**Shape, square

Description automatically generatedTRƯỜNG ĐẠI HỌC QUY NHƠN**

KHOA CÔNG NGHỆ THÔNG TIN

-----🙞🙜🕮🙞🙜-----

A blue and white logo

Description automatically generated with low confidence

**HỌC PHẦN: THỰC HÀNH LÀM VIỆC NHÓM**

**ĐỀ CƯƠNG CHI TIẾT VÀ KẾ HOẠCH LÀM VIỆC NHÓM**

***Giảng viên :*** Th.S Lê Thị Xinh

***Nhóm sinh viên:*** Nguyễn Trương Quỳnh Mai - CNTT K43A

Dương Thành Nghĩa - CNTT K44A

Phạm Hồng Thiện - CNTT K44A

Phạm Văn Đồng - CNTT K44A

***Mã lớp học phần:*** 231105013615

***Bình Định, Tháng 9 năm 2023***

**MỤC LỤC**

[I. Thành viên nhóm: 3](#_Toc9690)

[II. Đề tài, phân chia công việc và đề cương chi tiết: 3](#_Toc9049)

[1. Đề tài: 3](#_Toc20954)

[2. Kế hoạch phân chia công việc: 3](#_Toc3402)

[3. Đề cương chi tiết: 4](#_Toc4683)

*[a. Tên đề tài:](#_Toc30922)* [4](#_Toc30922)

*[b. Đặt vấn đề:](#_Toc21117)* [4](#_Toc21117)

*[c. Mục đích và mục tiêu nghiên cứu:](#_Toc5312)* [4](#_Toc5312)

*[d. Đối tượng cần nghiên cứu:](#_Toc15119)* [4](#_Toc15119)

*[e. Phương pháp nghiên cứu:](#_Toc12104)* [4](#_Toc12104)

[f.](#_Toc20957) *[Nội dung nghiên cứu:](#_Toc20957)* [4](#_Toc20957)

1. **Thành viên nhóm:**

Nhóm 1 học phần Thực hành làm việc nhóm gồm có 4 thành viên như danh sách sau:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Mã sinh viên** | **Họ và tên** | **Lớp** | **Vai trò** |
| 1 | 4351050193 | Nguyễn Trương Quỳnh Mai | CNTT K43A | Nhóm trưởng |
| 2 | 4451050352 | Phạm Hồng Thiện | CNTT K44A | Thành viên |
| 3 | 4451050217 | Dương Thành Nghĩa | CNTT K44A | Thành viên |
| 4 | 4451050078 | Phạm Văn Đồng | CNTT K44A | Thành viên |

1. **Đề tài, phân chia công việc và đề cương chi tiết:**
2. **Đề tài:**

Nhóm 1 được phân công đề tài tìm hiểu về thuật toán Naive Bayes.

1. **Kế hoạch phân chia công việc:**

Theo như đề tài được phân công, nhóm 1 phân chia công việc cụ thể cho các thành viên như sau:

|  |  |  |
| --- | --- | --- |
| **Thời gian** | **Công việc** | **Người thực hiện** |
| Tuần 1 | Tìm hiểu chung về học máy | Tất cả các thành viên |
| Tuần 2 | Tìm hiểu học máy có giám sát và ứng dụng | Phạm Hồng Thiện |
| Tìm hiểu học máy không giám sát và ứng dụng | Nguyễn Trương Quỳnh Mai |
| Tìm hiểu học máy bán giám sát và ứng dụng | Dương Thành Nghĩa |
| Tìm hiểu học tăng cường và ứng dụng | Phạm Văn Đồng |
| Tuần 3-4 | Tìm hiểu thuật toán được phân công: Naive Bayes | Tất cả các thành viên |
| Tuần 5 | Tìm hiểu công cụ quản lý code github | Tất cả các thành viên |
| Tuần 6-7 | Ứng dụng Naive Bayes để phân loại văn bản - Xây dựng mô hình - Lưu mô hình | Nguyễn Trương Quỳnh Mai  Phạm Hồng Thiện |
| Tuần 8-9 | Xây dựng web - Đưa mô hình lên web để ứng dụng thực tế | Phạm Văn Đồng  Dương Thành Nghĩa |
| Tuần 10 | Kiểm tra và cải thiện lỗi | Tất cả các thành viên |
| Tuần 11 | Tổng hợp và viết báo cáo | Nguyễn Trương Quỳnh Mai  Phạm Hồng Thiện |

1. **Đề cương chi tiết:**
2. *Tên đề tài:*

Tìm hiểu và ứng dụng thuật toán Naive Bayes để phân loại văn bản

1. *Đặt vấn đề:*

Thuật toán Naive Bayes là một thuật toán học máy kinh điển dựa trên định lý Bayes về lý thuyết xác suất để đưa ra các phán đoán cũng như phân loại dữ liệu dựa trên các dữ liệu được quan sát và thống kê, được ứng dụng rất nhiều trong các lĩnh vực Machine learning dùng để đưa các dự đoán có độ chính xác cao, dựa trên một tập dữ liệu đã được thu thập.

Phân loại văn bản là một trong số những vấn đề mà chúng ta quan tâm hiện nay. Thay vì phân loại thủ công mất nhiều thời gian thì ta có thể ứng dụng các thuật toán học máy để máy cớ thể tự học và phân loại nhanh hơn cách thủ công, dẫn đến hiệu quả và tiết kiệm được nhiều thời gian hơn.

1. *Mục đích và mục tiêu nghiên cứu:*

Trong học phần này chúng em sẽ tìm hiểu về các thuật toán học máy và chú trọng vào thuật toán Naive Bayes. Sau đó ứng dụng Naive Bayes vào phân loại các văn bản thực tế.

1. *Đối tượng cần nghiên cứu:*

Thuật toán Naive Bayes và ứng dụng.

1. *Phương pháp nghiên cứu:*

Nghiên cứu lý thuyết kết hợp với thực nghiệm.

1. *Nội dung nghiên cứu:*

Nội dung nghiên cứu thuật toán Naive Bayes thường bao gồm các bước sau:

\*\*Hiểu Thuật Toán Naive Bayes\*\*:

- Bắt đầu bằng việc tìm hiểu cơ bản về thuật toán Naive Bayes. Thuật toán này dựa trên nguyên tắc của Bayes và giả định đặc biệt (naive) rằng các đặc trưng độc lập với nhau trong mô hình.

\*\*Thu Thập Dữ Liệu\*\*:

- Thu thập dữ liệu phù hợp cho bài toán cụ thể muốn áp dụng Naive Bayes. Dữ liệu này cần phải có nhãn hoặc thông tin phân loại để có thể huấn luyện và kiểm tra mô hình.

\*\*Tiền Xử Lý Dữ Liệu\*\*:

- Rất quan trọng để làm sạch và tiền xử lý dữ liệu trước khi áp dụng Naive Bayes. Các bước tiền xử lý bao gồm loại bỏ dữ liệu nhiễu, chuẩn hóa dữ liệu, xử lý dữ liệu thiếu, và chuyển đổi dữ liệu thành dạng phù hợp cho thuật toán.

\*\*Chia Dữ Liệu\*\*:

- Chia dữ liệu thành tập huấn luyện và tập kiểm tra. Thường thì sẽ dùng một phần lớn dữ liệu để huấn luyện mô hình và một phần nhỏ để kiểm tra hiệu suất của mô hình.

\*\*Xây Dựng Mô Hình\*\*:

- Áp dụng thuật toán Naive Bayes trên tập huấn luyện để xây dựng mô hình. Cần chọn phiên bản phù hợp của Naive Bayes cho bài toán, chẳng hạn như Gaussian Naive Bayes, Multinomial Naive Bayes, hoặc Bernoulli Naive Bayes tùy thuộc vào loại dữ liệu và mục tiêu phân loại.

\*\*Huấn Luyện Mô Hình\*\*:

- Sử dụng tập huấn luyện để điều chỉnh các tham số của mô hình. Trong Naive Bayes, điều này liên quan đến tính toán các xác suất có điều kiện và xác suất tiên nghiệm dựa trên dữ liệu huấn luyện.

\*\*Đánh Giá Mô Hình\*\*:

- Sử dụng tập kiểm tra để đánh giá hiệu suất của mô hình bằng các độ đo như độ chính xác, độ nhạy (recall), độ chính xác dự đoán (precision), F1-score, và ma trận nhầm lẫn (confusion matrix).

\*\*Tối ưu Hóa và Điều Chỉnh Mô Hình (nếu cần)\*\*:

- Dựa trên kết quả đánh giá, có thể điều chỉnh mô hình bằng cách thay đổi tham số, thử nghiệm các phiên bản khác nhau của Naive Bayes, hoặc thực hiện các biện pháp khác để tối ưu hóa mô hình.

\*\*Sử Dụng Mô Hình\*\*:

- Sau khi đã huấn luyện và đánh giá mô hình, có thể sử dụng nó để dự đoán nhãn hoặc phân loại cho các dữ liệu mới.

\*\*Kiểm Tra và Cải Tiến\*\*:

- Liên tục kiểm tra và cải tiến mô hình theo thời gian khi có dữ liệu mới hoặc cần thiết.