

TRƯỜNG ĐẠI HỌC NAM CẦN THƠ
KHOA KỸ THUẬT - CÔNG NGHỆ



PHẠM GIA MINH

Đơn vị thực tập:

**CÔNG TY TNHH MTV CÔNG NGHỆ
KỸ THUẬT TIỀN PHONG**

TÊN ĐỀ TÀI:

**ỨNG DỤNG THUẬT TOÁN NAIVE BAYES
ĐỂ DỰ ĐOÁN BỆNH TIỂU ĐƯỜNG**

ĐỒ ÁN THỰC TẬP

Ngành: Công nghệ thông tin

Mã số ngành: 7480201

Cần Thơ, tháng 05 năm 2021

TRƯỜNG ĐẠI HỌC NAM CẦN THƠ
KHOA KỸ THUẬT - CÔNG NGHỆ



PHẠM GIA MINH
177754

TÊN ĐỀ TÀI:
ỨNG DỤNG THUẬT TOÁN NAIVE BAYES
ĐỂ DỰ ĐOÁN BỆNH TIỂU ĐƯỜNG

ĐỒ ÁN THỰC TẬP
Ngành: Công nghệ thông tin
Mã số ngành: 7480201

CÁN BỘ HƯỚNG DẪN
THS. TRƯƠNG HÙNG CHEN

Cần Thơ, tháng 05 năm 2021

CHẤP THUẬN CỦA HỘI ĐỒNG

Đồ án thực tập “Ứng dụng thuật toán Naive Bayes để dự đoán bệnh tiểu đường”, do sinh viên “Phạm Gia Minh” thực hiện dưới sự hướng dẫn của ThS. Trương Hùng Chen. Đồ án đã báo cáo và được Hội đồng chấm thông qua ngày.....tháng.....năm.....

Ủy viên

Thư ký

GHI CHÚC DANH, HỌ, TÊN

GHI CHÚC DANH, HỌ, TÊN

Phản biện 1

Phản biện 2

GHI CHÚC DANH, HỌ, TÊN

GHI CHÚC DANH, HỌ, TÊN

Cán bộ hướng dẫn

Chủ tịch Hội đồng

GHI CHÚC DANH, HỌ, TÊN

GHI CHÚC DANH, HỌ, TÊN

LỜI CẢM TẠ

Để đồ thực tập này đạt kết quả tốt đẹp, tôi đã nhận được sự hỗ trợ, giúp đỡ của các thầy cô và bạn bè. Với tình cảm sâu sắc, chân thành, cho phép tôi được bày tỏ lòng biết ơn sâu sắc đến tất cả các thầy cô và bạn bè đã tạo điều kiện giúp đỡ trong quá trình học tập và nghiên cứu đề tài.

Trước hết tôi xin gửi tới các thầy cô khoa Công nghệ thông tin trường Đại học Nam Cần Thơ lời chào trân trọng, lời chúc sức khỏe và lời cảm ơn sâu sắc. Với sự quan tâm, dạy dỗ, chỉ bảo tận tình chu đáo của thầy cô, đến nay tôi đã có thể hoàn thành đồ án thực tập với đề tài: “Ứng dụng thuật toán Naive Bayes để dự đoán bệnh tiểu đường”.

Đặc biệt tôi xin gửi lời cảm ơn chân thành nhất đến ThS. Trương Hùng Chen đã quan tâm giúp đỡ, hướng dẫn tận tình tôi hoàn thành tốt đồ án thực tập này trong thời gian vừa qua.

Tôi xin bày tỏ lòng biết ơn đến lãnh đạo Trường Đại học Nam Cần Thơ, các thầy cô khoa công nghệ thông tin đã trực tiếp và gián tiếp giúp đỡ tôi trong suốt quá trình học tập và nghiên cứu đề tài.

Với điều kiện thời gian cũng như kinh nghiệm còn hạn chế, đồ án này không thể tránh được những thiếu sót. Tôi rất mong nhận được sự chỉ bảo, đóng góp ý kiến của các thầy cô để tôi có điều kiện bổ sung, nâng cao ý thức của mình, phục vụ tốt hơn công tác thực tế sau này.

Tôi xin chân thành cảm ơn!

Cần Thơ, ngày 17 tháng 05 năm 2021.

Nhóm thực hiện.

Sinh viên 1

Sinh viên 2

(kí tên và ghi rõ họ tên)

(kí tên và ghi rõ họ tên)

LỜI CAM KẾT

Tôi xin cam kết đồ án thực tập này được hoàn thành dựa trên các kết quả nghiên cứu của tôi trong khuôn khổ của đề tài báo cáo “Ứng dụng thuật toán Naive Bayes để dự đoán bệnh tiểu đường” và các kết quả này chưa được dùng cho bất cứ đồ án cùng cấp nào trước đó.

Cần Thơ, ngày 17 tháng 05 năm 2021.

Nhóm thực hiện.

NHẬN XÉT, ĐÁNH GIÁ CỦA GIẢNG VIÊN



A series of horizontal dotted lines for writing the review.

MỤC LỤC

CHƯƠNG 1: GIỚI THIỆU CÔNG TY	1
1.1 Giới thiệu công ty.....	1
1.2 Trụ sở chính	1
1.3 Người đại diện pháp luật.....	1
1.4 Thông tin liên hệ	1
1.5 Thông tin cơ bản của công ty	1
1.6 Ngành nghề kinh doanh	2
CHƯƠNG 2: NGHIÊN CỨU ĐỀ TÀI	4
3.2 Lý do chọn đề tài.....	4
3.3 Tình hình bệnh tiểu đường hiện nay	4
3.4 Dữ liệu dự đoán bệnh tiểu đường.....	5
CHƯƠNG 3: CƠ SỞ LÝ THUYẾT	7
3.1 Giới thiệu về các kiểu phân loại và một số thuật toán phân loại	7
3.2 Thực nghiệm	8
3.3 Giới thiệu Naïve Bayes	11
3.3.1 Định nghĩa.....	11
3.3.2 Naïve Bayes với Python trong dự đoán bệnh tiểu đường	12
3.3.3 So sánh các kết quả thực nghiệm của Naïve Bayes	15
3.4 Những phần mềm sử dụng trong đề tài	18
3.4.1 Sublime Text 3	18
3.4.2 Qt Designer	18
CHƯƠNG 4: KẾT QUẢ ĐỀ TÀI.....	20
4.1 Kết quả tổng quát	20
4.2 Kết quả train cộng dồn dữ liệu.....	20
4.3 Kết quả train nhóm dữ liệu.....	21
4.4 Kết quả dự đoán bệnh tiểu đường	22
CHƯƠNG 5: HƯỚNG DẪN SỬ DỤNG	23
5.1 Hướng dẫn train dữ liệu	23
5.2 Hướng dẫn dự đoán.....	23
5.3 Mô hình use case.....	24
CHƯƠNG 6: TỔNG KẾT VÀ ĐÁNH GIÁ	25
6.1 Ưu điểm.....	25

6.2	Khuyết điểm	25
6.3	Hướng phát triển	25
	KẾT LUẬN	26
	TÀI LIỆU THAM KHẢO.....	27

DANH SÁCH HÌNH

<i>Hình 1: Ảnh chụp dữ liệu trên Github</i>	<i>5</i>
<i>Hình 2: Ảnh chụp dữ liệu trên excel</i>	<i>6</i>
<i>Hình 3: Biểu đồ độ chính xác Decision Tree</i>	<i>9</i>
<i>Hình 4: Biểu đồ độ chính xác Adaboost</i>	<i>10</i>
<i>Hình 5: Biểu đồ độ chính xác Naïve Bayes</i>	<i>11</i>
<i>Hình 6: Giao diện chính chương trình</i>	<i>20</i>
<i>Hình 7: Kết quả train kiểu cộng dồn dữ liệu</i>	<i>21</i>
<i>Hình 8: Kết quả train kiểu nhóm dữ liệu</i>	<i>22</i>
<i>Hình 9: Kết quả dự đoán bệnh nhân tiểu đường</i>	<i>22</i>
<i>Hình 10: Hướng dẫn train dữ liệu</i>	<i>23</i>
<i>Hình 11: Hướng dẫn dự đoán bệnh tiểu đường</i>	<i>24</i>

DANH SÁCH TỪ VIẾT TẮT

TNHH: Trách nhiệm hữu hạn

MTV: Một thành viên

NB: Naive Bayes

CMD: Command Prompt

KNN: K-Nearest Neighbours

SVM: Support vector machine