

TRƯỜNG ĐẠI HỌC NAM CẦN THƠ
KHOA CÔNG NGHỆ THÔNG TIN



VŨ HOÀNG KHANG
1810375

ỨNG DỤNG TRÍ TUỆ NHÂN TẠO
TRONG VIỆC GIẢI QUYẾT MỘT BỘ DỮ LIỆU
PHÂN LỚP
(MÁY HỌC TĂNG TIẾN
GIẢI QUYẾT BÀI TOÁN CHẨN ĐOÁN GẠO)

KHÓA LUẬN TỐT NGHIỆP ĐẠI HỌC
Ngành Công Nghệ Thông Tin
Mã số ngành: 748201

Cần Thơ, 06/2022

TRƯỜNG ĐẠI HỌC NAM CẦN THƠ
KHOA CÔNG NGHỆ THÔNG TIN

VŨ HOÀNG KHANG

1810375

ỨNG DỤNG TRÍ TUỆ NHÂN TẠO
TRONG VIỆC GIẢI QUYẾT MỘT BỘ DỮ LIỆU
PHÂN LỚP
(MÁY HỌC TĂNG TIẾN
GIẢI QUYẾT BÀI TOÁN CHẨN ĐOÁN GẠO)

KHÓA LUẬN TỐT NGHIỆP ĐẠI HỌC

Ngành Công Nghệ Thông Tin

Mã số ngành: 748201

CÁN BỘ HƯỚNG DẪN
TS. NGÔ HỒ ANH KHÔI

Cần Thơ, 06/2022

CHẤP THUẬN CỦA HỘI ĐỒNG

Khóa luận tốt nghiệp, do sinh viên Vũ Hoàng Khang thực hiện dưới sự hướng dẫn của giảng viên Ngô Hồ Anh Khôi. Khóa luận tốt nghiệp đã báo cáo và được Hội Đồng chấm khóa luận thông qua ngày ... tháng ... năm 2022.

Ủy Viên

Thư ký

TS. Ngô Hồ Anh Khôi

ThS. Trương Hùng Chen

Phản biện 1

Phản biện 2

ThS. Huỳnh Bá Lộc

ThS. Bùi Thị Diễm Trinh

Chủ tịch hội đồng

ThS. Nguyễn Văn Linh

LỜI CẢM ƠN

Trong khoảng thời gian nghiên cứu và tìm hiểu về luận văn “Ứng Dụng Trí Tuệ Nhân Tạo Trong Việc Giải Quyết Một Bộ Dữ Liệu Phân Lớp (Máy Học Tăng Tiến Giải Quyết Bài Toán Chẩn Đoán Gạo)”, em luôn được sự quan tâm, hướng dẫn và giúp đỡ tận tình của các những người thầy, người cô trong Công Nghệ Thông Tin đã giúp em hiểu biết được nhiều vấn đề và trang bị cho em những kiến thức, kỹ năng cần thiết.

Lời đầu tiên em xin được tỏ lòng biết ơn sâu sắc đến các thầy, cô trong Khoa Công Nghệ Thông Tin đã luôn tận tình giúp đỡ, tạo điều kiện thuận lợi để em có thể hoàn thành khóa luận này.

Đặc biệt em xin bày tỏ lòng biết ơn chân thành sâu sắc tới Giảng viên TS. Ngô Hồ Anh Khôi là người đã luôn tận tình giúp đỡ, hướng dẫn em hoàn thành tốt khóa luận này.

Cuối cùng em cũng xin được gửi lời cảm ơn đến những người bạn đã luôn gần gũi, giúp đỡ em hoàn thành thật tốt khóa luận này.

Tuy nhiên trong quá trình nghiên cứu đề tài, do kiến thức về đề tài này còn hạn chế nên em sẽ không tránh khỏi những thiếu sót khi tìm hiểu, đánh giá và trình bày về đề tài. Em rất mong nhận được sự quan tâm, góp ý của các thầy/ cô giảng viên bộ môn, để em có thêm những kinh nghiệm quý báu cho những công việc chuyên môn của em sau này.

Em xin chân thành cảm ơn!

Cần Thơ, ngày.... tháng.... năm 2022

Sinh viên thực hiện

Vũ Hoàng Khang

LỜI CAM ĐOAN

Em xin cam đoan khóa luận “Ứng Dụng Trí Tuệ Nhân Tạo Trong Việc Giải Quyết Một Bộ Dữ Liệu Phân Lớp (Máy Học Tăng Tiến Giải Quyết Bài Toán Chẩn Đoán Gạo)”, được trình bày trong quyển báo cáo này là khóa luận được em tìm hiểu và nghiên cứu độc lập, không có sự sao chép từ người khác. Đây là một sản phẩm em đã nỗ lực nghiên cứu trong quá trình học tập trong học kỳ vừa qua (học kỳ II 2021-2022) và những kết quả mang lại từ đề án này chưa được dùng cho bất kỳ đề án nào khác.

Cần Thơ, ngày... tháng... năm 2022

Sinh viên thực hiện

Vũ Hoàng Khang

NHẬN XÉT CỦA GIẢNG VIÊN PHẢN BIỆN 2

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

Cần Thơ, ngày ...tháng ...năm 2022

Giảng viên phản biện 2

ThS. Bùi Thị Diễm Trinh

MỤC LỤC

DANH MỤC HÌNH.....	i
DANH MỤC TỪ VIẾT TẮT.....	ii
TÓM TẮT	iii
CHƯƠNG 1: TỔNG QUAN ĐỀ TÀI.....	1
1.1 Mục tiêu đề tài.....	1
1.2 Tình hình gạo hiện nay.....	2
1.3 Phương pháp nghiên cứu.....	2
1.4 Phạm vi đề tài.....	2
1.5 Tổng kết	2
CHƯƠNG 2: CƠ SỞ LÝ THUYẾT	4
2.1 Tổng quan về phần mềm.....	4
2.2 Lịch sử các thuật toán máy học tăng tiến trong môi trường biến đổi	6
2.2.1 Sự tiến hóa khái niệm của lớp trong môi trường cố định (môi trường tĩnh)	10
2.2.2 Sự tiến hóa khái niệm của lớp trong môi trường không cố định ..	14
2.2.3 Bổ sung các khái niệm trong môi trường cố định.....	15
2.2.4 Thêm một số khái niệm trong môi trường không cố định	17
2.3 Giới thiệu iSVM gia tăng một lớp (mOC-iSVM).....	18
2.3.1 Phân loại theo SVM một lớp (OC-SVM)	18
2.3.1.1 Giai đoạn huấn luyện.....	19
2.3.1.2 Giai đoạn phân loại	19
2.3.2 Cấu trúc mOC-iSVM	20
2.3.2.1 Nguyên tắc của mOC-iSVM	20
2.3.2.2 Tối ưu hóa thông số.....	22
2.3.2.3 Thuật toán mOC-iSVM.....	22
2.3.2.4 Phiên bản mOC-iSVM.AP (SVM hỗ trợ theo tuổi)	24
2.4 Xây dựng hệ thống dựa trên yêu cầu.....	27
2.5 Những phần mềm dùng trong đề tài.....	30
2.5.1 Visual Studio Code.....	30
2.5.2 Django	31
2.5.3 Scikit-Learn	33

2.6	Tổng kết	34
CHƯƠNG 3: KẾT QUẢ ĐỀ TÀI		35
3.1	Mô tả phương pháp thực nghiệm	35
3.2	Kết quả Model thực nghiệm	39
3.3	Tổng kết.....	42
CHƯƠNG 4: GIAO DIỆN SẢN PHẨM		43
4.1	Giới thiệu về phần mềm	43
4.2	Hướng dẫn cài đặt.....	43
4.3	Giao diện đăng nhập.....	43
4.4	Giao diện chẩn đoán.	45
4.5	Giao diện trang mOC-iSVM.AP	46
4.6	Giao diện trang Cut File	46
4.7	Giao diện tìm số nu và gamma tốt nhất.....	47
4.8	Giao diện Vẽ biểu đồ.....	47
4.9	Giao diện trang test.....	48
CHƯƠNG 5: KẾT LUẬN		49
TÀI LIỆU THAM KHẢO		50

DANH MỤC HÌNH

Hình 1: Bộ dữ liệu thô.....	4
Hình 2: Dữ liệu đã qua xử lý.....	6
Hình 3: Bốn kiểu tiến hóa khái niệm: Sudden concept drift (a), Gradual concept drift (b), Recurring concept drift (c)	9
Hình 4: Huấn luyện với phương pháp Syed.....	11
Hình 5: Minh họa thuật toán OC-SVM.....	20
Hình 6: Phân loại dữ liệu (ký hiệu n) theo mOC-iSVM	21
Hình 7: Sơ đồ quy trình huấn luyện mOC-iSVM	22
Hình 8: Mô hình mOC-iSVM.AP với $\alpha = 1$, α là tuổi	26
Hình 9: Mô hình hệ thống	30
Hình 10: Ma trận dương tính, âm tính	35
Hình 11: Biểu đồ phần trăm trung bình của các thuật toán thực nghiệm theo tuổi (nhóm dữ liệu).....	39
Hình 12: Biểu đồ kết quả Model thực nghiệm theo độ tuổi (nhóm dữ liệu).....	41
Hình 13: Giao diện đăng nhập	44
Hình 14: Giao diện chẩn đoán gao	45
Hình 15: Giao diện chức năng mOC-iSVM.AP	46
Hình 16: Giao diện chức năng Cut File	46
Hình 17: Giao diện Find best gamma và nu.....	47
Hình 18: Giao diện Draw Chart	47
Hình 19: Giao diện trang Test.....	48

DANH MỤC TỪ VIẾT TẮT

AI : Artificial Intelligence

SVM : Support Vector Machines

mOC-iSVM : multi One-Class incremental Support Vector Machines

mOC-iSVM.AP : multi One-Class incremental SVM with Age Pruning

VFDT : Very Fast Decision Tree

ART : Adaptive Resonance Theory

FIS : Fuzzy Inference Systems

NCL : Negative Correlation In Incremental Learning

Fixed Size NC : Fixed Size Negative Correlation

GNCL : Growing Negative Correlation In Incremental Learning

K-NN : K-Nearest Neighbors

RN Radius Neighbors

OC-SVM : One-Class Support Vector Machines

TPR : True Positive Rate

FPR : False Positive Rate

TNR : True Negative Rate

FNR : False Negative Rate

MLP: Multi-layer Perceptron

BernoulliNB : Bernoulli Naive Bayes

LDA : Linear Discriminant Analysis

TÓM TẮT

Hiện nay với sự phát triển mạnh mẽ của khoa học công nghệ, thì sự hoà nhập của các thiết bị công nghệ vào đời sống của từng người đã không phải là chuyện của khoa học viễn tưởng nữa rồi. Mỗi người ở thời đại này đều luôn có bên mình một chiếc smartphone mà mọi thứ đều xuất hiện ở trên đó. Cùng với sự phát triển như giông bão của khoa học công nghệ, thì trí tuệ nhân tạo cũng là một trong những sự phát triển quan trọng của thế giới ngày nay. Trí tuệ nhân tạo là gì? Trí tuệ nhân tạo (AI) là một lĩnh vực nghiên cứu trong đó máy tính được ứng dụng để bắt chước trí thông minh của con người [1]. Máy học là một trường con của AI trong đó các phương pháp tiếp cận thống kê và toán học được áp dụng để cải thiện hiệu suất của máy tính. Trí tuệ nhân tạo (AI) đã đóng góp đáng kể vào việc giải quyết một loạt các vấn đề nông nghiệp trong thập kỷ qua. Học sâu, một lĩnh vực con của AI có tính linh hoạt cao và hỗ trợ trích xuất tính năng tự động, đang ngày càng được ứng dụng nhiều hơn trong các lĩnh vực nghiên cứu nông nghiệp.

Keywords— Ứng Dụng Trí Tuệ Nhân Tạo Trong Việc Giải Quyết Một Bộ Dữ Liệu Phân Lớp, Máy Học Tăng Tiến Giải Quyết Bài Toán Chẩn Đoán Gạo, mOC-iSVM.AP, mOC-iSVM.