TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO CUỐI KỲ NHẬP MÔN HỌC MÁY**

**TÌM HIỂU VỀ PHƯƠNG PHÁP OPTIMIZER, CONTINUAL LEARNING VÀ TEST PRODUCTION TRONG MÔ HÌNH HỌC MÁY**

*Người hướng dẫn*: **THẦY LÊ ANH CƯỜNG**

*Người thực hiện*: **NGUYỄN THỊ DIỄM SƯƠNG – 52000129**

Lớp **: 20050201**

Khoá  **: 24**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO GIỮA KỲ NHẬP MÔN HỌC MÁY**

**TÌM HIỂU VỀ PHƯƠNG PHÁP OPTIMIZER, CONTINUAL LEARNING VÀ TEST PRODUCTION TRONG MÔ HÌNH HỌC MÁY**

*Người hướng dẫn*: **THẦY LÊ ANH CƯỜNG**

*Người thực hiện*: **NGUYỄN THỊ DIỄM SƯƠNG – 52000129**

Lớp **: 20050201**

Khoá  **: 24**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

LỜI CẢM ƠN

Trong thời gian hoàn thành bài báo cáo vừa qua, em đã nhận được rất nhiều sự giúp đỡ, hướng dẫn và hỗ trợ tận tình từ quý thầy cô và các bạn. Em cảm thấy vô cùng biết ơn và muốn gửi những lời cảm ơn sâu sắc đến thầy cô, bạn bè và gia đình đã giúp bài báo cáo của em đạt kết quả tốt như hiện nay.

Đặc biệt, em muốn bày tỏ sự biết ơn sâu sắc và lòng kính trọng đến thầy Lê Anh Cường, thầy là người đã giảng dạy em môn Nhập môn Học máy trong suốt học kỳ vừa qua. Trong quá trình học tập, thầy đã truyền đạt cho em vô vàn kiến thức hay và bổ ích, giúp em có được cơ sở lý thuyết vững vàng để em vượt qua bài báo cáo này dễ dàng hơn.

Bài báo cáo này cũng không thể tránh khỏi những thiếu sót và hạn chế, em rất mong quý thầy cô sẽ bỏ qua cho em và chỉ bảo thêm để giúp em có điều kiện bổ sung và làm tốt hơn trong những bài báo cáo sau này.

Em xin kính chúc quý thầy, quý cô và quý nhà trường luôn mạnh khỏe, hạnh phúc và ngày một thành công hơn trong sự nghiệp trồng người của mình.

Chúng em xin chân thành cảm ơn!

**ĐỒ ÁN ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi và được sự hướng dẫn khoa học của ThS Lê Anh Cường. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong luận văn còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung luận văn của mình.** Trường đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày 20 tháng 12 năm 2023*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

*Sương*

*Nguyễn Thị Diễm Sương*

PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN

**Phần xác nhận của GV hướng dẫn**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

**Phần đánh giá của GV chấm bài**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

TÓM TẮT

**1. Vấn đề nghiên cứu:**

Bài làm gồm có 2 chương:

* Chương 1: Tìm hiểu, so sánh các phương pháp Optimizer trong huấn luyện mô hình học máy.
* Chương 2: Tìm hiểu về Continual Learning và Test Production khi xây dựng một giải pháp học máy để giải quyết một bài toán nào đó.

**2. Các hướng tiếp cận**

* Lý thuyết.
* Thực hành.

**3. Cách giải quyết vấn đề**

Xem lại những nội dung đã học qua slide bài giảng, các kiến thức được ghi chép lại trong quá trình học và nghiên cứu thêm các video bài giảng trên mạng. Vận dụng chúng vào để giải quyết các nội dung trong đề thi.

**4. Một số kết quả đạt được**

Ôn lại được những kiến thức đã học, nắm vững các lý thuyết và phương pháp làm bài môn Nhập môn học máy. Rèn luyện tư duy logic cho việc học tập các môn học sau.

MỤC LỤC

[LỜI CẢM ƠN 1](#_Toc148885754)

[PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN ii](#_Toc148885755)

[TÓM TẮT 1](#_Toc148885756)

[MỤC LỤC 2](#_Toc148885757)

[DANH MỤC CÁC BẢNG BIỂU, HÌNH VẼ, ĐỒ THỊ 4](#_Toc148885758)

[CHƯƠNG 1 - TRÌNH BÀY CÁC KHÍA CẠNH CỦA MÔ HÌNH HỌC MÁY 7](#_Toc148885759)

[1.1 Mô hình KNN: 7](#_Toc148885760)

[1.1.1 Mục tiêu của việc tạo ra mô hình: 7](#_Toc148885761)

[1.1.2 Phương pháp giải thuật, tiêu chí để học mô hình 8](#_Toc148885762)

[1.1.3 Loại bài toán và dữ liệu phù hợp với mô hình, các ưu nhược điểm của mô hình: 11](#_Toc148885763)

[1.2 Linear Regression 14](#_Toc148885764)

[1.2.1 Mục tiêu của việc tạo ra mô hình: 14](#_Toc148885765)

[1.2.2 Phương pháp giải thuật, tiêu chí để học mô hình 15](#_Toc148885766)

[1.2.3 Loại bài toán và dữ liệu phù hợp mô hình, ưu nhược điểm 17](#_Toc148885767)

[1.3 Naive Bayes classifiers 20](#_Toc148885768)

[1.3.1 Mục tiêu của việc tạo ra mô hình: 20](#_Toc148885769)

[1.3.2 Các phương pháp, giải thuật và tiêu chí học của mô hình Naive Bayes Classification (NBC): 20](#_Toc148885770)

[*1.3.2.1 Tìm hiểu về đính lý Bayes:* 20](#_Toc148885771)

[*1.3.2.2 Naive Bayes Classifier* 21](#_Toc148885772)

[*1.3.2.3* *Các kiểu mô hình Naive Bayes Classifier:* 22](#_Toc148885773)

[1.3.3 Mô hình Naive Bayes Classifier phù hợp cho loại bài toán và dữ liệu nào, ưu nhược điểm của mô hình này: 24](#_Toc148885774)

[1.4 Mô hình Decision Tree 26](#_Toc148885775)

DANH MỤC CÁC BẢNG BIỂU, HÌNH VẼ, ĐỒ THỊ

**DANH MỤC HÌNH**

[Hình 1. 1 Mục tiêu của việc tạo ra mô hình kNN 7](#_Toc148885811)

[Hình 1. 2 Các hàm khoảng cách cơ bản 8](#_Toc148885812)

[Hình 1. 3 Linear Regression 14](#_Toc148885813)

[Hình 1. 4: Công thức Linear Regression 15](#_Toc148885814)

[Hình 1. 5 Sơ đồ Linear Regression 15](#_Toc148885815)

[Hình 1. 6 Công thức tính MSE 16](#_Toc148885816)

[Hình 1. 7 Naive Bayes Classifier examples 22](#_Toc148885817)

[Hình 1. 8 Công thức Gausian Naive Bayes 24](#_Toc148885818)

[Hình 1. 9 Công thức giải quyết vấn đề zero 25](#_Toc148885819)

[Hình 1. 10 Minh họa Decision Tree 26](#_Toc148885820)

[Hình 1. 11 Decision Tree là gì? 27](#_Toc148885821)

[Hình 1. 12 Công thức Entropy cho một thuộc tính 28](#_Toc148885822)

[Hình 1. 13 Công thức Entropy cho nhiều thuộc tính 28](#_Toc148885823)

[Hình 1. 14 Công thức Information Gain 29](#_Toc148885824)

[Hình 1. 15 Công thức Information Gain đơn giản hơn 29](#_Toc148885825)

[Hình 1. 16 Công thức Gini Index 29](#_Toc148885826)

[Hình 1. 17 Công thức Gain Ratio 30](#_Toc148885827)

[Hình 1. 18 Công thức Reduction in variance 31](#_Toc148885828)

[Hình 1. 19 Công thức Chi-Square 32](#_Toc148885829)

[Hình 1. 20 Pruning 33](#_Toc148885830)

CHƯƠNG 1 - TÌM HIỂU, SO SÁNH CÁC PHƯƠNG PHÁP OPTIMIZER TRONG HUẤN LUYỆN MÔ HÌNH HỌC MÁY