**Báo cáo bài tập lớn**

**Nhóm 29**

*Link github của project:*[*https://github.com/Halcolt/AI/tree/AI\_1*](https://github.com/Halcolt/AI/tree/AI_1)

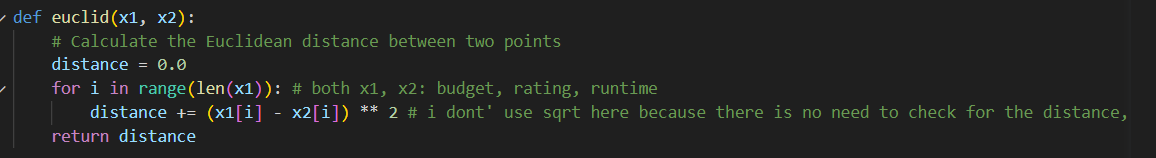
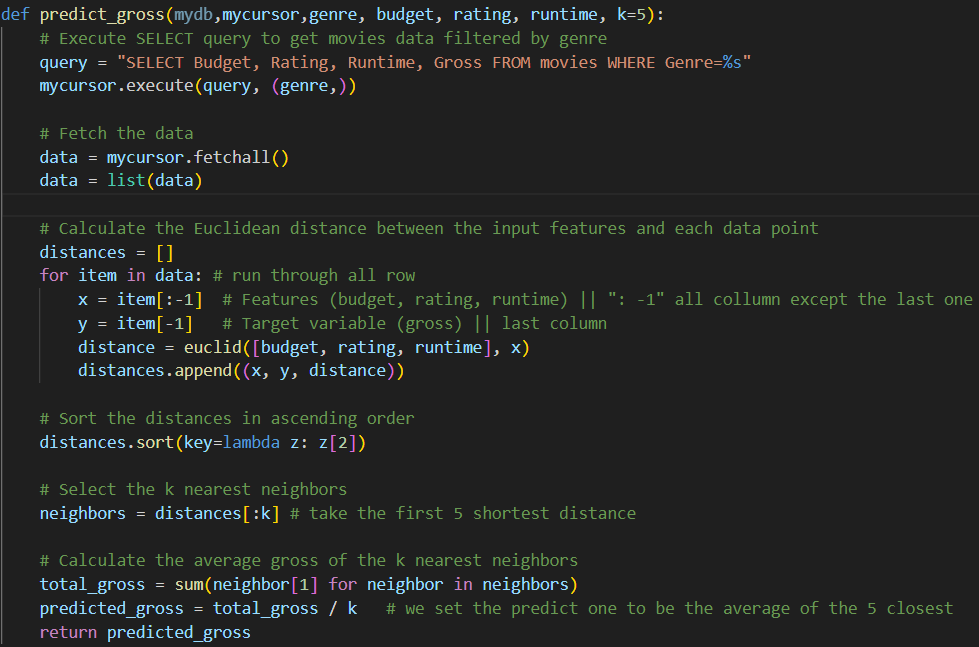
1. **Giới thiệu đề tài**

* Đề tài: Phân tích và dự đoán dữ liệu về thông tin phim ảnh
* Mục tiêu: từ những thông số có sẵn (Budget, thể loại, doanh thu sau tuần đầu tiên,...) dự đoán được doanh thu cuối, đưa ra đánh giá phim và từ đó có thể sắp xếp lịch chiếu phù hợp.

1. **Thuật toán**

* Sử dụng thuật toán k-nearest neighbors(Knn)
* Thuật toán euclid để tính khoảng cách cho Knn

1. **Tiến độ**

* Lấy dataset từ kaggle:: <https://www.kaggle.com/datasets/thedevastator/movie-gross-and-ratings-from-1989-to-2014?fbclid=IwAR08WEK1ixVa98icNk3-oYpNaHQt4Ufglg9K6aytlZ18wJV34-SY5Kr_dzk>
* Đưa dataset có sẵn lên database
* Kết nối tới database
* Thuật toán euclid:  
  
* Hàm nhận dữ liệu từ database và từ đó đưa ra dự đoán dựa vào giá trị trung bình của k nearest neighbors(trong trường hợp này k=5)
* 

1. **công nghệ sử dụng**

* Python , Mysql

1. **Dự kiến tính năng bổ sung**

* Cải tiến độ chính xác bằng cách thêm tỉ lệ lạm phát
* Show biểu đồ dự đoán