# 1. Giới thiệu

## 1.1 Lý do chọn đề tài

Nhận dạng là bài toán xuất hiện cách đây khá lâu và vẫn luôn thu hút được nhiều sự quan tâm, nghiên cứu. Đặc biệt là trong vài thập niên gần đây, do sự thúc đẩy của quá trình tin học hoá trong mọi lĩnh vực, bài toán nhận dạng không còn dừng lại ở mức độ nghiên cứu nữa mà nó trở thành một lĩnh vực đáp dụng vào thực tế.

Cùng với mong muốn mở rổng kiến thức về lĩnh vực này vì vậy nhóm đã chọn đề tài nhận dạng ký tự trong ảnh làm đề tài để thực hiện.

## 1.2 Lợi ích của đề tài

Hiểu được Machine Learning là gì.

Áp dụng Machine Learning vào thực tế.

Cụ thể hơn trong đề tài này giúp ta hiểu hơn về thuật toán k-Nearest Neighbor.

# 2. Mục tiêu

Nhận dạng được các ký tự trong ảnh.

Ví dụ: 

Output: Based

# 3. Lý thuyết

## 3.1 Xử lý ảnh

### **3.1.1 Định nghĩa**

Xử lý hình ảnh là việc chọn lọc những thông tin mong muốn từ những bức ảnh.

Để chọn lọc những thông tin mong muốn được chứa trong bức ảnh, phương pháp cơ bản được dùng trong xử lý hình ảnh là loại bỏ những phần thông tin ảnh không mong muốn.

Xử lý ảnh gồm 4 lĩnh vực chính: xử lý nâng cao chất lựong ảnh, nhận dạng ảnh, nén ảnh và truy vấn ảnh.

### **3.1.2 Các bước xử lý ảnh cơ bản**

Bước 1: Nhập hình ảnh bằng máy quét quang học hoặc bằng cách chụp ảnh kỹ thuật số.

Bước 2: Phân tích và xử lý hình ảnh, bao gồm nén dữ liệu và tăng cường chi tiết ảnh, không giống như chế độ nhận diện hình ảnh vệ tinh của mắt người.

Bước 3: Kết quả là ảnh đầu ra có thể thay đổi dựa trên phân tích hình ảnh.

## 3.2 Thuật toán k-Nearest Neighbor

### **3.2.1 k-Nearest Neighbor là gì?**

Là một trong những thuật toán supervised-learning đơn giản nhất trong Machine Learning.

Ý tưởng của KNN là tìm ra output của dữ kiệu dựa trên thông tin của những dữ liệu training gần nó nhất.

### **3.2.2 Các thức làm việc của kNN**

Đầu tiên chúng ta phải chuẩn bị một tập huấn luyện (training set) mà tất cả các dữ liệu trong tập đó đều biết trước được thuộc loại nào.

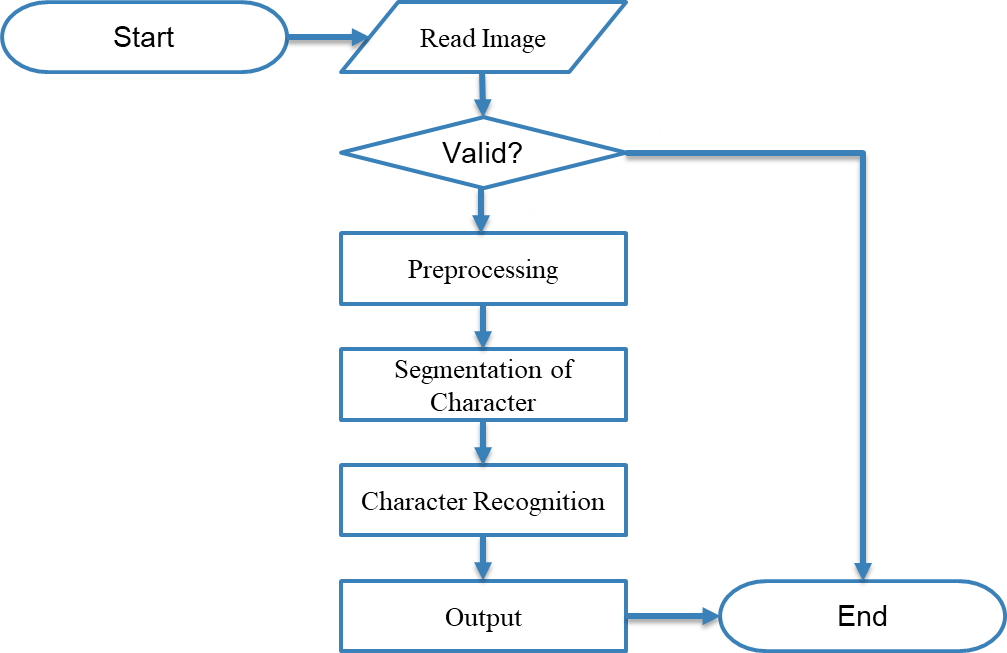
Người dùng sẽ đưa vào một dữ liệu chưa biết được thuộc loại nào.

kNN sẽ so sánh dữ liệu đó với tất cả dữ liệu trọng tập huấn luyện và chọn ra k dữ liệu gần giống nhất.

Trong k dữ liệu đó, kNN sẽ xem xét xem loại nào là loại chiếm đa số.

Đưa ra kết luận rằng tập dữ liệu cần xác định thuộc loại đó.

# 4. Flowchart



# 5. Block Diagram

