BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC CẦN THƠ**

**TRƯỜNG CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

**🙞 🕮 🙜**



**BÁO CÁO**

**KHAI KHOÁNG DỮ LIỆU(CT31202)**

**Đề tài**

**MÔ HÌNH DỰ ĐOÁN PHÂN LOẠI GẠO**

**NHÓM BÁO CÁO**

|  |  |
| --- | --- |
| **MSSV** | **HỌ TÊN** |
| **B2106839** | **Cao Hoàng Khải** |
| **B2113345** | **Bùi Thị Quyền Trân** |

**HK 1, NH 2024 – 2025**

Cần Thơ, 11/2024

**NHẬN XÉT CỦA GIẢNG VIÊN**

**---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------**

**LỜI CẢM ƠN**

|  |
| --- |
| Cần Thơ, ngày 29 tháng 11 năm 2024 |
| Người viết |
| Bùi Thị Quyền Trân, Cao Hoàng Khải |

**MỤC LỤC**

**DANH MỤC HÌNH**

**DANH MỤC BẢNG**

**TÓM TẮT**

# PHẦN I: PHẦN GIỚI THIỆU

## 1. Đặt vấn đề

## 2. Mục tiêu đề tài

## 3. Phạm vi nghiên cứu

## 4. Phương pháp nghiên cứu

## 5. Nội dung nghiên cứu

## 6. Kết quả đạt được

## 7. Bố cục bài báo cáo

# PHẦN NỘI DUNG

# CHƯƠNG 1: MÔ TẢ BÀI TOÁN

## 1. Mô tả chi tiết bài toán

## 2. Vấn đề và giải pháp liên quan đến bài toán

### 2.1 Phân tích trực quan

### 2.2 Mô hình phân lớp

# CHƯƠNG 2: TRỰC QUAN HÓA, XỬ LÝ TẬP DỮ LIỆU

## 1. Trực quan hóa tập dữ liệu

### 1.1 Tổng quan về tập dữ liệu

- Dữ liệu thu nhập từ:

<https://github.com/ltdaovn/dataset/raw/master/Rice2024.xlsx>

- Tập dữ liệu chứa 3810 dữ liệu về giống gạo và 10 cột thuộc tính là “Id”, “Nickname”, “Class”, “Area”, “Perimeter”, “Major\_Axis\_Lenght”, “Minor\_Axis\_Length”, “Eccentricity”, “Convex\_Area” và “Extent”.

* Cột “Id” là thứ tự về gạo.
* Cột tiếp theo là “Nickname” ( Tên ): tên gạo
* Cột “Class” ( Lớp ): loại của gạo ( có 2 lớp C và O )
* Cột “Area” ( Diện tích ): Trả về trọng lượng điểm ảnh trong ranh giới của hạt gạo.
* Cột “Perimeter” ( Chu vi ): Chu vi khoảng cách giữa các điểm ảnh quanh ranh giới hạt gạo.
* Cột “Major\_Axis\_Length” ( Độ dài trục lớn ): Đường dài nhất có thể vẽ trên hạt gạo (khoảng cách trục chính).
* Cột “Minor\_Axis\_Length” ( Độ dài trục nhỏ ): Đường dài nhất có thể vẽ trên hạt gạo (khoảng cách trục phụ).
* Cột “Eccentricity” ( Độ lệch tâm ): Đo độ tròn của hình ellipse tương ứng với hạt gạo
* Cột “Convex\_Area” ( Diện tích lồi ): Số lượng điểm ảnh của vỏ lồi nhỏ nhất bao quanh hạt gạo.
* Cột “Extent” ( Tỷ lệ ): tỷ lệ giữa vùng hạt gạo và các điểm của hộp giới hạn.

### 1.2 Trực quan hóa tập dữ liệu

Hình 1. Tổng quan về dữ liệu