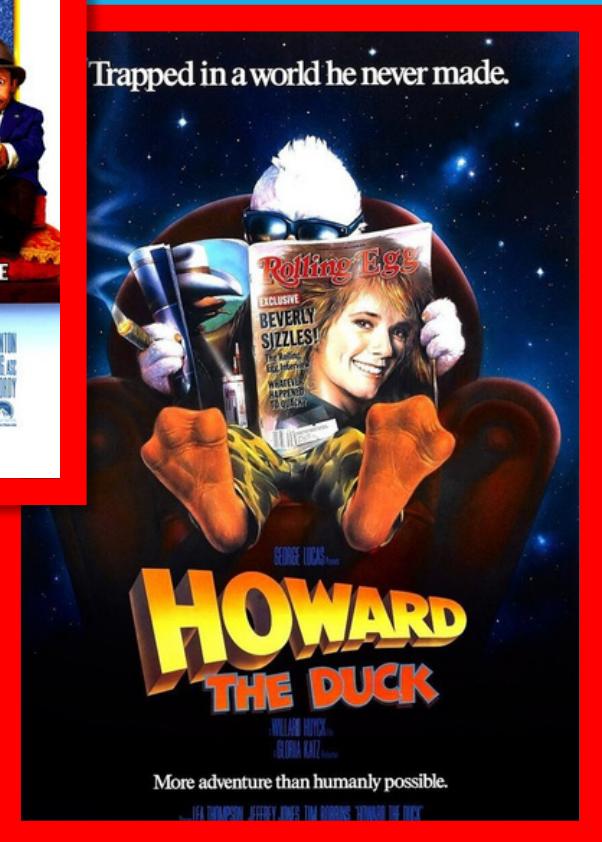
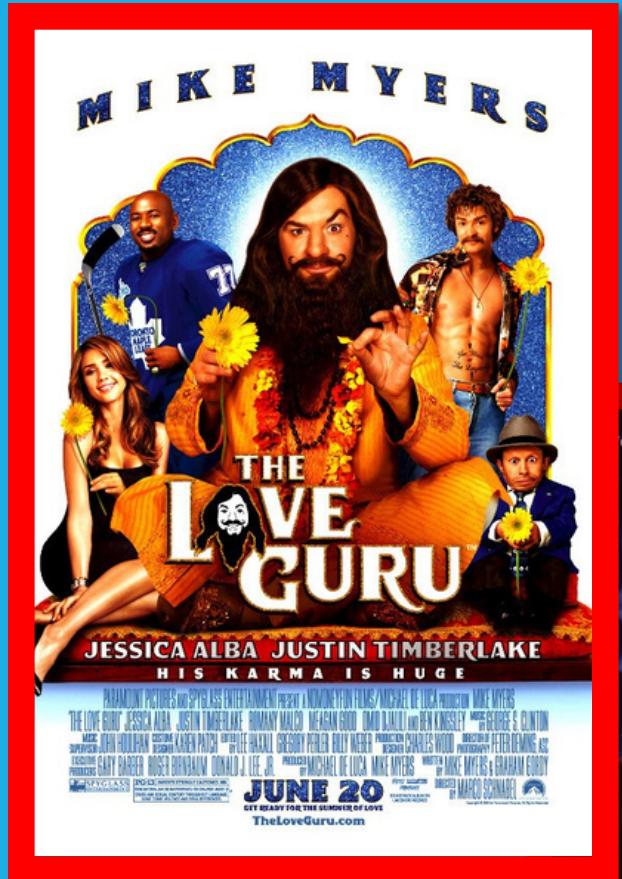




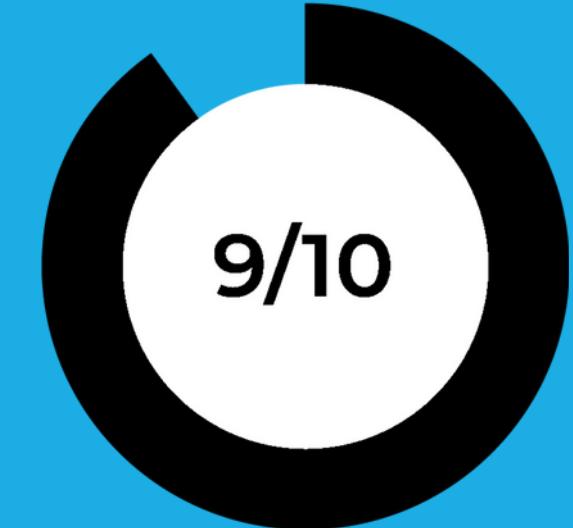
Hệ thống dự đoán điểm đánh giá phim

Tên thành viên :
Dương Văn Chính
Nguyễn Thành Long
Lê Đăng Trường Đạt



THU THẬP DỮ LIỆU PHIM

- Mỗi năm lấy 500 bộ phim có doanh thu cao nhất trên toàn thế giới
- Trung bình mỗi bộ phim có điểm IMDB cao hơn 5



→ 12161 phim từ năm
1991 đến năm 2020



TRỰC QUAN HOÁ DỮ LIỆU



Những đặc trưng của dữ liệu

CHI PHÍ
SẢN XUẤT

SỐ NGƯỜI
ĐÁNH GIÁ

ĐẠO DIỄN

ĐIỂM
MCORE

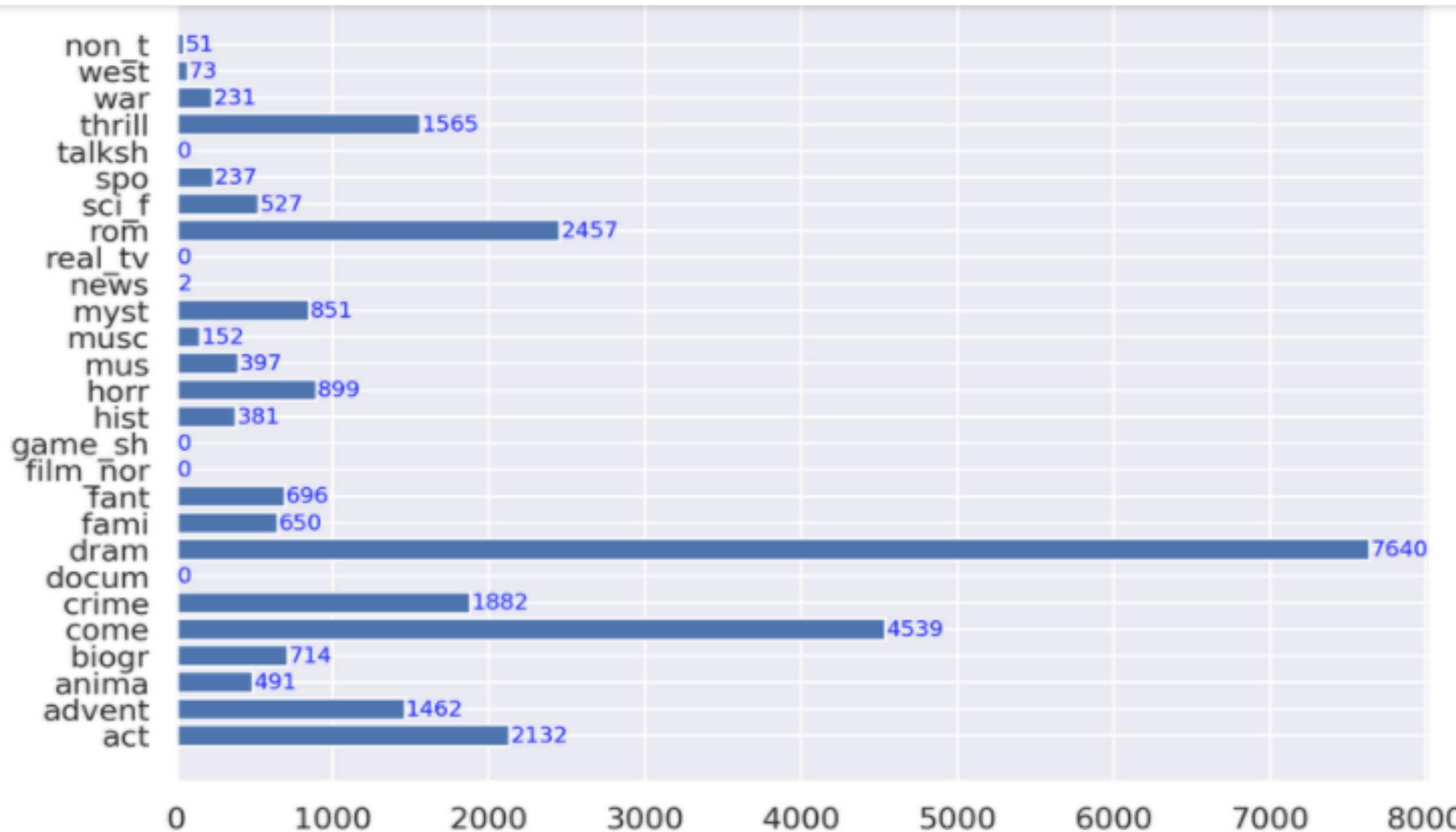
NĂM XUẤT
BẢN

THỂ LOẠI

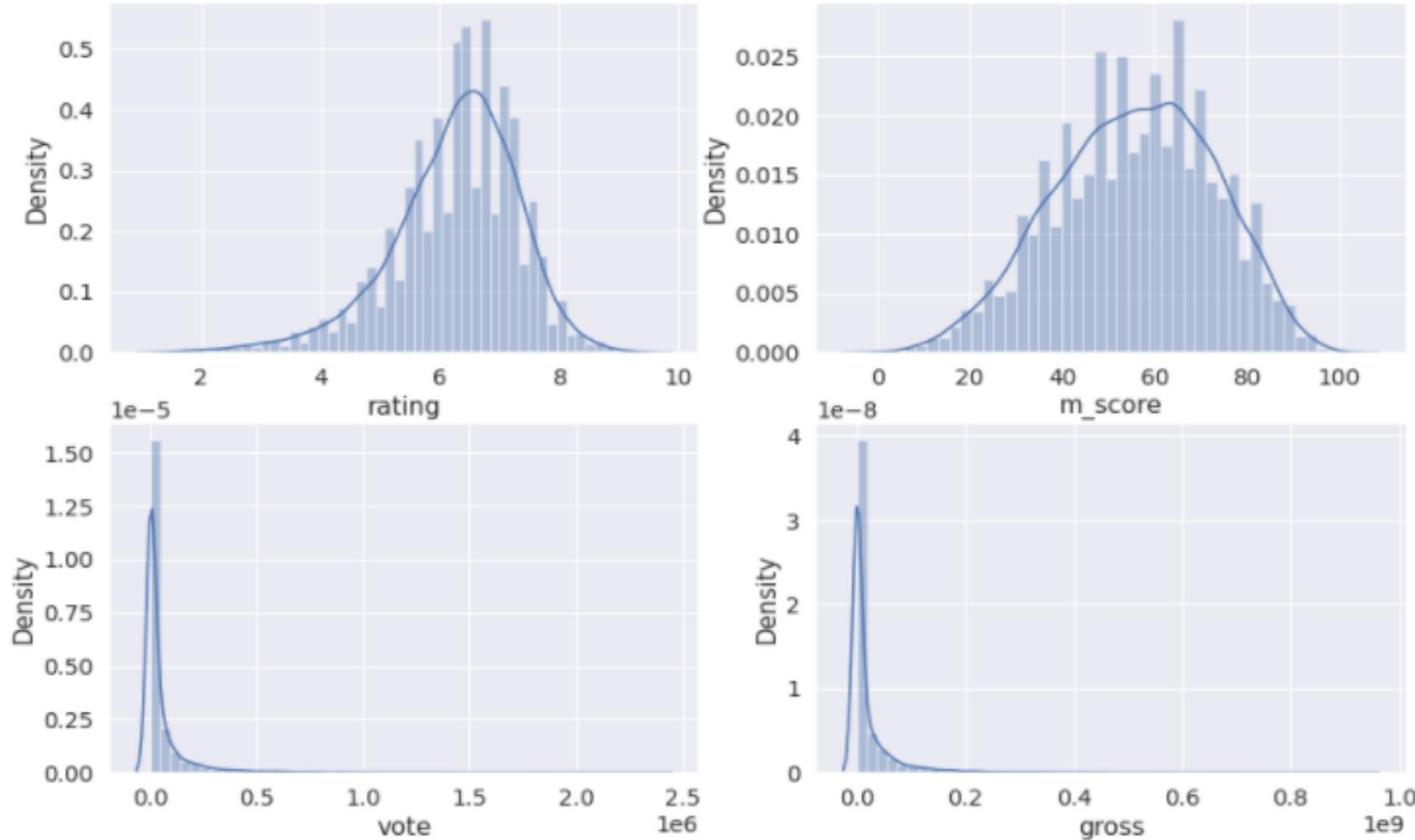
TÊN PHIM

ĐẠO DIỄN

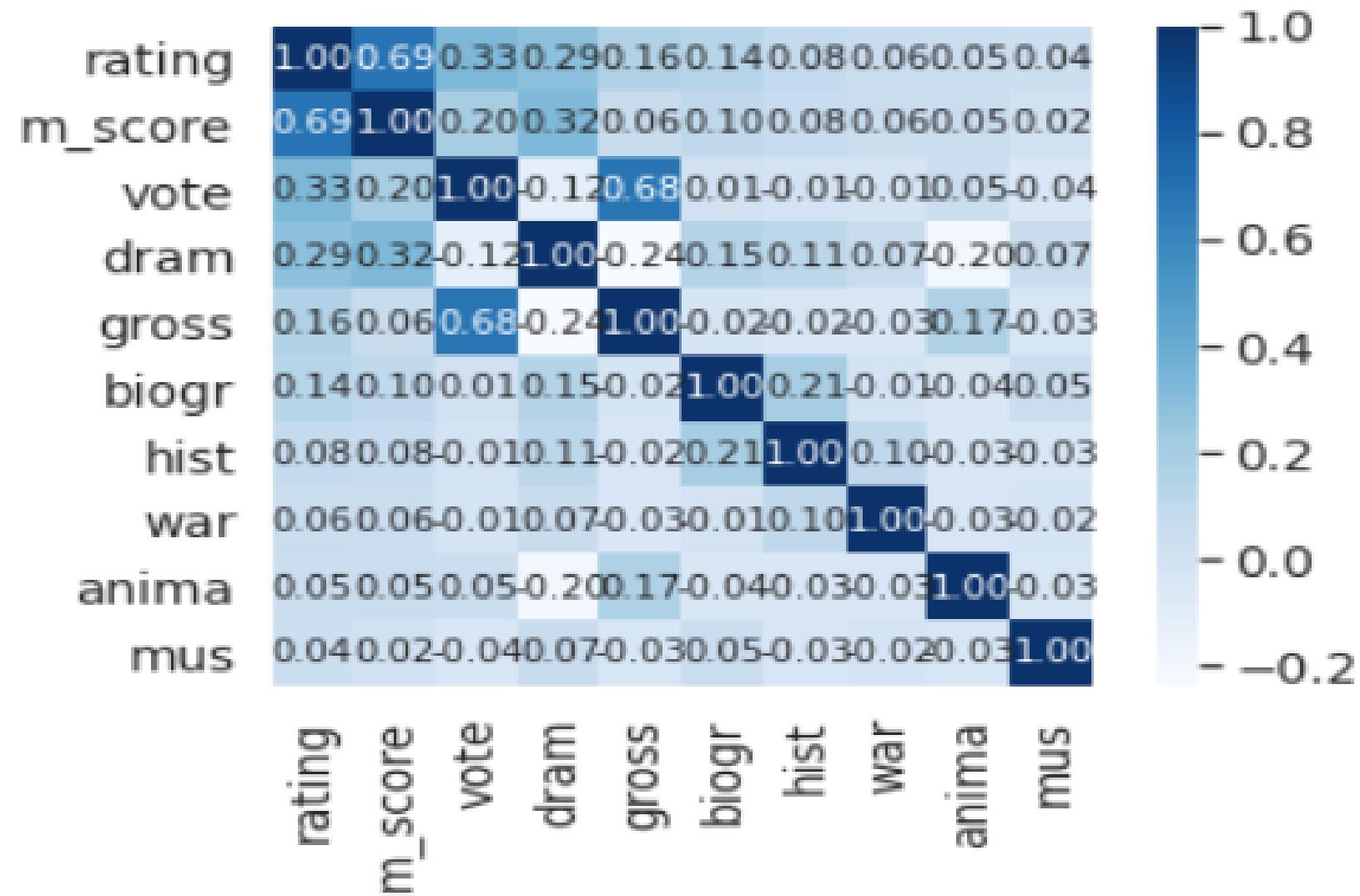
Thống kê số lượng phim qua thể loại



Thống kê mật độ giá trị của các đặc trưng số



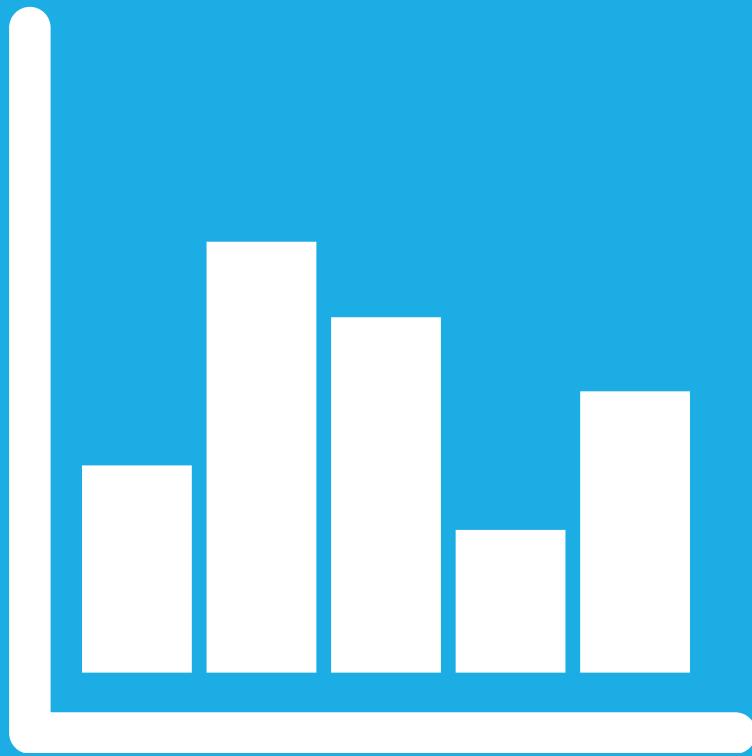
10 đặc trưng có độ tương quan cao nhất với ‘rating’



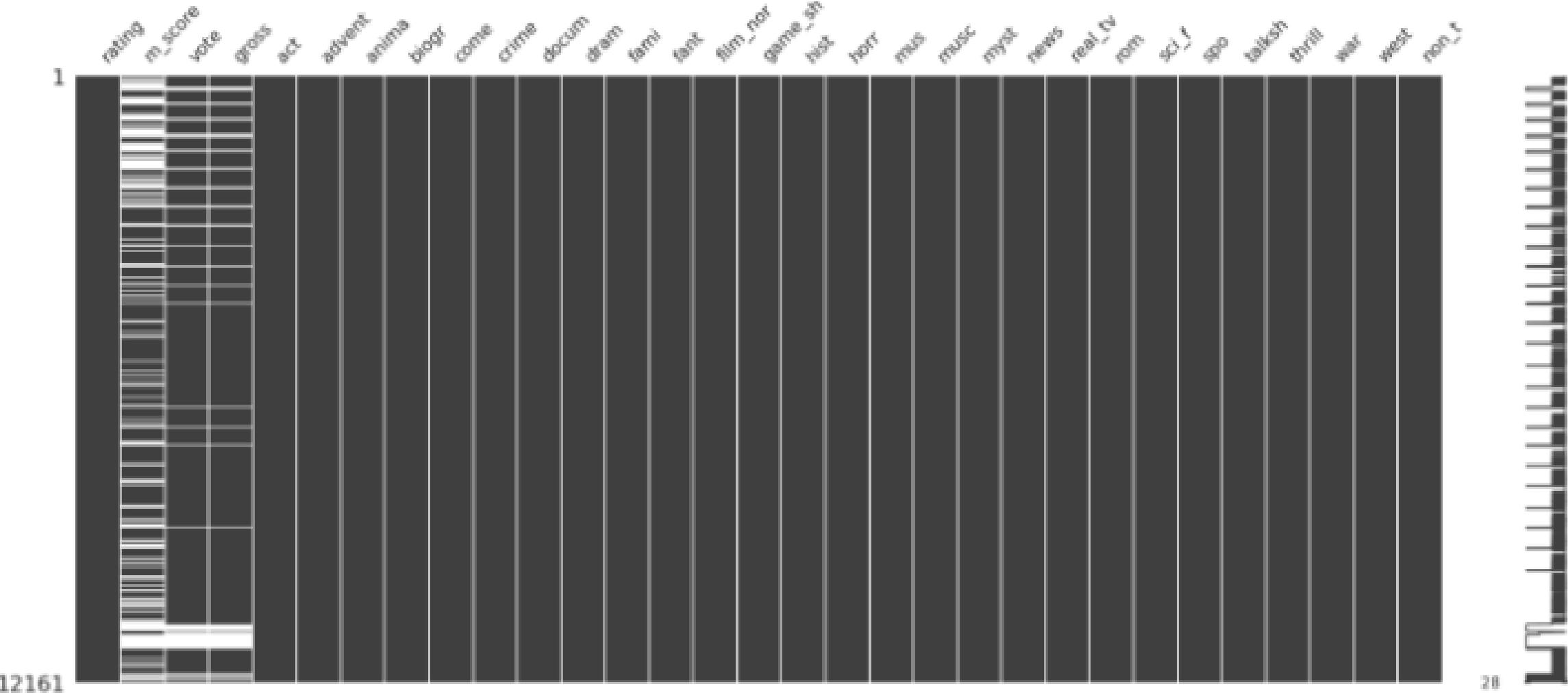
TRÍCH TRỌN ĐẶC TRƯNG

	rating	m_score	vote
0	8.6	85.0	1294962.0
1	8.5	75.0	1008750.0
2	8.0	95.0	421130.0
3	6.8	52.0	238177.0
4	6.9	51.0	180551.0

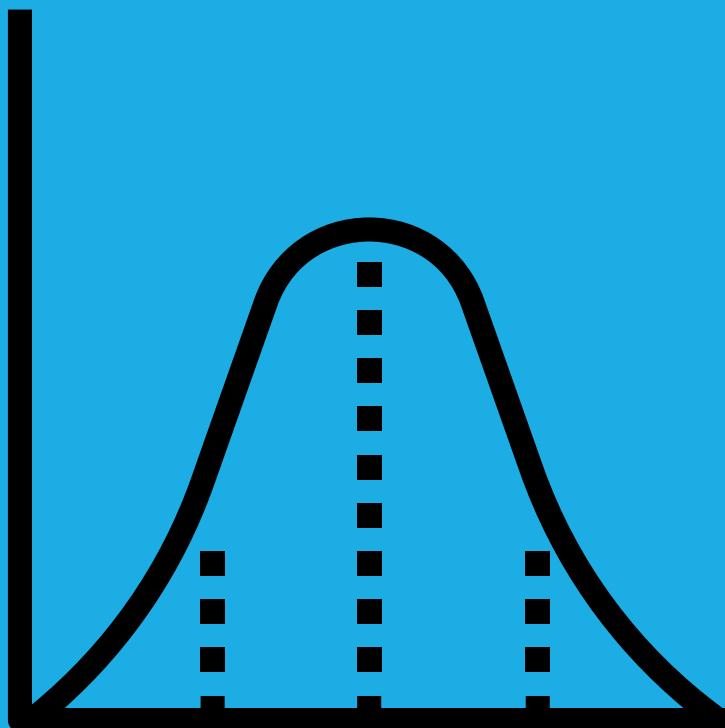
XỬ LÝ DỮ LIỆU TRỐNG



Trực quan dữ liệu trống dựa trên ma trận



CHUẨN HOÁ DỮ LIỆU



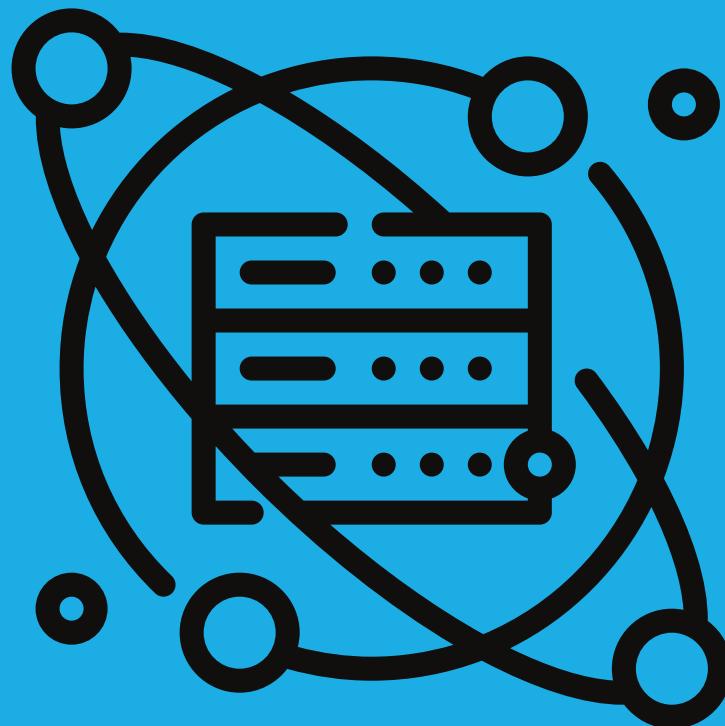
Dữ liệu được chuẩn hóa bằng phương pháp min-max



```
min_max = MinMaxScaler()
df_train = pd.DataFrame(min_max.fit_transform(df_train),columns=df_train.columns)
df_train.head()
```

	rating	m_score	vote
0	0.911392	0.848485	0.951069
1	0.898734	0.747475	0.931068
2	0.835443	0.949495	0.861118
3	0.683544	0.515152	0.815479
4	0.696203	0.505051	0.793298

MÔ HÌNH HOÁ DỮ LIỆU

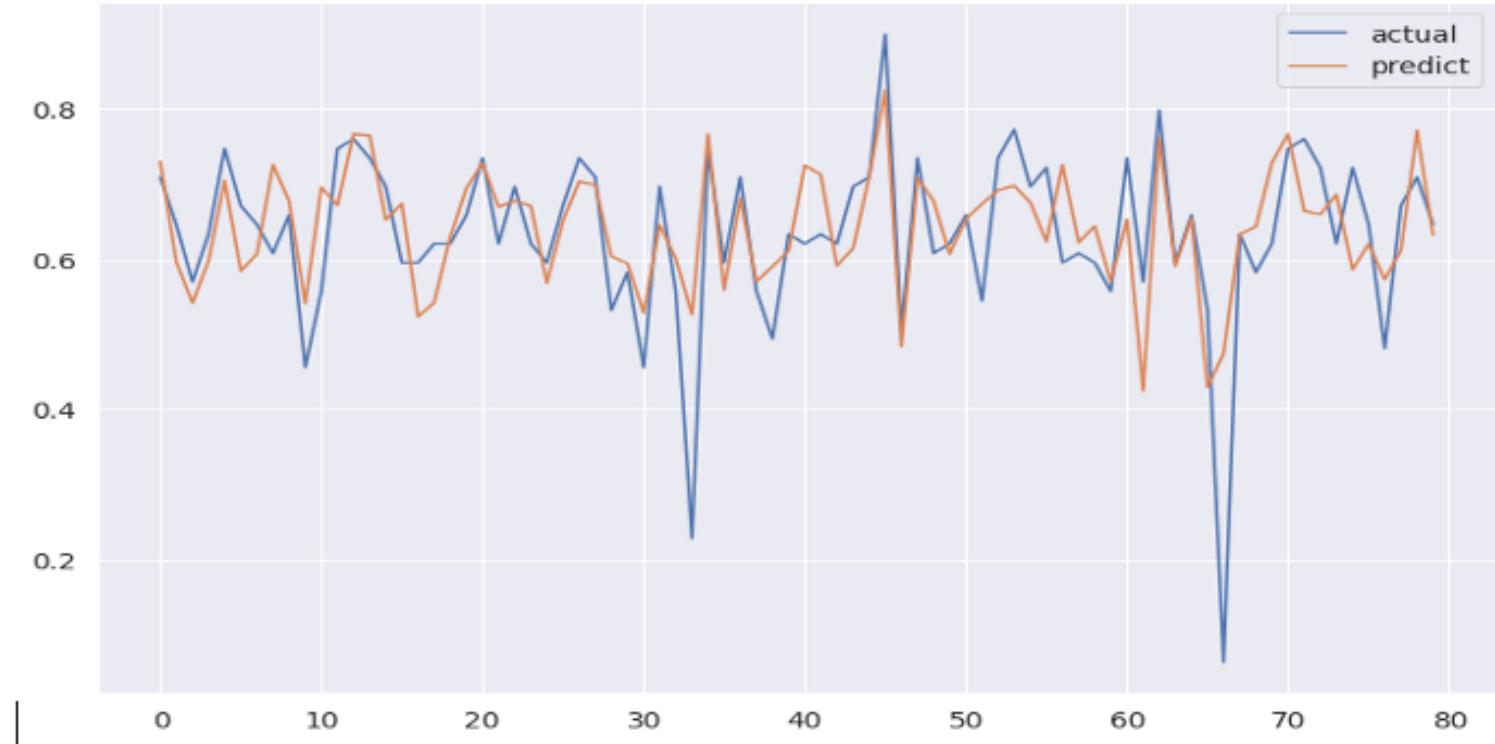


LINEAR REGRESSION MODEL

R² ≈ 0.54

RMSE ≈ 0.077

MSE ≈ 0.006



So sánh đầu ra dự đoán và đầu ra
trên tập dữ liệu (tập test) (80 mẫu đầu tiên)

NEURAL NETWORK

R2 ≈ 0.56

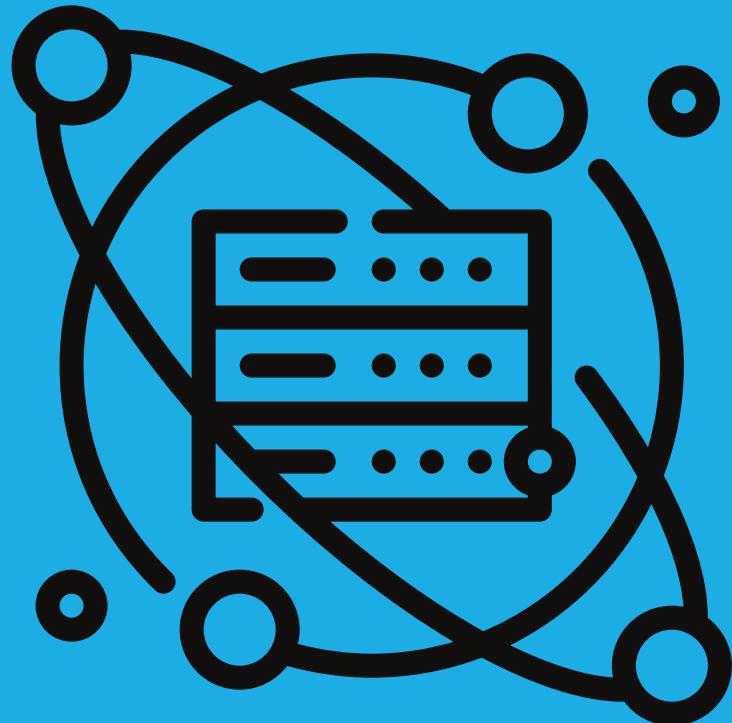
RMSE ≈ 0.075

MSE ≈ 0.005



So sánh đầu ra dự đoán và đầu ra
trên tập dữ liệu (tập test) (80 mẫu đầu tiên)

ĐÁNH GIÁ MÔ HÌNH



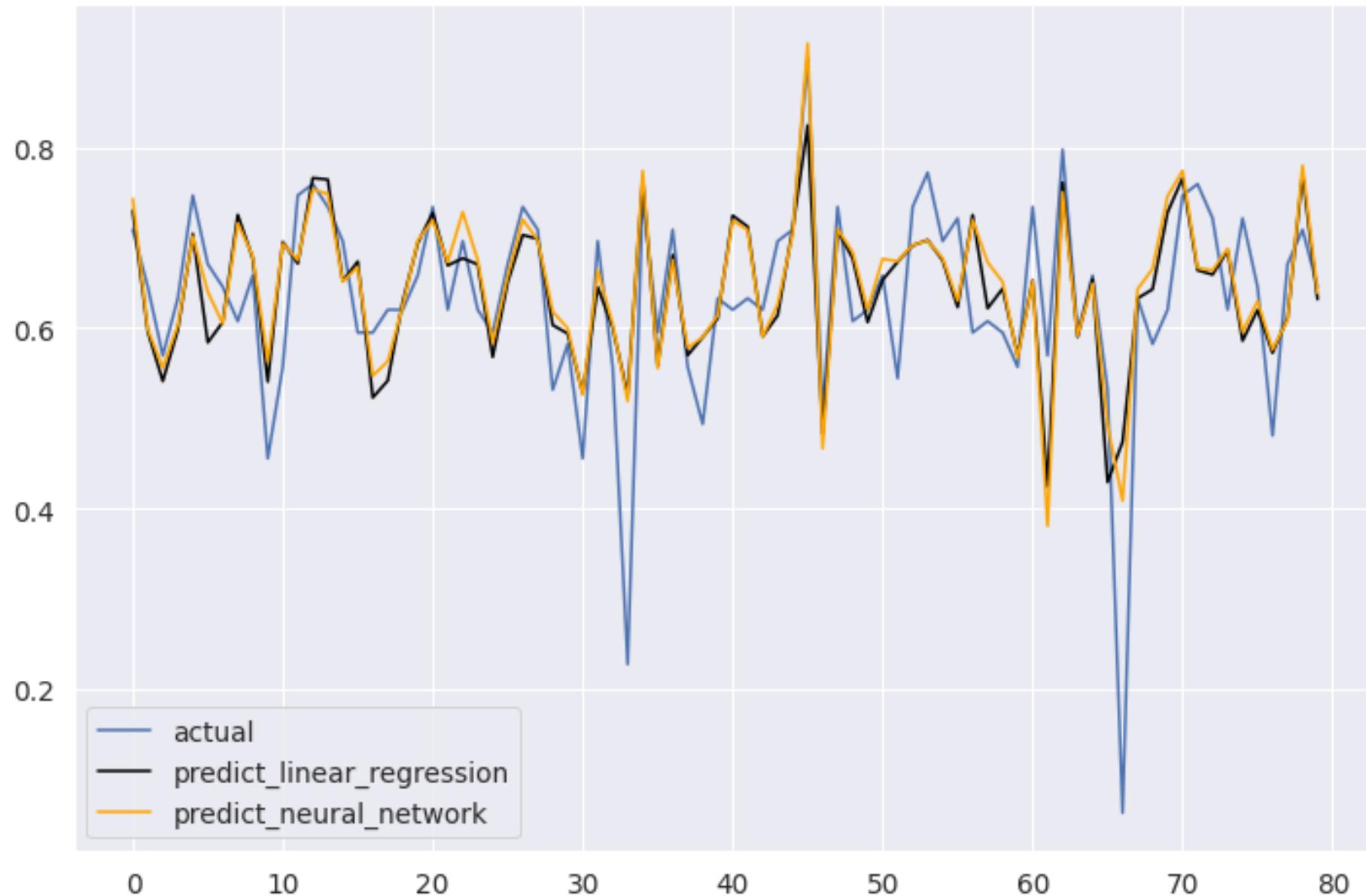
BẢNG SO SÁNH CÁC ĐỘ ĐO METRICS

BẢNG SO SÁNH CÁC ĐỘ ĐO METRICS

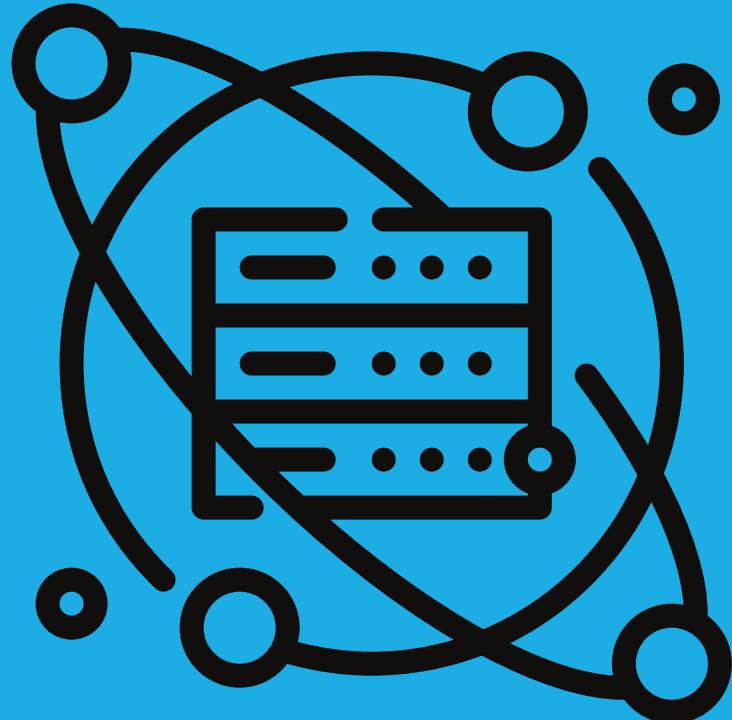
	Linear Regression	Linear Regression (NN)
MSE	0.0059979503995608915	0.005682486500773637
RMSE	0.07744643568015827	0.07538226914051896
R2_score	0.5393099588876914	0.5635400819831391

Bảng 2: So sánh các độ đo metrics

ĐỒ THỊ SO SÁNH HAI MÔ HÌNH



KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN



KẾT LUẬN

- Học cách thu thập dữ liệu
- Trực quan hóa dữ liệu
- Xử lý dữ liệu
- Mô hình hóa huấn luyện

HƯỚNG PHÁT TRIỀN

- Thu thập nhiều dữ liệu với nhiều đặc trưng hơn
- Tìm hiểu và chọn ra mô hình để huấn luyện phù hợp hơn

