**ĐẠI HỌC SƯ PHẠM KỸ THUẬT THÀNH PHỐ HỒ CHÍ MINH**

**KHOA ĐÀO TẠO CHẤT LƯỢNG CAO**

**NGÀNH CÔNG NGHỆ THÔNG TIN**

---------------o0o---------------



**ĐỒ ÁN 1**

**TÌM HIỂU THUẬT TOÁN K-NEAREST NEIGHBOR**

GVHD: TS.Lê Văn Vinh

SVTH: Nguyễn Tiến Đat 16110048

Nguyễn Văn Đạt 16110050

Hồ Chí Minh, tháng 10 năm 2018

Nhận xét giáo viên :………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………….………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………….………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………………………………………………

# LỜI CẢM ƠN

Nay em xin được phép gửi lời cảm ơn chân thành này đến thầy Lê Văn Vinh, người đã trực tiếp hỗ trợ em trong suốt quá trình định hướng chọn đề tài, hướng dẫn, nhận xét và góp ý cũng như cung cấp tài liệu tham khảo. Nếu không có những lời hướng dẫn, những kinh nghiệm thực tiễn của thầy thì em nghĩ rằng bài thu hoạch này sẽ khó có thể hoàn thiện, em xin cảm ơn thầy. Bài thu hoạch được thực hiện trong khoảng thời gian gần 4 tuần. Khoảng thời gian có hạn, cùng với kiến thức còn hạn chế và còn nhiều bỡ ngỡ khác do đó thiếu sót là điều không thể tránh khỏi nên em rất mong nhận được những ý kiến đóng góp quý báo của các quý Thầy Cô để kiến thức của em được hoàn thiện hơn sau này. Em xin chân thành cảm ơn.

Thành phố Hồ Chí Minh, ngày tháng 10 năm 2018

Sinh viên thực hiện

(ký và ghi họ tên)

……………………..

**MỤC LỤC**

***PHẦN 1: TÌM HIỂU THUẬT TOÁN K-NEAREST NEIGHBOR***

# *GIỚI THIỆU THUẬT TOÁN K-NEAREST NEIGHBOR*

# *Ý NGHĨA CỦA THUẬT TOÁN K-NEAREST NEIGHBOR*

1. ***VÍ DỤ MINH HỌA***

***PHẦN 2: PROJECT***

1. **THIẾT KẾ GIAO DIỆN**
2. **CODE**

***PHẦN 3: ĐÁNH GIÁ VÀ PHÂN CÔNG***

1. **PHÂN CÔNG**
2. **ĐÁNH GIÁ**
3. **KHUYẾT ĐIỂM**

***PHẦN 1: TÌM HIỂU THUẬT TOÁN K-NEAREST NEIGHBOR***

# *GIỚI THIỆU THUẬT TOÁN K-NEAREST NEIGHBOR*

K-Nearest Neighbors algorithm (K-NN) được sử dụng rất phổ biến trong lĩnh vực Data Mining. K-NN là phương pháp để phân lớp các đối tượng dựa vào khoảng cách gần nhất giữa đối tượng cần xếp lớp (Query point) và tất cả các đối tượng trong Training Data.

Một đối tượng được phân lớp dựa vào K láng giềng của nó. K là số nguyên dương được xác định trước khi thực hiện thuật toán. Người ta thường dùng khoảng cách Euclidean để tính khoảng cách giữa các đối tượng.

K-nearest neighbor là một trong những thuật toán supervised-learning đơn giản nhất (mà hiệu quả trong một vài trường hợp) trong Machine Learning. Khi training, thuật toán này không học một điều gì từ dữ liệu training (đây cũng là lý do thuật toán này được xếp vào loại lazy learning), mọi tính toán được thực hiện khi nó cần dự đoán kết quả của dữ liệu mới. K-nearest neighbor có thể áp dụng được vào cả hai loại của bài toán Supervised learning là Classification và Regression. KNN còn được gọi là một thuật toán Instance-based hay Memory-based learning.

1. ***MÔ TẢ THUẬT TOÁN K-NEAREST NEIGHBOR***

**Thuật toán K-NN được mô tả như sau:**

1.    Xác định giá trị tham số K (số láng giềng gần nhất)

2.    Tính khoảng cách giữa đối tượng cần phân lớp (Query Point) với tất cả các đối tượng trong training data (thường sử dụng khoảng các Euclidean)

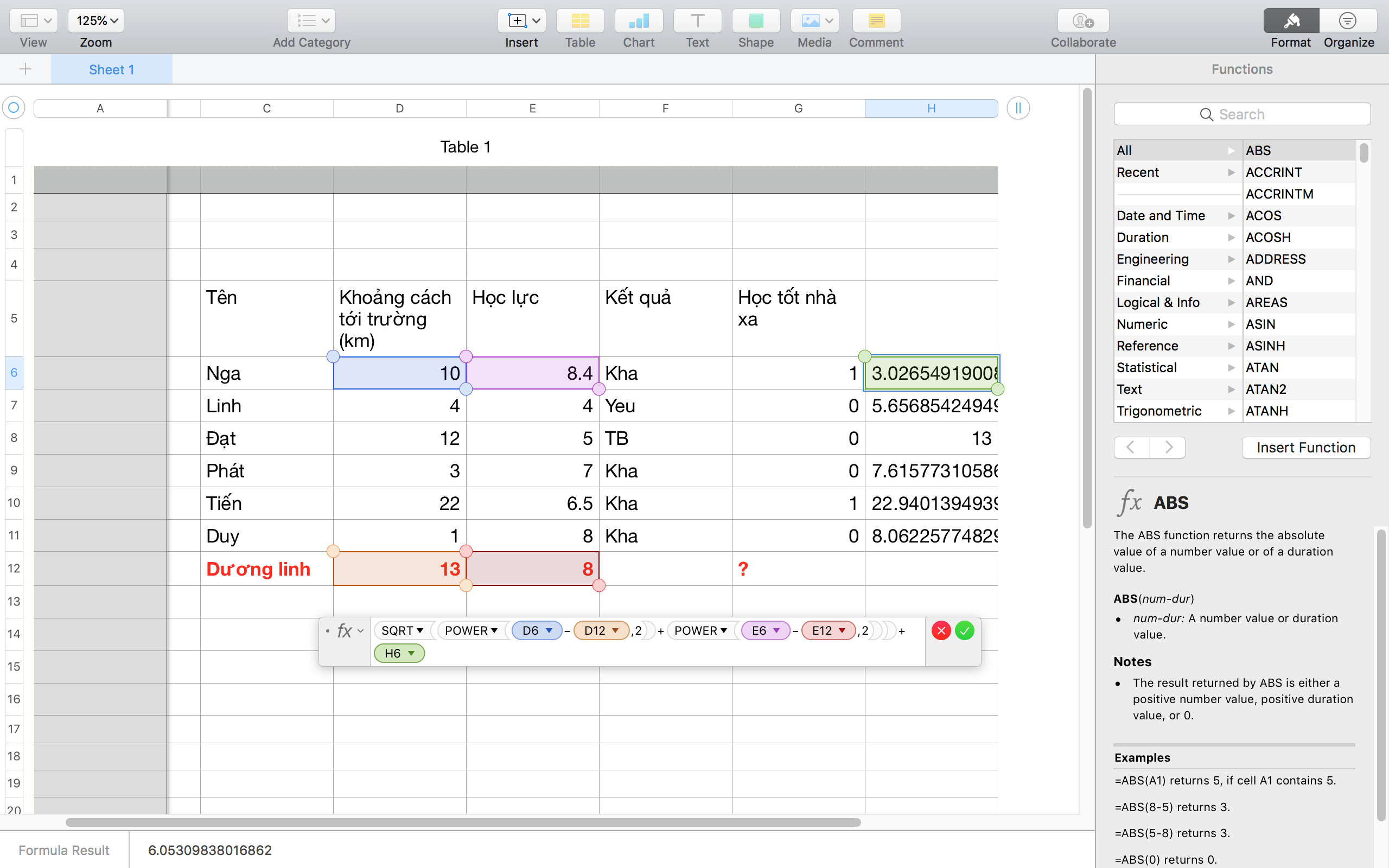
3.    Sắp xếp khoảng cách theo thứ tự tăng dần và xác định K láng giềng gần nhất với Query Point

4.    Lấy tất cả các lớp của K láng giềng gần nhất đã xác định

5.    Dựa vào phần lớn lớp của láng giềng gần nhất để xác định lớp cho Query Point

1. ***VÍ DỤ MINH HỌA***

Ta có bảng sau có số liệu về học lực và khoảng cách từ nhà đến trường nên ta dễ dàng tính được học lực và bạn ấy có khó khăn trong học tập hây không khi cho 1 bạn mới là bạn Dương Linh vào thì ta sẽ từ dữ liệu sẵn có tính ra được 1 chỉ số bên cột cuối cùng dùng công thức Euclidean và ta có thể dự đoán khi bạn Nga có chỉ số thấp nhất gần Dương Linh nhất và bạn ấy là người gặp khó khăn trong học tập vậy ta cũng có thể nhận định rằng Dương Linh cũng gặp khó khăn trong học tập.



***PHẦN 2: PROJECT***

**Thiết kế giao diện:**

|  |  |
| --- | --- |
| **Thiết kế giao diện hiển thị** | **Người thực hiện** |
|  | **Nguyễn Văn Đạt** |

**Code**

|  |  |
| --- | --- |
| **Phương thức** | **Người thực hiện** |
| Tạo đối tượng Điểm | **Nguyễn tiến Đạt** |
| Khởi tạo danh sách các đối tượng Điểm ban đầu | **Nguyễn Tiến Đạt** |
| Phân loại các Điểm thuộc các lớp và Điểm mới  vào sẽ thuộc các lớp nào dựa công thức Euclidean . | **Nguyễn Tiến Đạt** |
| Thêm 1 điểm mới vào khi biết 2 toạ độ của điểm đó | **Nguyễn Tiến Đạt** |
| Vẽ các điểm | **Nguyễn Văn Đạt** |

***PHẦN 3: ĐÁNH GIÁ VÀ PHÂN CÔNG***

1. **PHÂN CÔNG**

|  |  |  |
| --- | --- | --- |
| **TÊN SV** | **ĐÓNG GÓP** | **CÔNG VIỆC** |
| Nguyễn Tiến Đạt | **70%** | **Xây dựng ý tưởng,xây dựng giao diện ,Viết code** |
| Nguyễn Văn Đạt | **30%** | **Xây dựng giao diện hiển thị,viết báo cáo.** |

1. **ĐÁNH GIÁ**

Ứng dụng chạy nhanh, ổn định

1. **KHUYẾT ĐIỂM**

Ứng dụng chưa có nhiều cơ sở dữ liệu để minh hoạ đủ thuật toán.

***PHẦN 4: TÀI LIỆU THAM KHẢO***

1. Tham khảo từ thầy.
2. Tham khảo từ Internet

Các web đã tham khảo:

1. <https://vi.wikipedia.org/wiki/Gi%E1%BA%A3i_thu%E1%BA%ADt_k_h%C3%A0ng_x%C3%B3m_g%E1%BA%A7n_nh%E1%BA%A5t>