**Tên bài TH: Dự đoán khả năn bị béo phì dựa trên thuật toán KNN**

**Họ tên sinh viên**: Nguyễn Minh Tiến

**Mã số SV**: 2018606457

**Lớp**: KTMT2-K13

**Ngày TH**: 22/11/2021

**1. Mô tả nhiệm vụ bài TH**

*1.1 Giới thiệu ngôn ngữ lập trình*

Python là một ngôn ngữ lập trình thông dịch (interpreted), hướng đối tượng (object-oriented), và là một ngôn ngữ bậc cao (high-level) ngữ nghĩa động (dynamic semantics). Python hỗ trợ các module và gói (packages), khuyến khích chương trình module hóa và tái sử dụng mã. Trình thông dịch Python và thư viện chuẩn mở rộng có sẵn dưới dạng mã nguồn hoặc dạng nhị phân miễn phí cho tất cả các nền tảng chính và có thể được phân phối tự do.

*1.2 Chức năng ứng dụng*.

Phân cụm dựa trên các thuộc tính dữ liệu

**2. Xây dựng một chương trình ứng dụng**

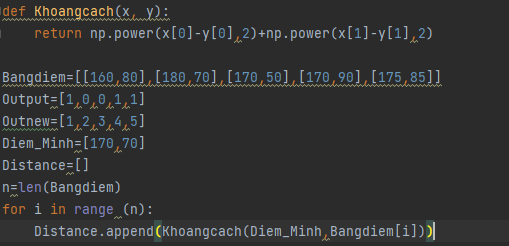
*2.1 Ý tưởng và cách thực hiện*

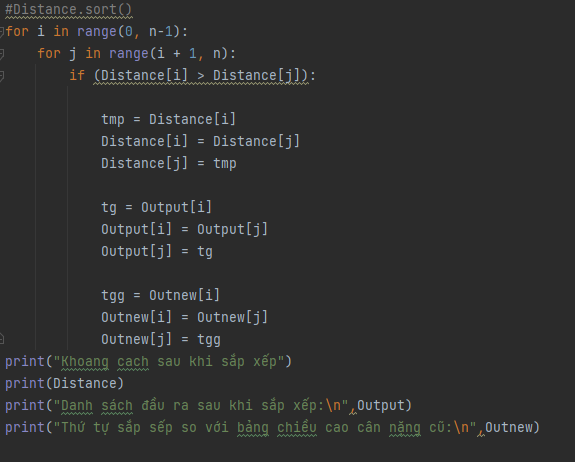
-dữ liệu đầu vào nhỏ

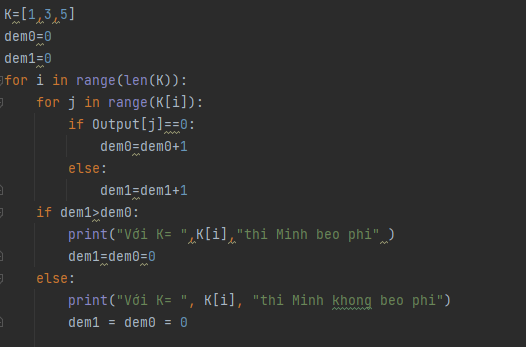
-chọn k =1,3,5

-tính toán khoảng cách và phân loại

*2.2 Chương trình và cách hoạt động*







Cách hoạt động:

-Sử dụng 1 mảng chứa dữ liệu đầu vào và 1 mảng chứa nhãn

-tính khoảng cách từ điểm cần phân biệt đến các mẫu

-xác định K

-phân lớp

*2.3 Các hàm trong chương trình và các diễn giải quan trọng cho chức năng các hàm sử dụng*

a. dữ liệu đầu vào

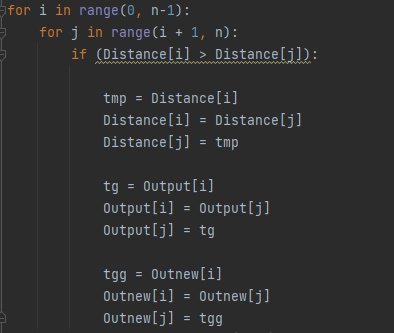
  
b. dữ liệu phân loại



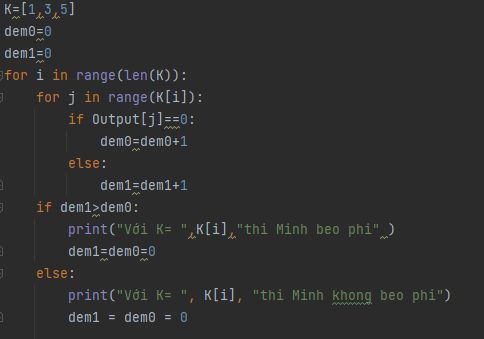
c. tính khoảng cách từ điểm truy vấn đến các mẫu



d. sắp xếp các mẫu theo khoảng cách



e. phân loại với k khác nhau

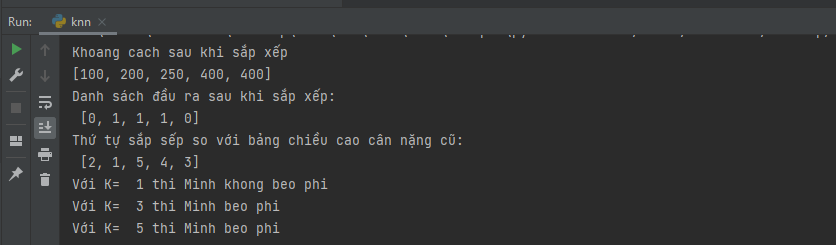


**3. Phân tích kết quả.**

-dữ liệu đầu vào đơn giảm gồm 5 mẫu , 2 thuộc tính và 2 nhãn

-sau khi tính khoảng cách từ điểm truy vẫn đến các mẫu ta sắp xếp lại thứ tự các mẫu và các nhãn theo khoảng cách

-đưa ra kết quả với k khác nhau do mẫu dữ liệu đầu vào ít .



Ưu điểm :

-quá trình tính toán đơn giản

-dễ dàng dự đoán kết quả đầu ra

-Không cần giả sử gì về phân phối của các class..

Nhược điểm:

-thuật toán phụ thuộc nhiều vào hệ số K, nếu chọn K không hợp lý sẽ ảnh hưởng đến hiệu suất phân lớp

-độ phức tạp của bài toán tăng lên theo số K