Mô tả dự án:
**Dữ liệu đầu vào**: start time, end time, number of ingredients được lưu trữ trong file ./Data/Data_AIL.xlsx
**Dữ liệu đầu ra:** viewer feeling of youtuber's style
**Link file project**: Github

## I. Xử lí dữ liệu

Dữ liệu đầu vào đến từ link, dữ liệu ban đầu có rất nhiều chỗ annotation cần được xử lí lại, Sau khi thực hiện thao tác loại bỏ các dữ liệu không tốt, thì em thực hiện standardize các giá trị của từng thuộc tính (feature) mà em có bằng cách chia cho giá trị lớn nhất của thuộc tính đó

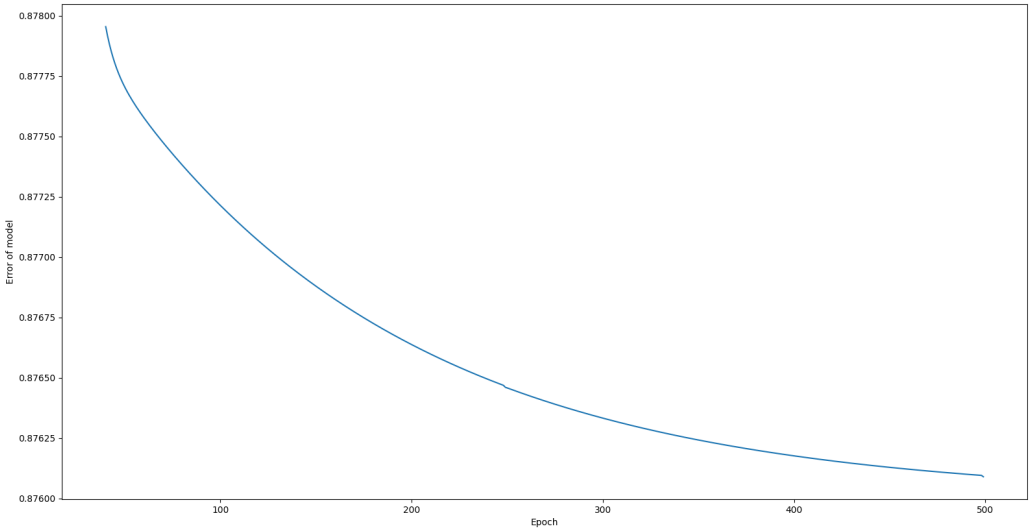
## II. Model

Ở đây em sẽ sử dụng model Linear Regression vì những lí do sau

- Đầu tiên Linear Regression là thuật toán đơn giản nhất cũng như dễ dàng thực thi
- Điều ta mong muốn là dự đoán được viewer feeling nằm trong khoảng giá trị từ [0, 5].
- Linear Regression sẽ giúp ta dự đoán được các giá trị liên tục trong đoạn từ [0, 5], trong khi bộ dữ liệu của ta chỉ cho feeling của viewer là các giá trị rời rạc 0, 1, 2, 3, 4, 5. Điều này có thể sử dụng 5 model Logistic regression -> mức độ phức tạp của việc thực thi cũng như xử lí dữ liệu sẽ cao hơn, nhưng việc đảm bảo về mức độ chính xác tốt thì không được đảm bảo
- Việc sử dụng Linear Regression giúp ta đưa ra được các giá trị liên tục trong khoảng từ [0, 5] để giúp ta có thể dự đoán viewer feeling tốt hơn, cũng như giảm bớt được số lượng parameter để tăng tốc độ thực thi của thuật toán. Nếu giá trị model đưa ra được là 2.7 thì ta có thể hiểu là video của ta vẫn chưa được tốt lắm để ta đánh giá nó với thang điểm 3/5

## III. Train

Model của em được huấn luyện với số lần huấn luyện là 500, learning rate là 0.1, và sau lần huấn luyện thứ 250 thì learning rate sẽ giảm xuống 1 lượng bằng 0.01 learning rate ban đầu
Hàm loss function em sử dụng là Mean Square Error (MSE).
Và ở đây em chia tập dữ liệu của em ra dùng 70% để thực hiện quá trình huấn luyện. 30% cho việc thử nghiệm lại độ chính xác của model



Quá trình thực hiện tối ưu của ta diễn ra khá là tốt nhưng vẫn không đạt được độ chính xác cao

## IV. Accuracy

Trong quá trình tối ưu hóa các giá trị của model thì độ sai số của dự đoán là 0.837 so với dữ liệu thực tế

Và độ chính xác khi thực hiện thử nghiệm lại thì sai số của dự đoán với dữ liệu thực tế trung bình là 1.0