

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



**PHẠM HỮU DƯỢC**

**ỨNG DỤNG PHƯƠNG PHÁP DECISION TREE  
ĐỂ NHẬN DIỆN CHỮ SỐ VIẾT TAY**

**ĐỒ ÁN THỰC TẬP**

**Ngành Công nghệ Thông tin**

**Mã số Ngành: 7480201**

05 – 2021

**TRƯỜNG ĐẠI HỌC NAM CẦN THƠ  
KHOA KỸ THUẬT – CÔNG NGHỆ**

**PHẠM HỮU DƯỢC**

**MSSV: 177088**

**ỨNG DỤNG PHƯƠNG PHÁP DECISION TREE  
ĐỂ NHẬN DIỆN CHỮ SỐ VIẾT TAY**

**ĐỒ ÁN THỰC TẬP**

**Ngành Công nghệ Thông tin**

**Mã số Ngành: 7480201**

**GIẢNG VIÊN HƯỚNG DẪN**

**TS. NGÔ HỒ ANH KHÔI**

05-2021

## **CHẤP THUẬN CỦA HỘI ĐỒNG**

Đồ án thực tập cuối khóa “cài đặt giải thuật decision tree để nhận dạng chữ số viết tay” do sinh viên Phạm Hữu Dược thực hiện dưới sự hướng dẫn của TS. Ngô Hồ Anh Khôi. Đồ án thực tập đã báo cáo và được hội đồng chấm đồ án 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**

**NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN**

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Cần Thơ, Ngày.....tháng.....năm 2021

**Giáo viên hướng dẫn**  
(Ký tên)

**TS. Ngô Hồ Anh Khôi**

# **NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN**

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Cần Thơ, Ngày....tháng.....năm 2021

**Giáo viên phản biện**

(Ký tên)

**Th.S Huỳnh Bá lộc**

## LỜI CẢM ƠN

Trong thời gian thực tập cuối khóa(CNTT) lần này, em đã nhận được sự giúp đỡ nhiệt tình từ các thầy cô để em hoàn thành thực tập cuối khóa(CNTT) kịp thời gian đã quy định. Vì thế, cho phép em gửi lời cảm ơn sâu sắc đến các thầy cô giảng viên khoa kỹ thuật – công nghệ trường Đại học Nam Cần Thơ đã dạy bảo và trang bị cho em những kiến thức vô cùng hữu ích để em có cơ sở vững chắc hoàn thành đề án thực tập lần này. Đặc biệt em xin gửi lời chúc sức khỏe và lời cảm ơn chân thành nhất tới giảng viên TS. Ngô Hồ Anh Khôi thầy đã giúp đỡ và chỉ bảo tận tình để từ đó em định hướng được mục tiêu và hoàn thành tốt thực tập cuối khóa(CNTT) lần này.

Mặc dù em đã cố gắng và nỗ lực rất nhiều nhưng do đây là em làm đề án thực tập nên kinh nghiệm là một trở ngại đối với em nên đề án thực tập lần này không tránh được những thiếu sót và hạn chế. Em rất mong nhận được thông cảm, những nhận xét và chỉ bảo lại của thầy cô để em kịp bổ sung kiến thức và cố gắng làm tốt hơn cho công việc sau này.

Em xin chân thành cảm ơn!

Cần Thơ, ngày ..... tháng ..... năm 2021

**Người thực hiện**

## **LỜI CAM ĐOAN**

Tôi xin cam kết khóa luận này được hoàn thành dựa trên các kết quả nghiên cứu của tôi và các kết quả nghiên cứu này chưa được dùng cho bất cứ khóa luận cùng cấp nào khác.

Cần Thơ, ngày ..... tháng ..... năm 2021

**Người thực hiện**

## MỤC LỤC

<i>CHẤP THUẬN CỦA HỘI ĐỒNG</i> .....	<i>i</i>
<i>NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN</i> .....	<i>ii</i>
<i>NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN</i> .....	<i>iii</i>
<i>LỜI CẢM ƠN</i> .....	<i>iv</i>
<i>LỜI CAM ĐOAN</i> .....	<i>v</i>
<i>MỤC LỤC</i> .....	<i>vi</i>
<i>DANH SÁCH BẢNG</i> .....	<i>viii</i>
<i>DANH SÁCH HÌNH</i> .....	<i>ix</i>
<i>CHƯƠNG 1: GIỚI THIỆU NƠI THỰC TẬP</i> .....	<i>1</i>
<i>1.1 Giới thiệu về công ty</i> .....	<i>1</i>
<i>1.2 Thông tin về công ty</i> .....	<i>1</i>
<i>1.3 Trụ sở chính</i> .....	<i>1</i>
<i>1.4 Người đại diện theo pháp luật:</i> .....	<i>1</i>
<i>1.5 Thông tin liên hệ</i> .....	<i>1</i>
<i>CHƯƠNG 2: GIỚI THIỆU</i> .....	<i>2</i>
<i>2.1 Đặt vấn đề nghiên cứu</i> .....	<i>2</i>
<i>2.2 Mục tiêu nghiên cứu</i> .....	<i>3</i>
<i>2.3 Giới thiệu về bộ cơ sở dữ liệu</i> .....	<i>3</i>
<i>2.4 Phương pháp nghiên cứu</i> .....	<i>5</i>
<i>2.4.1 Phương pháp nghiên cứu lý thuyết</i> .....	<i>5</i>
<i>2.4.2 Phương pháp nghiên cứu thực nghiệm</i> .....	<i>5</i>
<i>2.4.3 Phương pháp điều tra</i> .....	<i>5</i>
<i>CHƯƠNG 3: CƠ SỞ LÝ LUẬN</i> .....	<i>6</i>
<i>3.1 Cơ sở lý luận</i> .....	<i>6</i>
<i>3.2 Giới thiệu về giải thuật Decision Tree</i> .....	<i>6</i>
<i>3.2.1 Giới thiệu chung</i> .....	<i>6</i>
<i>3.2.2 Cây quyết định C4.5</i> .....	<i>9</i>
<i>3.2.3 Hàm số entropy</i> .....	<i>11</i>
<i>3.2.4 Thuật toán trong C4.5</i> .....	<i>12</i>
<i>3.2.5 Điều kiện dừng</i> .....	<i>16</i>



3.2.6 Pruning.....	16
3.2.7 Tri thức định dạng.....	17
3.2.8 Lập trình Python cho C4.5.....	17
3.3 Giới thiệu về ngôn ngữ Python .....	17
<b>CHƯƠNG 4: GIẢI THUẬT DECISION TREE TRONG NHẬN DẠNG CHỮ SỐ VIẾT TAY.....</b>	<b>20</b>
4.1 Phương pháp nhận dạng Decision Tree .....	20
4.2 Quá trình nhận dạng chữ số viết tay.....	21
4.2.1 Đưa ảnh vào.....	21
4.2.2 Tiền xử lý.....	21
4.2.3 Sử dụng Decision Tree để nhận dạng.....	21
4.3 Sơ đồ Use case chương trình .....	22
<b>CHƯƠNG 5: THỰC NGHIỆM VÀ KẾT QUẢ.....</b>	<b>23</b>
5.1 Kết quả nghiên cứu .....	23
5.2 Giao diện.....	24
5.3 Hướng dẫn sử dụng.....	25
5.4 Hướng dẫn cài đặt.....	36
<b>CHƯƠNG 6: KẾT LUẬN.....</b>	<b>37</b>

## DANH SÁCH BẢNG

<i>Bảng 5.1: Bảng Parameter</i> .....	23
<i>Bảng 5.2: Bảng so sánh các giải thuật</i> .....	24

## DANH SÁCH HÌNH

<i>Hình 2.1: Giới thiệu bộ dữ liệu mnist.....</i>	<i>4</i>
<i>Hình 2.2: Ảnh về pixel trong mnist.....</i>	<i>4</i>
<i>Hình 3.1: Các nút của cây nhị phân.....</i>	<i>9</i>
<i>Hình 3.2: Ước lượng trên cây quyết định .....</i>	<i>10</i>
<i>Hình 3.3: Biểu đồ .....</i>	<i>12</i>
<i>Hình 3.4: Mô tả cách tính information gain .....</i>	<i>14</i>
<i>Hình 4.1: Sơ đồ use case chương trình .....</i>	<i>22</i>
<i>Hình 5.1: Giao diện kết quả models test .....</i>	<i>23</i>
<i>Hình 5.2: Giao diện của chương trình .....</i>	<i>24</i>
<i>Hình 5.3: Sơ đồ Use case hướng dẫn sử dụng .....</i>	<i>25</i>
<i>Hình 5.4: Giao diện dùng để train .....</i>	<i>26</i>
<i>Hình 5.5: Giao diện sau khi train.....</i>	<i>27</i>
<i>Hình 5.6: Giao diện test .....</i>	<i>28</i>
<i>Hình 5.7: Giao diện test file model .....</i>	<i>29</i>
<i>Hình 5.8: Giao diện tỷ lệ % đúng của list models.....</i>	<i>30</i>
<i>Hình 5.9: Giao diện report của models.....</i>	<i>31</i>
<i>Hình 5.10: Giao diện test một model .....</i>	<i>32</i>
<i>Hình 5.11: Giao diện sau khi test 1 model.....</i>	<i>33</i>
<i>Hình 5.12: Giao diện biểu đồ report test một model .....</i>	<i>34</i>
<i>Hình 5.13: Giao diện nhận dạng chữ số viết tay.....</i>	<i>35</i>
<i>Hình 5.14: Sơ đồ Use case hướng dẫn cài đặt.....</i>	<i>36</i>

## **DANH MỤC TỪ VIẾT TẮT**

DT	Decision Tree
CNTT	Công nghệ thông tin
MNIST	Modified National Institute of Standards and Technology