| **TRƯỜNG ĐẠI HỌC GTVT**  **KHOA CÔNG NGHỆ THÔNG TIN** | **CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM**  **Độc lập - Tự do - Hạnh phúc** |
| --- | --- |

*Hà Nội, ngày 17 tháng 02 năm 2025*

# ĐỀ CƯƠNG ĐỒ ÁN TỐT NGHIỆP

**Họ và tên sinh viên:**

Mã SV : 211210438 Lớp: CNCNTT1 Khóa: 62

Số điện thoại : 0865636911 Email: ktcopx@gmail.com

Ngành : Công nghệ thông tin Hệ: Chính quy

**Giảng viên (cán bộ) hướng dẫn**: TS. Nguyễn Quốc Tuấn

Đơn vị công tác : Khoa Công nghệ thông tin, Trường Đại học Giao thông vận tải

Số điện thoại : 0912228980 Email: nqtuan@utc.edu.vn

**Tên đề tài**: Ứng dụng khai phá dữ liệu dự đoán kết quả học tập của sinh viên

1. **Nội dung, phạm vi của đề tài**

* **Nội dung đề tài:**
* Đề tài tập trung vào việc ứng dụng các kỹ thuật khai phá dữ liệu và học máy để dự báo kết quả học tập của sinh viên. Cụ thể, nghiên cứu sẽ tìm hiểu và áp dụng các thuật toán như cây quyết định, Naive Bayes, hồi quy Logistic,… nhằm khai thác dữ liệu sinh viên thu thập từ nhiều nguồn khác nhau. Mục tiêu là khám phá tri thức hữu ích từ dữ liệu này để nâng cao hiệu quả giảng dạy và học tập.
* Đề tài này sẽ nghiên cứu việc ứng dụng các kỹ thuật khai thác dữ liệu để phân tích và dự đoán kết quả học tập trong môi trường giáo dục. Các khía cạnh chính sẽ bao gồm:
* Nghiên cứu các khái niệm cơ bản và phương pháp khai thác dữ liệu.
* Phân tích vai trò của dữ liệu trong giáo dục và cách thức thu thập dữ liệu sinh viên.
* Tìm hiểu các thuật toán phân lớp.
* Phân tích hiệu suất của các mô hình thông qua các tiêu chí đánh giá khác nhau.
* Xem xét các ứng dụng thực tiễn của nghiên cứu trong giáo dục.
* Đề xuất hướng phát triển cho các công cụ và phương pháp hỗ trợ giảng dạy và học tập trong tương lai.
* **Phạm vi đề tài:** Khoa CNTT - Trường DH GTVT

1. **Công nghệ, công cụ và ngôn ngữ lập trình**

* Công cụ lập trình: Python.
* Pandas, numpy: Tiền xử lý dữ liệu.
* Matplotlib, seaborn: Trực quan hóa dữ liệu.
* Scikit-learn, GridSearchCV: xây dựng mô hình học máy.
* Flask, FastAPI: xây dựng ứng dụng giao diện người dùng.
* Lưu trữ, làm sạch và truy vấn dữ liệu: Excel, MSSQL.
* Công cụ kiểm thử: Postman.

1. **Các kết quả chính dự kiến đạt được**

* **Hiểu biết sâu sắc về kỹ thuật khai phá dữ liệu:** Nắm vững các khái niệm cơ bản và phương pháp khai thác dữ liệu trong giáo dục.
* **Phân tích vai trò của dữ liệu**: Làm rõ tầm quan trọng của dữ liệu trong việc dự đoán kết quả học tập, cùng với các nguồn dữ liệu có thể thu thập.
* **Áp dụng thuật toán phân lớp**: Thực hành và so sánh các thuật toán như cây quyết định, Naive Bayes, và hồi quy Logistic trong việc phân tích kết quả học tập.
* **Đánh giá hiệu suất mô hình**: Phát triển các tiêu chí để đánh giá hiệu suất của các mô hình dự đoán, từ đó rút ra kết luận về độ chính xác và khả năng áp dụng.
* **Ứng dụng thực tiễn**: Đề xuất các ứng dụng cụ thể của nghiên cứu này trong môi trường giáo dục nhằm cải thiện chất lượng giảng dạy và học tập.
* **Kế hoạch phát triển sản phẩm**: Đề xuất hướng phát triển cho các công cụ và phương pháp hỗ trợ giảng dạy và học tập dựa trên kết quả nghiên cứu, nhằm tối ưu hóa quy trình giáo dục.
* **Triển khai mô hình**: Thực hiện triển khai các mô hình dự đoán vào ứng dụng thực tế, bao gồm việc phát triển giao diện người dùng và tích hợp vào hệ thống quản lý học tập, giúp giáo viên và sinh viên dễ dàng truy cập và sử dụng.

1. **Kế hoạch thực hiện đề tài**

| **STT** | **Nội dung công việc** | **Thời gian dự kiến** | **Ghi chú** |
| --- | --- | --- | --- |
| 1 | Nghiên cứu tài liệu:  - Tìm hiểu lý thuyết về khai phá dữ liệu và học máy.  - Nghiên cứu các thuật toán phân lớp. | 03/03/2025-10/03/2025 |  |
| 2 | Thu thập và chuẩn bị dữ liệu:  - Xác định nguồn dữ liệu và phương pháp thu thập.  - Tiến hành thu thập và làm sạch dữ liệu. | 11/03/2025-24/03/2025 |  |
| 3 | Phân tích dữ liệu và áp dụng mô hình:  - Thực hiện phân tích dữ liệu ban đầu.  - Áp dụng các thuật toán như cây quyết định, Naive Bayes,... | 25/03/2025-09/04/2025 |  |
| 4 | Đánh giá hiệu suất mô hình:  - Xây dựng tiêu chí đánh giá và kiểm tra mô hình.  - So sánh kết quả và cải thiện mô hình. | 10/04/2025-21/04/2025 |  |
| 5 | Xây dựng ứng dụng dự báo:  - Thiết kế giao diện người dùng.  - Áp dụng mô hình vào dự báo kết quả học tập sinh viên. | 22/04/2025-02/05/2025 |  |
| 6 | Viết báo cáo và hoàn thiện đồ án | 03/05/2025-21/05/2025 |  |

| **Trưởng Khoa**  *(Ký và ghi rõ họ tên)* | **Trưởng Bộ môn**  *(Ký và ghi rõ họ tên)* | **Giảng viên hướng dẫn**  *(Ký và ghi rõ họ tên)* | **Sinh viên thực hiện**  *(Ký và ghi rõ họ tên)* |
| --- | --- | --- | --- |