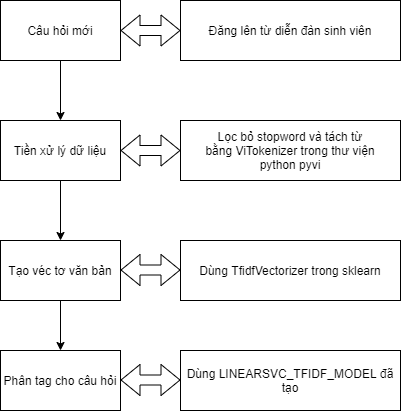
##### *Hình 2.4: Qúa trình chọn câu hỏi-câu trả lời tương tự trên diễn đàn*

Câu hỏi mới được đưa vào hệ thống sẽ được tách câu, tách từ, loại bỏ các hư từ và loại bỏ các cụm từ xuất hiện nhiều nhưng không có ý nghĩa, để còn lại các từ cần thiết (từ khoá) cho việc phân loại và tìm câu tương tự.

Các từ khóa thu được trong giai trước sẽ được sử dụng để xây dụng vector đặc trưng, sau đó xác định các cụm chứa các câu hỏi tương tự nhất. Vector bài viết/câu hỏi sẽ được so khớp với tất cả các vector câ u hỏi trong các cụm đó theo độ đo tương tự cosin. Các giá trị tương tự này được xếp hạng, và hệ thống chọn nQ câu hỏi có giá trị tương tự cao nhất đứa vào giai đoạn tiếp theo.

nQ câu hỏi này được chuyển sang giai đoạn so khớp câu trả để tìm nQ câu trả lời tương ứng. Trong giai đoạn này, vector truy vấn sẽ được so khớp với vector của các câu trả tìm được. Một chiến lược xếp hạng được sử dụng để câu trả lời tốt nhất.

#### 2.3.1.4 Giai đoạn phân tag cho câu hỏi



##### *Hình 2.5: Quá trình phân tag cho câu hỏi*

Các từ khóa thu được sau giai đoạn tiền xử lý sẽ được chuyển vào giao đoạn tiếp để tạo véc tơ văn bản sau đó dùng LINEARSVC\_TFIDF\_MODEL đã tạo để dự đoán, tính toán tag cho câu hỏi.

### 2.3.2 Bài toán dự đoán điểm.

#### 2.3.2.1 Đặt vấn đề.

Sinh viên khi học tập tại trường Đại học GTVT phân hiệu tại TP.HCM luôn mong muốn có thành tích cao và ra trường đúng hạn cần phải đặt mục tiêu tính toán điểm, chiến thuật dự đoán điểm hợp lý. Nhưng việc tính toán điểm mục tiêu đặt ra với toàn bộ chương trình sẽ rất khó khăn vì số điểm mục tiêu không giống với thực tế nên việc phải tính toán từ đầu tốn nhiều thời gian và công sức.

Mặc dù có nhiều tài liệu về việc dự đoán kết quả học tập của học sinh, sinh viên nhưng các nghiên cứu đối với các chương trình đại học còn ít vì một số khác biệt so với các chương trình khác. Thứ nhất là đối với chương trình đại học, mỗi sinh viên có nền tảng khác nhau, ngành nghề khác nhau, được chọn những môn học khác nhau dẫn đến khó thống kê hết toàn bộ. Thứ hai là một số môn học không có nhiều thông tin để có thể đưa ra dự đoán chính xác. Ngoài ra còn vì một số lý do nhỏ khác.

Đối với trong nước, theo tìm hiểu của nhóm chúng em thì việc ứng dụng Machine Learning vào việc dự đoán kết quả học tập cho sinh viên còn chưa được chú trọng nhiều. Phần lớn việc ứng dụng Machine Learning tập trung nhiều vào việc phân tích, xử lý hình ảnh, nhận dạng khuôn mặt và khai phá dữ liệu. Đây cũng điều kiện thúc đẩy nhóm em tìm hiểu đề tài này

#### 2.3.2.2 Giải pháp

- Qua tìm hiểu thuật toán hồi quy tuyến tính thích hợp với bài toán dự đoán điểm sinh viên dự vào các môn liên quan, ví dụ: Dựa vào điểm của môn tin học đại cương và môn lập trình nâng cao sẽ dự đoán ra điểm môn lập trình hướng đối tượng

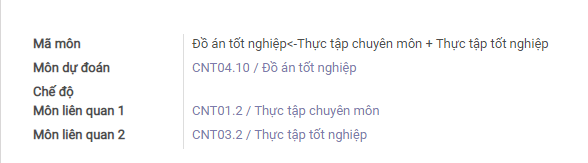
- Xây dựng hệ thống giúp sinh viên dự đoán điểm, quản lý mục tiêu của mình. Hệ thống sẽ cho ra 3 loại điểm tích lũy:

+ Điểm tích lũy hiện tại: Hệ thống tính điểm tích lũy hiện tại của sinh viên

+ Điểm tích lũy mục tiêu: Sinh viên có thể cập nhật điểm mục tiêu của từng môn trong chương trình học từ đó hệ thống sẽ cho ra điểm tích lũy mục tiêu.

+ Điểm tích lũy dự đoán: Sinh viên chọn công thức dự đoán trên những môn để hệ thống dự đoán.

- Công thức dự đoán được giảng viên tạo ra bằng các chọn 1 hoặc 2 môn liên quan và một môn đích. Về thời gian sau có thể tích độ tin cậy của công thức bằng nhưng kết quả điểm thực của sinh viên.



*Hình 2.6: Công thức dự đoán điểm*

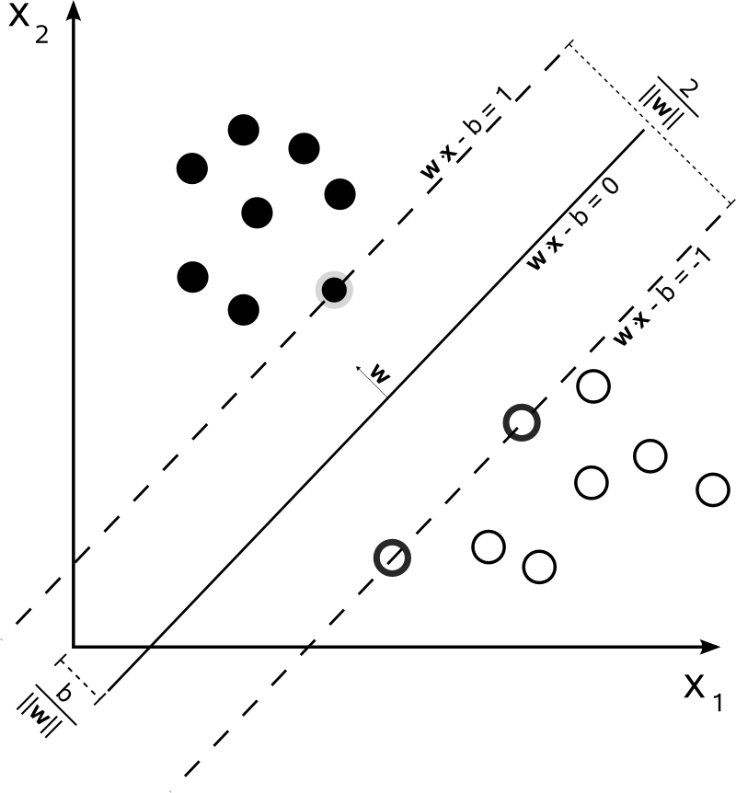
**CHƯƠNG 3: PHÂN TÍCH VÀ THIẾT KẾ HỆ THỐNG**

**3.1 Các phương pháp phân loại văn bản bằng máy học:**

**3.1.1 Phương pháp SVM (Suport Vec-tơ Machine):**

Máy học véc-tơ hỗ trợ (SVM) là một giải thuật học máy được xây dựng dựa trên lý thuyết học thống kê do Vapnik và Chervonenkis xây dựng [3]. Có nhiều tiềm năng phát triển về mặt lý thuyết cũng như ứng dụng trong thực tiễn.

Ý tưởng của phương pháp này là cho trước một tập huấn luyện được biểu diễn trong không gian véc-tơ, trong đó mỗi văn bản được xem như là một điểm trong không gian này. Phương pháp này tìm ra một mặt siêu phẳng h quyết định tốt nhất có thể chia các điểm trên không gian này thành hai lớp riêng biệt tương ứng, tạm gọi là lớp + (dương) và lớp – (âm). Như vậy, bộ phân loại SVM là một mặt siêu phẳng tách các mẫu thuộc lớp dương ra khỏi cách mẫu thuộc lớp âm với độ chênh lệch lớn nhất. Độ chênh lệch này hay còn gọi là khoảng cách biên được xác định bằng khoảng cách giữa mẫu dương và mẫu âm gần mặt siêu phẳng nhất (hình). Khoảng cách này càng lớn các mẫu thuộc hai lớp càng được phân chia rõ ràng, nghĩa là sẽ đạt được kết quả phân loại tốt. Mục tiêu của thuật toán SVM là tìm được khoảng cách biên lớn nhất để tạo được kết quả phân loại tốt.

Siêu phẳng với lề cực đại cho một SVM phân tách dữ liệu thuộc hai lớp

Phương trình mặt siêu phẳng chứa véc-tơ x trong không gian đối tượng như sau:

Trong đó w là véc-tơ trọng số, b là độ dịch. Hướng và khoảng cách từ gốc tọa độ đến mặt siêu phẳng thay đổi khi thay đổi w và b. Bộ phân loại SVM được định nghĩa như sau:

Trong đó:

Gọi yi mang giá trị +1 hoặc -1. Nếu yi = +1 thì x thuộc về lớp dương, ngược lại yi = -1 thì x thuộc về lớp âm.

Hai mặt siêu phẳng tách các mẫu thành hai phần được mô tả bởi các phương trình:

Bằng hình học ta có thể tính khoảng cách giữa hai mặt siêu phẳng này là . Để khoảng cách biên là lớn nhất, ta phải tìm giá trị nhỏ nhất của . Đồng thời ngăn chặn các điểm dữ liệu rơi vào vùng bên trong biên, chúng ta thêm ràng buộc sau:

Có thể gộp lại thành:

Khi đó để tìm mặt siêu phẳng h ta sẽ giải bài toán tìm Min||w|| với w và b thỏa điều kiện sau:

**3.1.2 Phương pháp NAIVE BAYES:**

Naive Bayes đã được nghiên cứu rộng rãi từ những năm 1950. Được dùng lần đầu tiên trong lĩnh vực phân loại vào đầu những năm 1960. Sau đó nó trở nên phổ biến và được sử dụng rộng rãi trong lĩnh vực này cho đến ngày nay. Ý tưởng cơ bản của cách tiếp cận này là sử dụng xác suất có điều kiện giữa từ hoặc cụm từ và chủ đề để dự đoán xác suất chủ đề của một tập tin cần phân loại. Điểm quan trọng của phương pháp này chính là ở chỗ giả định rằng sự xuất hiện của tất cả các từ trong tập tin đều độc lập với nhau. Ví dụ một loại trái cây có thể được cho là quả táo nếu nó đỏ, tròn và đường kính là 10cm. Giải thuật Naive Bayes sẽ cho rằng mỗi tính năng này đều đóng góp một cách độc lập để xác suất trái cây này là quả táo bất kể sự hiện diện hay vắng mặt của các tính năng khác.

Thuật toán Naive Bayes dựa trên định lý Bayes được phát biểu như sau:

Trong đó:

là xác xuất X thuộc lớp Y.

là xác suất một phần tử thuộc lớp Y, và phần tử đó có đặc điểm X.

Y) xác suất xảy ra lớp Y, mức độ thường xuyên lớp Y xuất hiện trong tập dữ liệu

xác suất xảy ra lớp X

Ví dụ: Giả sử ta có hai lớp Y1 = Nam, Y2 = nữ. Và một người không biết giới tính là Phương, X = Phương. Việc xác định Phương là Nam hay Nữ tương đương với việc so sánh xác suất P(Nam/Phương) và P(Nữ/Phương). Theo thuật toán Naive Bayes ta có công thức như sau:

Trong đó : : xác suất Phương là nam

: xác suất những người phái nam có tên Phương.

P(Nam) : xác suất phái nam trong tập dữ liệu.

P(Phương) : xác suất tên Phương trong tập dữ liệu.

Tương tự ta có :

Giả sử ta có bẳng dữ liệu tên và giới tính như sau :

|  |  |
| --- | --- |
| **Tên** | **Giới Tính** |
| Phương | Nam |
| Nga | Nữ |
| Hồng | Nữ |
| Nam | Nam |
| Phương | Nữ |
| Phương | Nữ |
| Tiến | Nam |
| Giang | Nữ |
| Tùng | Nam |
| Đài | Nữ |

Như vậy Phương là Nữ có xác suất cao hơn nên Phương được phân vào lớp nữ khi phân loại.

Áp dụng vào bài toán phân loại văn bản :

Tập dữ liệu đã được véc-tơ hóa

Tập các lớp

Các thuộc tính độc lập đôi một với nhau.

Khi đó ta có :

Với xác suất thuộc tính thứ k mang giá trị khi đã biết X thuộc phân lớp i.

Ưu điểm của giải thuật này đó là việc giả định rằng sự xuất hiện của tất cả các từ trong tập tin đều độc lập với nhau làm cho việc tính toán Naive Bayes hiệu quả và nhanh chóng vì không sử dụng việc kết hợp các từ để đưa ra phán đoán chủ đề. Một lợi thế nữa của Naive Bayes là nó chỉ đòi hỏi một lượng nhỏ dữ liệu huấn luyện để ước lượng các tham số cần thiết để phân loại. Bởi vì các biến được giả định độc lập với nhau, nên chỉ có các phương sai của các biến cho mỗi lớp cần phải được xác định và không phải là toàn bộ ma trận hiệp phương sai. Tuy nhiên nhược điểm của phương pháp này cũng chính là giả định đó, vì nó rất khó xảy ra trong thực tế.

**3.1.3 Phương pháp cây quyết định (classification and regression trees)**

Học máy cây quyết định là sử dụng mô hình cây quyết định để dự đoán kết quả về giá trị mục tiêu của một sự vật, hiện tượng. Nghĩa là ánh xạ từ các quan sát của một sự vật, hiện tượng đến các kết luận về giá trị mục tiêu của sự vật, hiện tượng. Cây quyết định là một trong những cách tiếp cận được sử dụng rộng rãi trong thống kê, khai phá dữ liệu và học máy. Nó là một trong những kĩ thuật thành công nhất trong việc học máy phân loại. Trong mô hình cây phân loại, các nút lá là các phân lớp, các nhánh là các liên từ, tính năng dẫn đến các lớp đó.

Ví dụ: cây quyết định phân lớp mức lương

Age ?

Salary?

Salary

bad

good

bad

good

Có nhiều thuật toán xây dựng cây quyết định như CLS, ID3, C4.5,CART… nhưng nhìn chung quá trình xây dựng cây quyết định đều được chia thành ba giai đoạn cơ bản:

Xây dựng cây: thực hiện chia một cách đệ quy tập mẫu dữ liệu huấn luyện cho đến khi các mẫu ở mỗi nút lá thuộc cùng một lớp.

Cắt tỉa cây: nhằm tối ưu hóa cây. Công việc chính là trộn một cây con vào trong một nút lá.

Đánh giá cây: đánh giá độ chính xác của cây kết quả.Tiêu chí đánh giá là phần trăm số mẫu phân lớp đúng trên tổng số mẫu đưa vào.

Việc chọn thuật toán nào để có hiệu quả phân lớp cao tuy thuộc vào rất nhiều yếu tố, trong đó cấu trúc dữ liệu ảnh hưởng rất lớn đến kết quả của các thuật toán. Chẳng hạn như thuật toán ID3 và CART cho hiệu quả phân lớp rất cao đối với các trường dữ liệu số (quantitative value) trong khi đó các thuật toán như J48, C4.5 có hiệu quả hơn đối với các dữ liệu Qualititive value (ordinal, binary, nominal). Trong phần này chúng ta chỉ đi vào tìm hiểu giải thuật CART (Classification and Regression Tree).

Giải thuật CART [11] chấp nhận sự tham lam (nonbacktracking) cách tiếp cận cây quyết định được xây dựng từ trên xuống một cách đệ quy, bắt đầu với một bộ dữ liệu huấn luyện tập và các nhãn lớp của họ. Hầu hết giải thuật cây quyết định đều theo cách tiếp cận từ trên xuống. Tập dữ liệu huấn luyện được phân vùng một cách đệ quy thành tập hợp con nhỏ hơn trong lúc cây được xây dựng.

Đối với các phương pháp phân loại khác, cây quyết định tương đối dễ hiểu, đòi hỏi mức tiền xử lý dữ liệu đơn giản. Tuy nhiên hiệu quả phân lớp của cây quyết định, phụ thuộc rất nhiều vào huấn luyện (training) data.

**3.1.4 K- NEAREST NEIGHBOR (KNN):**

K- Nearest Neighbor hay còn gọi là K láng giềng gần nhất là phương pháp phân loại dựa trên hướng tiếp cận thống kê. Nó là một trong những phương pháp tốt nhất từ thời kì đầu của phân loại văn bản.

Ý tưởng của phương pháp này là khi cần phân loại một văn bản mới, thuật toán sẽ tính toán khoảng cách của tất cả các văn bản trong tập huấn luyện đến văn bản này để tìm ra tập K láng giềng gần nhất. Sau đó dùng khoảng cách này đánh trọng số cho tất cả các chủ đề. Khi đó trọng số của một chủ đề chính là tổng khoảng cách tất cả các văn bản nằm trong tập K láng giềng có cùng chủ đề. Những chủ đề mà không xuất hiện bất kì văn bản nào trong tập K láng giềng thì có trọng số bằng 0. Sau đó các chủ đề sẽ được sắp xếp theo giá trị trọng số giảm ần, chủ đề có trọng số cao sẽ được chọn làm chủ đề của văn bản phân loại.

Để tính khoảng cách có thể áp dụng các công thức như độ đo Cosin(4.8), độ đo Euclid (4.9), hay công thức Manhattan:

Trong đó:

là véc-tơ đặc trưng đại điện cho văn bản thứ j trong tập huấn luyện.

) là véc-tơ đặc trưng của văn bản mới cần phân loại.

**3.1.5 LINEAR LEAST SQUARE FIT (LLSF):**

Linear Least Square Fit là phương pháp phân loại dựa trên cách tiếp cận ánh xạ. LLSF sử dụng phương pháp hồi quy để học từ tập huấn luyện và các chủ đề có sẵn. Tập huấn luyện được biểu diễn dưới dạng một cặp véc-tơ đầu vào và đầu ra như sau:

Véc-tơ đầu vào một văn bản bao gồm các từ và trọng số

Véc-tơ đầu ra gồm các chủ đề cùng với trọng số nhị phân của văn bản ứng với véc-tơ đầu vào.

Giải phương trình các cặp véc-tơ đầu vào/ đầu ra, ta sẽ được ma trận đồng hiện của hệ số hồi quy của từ và chủ đề (matrix of wordcategory regression coefficients). Phương pháp này sử dụng công thức:

Trong đó:

A, B là ma trận đại diện tập dữ liệu huấn luyện (các cột trong ma trận tương ứng là các véc-tơ đầu vào và đầu ra).

FLS là ma trận kết quả chỉ ra một ánh xạ từ một văn bản bất kỳ vào véc-tơ của chủ đề đã gán trọng số.

Nhờ vào việc sắp xếp trọng số của các chủ đề, ta được một danh sách chủ đề có thể gán cho văn bản cần phân loại. Nhờ đặt ngưỡng lên trọng số của các chủ đề mà ta tìm được chủ đề thích hợp cho văn bản đầu vào. Hệ thống tự động học các ngưỡng tối ưu cho từng chủ đề, giống với KNN. Mặc dù LLSF và KNN khác nhau về mặt thống kê, nhưng ta vẫn tìm thấy điểm chung ở hoạt động của hai phương pháp là việc học ngưỡng tối ưu

**3.2 Các thông số đánh giá giải thuật:**

**3.3 Phương pháp xây dựng trả lời câu hỏi tự động:**



## 3.1 Biểu đồ usecase của hệ thống

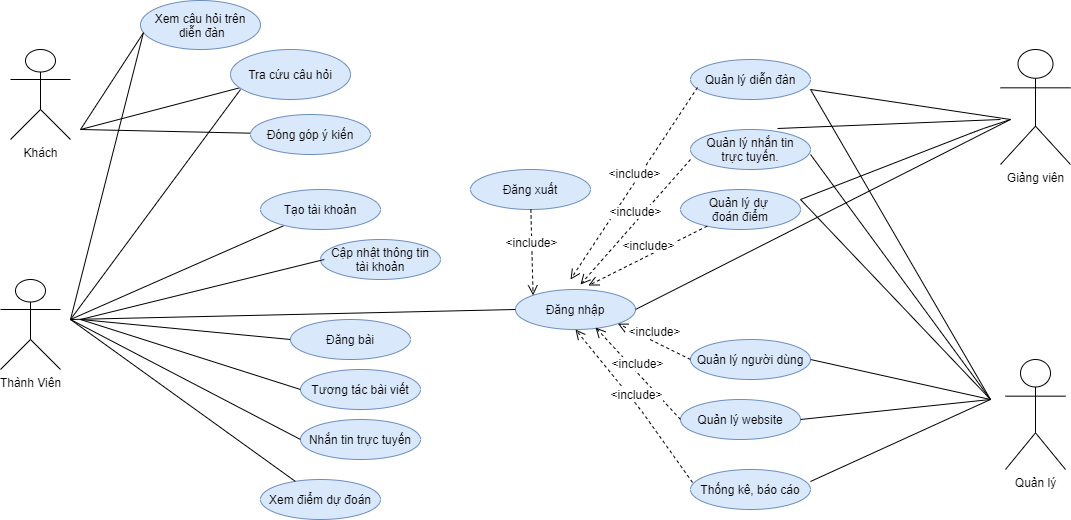
Biểu đồ usecase biểu diễn sơ đồ chức năng của hệ thống. Từ các yêu cầu của hệ thống, biểu đồ usecase chỉ ra hệ thống cần thực hiện những điều gì để đáp ứng nhu cầu của người sử dụng hệ thống.

Người quản trị trang web (admin) thông qua đăng nhập để thực hiện các chức năng quản lý, thống kê báo cáo và các chức năng khác trên website.

Giảng viên đăng nhập vào hệ thống có thể quản lý diễn đàn, quản lý nhắn tin trực tuyến, dự đoán điểm.

Khách là những người không có gmail sinh viên của trường truy cập vào trang web có thể thực hiện các chức năng như: Xem câu hỏi trên diễn đàn, tra cứu câu hỏi, đóng góp ý kiến

Đối với thành viên là sinh viên có tài khoản gmail do trường đại học Giao Thông Vận Tải Phân Hiệu Tại Tp. Hồ Chí Minh cung cấp sau khi đăng nhập có thể thực hiện đăng các câu hỏi lên diễn đàn, tương tác với bài viết, thực hiện chức năng chat trực tuyến, xem điểm dự đoán. Cũng như có thể cập nhật thông tin tài khoản.

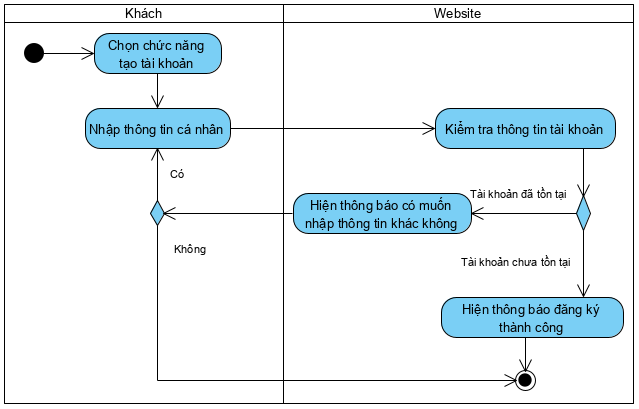


##### *Hình 3.1: Sơ đồ usecase tổng quan của hệ thống.*



## 3.2 Biểu đồ hoạt động

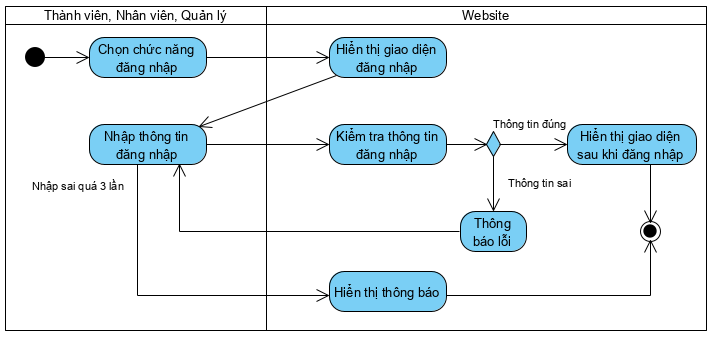
### 3.2.1 Tạo tài khoản



Thành Viên

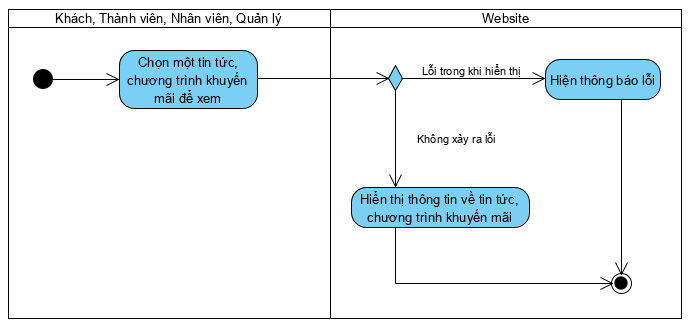
##### *Hình 3.2: Sơ đồ hoạt động tạo tài khoản.*

### 3.2.2 Đăng nhập



##### *Hình 3.3: Hình Sơ đồ hoạt động đăng nhập*

### 3.2.3 Xem thông tin trên web



Hiển thị thông tin các bài đăng và chức năng của diễn đân

##### *Hình 3.4: Sơ đồ hoạt động xem thông tin trên trang web*

### 3.2.4 Thao tác với bài viết



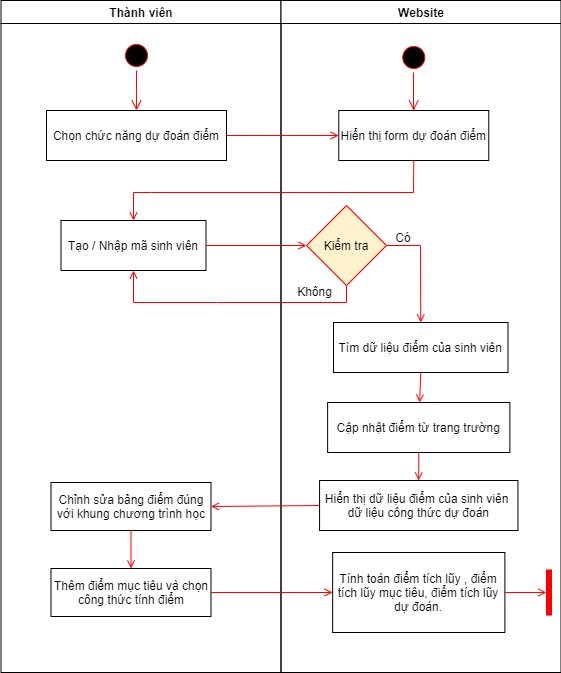
##### *Hình 3.5: Sơ đồ hoạt động thao tác với bài viết*

### 3.2.5 Tìm kiếm bài viết câu hỏi



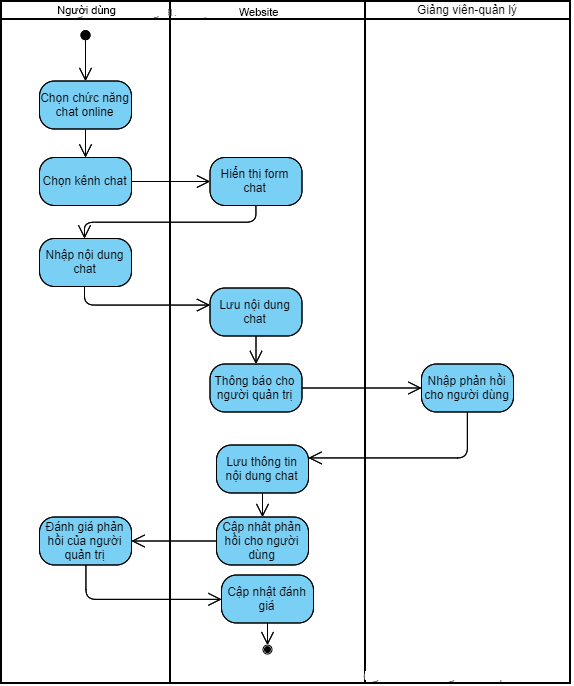
##### *Hình 3.6: Sơ đồ tìm kiếm bài viết câu hỏi.*

### 3.2.6 Dự đoán điểm



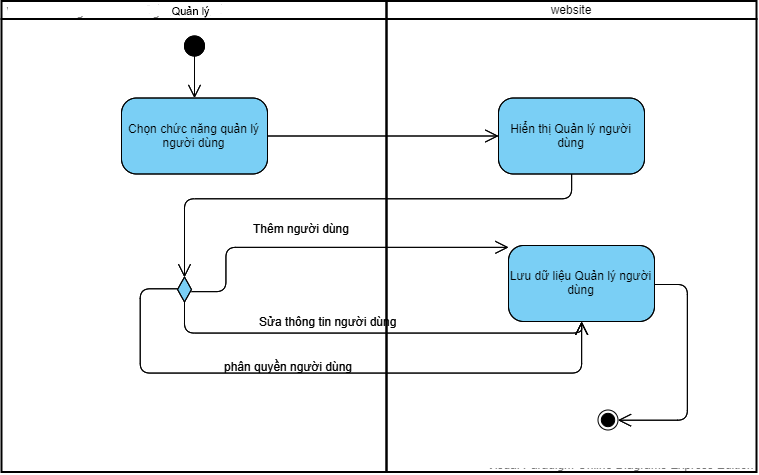
##### *Hình 3.7: Sơ đồ hoạt động dự đoán điểm.*

### 3.2.7 Hỗ trợ chat trực tuyến



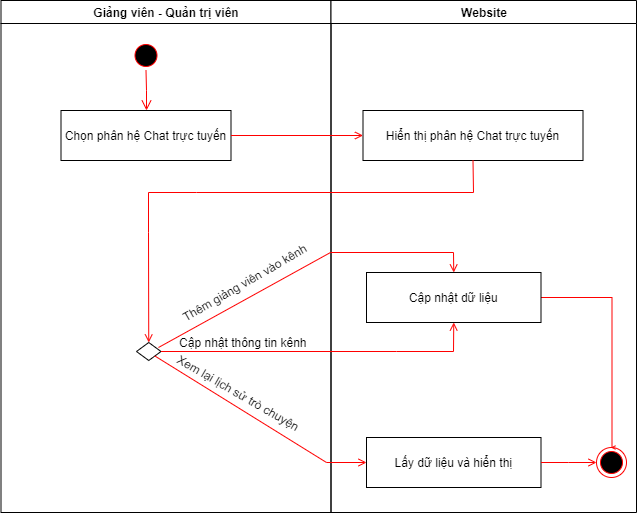
##### *Hình 3.8 Sơ đồ hoạt động chat trực tuyến*

### 3.2.8 Quản lý người dùng



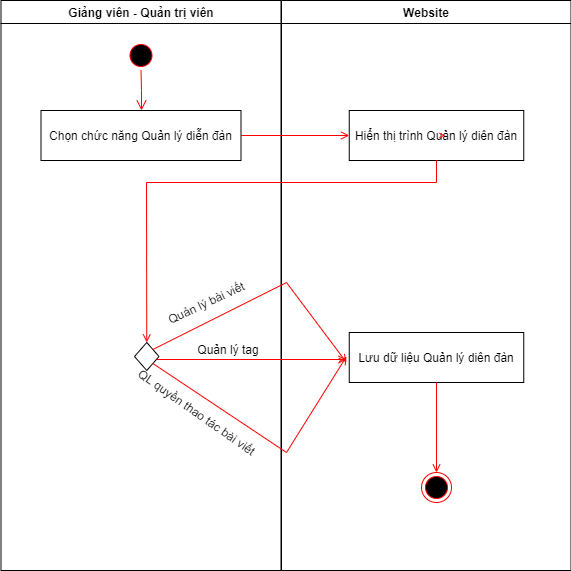
##### *Hình 3.9: Sơ đồ hoạt động quản lý người dùng.*

### 3.2.9 Quản lý hỗ trợ trực tuyến



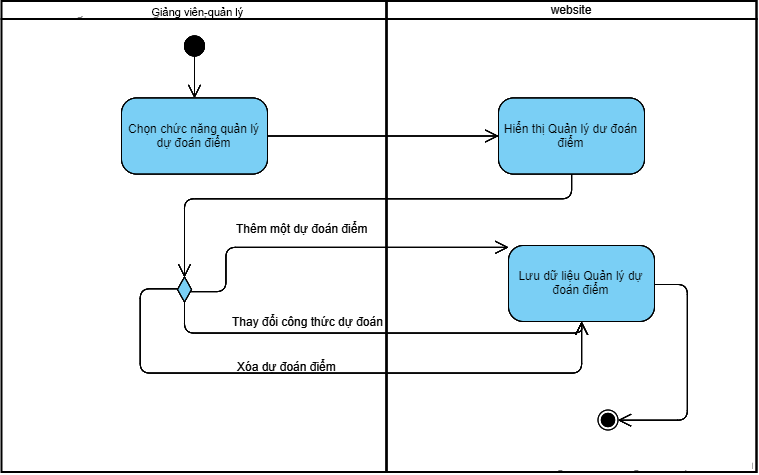
##### *Hình 3.10: Sơ đồ hoạt động hỗ trợ trực tuyến*

### 3.2.10 Quản lý diễn đàn



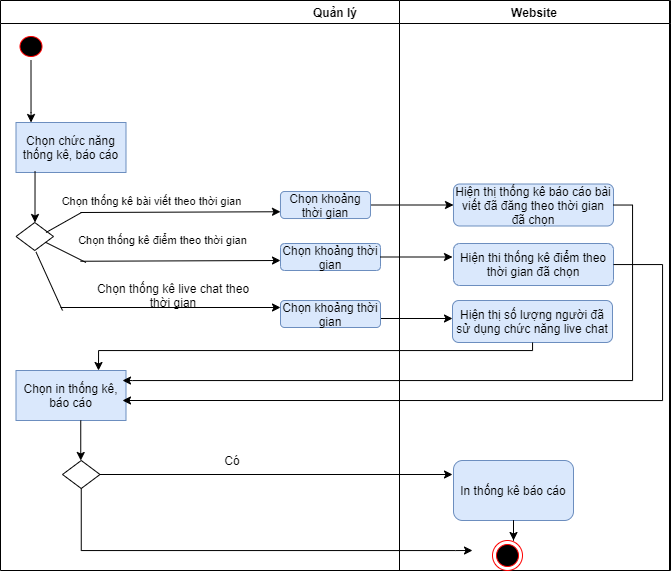
##### *Hình 3.11: Sơ đồ hoạt động quản lý diễn đàn.*

3.2.11 Quản lý dự đoán điểm



##### *Hình 3.12: Sơ đồ hoạt động quản lý dự đoán điểm.*

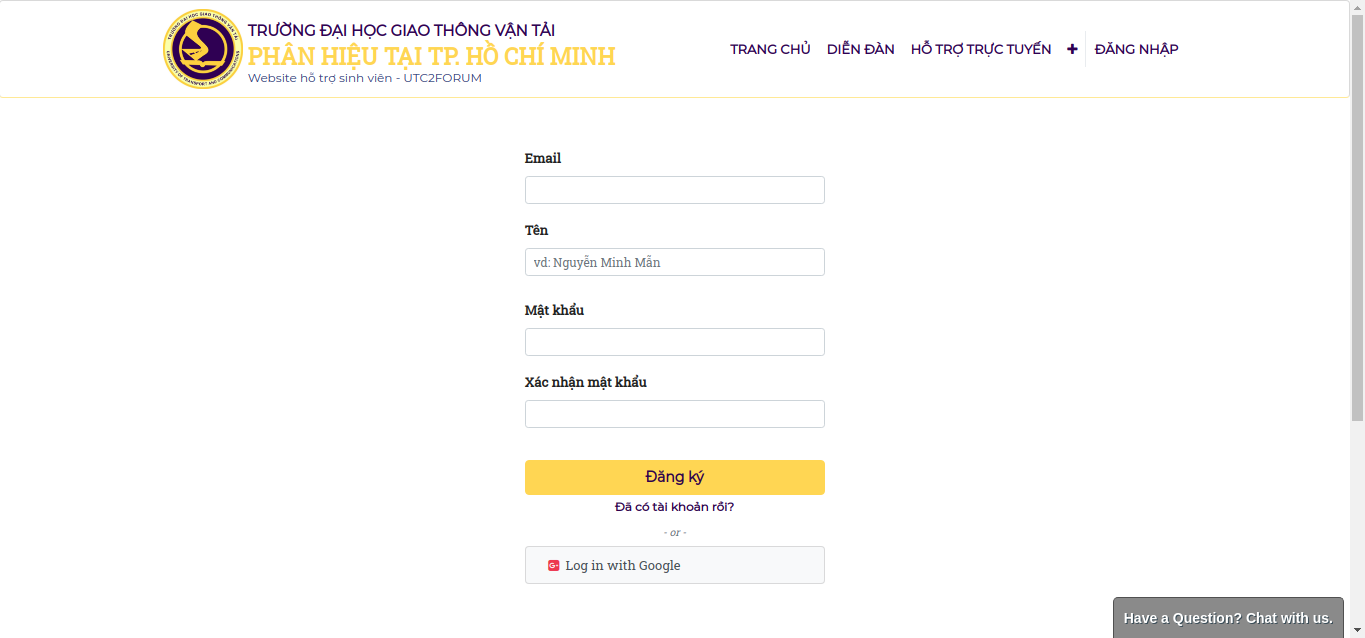
### 3.5.11 Báo cáo thống kê.



##### *Hình 3.13: Sơ đồ hoạt động quản lý báo cáo thống kê*

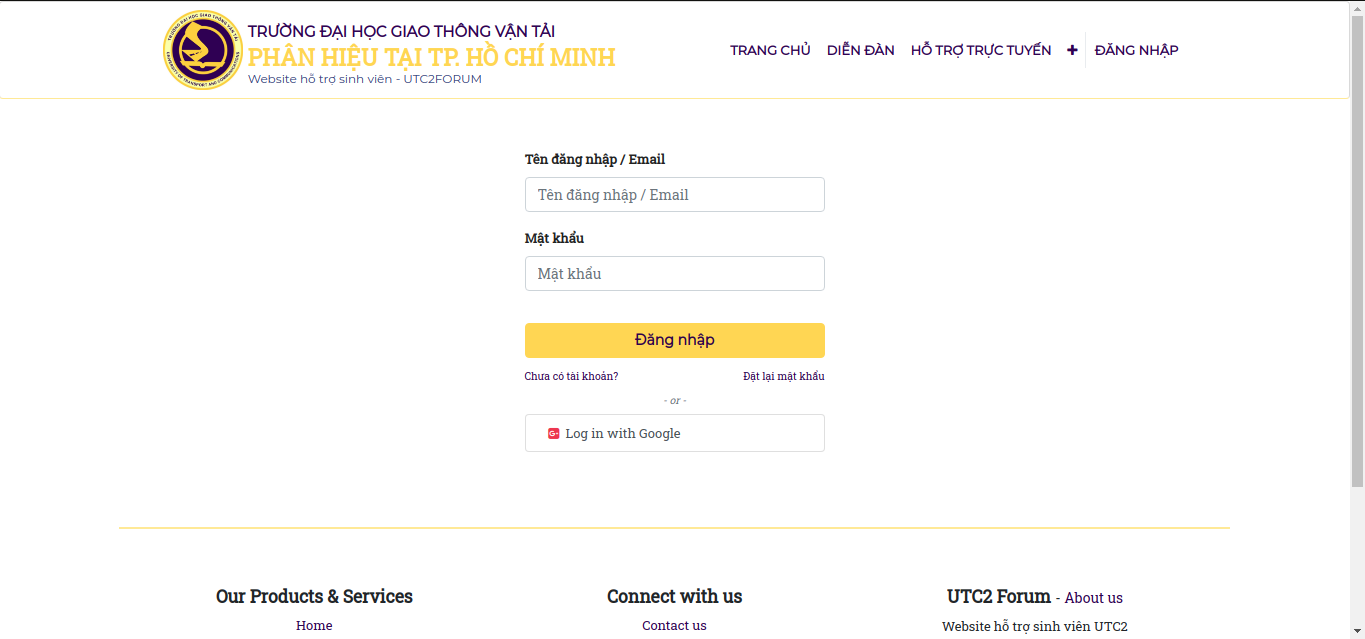
# CHƯƠNG 4: XÂY DỰNG CHƯƠNG TRÌNH

## 4.1 Tạo tài khoản và đăng nhập



##### *Hình 4.1: Giao diện tạo tài khoản đăng nhập*

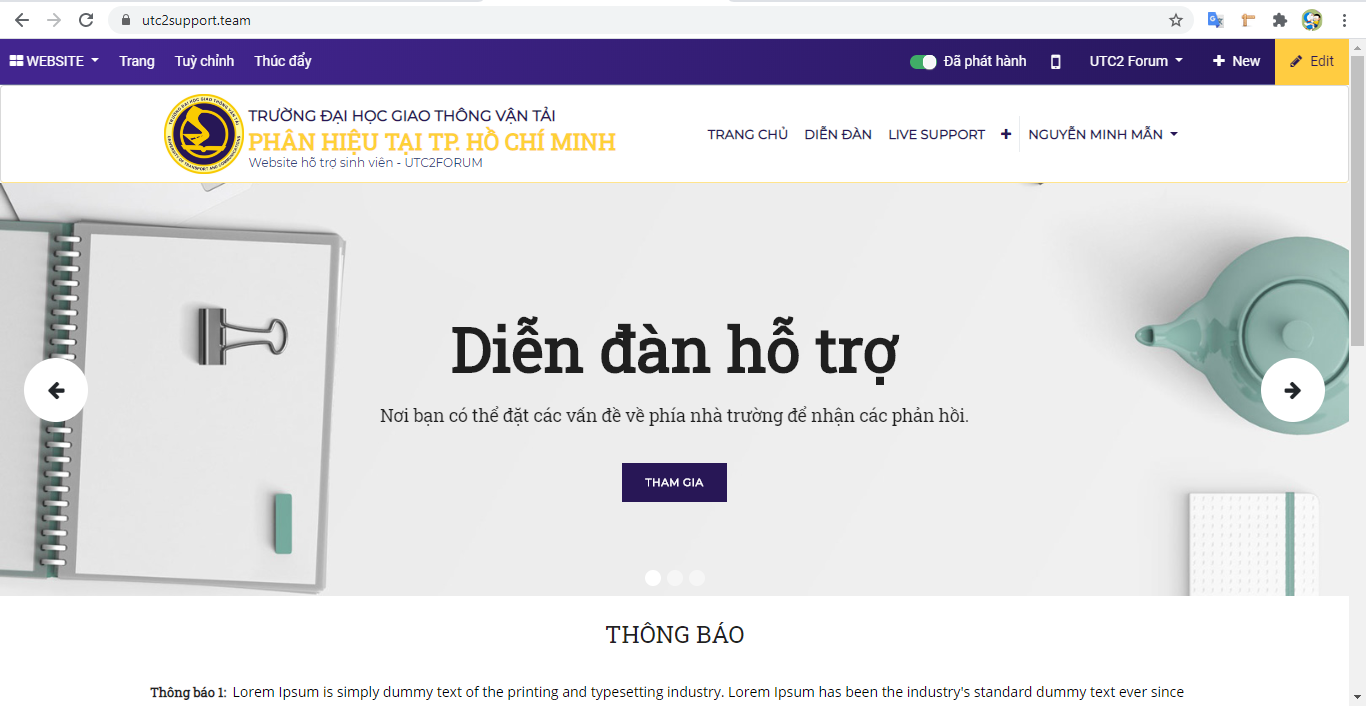
Đối với các sinh viên trường đại học Giao Thông Vận Tải phân hiệu tại TP. Hồ Chí Minh sẽ được cấp một tài khoản gmail. Các bạn sinh viên sử dụng gmail đã được trường cấp để tạo tài khoản đăng nhập.



##### *Hình 4.2: Giao diện đăng nhập*

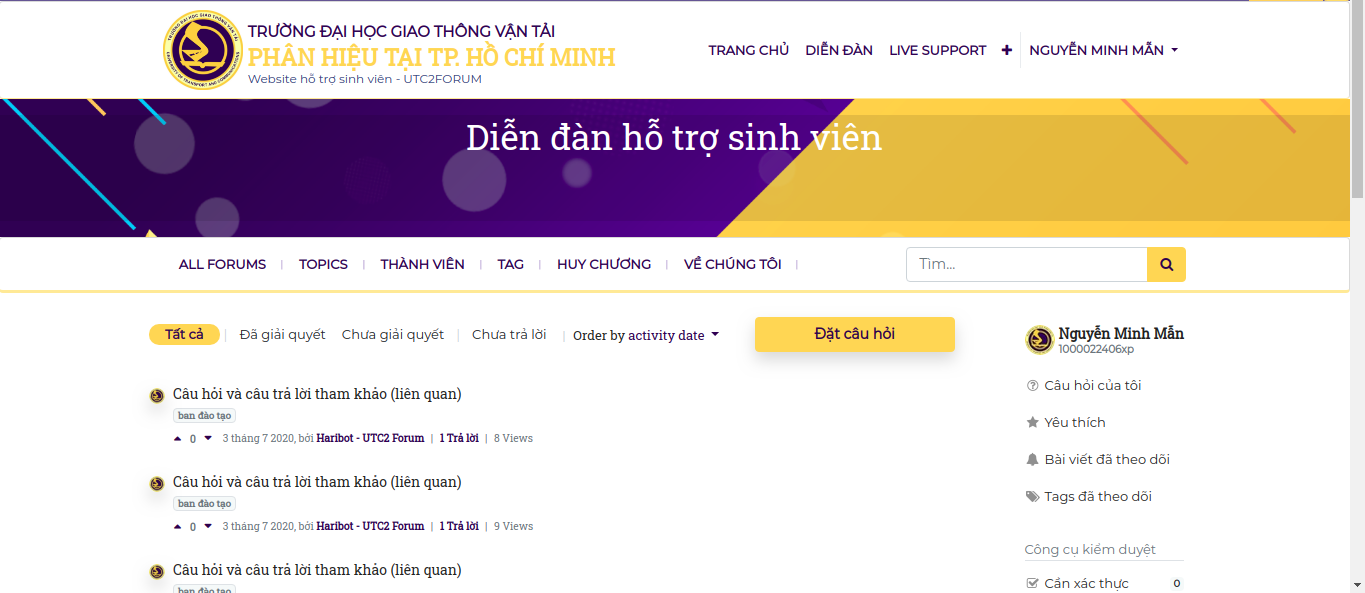
Sau khi đã có tài khoản, người dùng có thể đăng nhập để vào trang web, ngoài ra có thể đăng nhập với tài khoản google

## 4.2 Giao diện trang chủ website



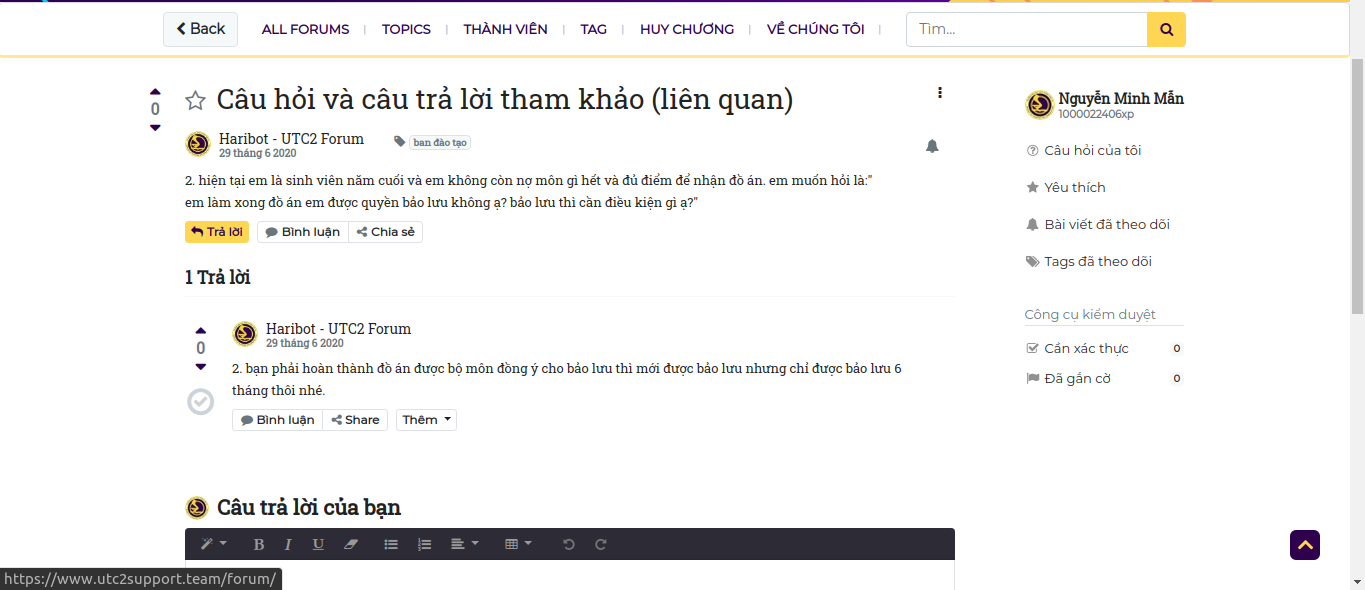
##### *Hình 4.3: Giao diện trang chủ của website*

## 4.3 Tương tác bài viết



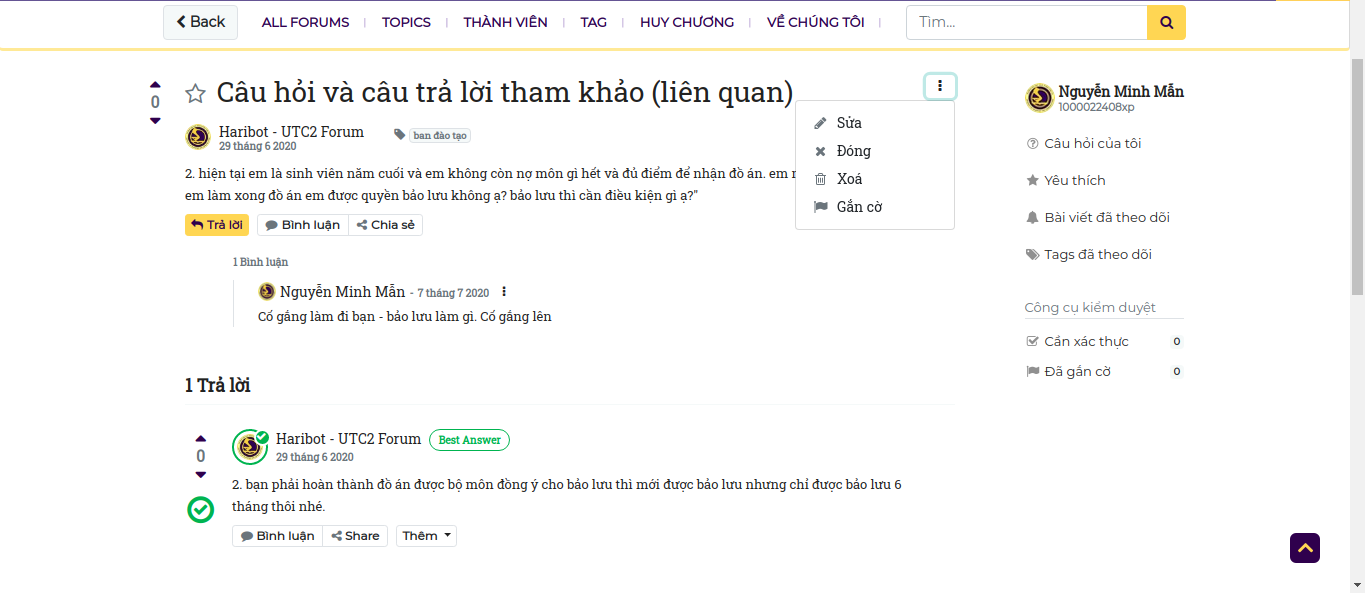
##### *Hình 4.4: Giao diện đặt câu hỏi trên diễn đàn.*

Sau khi đăng nhập vào trang web, người dùng có thể theo dõi các câu hỏi trên diễn đàn, cũng như có thể đăng các câu hỏi.



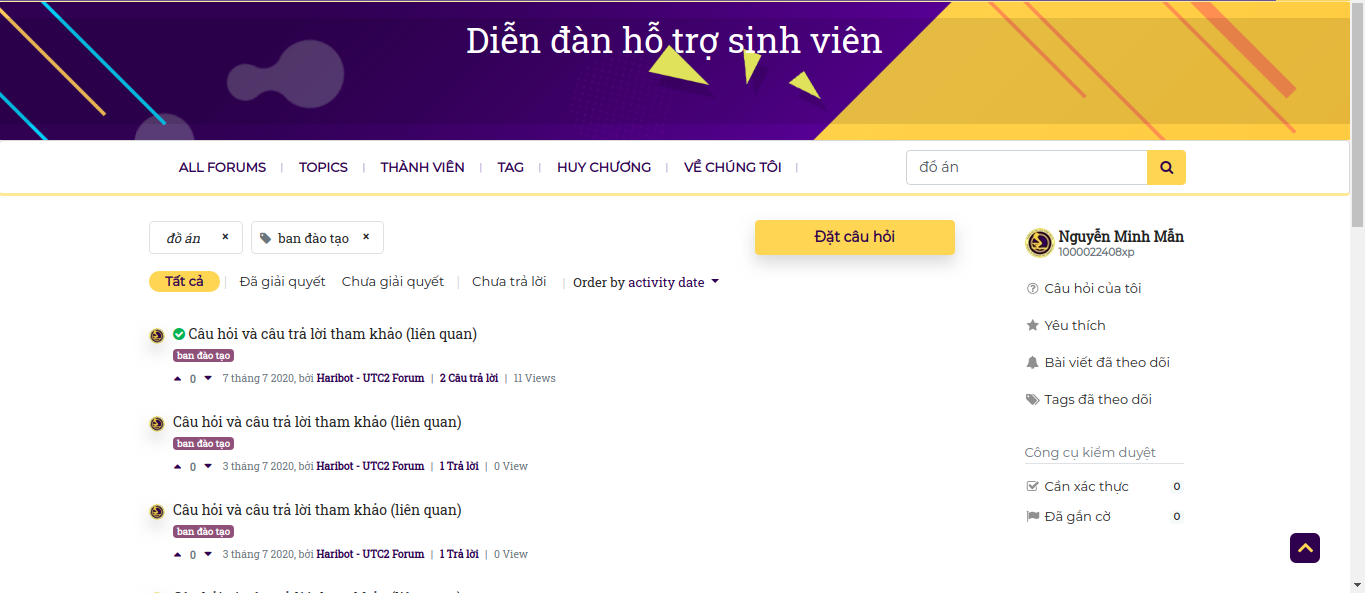
##### *Hình 4.5: Giao diện trả lời câu hỏi*

Sau khi đăng câu hỏi lên diễn đàn người dùng sẽ nhận được các đáp án phù hợp với câu hỏi của mình, ngoài ra các người dùng có thể nhận thêm được các câu trả lời từ giảng viên, các bạn sinh viên trong trường.



##### *Hình 4.6: Các chức năng thao tác với câu hỏi*

Người dùng có thể tương tác với bài viết như: Sửa, đóng, xóa, gắn cờ.



##### *Hình 4.7: Giao diện chức năng tìm kiếm câu hỏi.*

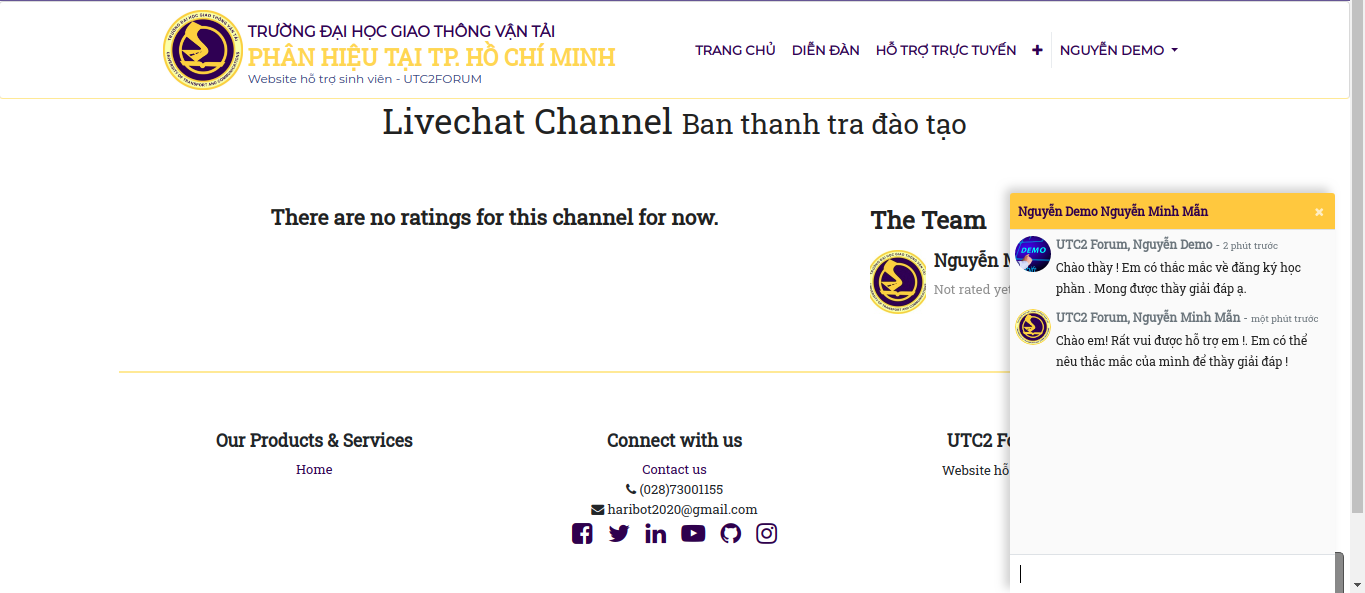
Người dùng chọn tìm kiếm câu hỏi và gõ các từ khóa cần tìm để tìm ra câu hỏi mà mình mong muốn

## 4.4 Chat trực tuyến.



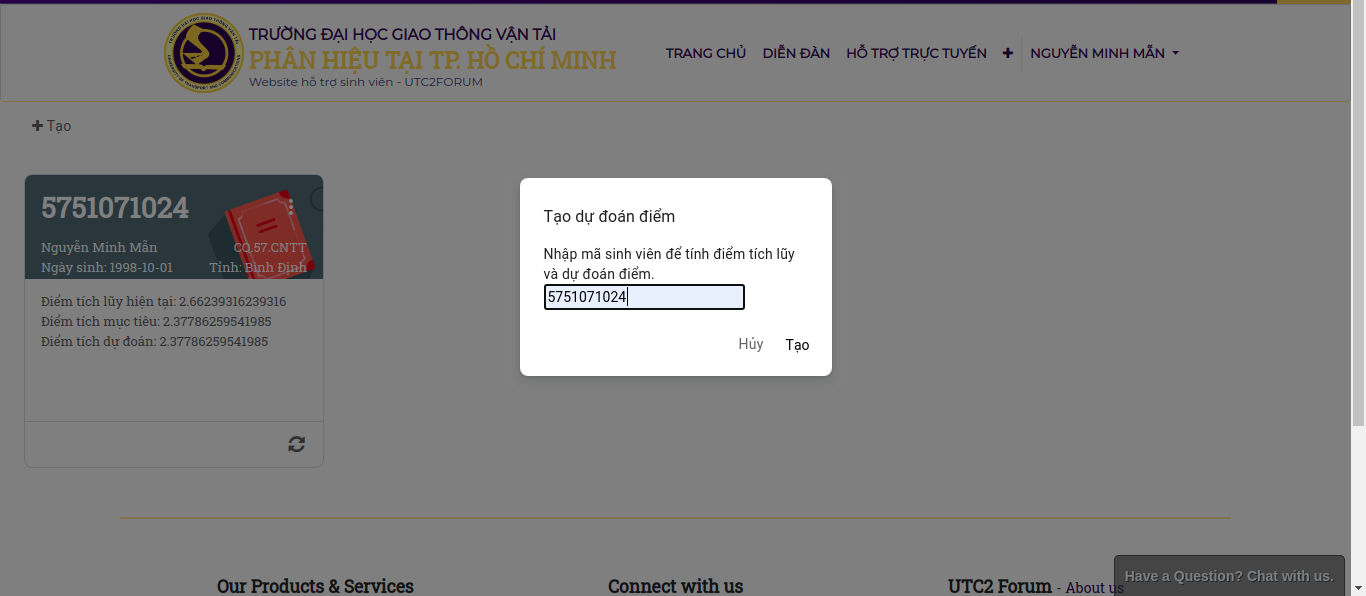
##### *Hình 4.8: Giao diện chức năng chat trực tuyến*

Khi người dùng chọn chức năng chat trực tuyến sẽ hiện ra kênh chat hỗ trợ trực tuyến, với các phòng ban như: Hỗ trợ- tư vấn, BQL KTX… Người dùng chọn phòng ban muốn hỏi và thực hiện chat.



##### *Hình 4.9: Giao diện người dùng sử dụng chức năng chat trực tuyến.*

## 4.4 Dự đoán điểm



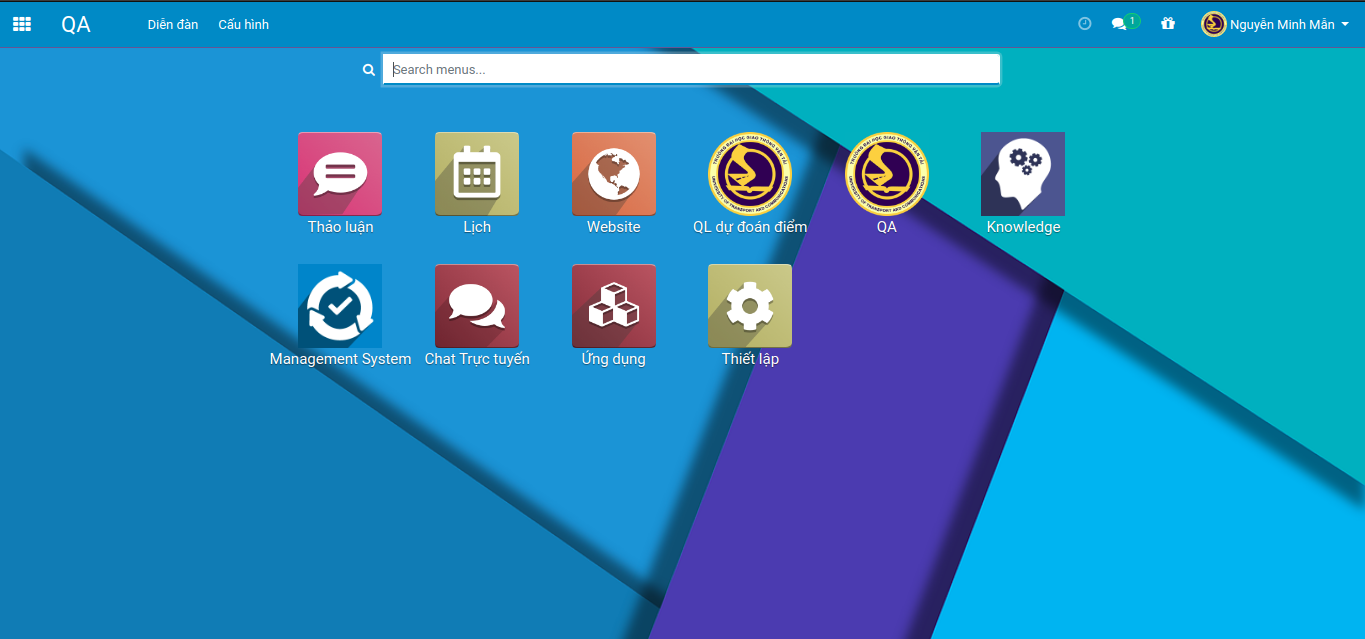
##### *Hình 4.10: Giao diện chức năng dự đoán điểm*

Khi người dùng chọn chức năng dự đoán điểm sẽ hiện ra thông tin yêu cầu nhập mã sinh viên để theo dõi điểm cũng như dự đoán điểm số



##### *Hình 4.11: Giao diện bảng điểm sau khi người dùng chọn chức năng dự đoán điểm.*

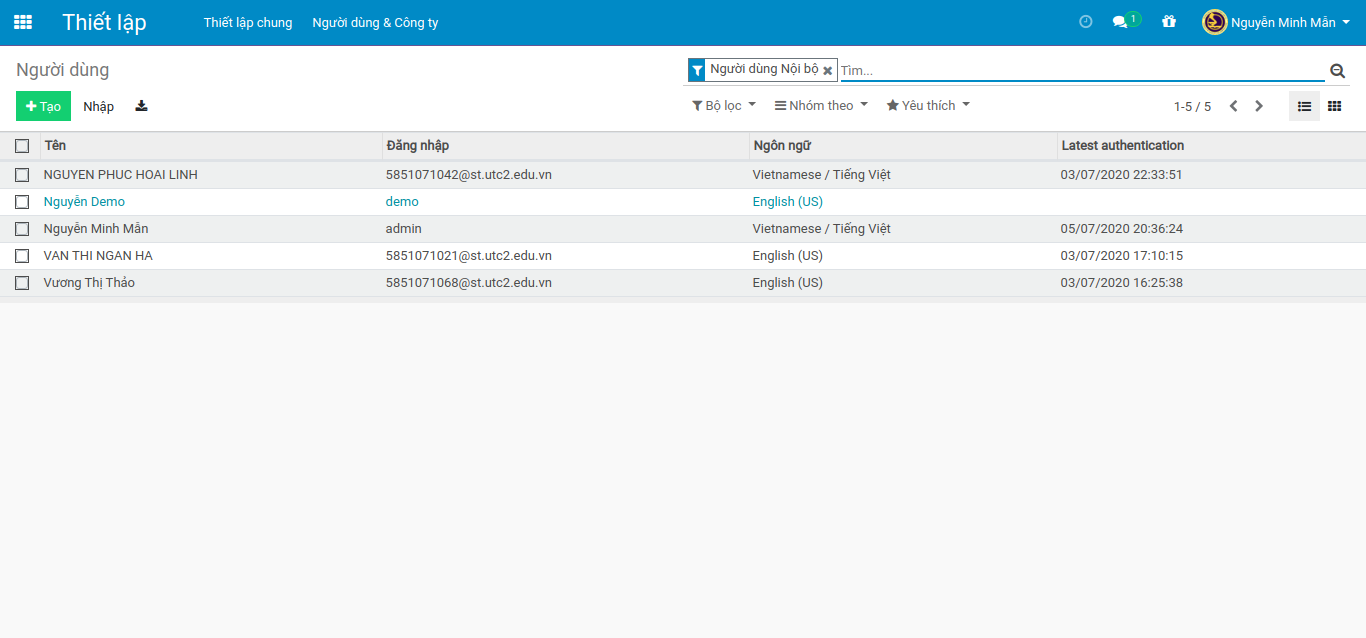
## 4.5 Nhóm chức năng quản lý



##### *Hình 4.12: Giao diện nhóm chức năng quản lý*

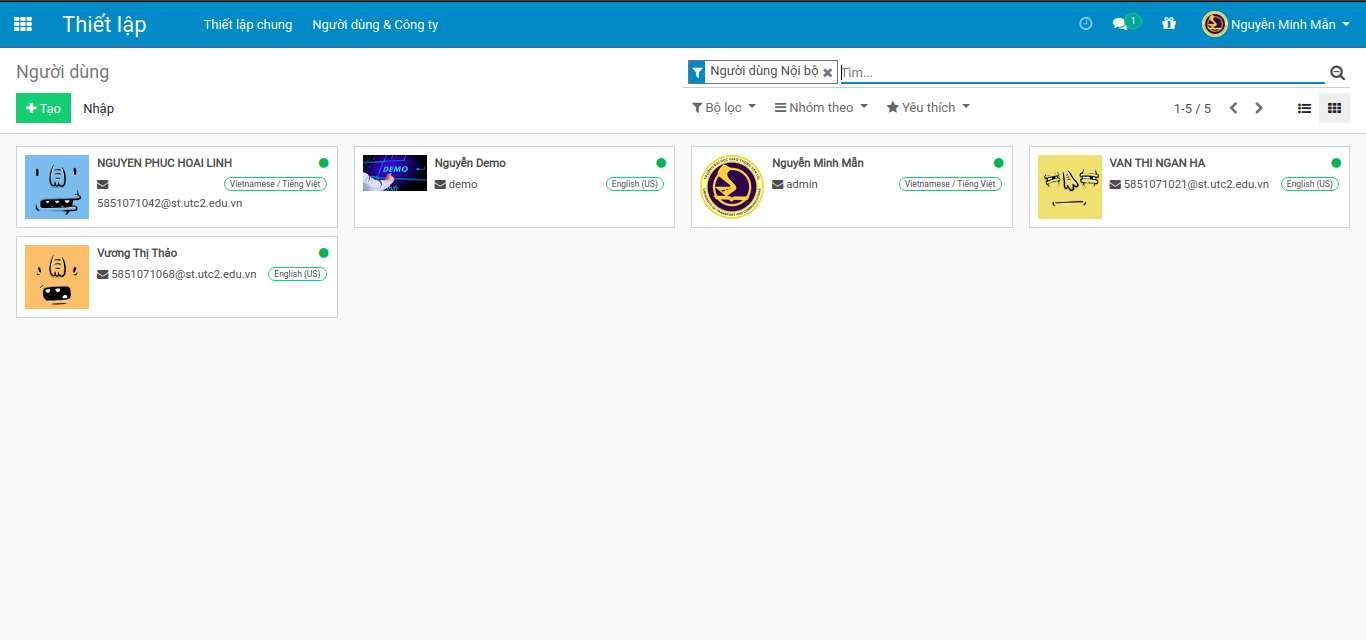
Giảng viên, quản lý(admin) của trang web sẽ có quyền thực hiện các chức năng quản lý website như: Quản lý dự đoán điểm, thiết lập cài đặt cho trang web, quản lý dự đoán điểm, quản lý người dùng, …

### 4.5.1 Quản lý người dùng



##### *Hình 4.13: Giao diện quản lý người dùng*

Chức năng quản lý người dùng do quản ly(admin) của trang sử dụng, dùng để phân quyền, thêm quyền khi muốn thêm người quản lý trang web.

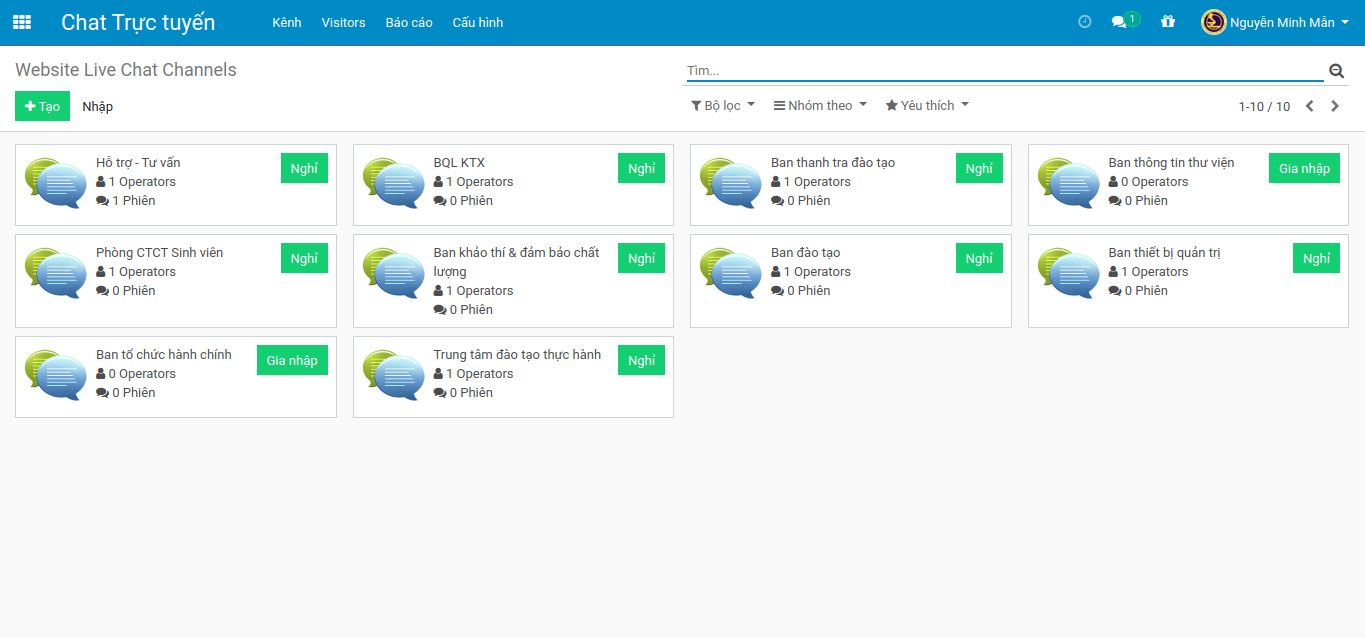
**

##### *Hình 4.14: Giao diện hiện thị người dùng truy cập trang web*

Quản lý của trang có thể quản lý số người đã truy cập và có tài khoản trong trang web.

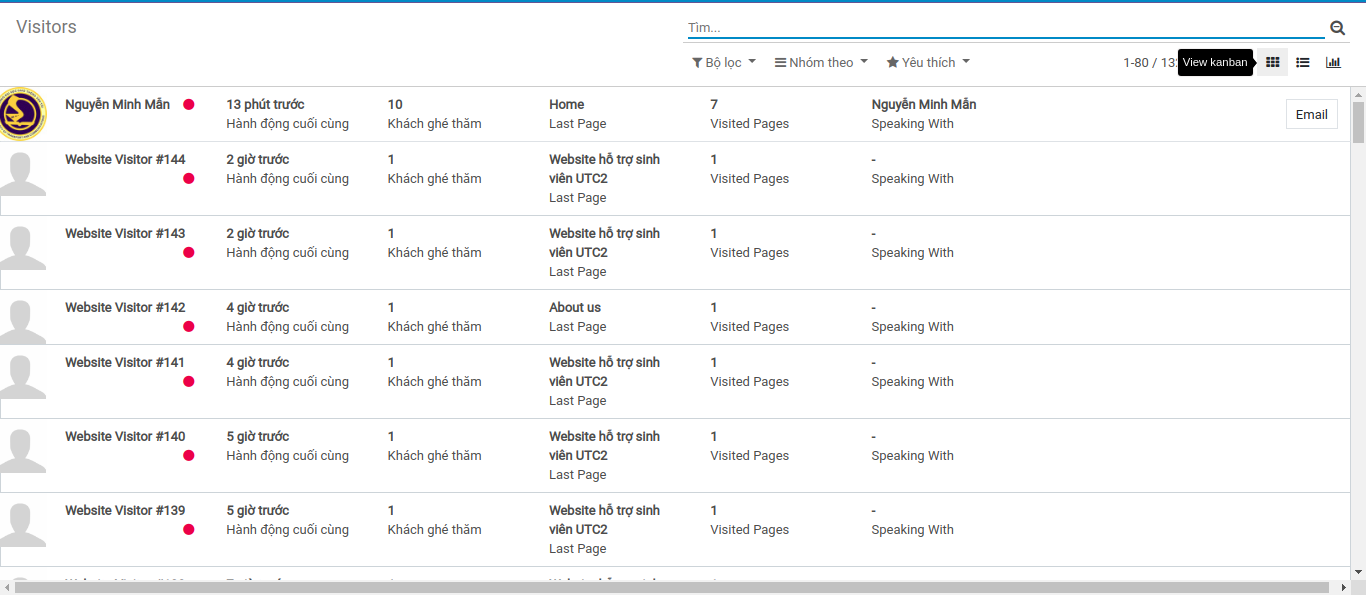
Nếu được cấp quyền admin sẽ hiện thị tài khoản admin, nếu chỉ là thành viên sẽ hiện tên gmail sinh viên truy cập.

### 4.5.2 Quản lý chat trực tuyến.



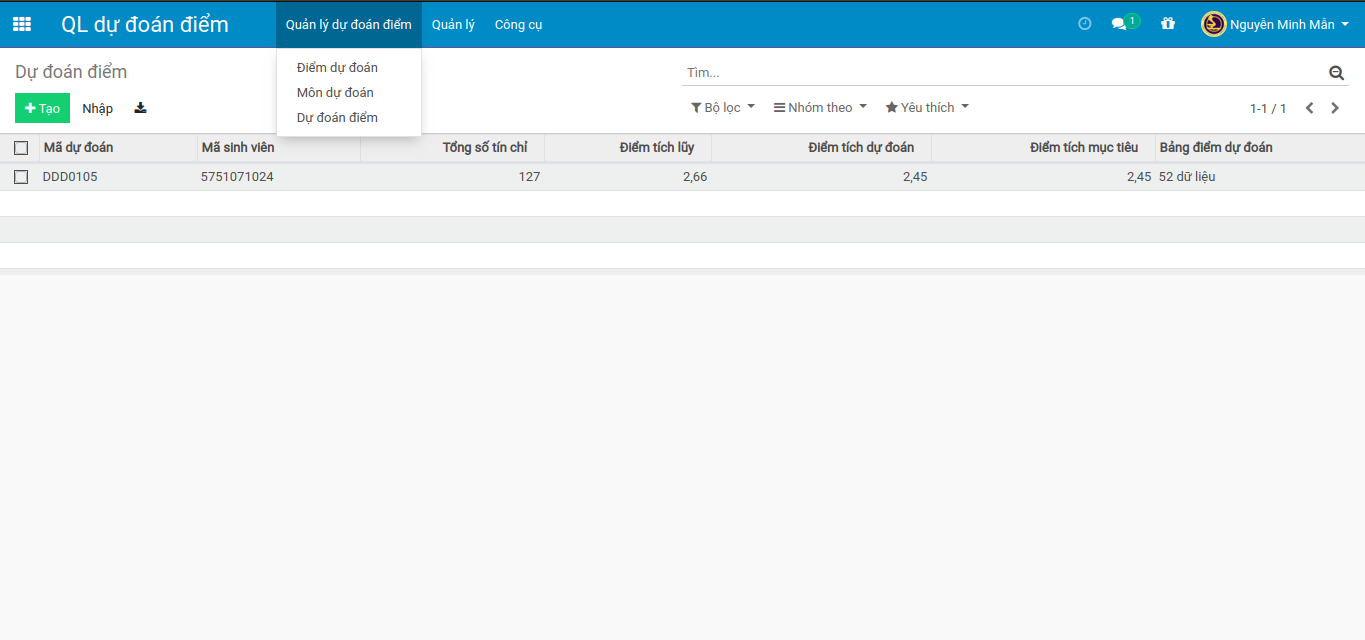
##### *Hình 4.15: Giao diện quản lý chat trực tuyến.*

Quản lý có thể thoi dõi các kênh chat, theo dỗi có người đang hoạt động trong các phòng chat. Số lượng tài khoản nhắn tin, lịch sử chat, …



##### *Hình 4.16: Giao diện theo dõi lịch sử chat.*

### 4.5.3 Quản lý dự đoán điểm



##### *Hình 4.17: Giao diện quản lý điểm*

# 

# CHƯƠNG 5: KẾT LUẬN

* 1. **Kết quả đạt được**

Sau quá trình tìm hiểu và thực hiện, đề tài đã phần nào đáp ứng được những yêu cầu đã đặt ra ở mục tiêu đề ra:

* Xây dựng được website hỗ trợ sinh viên UTC2 có đầy đủ các chức năng cần thiết hỗ trợ sinh viên như:
  + Cho phép các bạn sinh viên đăng câu hỏi để được giải đáp các thắc mắc, cũng như có thể tương tác đối với bài đăng. Có thể tìm kiếm các câu hỏi theo từ khóa cần tìm.
  + Sinh viên có thể thực hiện chức năng chat trực tuyến với các bộ phận phòng ban
  + Sinh viên có thể sử dụng chức năng dự đoán điểm
* Hệ thống chạy ổn định với giao diện thân thiện, dễ sử dụng và tương thích với cả máy tính và điện thoại di động, đồng thời website cũng thực hiện tốt trên các trình duyệt web phổ biến như hiện nay như: Chrome, Fifox, Opera.
  1. **Nhược điểm**

Do thời gian và kinh phí có hạn nên đề tài cũng còn tồn tại một số hạn chế như:

+ Một số chức năng chưa chạy ổn định như chức năng dự đoán điểm. Chức năng dự đoán điểm còn sai trong một vài trường hợp.

+ Chưa kết nối được toàn bộ dữ liệu thật tại Phân hiệu.

* 1. **Hướng phát triển**

Từ những hạn chế trên, trong thời gian tới nhóm em sẽ hoàn thiện tất cả các thiếu sót, những hạn chế mà trang web còn gặp phải để trang web có thể được nhà trường sử dụng vào thực tế, từ đó có thể giúp đỡ được nhiều bạn sinh viên.

# TÀI LIỆU THAM KHẢO

|  |  |
| --- | --- |
| [1] | Odoo, "ITPlus," [Online]. Available: http://itplus-academy.edu.vn/odoo-la-gi-tai-sao-doanh-nghiep-lai-chon-phan-mem-nay.html. [Accessed 03 05 2020]. |
| [2] | [Online]. Available: https://devteam.mobi/odoo-la-gi-tim-hieu-chi-tiet-ve-phan-mem-odoo/. |
| [3] | M. Nguyen, "Odoo," [Online]. Available: https://minhng.info/odoo/odoo-la-gi.html. [Accessed 01 05 2020]. |
| [4] | Chaupm, "Postpresql," [Online]. Available: https://tech.bizflycloud.vn/postgresql-la-gi-tim-hieu-ve-co-so-du-lieu-ma-nguon-mo-tien-tien-nhat-the-gioi-20180919175924611.htm. [Accessed 03 05 2020]. |
| [5] | W. Foundation, "Python," [Online]. Available: https://vi.wikipedia.org/wiki/Python\_(ng%C3%B4n\_ng%E1%BB%AF\_l%E1%BA%ADp\_tr%C3%ACnh). [Accessed 01 05 2020]. |
| [6] | V. M. Sebastian Raschka, "Python," in *Python Machine Learning - Third Edition*, 2019. |
| [7] | F. Wikipedia, "JavaScript," [Online]. Available: https://vi.wikipedia.org/wiki/JavaScript. [Accessed 01 05 2020]. |
| [8] | J. L. E. a. Joseph Labrecque, "JavaScript," in *The JavaScript Workshop*, 2019. |
| [9] | W. Foundation, "XML," [Online]. Available: https://vi.wikipedia.org/wiki/XML. [Accessed 07 05 2020]. |
| [10] | J. Robbins, Learning Web Design: A Beginner's Guide to HTML, CSS, JavaScript, and Web Graphics, 2020. |