TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN CUỐI KÌ**

**XỬ LÝ DỮ LIỆU LỚN**

**Phát hiện tin tức giả tiếng Việt**

*Người hướng dẫn*: **TS. BÙI THANH HÙNG**

*Người thực hiện*: **NGUYỄN HOÀNG MINH THƯ – 518H0061**

**MẠC THUẬN ĐẠT – 518H0606**

*Nhóm:* **17**

*Khóa:* **22**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2021**

**ĐỒ ÁN ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Chúng tôi xin cam đoan đây là sản phẩm đồ án của riêng chúng tôi và được thực hiện dưới sự hướng dẫn của TS. Bùi Thanh Hùng. 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 đồ án 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 chúng tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình.** Trường đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do chúng tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày tháng năm*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

*Nguyễn Hoàng Minh Thư*

*Mạc Thuận Đạt*

# LỜI CẢM ƠN

Đây là phần tác giả tự viết ngắn gọn, thể hiện sự biết ơn của mình đối với những người đã giúp mình hoàn thành Đồ án. Tuyệt đối không sao chép theo mẫu những “lời cảm ơn” đã có.

# 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)

# MỤC LỤC

[LỜI CẢM ƠN 3](#_Toc92200337)

[PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN 4](#_Toc92200338)

[MỤC LỤC 5](#_Toc92200339)

[DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT 6](#_Toc92200340)

[DANH MỤC CÁC HÌNH VẼ 7](#_Toc92200341)

[DANH MỤC CÁC BẢNG 8](#_Toc92200342)

# DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT

**CÁC KÝ HIỆU**

*f Tần số của dòng điện và điện áp (Hz)*

*p Mật độ điện tích khối (C/m3)*

**CÁC CHỮ VIẾT TẮT**

CSTD Công suất tác dụng

MF Máy phát điện

# DANH MỤC CÁC HÌNH VẼ

**Ẽ**[Hình 1. Cây cấu trúc HTML 5](#_Toc87559285)

[Hình 2. Biểu đồ thống kê số lượng câu nói nổi tiếng 6](#_Toc87559286)

[Hình 3. Thống kê độ tuổi tác giả 7](#_Toc87559287)

# DANH MỤC CÁC BẢNG

[Bảng 1. Mô tả về bộ dữ liệu 8](#_Toc87559298)

PHÁT HIỆN TIN TỨC GIẢ TIẾNG VIỆT

## Giới thiệu về bài toán

Bài toán này là bài toán gì

Bài toán này có ý nghĩa gì?

## Phân tích yêu cầu của bài toán

1.2.1 Yêu cầu của bài toán

Bài toán này có các yêu cầu gì, trình bày cụ thể các yêu cầu đó

1.2.2 Các phương pháp giải quyết bài toán

Các phương pháp giải quyết bài toán này như thế nào? Trình bày được cụ thể các phương pháp giải quyết bài toán, mỗi phương pháp giải quyết bài toán đều phải nêu được tên của bài báo có liên quan (tiêu đề bài báo, tác giả, năm), phương pháp giải quyết của bài báo đó, dữ liệu thực nghiệm, kết quả đạt được và các hạn chế nếu có

Gợi ý: Nên tìm kiếm các bài báo survey để có cái nhìn tổng quát về vấn đề cần giải quyết và các phương pháp giải quyết bài toán

1.2.3 Phương pháp đề xuất giải quyết bài toán

Đưa ra nhận định tổng quát về hướng giải quyết bài toán và lý giải tại sao lựa chọn hướng giải quyết này.

1.3 Phương pháp giải quyết bài toán

1.3.1 Mô hình tổng quát

Vẽ được mô hình tổng quát giải quyết bài toán và trình bày tổng quát về từng phần trong mô hình này

1.3.2 Đặc trưng của mô hình đề xuất

Trình bày cụ thể chi tiết từng thành phần trong mô hình đề xuất trên.

1.4 Thực nghiệm

1.4.1 Dữ liệu

Dữ liệu được lấy ở đâu, từ nguồn nào

Dữ liệu gồm có những gì, mô tả chi tiết về dữ liệu

1.4.2 Xử lý dữ liệu

Có phải tiền xử lý dữ liệu không? Tại sao

Xử lý dữ liệu như thế nào?

1.4.3 Công nghệ sử dụng

(Phần này trình bày không quá 3-6 dòng)

Trình bày vắn tắt về tên của Ngôn ngữ lập trình sử dụng ,

Các thư viện sử dụng để hiện thực bài toán,

Các công cụ sử dụng giải quyết bài toán

1.4.4 Cách đánh giá

Cách đánh giá bằng độ đo gì, độ đo đó được tính bằng công thức nào?

**1.5 Kết quả đạt được**

Trình bày các tham số cụ thể hóa các thực nghiệm: ví dụ như số epoch, số tầng, số lớp, tham số x, …

Trình bày các kết quả đạt được theo các độ đo ở trên và trên các bộ dữ liệu hay các phương pháp học máy, học sâu khác nhau.

So sánh với ít nhất 1 phương pháp khác, Vẽ được biểu đồ so sánh giữa các phương pháp theo các độ đo đã trình bày ở trên ví dụ như: Accuracy, MSE, RMSE, MAP, …. (hãy lựa chọn ít nhất 2 độ đo trong các độ đo phổ biến để đánh giá bài toán trên)

Giải thích tại sao đạt được kết quả đó, và kết quả khác nhau giữa các phương pháp trên là nguyên nhân nào?

1.6 Kết luận

Trình bày tóm tắt các kết quả đạt được của bài toán trên,

Trình bày được hạn chế của phương pháp giải quyết bài toán

Trình bày hướng phát triển trong tương lai

**TÀI LIỆU THAM KHẢO**

**Chú ý các tài liệu tham khảo phải được chỉ mục vào trong các Chương nếu có tham khảo, phải có ít nhất từ 7-15 tài liệu tham khảo chính thống: Sách, bài báo Hội nghị, tạp chí - journal, … và tài liệu này phải được đánh chỉ mục từ nhỏ tới lớn và gắn vào trong báo cáo.**

**Các hình ảnh tham khảo hay các phần tham khảo ít thì có thể đánh thành footnote ngay trong trang tham khảo**

1. Quách Ngọc Ân (1992), “Nhìn lại hai năm phát triển lúa lai”, *Di tuyền học ứng dụng*, 98(1), tr. 10-16.
2. Bộ nông nghiệp & PTNT (1996), *Báo cáo tổng kết 5 năm (1992-1996) phát triển lúa lai,* Hà Nội.
3. Nguyễn Hữu Đống, Đào Thanh Bằng, Lâm Quang Dụ, Phan Đức Trực (1997), *Đột biến –* *Cơ sở lý luận và ứng dụng,* Nhà xuất bản nông nghiệp, Viện khoa học kỹ thuật nông nghiệp Việt Nam, Hà Nội.
4. Nguyễn Thị Gấm (1996), *Phát hiện và đánh giá một số dòng bất dục đực cảm ứng nhiệt* *độ,* Luận văn thạc sĩ khoa học nông nghiệp, Viện khoa học kỹ thuật nông nghiệp Việt Nam, Hà Nội.

……….

1. Anderson J.E. (1985), The Relative Inefficiency of Quota, The Cheese Case, *American* *Economic Review*, 75(1), pp. 178-90.
2. Borkakati R. P.,Virmani S. S. (1997), Genetics of thermosensitive genic male sterility in Rice, *Euphytica* 88, pp. 1-7.
3. Boulding K.E. (1955), *Economics Analysis*, Hamish Hamilton, London.
4. Burton G. W. (1988), “Cytoplasmic male-sterility in pearl millet (penni-setum glaucum L.)”, *Agronomic Journal* 50, pp. 230-231.
5. Central Statistical Oraganisation (1995), *Statistical Year Book*, Beijing.
6. FAO (1971), *Agricultural Commodity Projections (1970-1980)*, Vol. II. Rome.
7. Institute of Economics (1988), *Analysis of Expenditure Pattern of Urban Households in* *Vietnam,* Departement pf Economics, Economic Research Report, Hanoi.

**PHỤ LỤC**

Phần này bao gồm những nội dung cần thiết nhằm minh họa hoặc hỗ trợ cho nội dung đồ án như số liệu, biểu mẫu, tranh ảnh. . . . nếu sử dụng những câu trả lời cho một *bảng câu hỏi thì bảng câu hỏi mẫu này phải được đưa vào phần Phụ lục ở dạng nguyên bản* đã dùng để điều tra, thăm dò ý kiến; **không được tóm tắt hoặc sửa đổi**. Các tính toán mẫu trình bày tóm tắt trong các biểu mẫu cũng cần nêu trong Phụ lục của luận văn. Phụ lục không được dày hơn phần chính của đồ án

**MỘT SỐ CHÚ Ý KHI VIẾT BÁO CÁO**

1. Thống nhất kích cỡ chữ, kiểu chữ trong toàn bộ báo cáo. Không tô màu chữ, chỉ dùng màu đen
2. Các công thức phải tự gõ và đánh số theo Chương, ví dụ 1.1, 2.1, 2.2, 2.3
3. Các hình và Bảng phải đánh số theo chương, ví dụ Hình 1.1, Hình 2.1, Bảng 3.1, Bảng 3.2
4. Các hình nếu lấy ở ngoài phải đề footnote chú thích nguồn ở dưới
5. Hình mô hình tổng quát phải tự vẽ bằng Word, không dán hình
6. Các tài liệu tham khảo phải đính vào luận văn theo thứ tự từ nhỏ tới lớn, bắt đầu từ 1, ít nhất phải từ 7-15 tài liệu tham khảo, lựa chọn các tài liệu tham khảo mới
7. Tóm tắt trình bày được các nội dung sau: giới thiệu, phương pháp làm, kết quả, nhận xét (không dùng hình, bảng ở mục này)

PHÁT HIỆN TIN TỨC GIẢ TIẾNG VIỆT

BÀI TOÁN

Phân tích yêu cầu bài toán

Phân tích yêu cầu bài toán: Phân tích được yêu cầu của bài toán là gì (0.5 điểm)

Phương pháp giải quyết bài toán

Phương pháp giải quyết: Trình bày được các phương pháp giải quyết bài toán. Giải thích lý do tại sao chọn phương pháp này, Vẽ được sơ đồ tổng quát giải quyết bài toán (2 điểm)

THU THẬP DỮ LIỆU

XỬ LÝ DỮ LIỆU

MÔ HÌNH SỬ DỤNG

ĐÁNH GIÁ KẾT QUẢ

HƯỚNG PHÁT TRIỂN

TỰ ĐÁNH GIÁ

(Với nhóm có 2 thành viên)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Câu | Nội dung | Điểm chuẩn | Tự chấm | Ghi chú |
| 1  (8.5) | 1.1 Giới thiệu về bài toán | 0.5 |  |  |
| 1.2 Phân tích yêu cầu của bài toán | 1.0 |  |  |
| 1.3 Phương pháp giải quyết bài toán | 1.5 |  |  |
| 1.4 Thực nghiệm | 4 |  |  |
| **1.5 Kết quả đạt được** | 1 |  |  |
| 1.6 Kết luận | 0.5 |  |  |
| 2 | Điểm nhóm | 0.5đ |  |  |
| 3 | Báo cáo (chú ý các chú ý 2,3,4,6 ở trang trước, nếu sai sẽ bị trừ điểm nặng) | 1đ |  |  |
| Tổng điểm | | |  |  |

TỰ ĐÁNH GIÁ

(Với nhóm có 1 thành viên)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Câu | Nội dung | Điểm chuẩn | Tự chấm | Ghi chú |
| 1  (9 đ) | 1.1 Giới thiệu về bài toán | 0.5 |  |  |
| 1.2 Phân tích yêu cầu của bài toán | 1.0 |  |  |
| 1.3 Phương pháp giải quyết bài toán | 1.5 |  |  |
| 1.4 Thực nghiệm | 4.5 |  |  |
| **1.5 Kết quả đạt được** | 1 |  |  |
| 1.6 Kết luận | 0.5 |  |  |
| 2 | Báo cáo (chú ý các chú ý 2,3,4,6 ở trang trước, nếu sai sẽ bị trừ điểm nặng) | 1đ |  |  |
| Tổng điểm | | |  |  |