

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM**

**KHOA ĐÀO TẠO CHẤT LƯỢNG CAO**



**ĐỒ ÁN 3**

**PHAN LOAI HÌNH ẢNH ĐA NHÃN**

|  |  |  |
| --- | --- | --- |
| **SVTH** | | **MSSV** |
| Lê Huy Hiệp | | 17110137 |
| Ngô Thị Trang | | 17110241 |
| **Khoá** | **2017 - 2021** | |
| **Ngành** | **CÔNG NGHỆ THÔNG TIN** | |
| **GVHD** | **TS. HUỲNH XUÂN PHỤNG** | |

**TP.HCM, ngày 20, tháng 1, năm 2020**

# LỜI CẢM ƠN

Lời đầu tiên, nhóm thực hiện xin được phép gửi lời cảm ơn chân thành đến khoa Đào tạo Chất Lượng Cao – Trường Đại học Sư phạm Kỹ thuật Thành phố Hồ Chí Minh đã tạo mọi điều kiện thuận lợi nhất cho nhóm thực hiện có cơ hội được tự do tiếp cận, tham khảo, mở rộng thêm kiến thức trong lĩnh vực Công nghệ thông tin nói chung và môn Đồ án 3 nói riêng.

Lời cảm ơn trân trọng nhất nhóm thực hiện xin chân thành gửi đến Thầy **Huỳnh Xuân Phụng** – người đã dùng mọi tâm huyết và tri thức của người Thầy, cùng đồng hành và trực tiếp giảng dạy, hướng dẫn và tạo mọi điều kiện thuận lợi giúp đỡ cho nhóm phát huy hết khả năng cũng như nâng cao kiến thức trong suốt quá trình học tập, đặc biệt là trong quá trình chuẩn bị và thực hiện đề tài môn học. Cảm ơn sự nhiệt tình của Thầy, là động lực vô cùng to lớn giúp nhóm kiên trì trong suốt quá trình thực hiện đề tài và khám phá ra những kiến thức mới đầy thú vị và bổ ích liên quan đến môn học.

|  |
| --- |
| TPHCM, ngày 20 tháng 1 năm 2020 |
| **Nhóm sinh viên thực hiện** |

# MỤC LỤC

[LỜI CẢM ƠN ii](#_Toc62167169)

[MỤC LỤC iii](#_Toc62167170)

[DANH MỤC CÁC HÌNH ẢNH, BIỂU ĐỒ iv](#_Toc62167171)

[Chương 1: GIỚI THIỆU VẤN ĐỀ 1](#_Toc62167172)

[Chương 2: TỔNG QUAN VỀ DỮ LIỆU 3](#_Toc62167173)

[Chương 3: PHÂN TÍCH DỮ LIỆU 4](#_Toc62167174)

[3.1. Số lượng phim của từng thể loại 4](#_Toc62167175)

[3.2. Số lượng thể loại của mỗi phim 5](#_Toc62167176)

[Chương 4: TÍNH TƯƠNG QUAN 6](#_Toc62167177)

[4.1. Heatmap 6](#_Toc62167178)

[4.2. Sự phân bố của các thể loại phim 7](#_Toc62167179)

[4.3. Số lượng thể loại phim trên từng thể loại 10](#_Toc62167180)

[Chương 5: MODEL 12](#_Toc62167181)

[5.1. Multi-label classification 12](#_Toc62167182)

[5.1.1. Multi-label classification 12](#_Toc62167183)

[5.1.2. CNN 12](#_Toc62167184)

[5.1.2.1. Giới thiệu 12](#_Toc62167185)

[5.1.2.2. Mạng nơ ron tích chập 14](#_Toc62167186)

[5.2. Xây đựng Model 15](#_Toc62167187)

[5.3. UnderFitting & OverFitting 17](#_Toc62167188)

[5.4. Kiểm tra kết quả 19](#_Toc62167189)

[Chương 6: TỔNG KẾT 22](#_Toc62167190)

[6.1. Hạn chế 22](#_Toc62167191)

[6.2. Hướng phát triển 22](#_Toc62167192)

[TÀI LIỆU THAM KHẢO 23](#_Toc62167193)

# DANH MỤC CÁC BẢNG

[Bảng 2.1. Các thể loại trong dữ liệu 3](#_Toc62167812)

# DANH MỤC CÁC HÌNH ẢNH, BIỂU ĐỒ

[Hình 1. Hình ảnh minh hoạ nhiều đối tượng 2](#_Toc62167663)

[Hình 2. Phân bố của số lượng phim trong từng thể loại 4](#_Toc62167664)

[Hình 3. Biểu đồ phân tích dữ liệu số lượng thể loại của mỗi phim 5](#_Toc62167665)

[Hình 4. Tương quan giữa các thể loại phim 6](#_Toc62167666)

[Hình 5. Biểu đồ tương quan các thể loại phim 7](#_Toc62167667)

[Hình 6. Biểu đồ tương quan các thể loại phim 8](#_Toc62167668)

[Hình 7. Biểu đồ tương quan các thể loại phim 9](#_Toc62167669)

[Hình 8. Biểu đồ số lượng thể loại của mỗi phim 10](#_Toc62167670)

[Hình 9. Biểu đồ số lượng thể loại của mỗi phim 11](#_Toc62167671)

[Hình 10. Mạng nơron của người (từ Rob Fergus) 13](#_Toc62167672)

[Hình 11. Mạng nơron nhiều tầng 13](#_Toc62167673)

[Hình 12. Kết nối giữa các tầng trong mạng nơ ron truyền thống 14](#_Toc62167674)

[Hình 13. Kết nối giữa các tầng trong mạng nơ ron tích chập 15](#_Toc62167675)

[Hình 14. Các bước cơ bản trong mạng neuron tích chập 15](#_Toc62167676)

[Hình 15. Model 1 16](#_Toc62167677)

[Hình 16. Model 2 16](#_Toc62167678)

[Hình 17. Model 3 17](#_Toc62167679)

[Hình 18. Đồ thị hàm mất mát Model 1&2 18](#_Toc62167680)

[Hình 19. Đồ thị hàm mất mát Model 3 19](#_Toc62167681)

[Hình 20. Poster dự đoán 20](#_Toc62167682)

[Hình 21. Kết quả 4 thể loại có khả năng cao nhất của Model 1 20](#_Toc62167683)

[Hình 22. Kết quả 4 thể loại có khả năng cao nhất của Model 2 20](#_Toc62167684)

[Hình 23. Kết quả 4 thể loại có khả năng cao nhất của Model 3 21](#_Toc62167685)

# Chương 1: GIỚI THIỆU VẤN ĐỀ

Những năm gần đây, AI - Artificial Intelligence (Trí Tuệ Nhân Tạo), và cụ thể hơn là Machine Learning (Học Máy hoặc Máy Học) nổi lên như một bằng chứng của cuộc cách mạng công nghiệp lần thứ tư. Machine Learning là một tập con của AI, là một lĩnh vực nhỏ của Khoa Học Máy Tính, nó có khả năng tự học hỏi dựa trên dữ liệu đưa vào mà không cần phải được lập trình cụ thể.

Machine learning được phân làm 3 loại chính: supervised learning (học có giám sát), unsupervised learning (học không có sự giám sát), reinforcement learing (học tăng cường). Phân loại hình ảnh nhiều lớp là một ví dụ điển hình cho supervised learning. Khi huấn luyện hình ảnh sẽ được gắn nhãn theo chủ đề cụ thể.

Ví dụ

|  |  |
| --- | --- |
| Điểm mặt 7 loài động vật dễ thương vô cùng tại “đảo thiên đường” Hokkaido |  Khám phá Nhật Bản | Samurai Tour | Gắn nhãn à Lớp Con thỏ |
| Gắn nhãn à Lớp Con gấu trúc | 12 loài vật “cực đẹp” nhưng tốt nhất không nên nuôi làm thú cưng |

Ở ví dụ trên là những tấm hình đơn giản và chỉ có một đối tượng, nhưng thực tế thì có rất nhiều tấm hình mà ở trong đó có rất nhiều đối tượng như ở tấm hình dưới đây



Hình 1. Hình ảnh minh hoạ nhiều đối tượng

Khi đó phân loại hình ảnh nhiều lớp sẽ không thể giải quyết được bài toán nữa. Để giải quyết được vấn đề trên nhóm đã quyết định nghiên cứu và xây dựng mô hình phân loại hình ảnh đa nhãn. Cụ thể là phân loại các hình ảnh poster của phim, từ một tấm poster có thể dự đoán được các thể loại của phim.

# Chương 2: TỔNG QUAN VỀ DỮ LIỆU

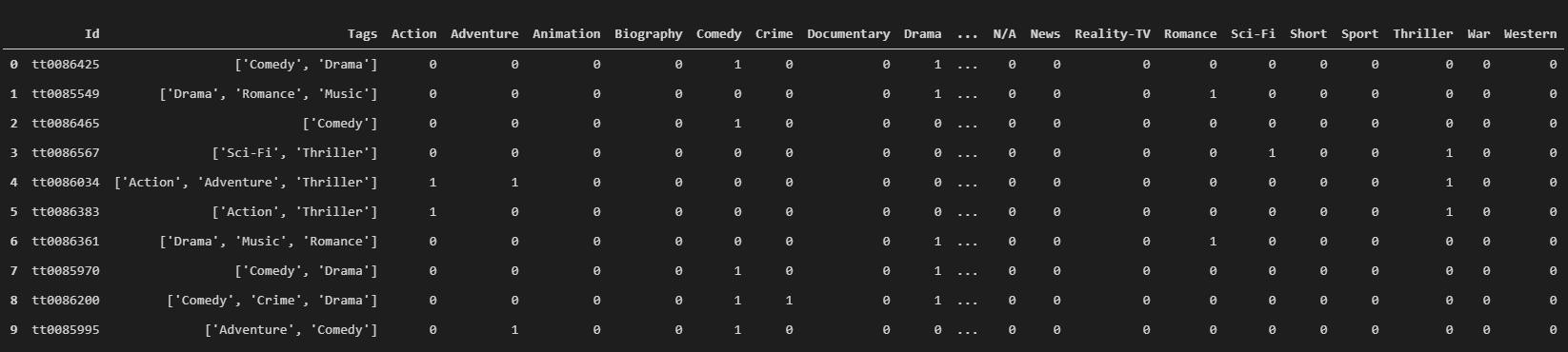
Dữ liệu có 27 cột trong đó 25 cột tương ứng với 25 thể loại phim của các poster. Đây sẽ là mục tiêu cần dự đoán của chúng ta. Nếu một bộ phim thuộc một thể loại nào đó, cột tương ứng sẽ có giá trị 1, ngược lại sẽ là 0. Hàng sẽ gồm:

* 2 cột đầu chứa thông tin của poster:
* Id: tên file poster.
* Gerne: Thể loại phim.
* 25 cột còn lại đại diện cho từng thể loại phim được tổng hợp ở bảng dưới:

Bảng 2.1. Các thể loại trong dữ liệu

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Action | Adventure | Animation | Biography | Comedy |
| Crime | Documentary | Drama | Family | Fantasy |
| History | Horror | Music | Musical | Mystery |
| N/A | News | Reality-TV | Romance | Sci-Fi |
| Short | Sport | Thriller | War | Western |

Dữ liệu sẽ có dạng (7254, 27). Dưới đây là một số dòng đầu của dữ liệu:

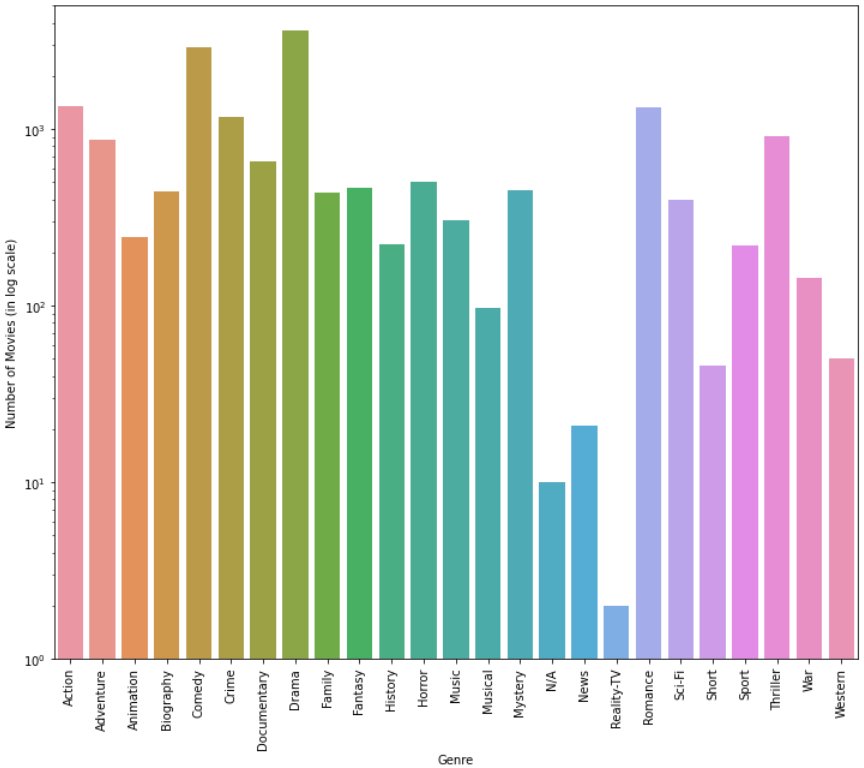


Mục tiêu của nhóm là dự đoán một bộ phim sẽ thuộc những thể loại nào dựa trên poster của nó. Dữ liệu có 7254 mẫu với 27 cột.

# Chương 3: PHÂN TÍCH DỮ LIỆU

## 3.1. Số lượng phim của từng thể loại

Đồ thị bên dưới cho thấy sự phân bố của số lượng phim trong mỗi thể loại trong dữ liệu:

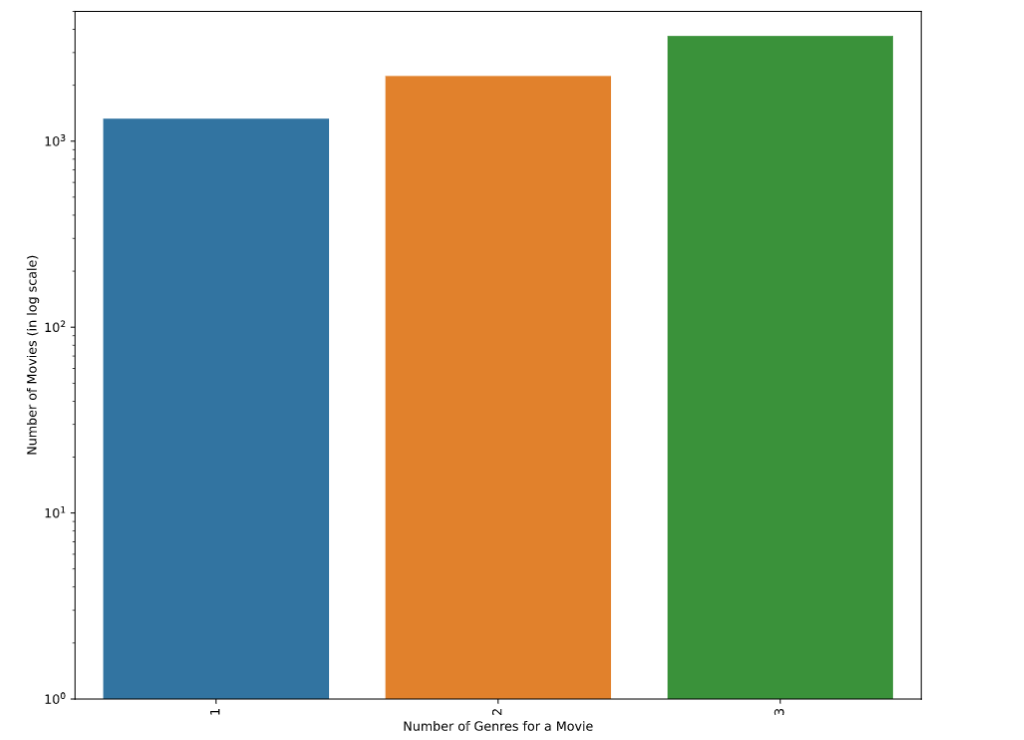


Hình 2. Phân bố của số lượng phim trong từng thể loại

Ta có thể dễ dàng nhận thấy:

* Thể loại có số lượng phim thấp nhất là Reality-TV.
* Thể loại có số lượng phim cao nhất là Drama, thứ 2 là Comedy.

## 3.2. Số lượng thể loại của mỗi phim

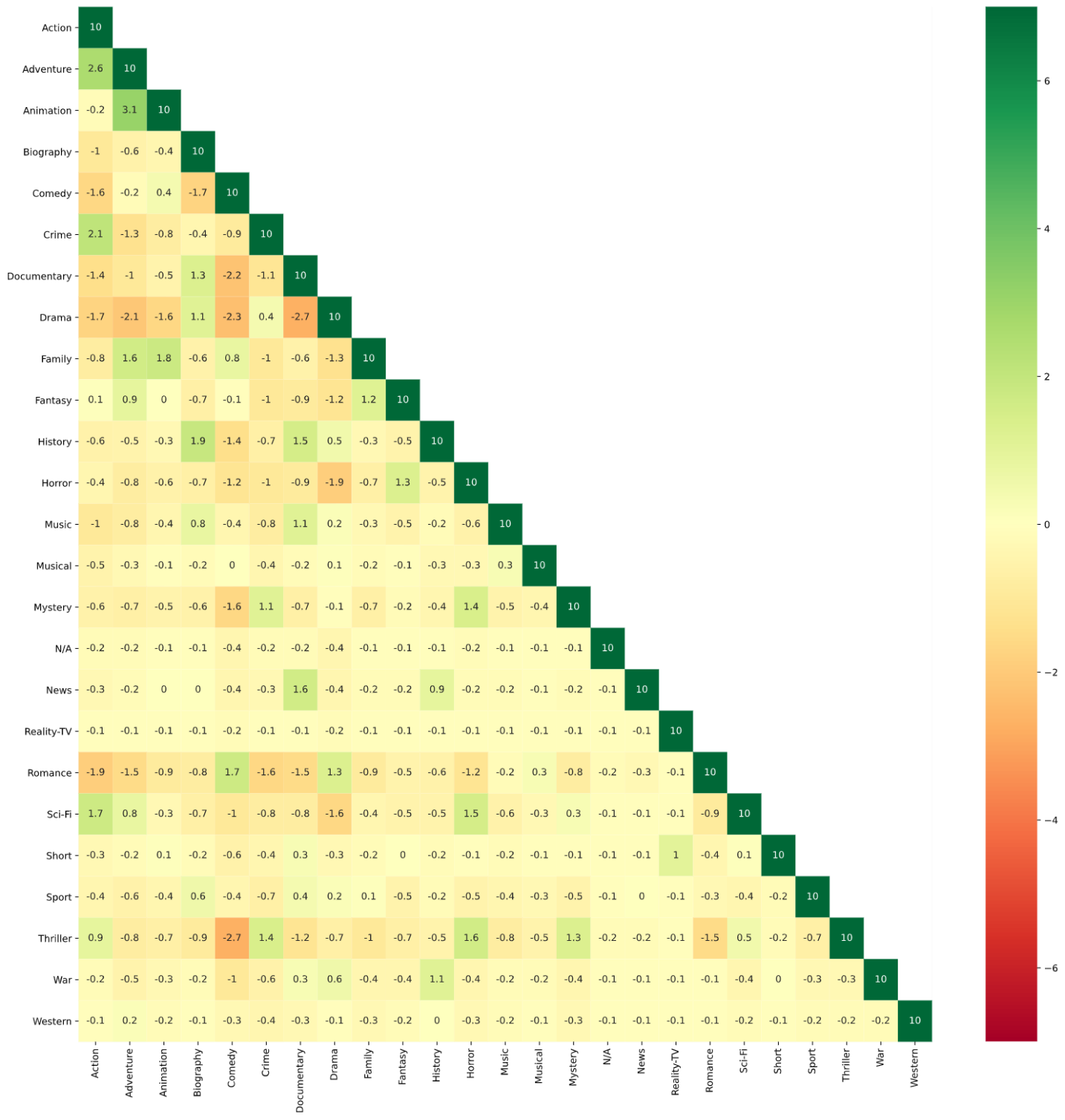


Hình 3. Biểu đồ phân tích dữ liệu số lượng thể loại của mỗi phim

# Chương 4: TÍNH TƯƠNG QUAN

## 4.1. Heatmap

Dưới đây là HeatMap về mối tương quan giữa tất cả các thể loại phim có trong tập dữ liệu:



Hình 4. Tương quan giữa các thể loại phim

Sau khi quan sát. Ta có thể thấy các một số thể loại phim dưới đây có mối tương quan đồng biến mạnh với nhau:

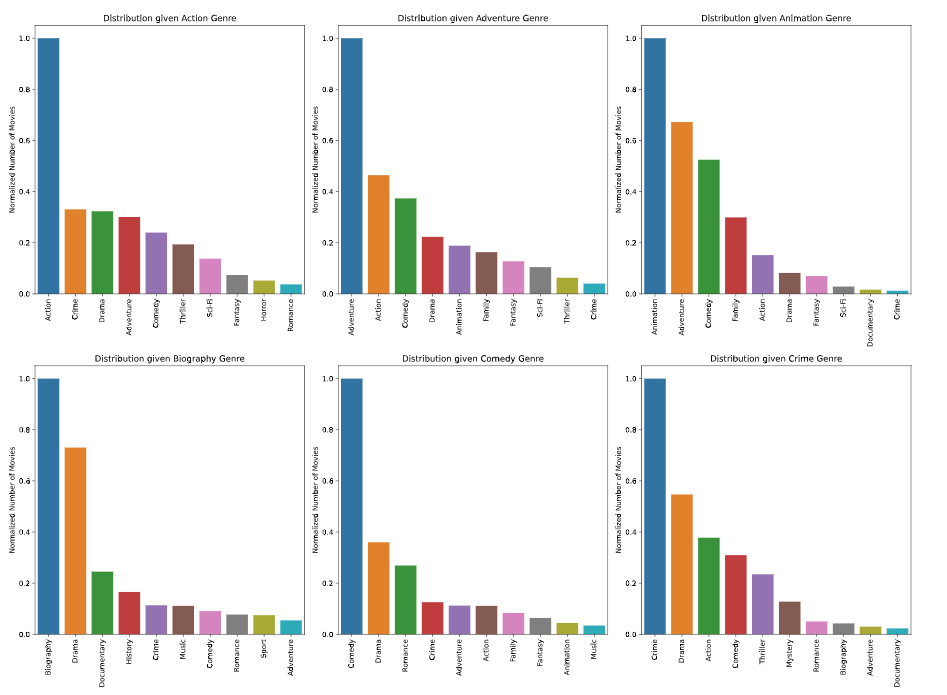
* Action, Adventure, Sci-Fi và Thriller.
* Animation, Fantasy và Family.
* Crime, Thriller, Mystery và Drama.
* Biography, Documentary và History.
* Drama và Romance.
* Game-show và Reality-TV.
* Horror, Thriller và Fantasy.
* Talk-show và News.
* War và History.

Các thể loại phim bên dưới cho thấy mối tương quan nghịch biến mạnh mẽ với nhau:

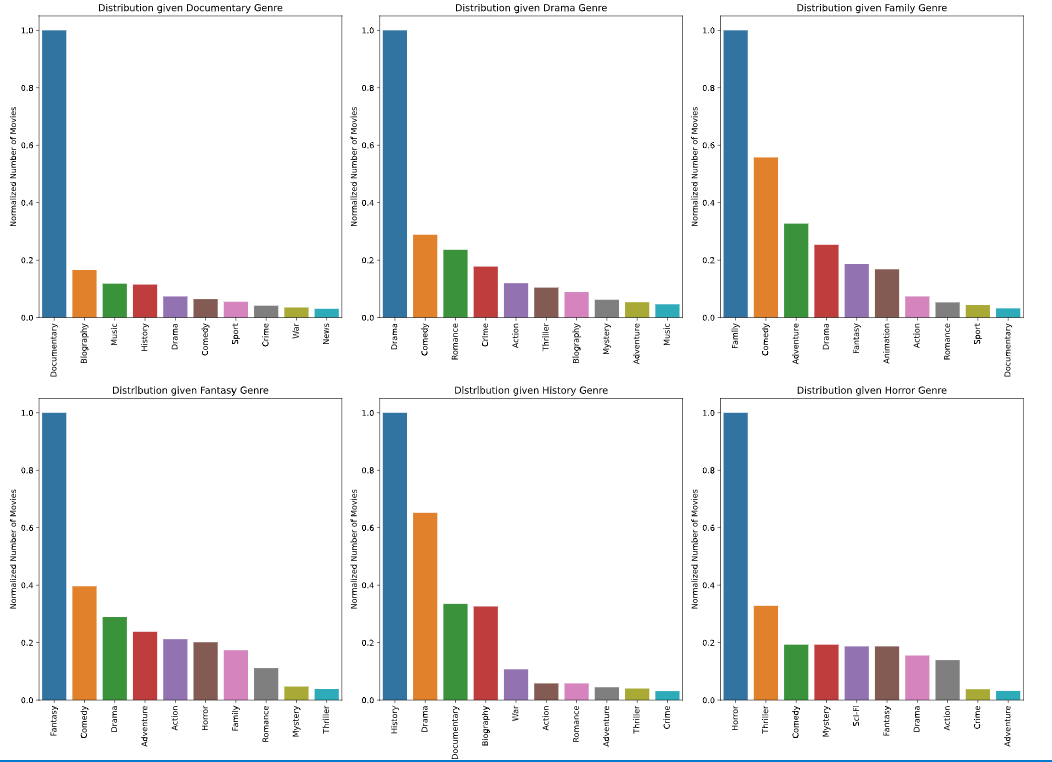
* Animation và Drama.
* Comedy with Documentary và Reality-TV.
* Documentary with Comedy, Drama và Romance.
* Drama with Animation, Reality-TV và Comedy.

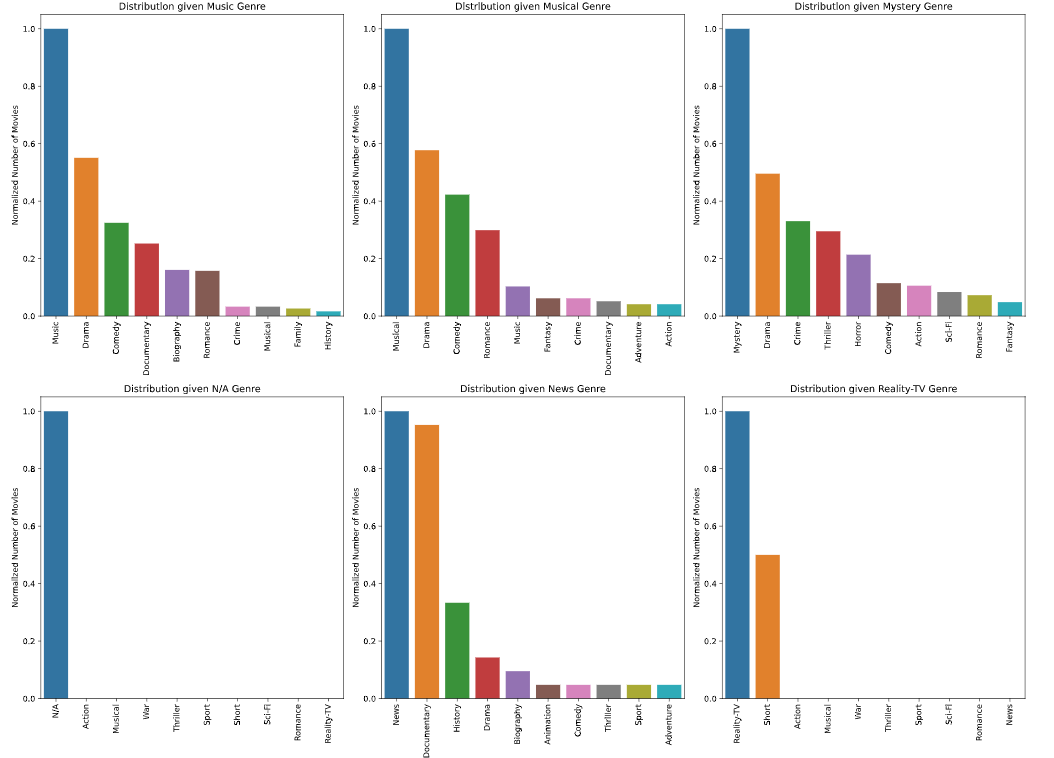
## 4.2. Sự phân bố của các thể loại phim

Xem xét nếu một bộ phim thuộc một thể loại nhất định, thì nó có thể thuộc các thể loại khác là gì..

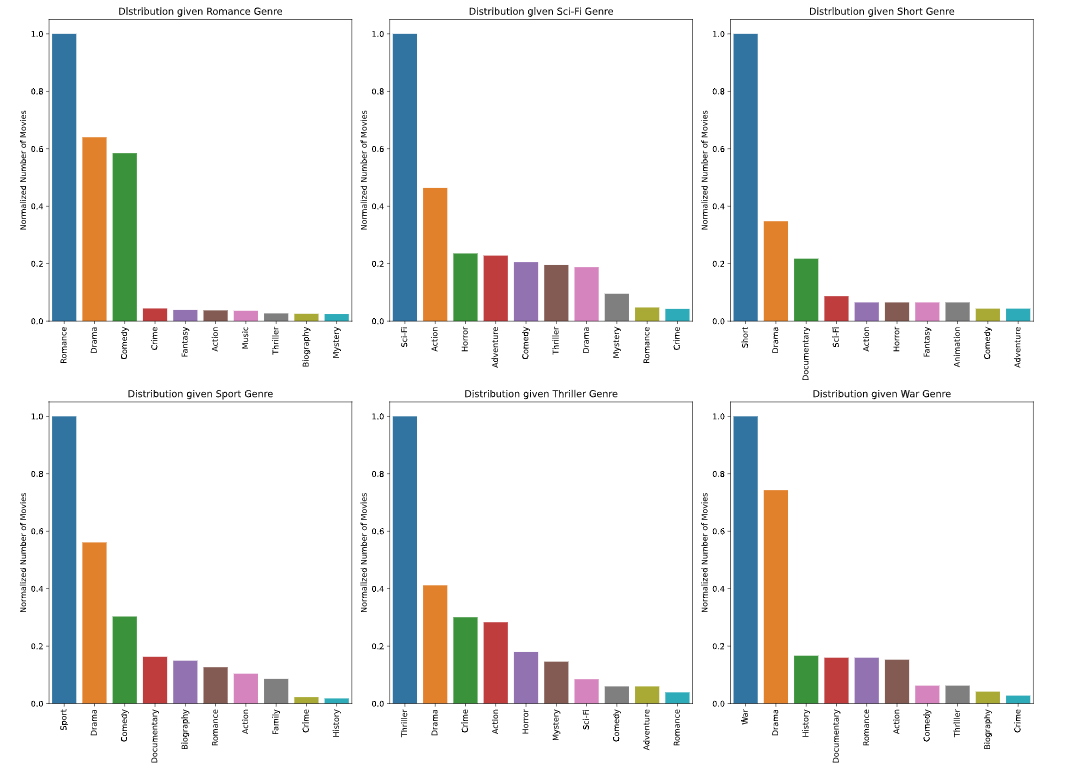


Hình 5. Biểu đồ tương quan các thể loại phim





Hình 6. Biểu đồ tương quan các thể loại phim





Hình 7. Biểu đồ tương quan các thể loại phim

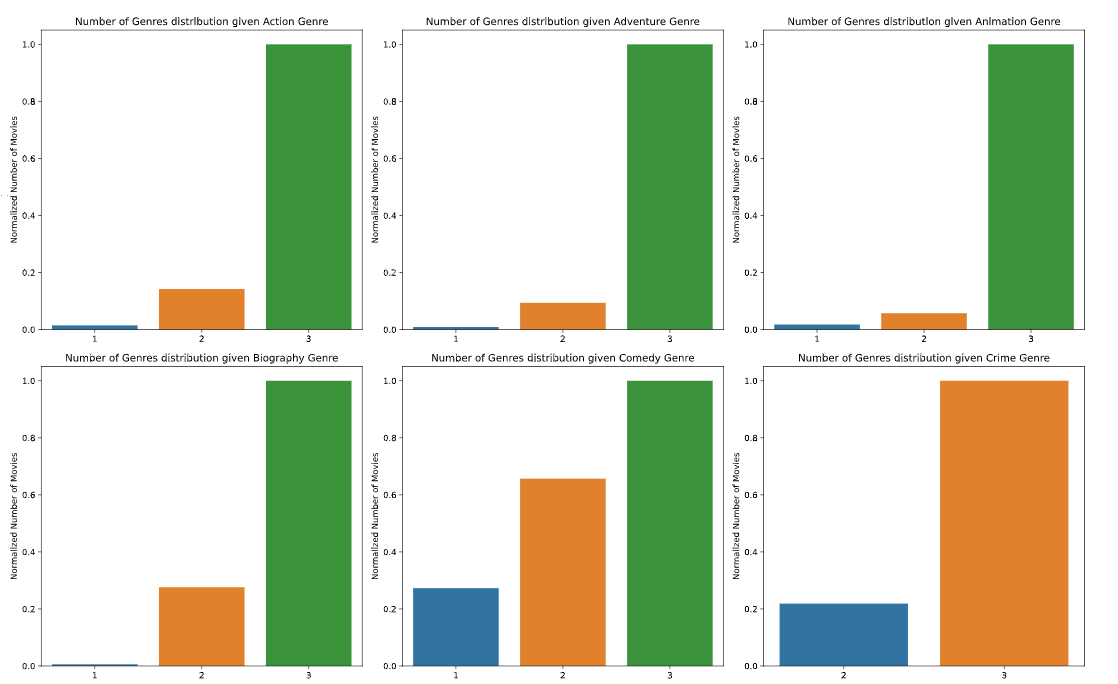
Từ các biểu đồ trên, chúng ta có thể thấy các mối tương quan sau:

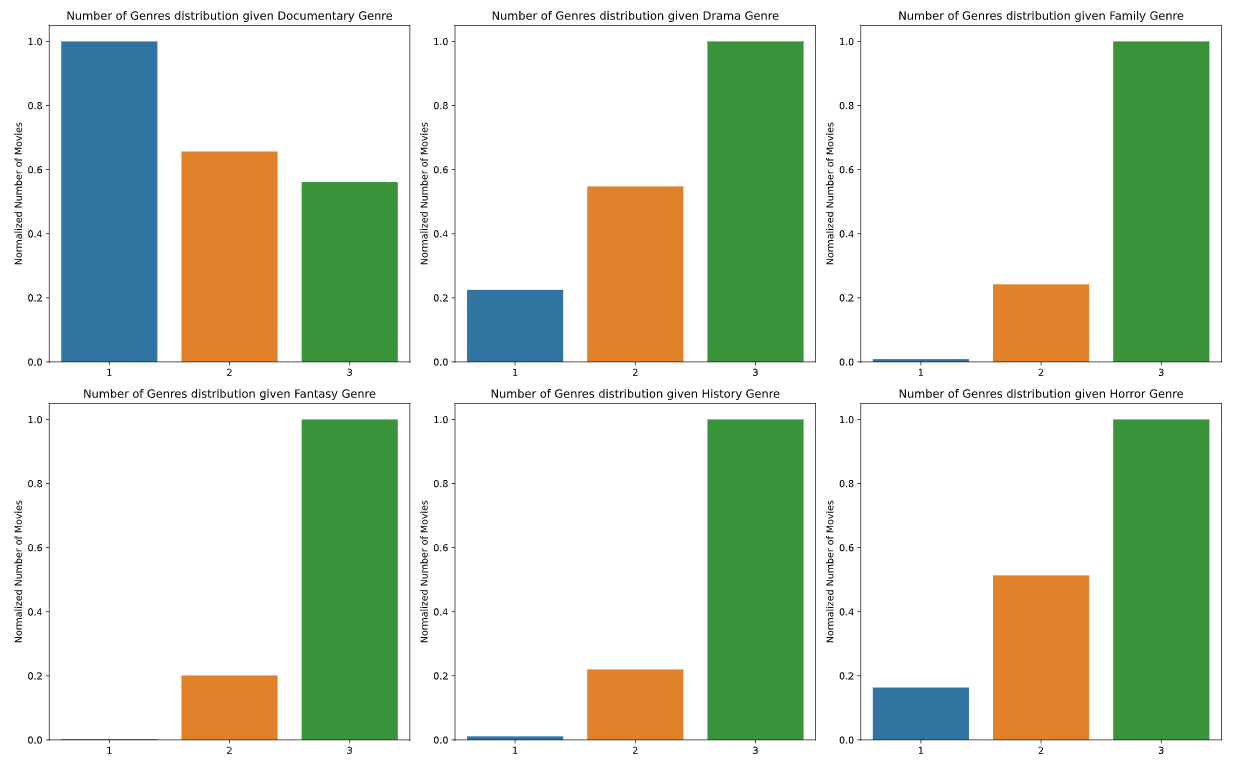
* Khoảng một nửa phim Adventure thuộc thể loại Action
* Trên 50% phim thuộc Animation movies cũng thuộc Adventure và Comedy
* 80% phim Game-Shows cũng thuộc thể loại Reality TV
* 65% phim Musical cũng thuộc thể loại Comedy
* 90% phim New cũng thuộc thể loại Documentary
* 50% phim Reality-TV cũng thuộc thể loại Short

## 4.3. Số lượng thể loại phim trên từng thể loại

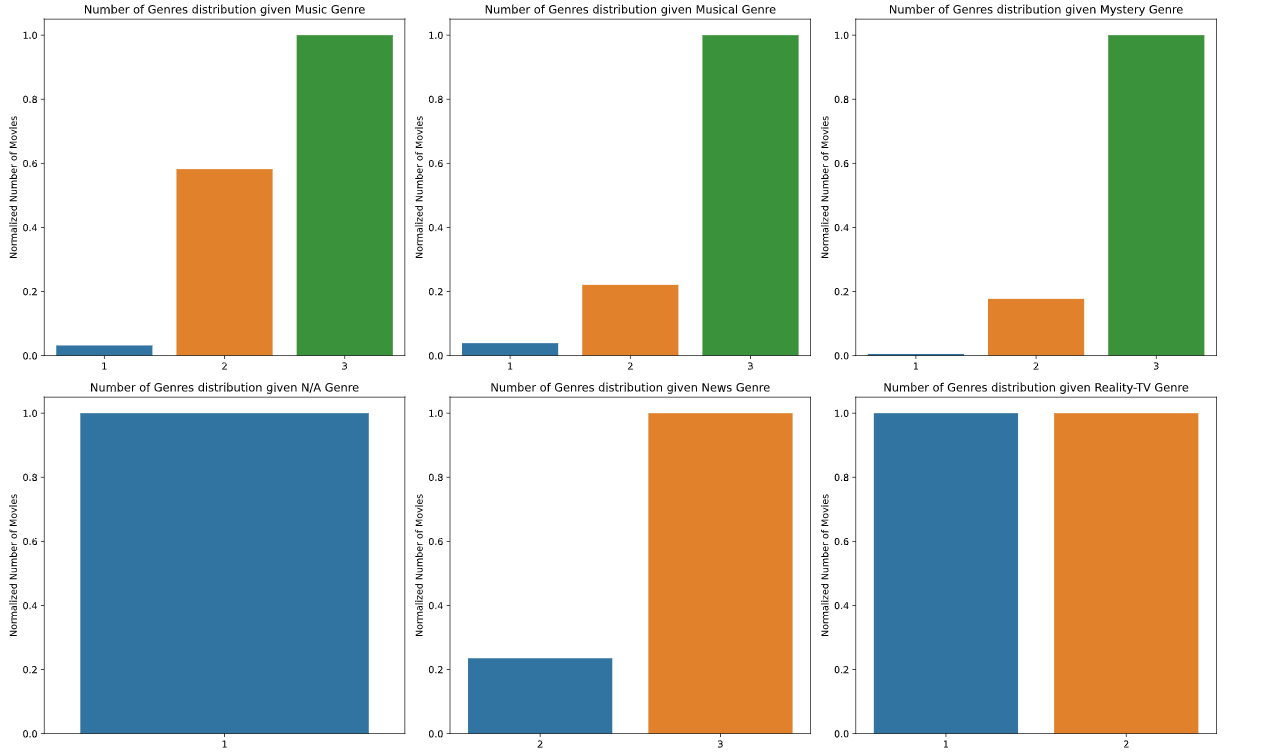
Ở đây có thể quan sát số lượng thể loại của mỗi phim theo từng thể loại:

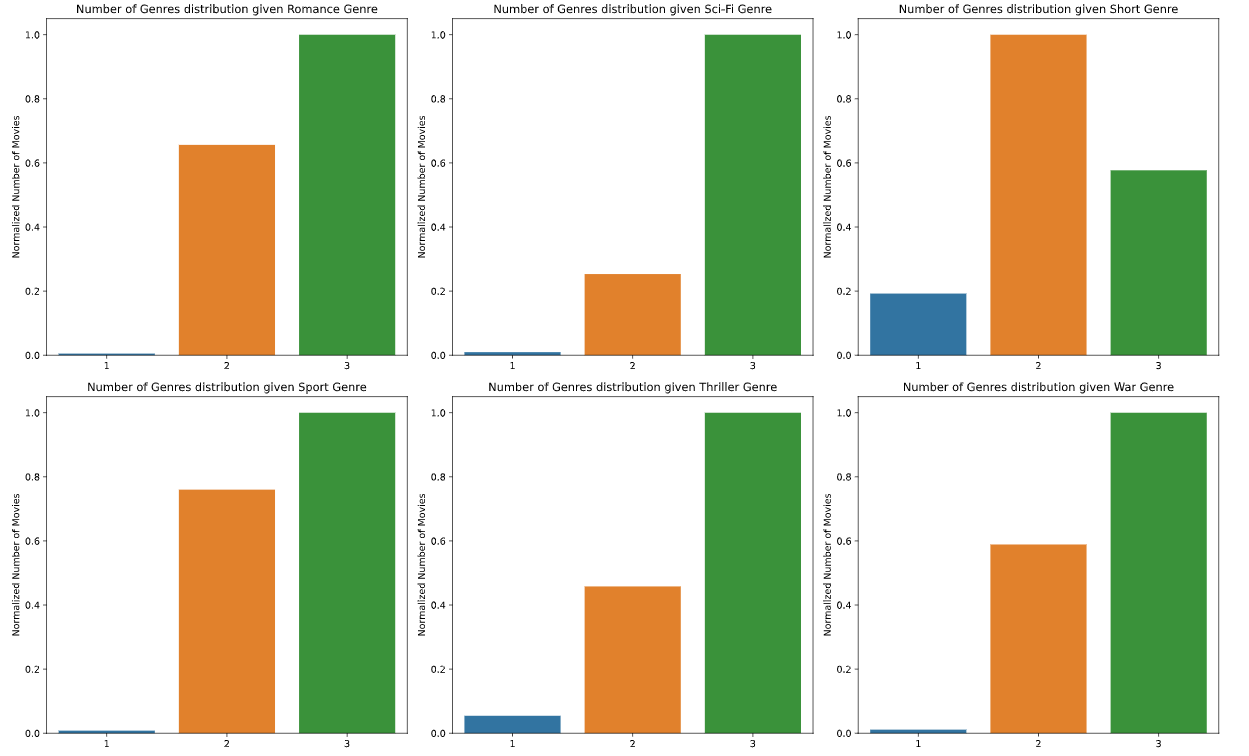
* Mọi bộ phim sẽ có tối đa 3 thể loại.
* Hầu hết các phim có 2 hoặc 3 thể loại.





Hình 8. Biểu đồ số lượng thể loại của mỗi phim







Hình 9. Biểu đồ số lượng thể loại của mỗi phim

# Chương 5: MODEL

## 5.1. Multi-label classification

### 5.1.1. Multi-label classification

Phân loại nhiều nhãn là một kỹ thuật phân tích văn bản AI tự động gắn nhãn (hoặc thẻ) văn bản để phân loại văn bản đó theo chủ đề. Điều này khác với phân loại nhiều lớp vì nhiều nhãn có thể áp dụng nhiều hơn một thẻ phân loại cho một văn bản.

**Phân biệt phân loại nhiều lớp và phân loại nhiều nhãn:**

Bộ phân loại nhiều lớp được thiết kế để cung cấp cho mỗi đoạn văn bản chỉ một thẻ danh mục (hoặc lớp).Ví dụ: nếu đang phân loại “Loại hình giải trí”, các thẻ phân loại có thể là: Sách, Phim, Chương trình truyền hình , v.v.

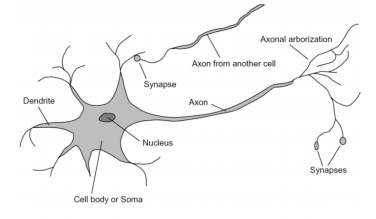
Trong ví dụ này, thẻ cho “Khi Harry gặp Sally” đối với phân loại nhiều lớp sẽ là: Phim.

Mặt khác với phân loại nhiều nhãn thì có thể xuất ra nhiều thẻ cùng một lúc. Có thể gắn thẻ “Phim” theo thể loại: Hài, Chính kịch, Lãng mạn, Kinh dị , v.v. Trong trường hợp này, “Khi Harry gặp Sally” có thể được gắn thẻ (hoặc nhiều nhãn) cả Hài kịch và Lãng mạn.

### 5.1.2. CNN

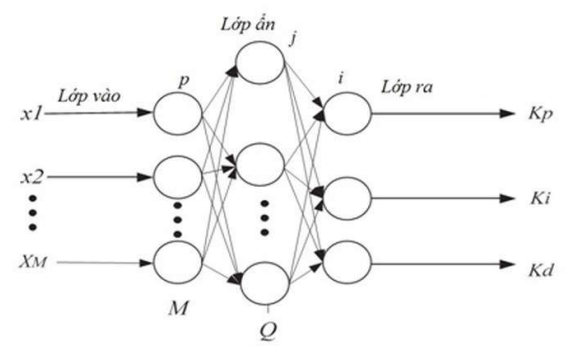
#### 5.1.2.1. Giới thiệu

Mạng nơ ron nhân tạo được thiết kế nhằm mô phỏng mạng neuron của bộ não người. Mạng neuron nhân tạo được cấu thành từ một tập các phần tử xử lý đơn giản được kết nốivới nhau. Mỗi phần tử xử lý này chỉ có thể thực hiện được một thao tác tính toán nhỏ, nhưng một mạng lưới các phần tử như vậy có một khả năng tính toán lớn hơn rất nhiều. Phần tử tính toán cơ bản của mạng nơ ron là một perceptron hay một nơ ron.



Hình 10. Mạng nơron của người (từ Rob Fergus)

Một nơ ron mô phỏng quá trình tính toán của bộ não con người. Dữ liệu được đưa tới các nơ ron thông qua các Dendrite vào Nucleus để tính toán. Tín hiệu ra được xuất ra ở dây Axon. Các nơ ron được liên kết với nhau thông qua các dây Synapse. Mạng nơ ron bao gồm rất nhiều phần tử như vậy liên kết với nhau.



Hình 11. Mạng nơron nhiều tầng

Hình 7 minh họa một mạng nơ ron truyền thẳng với 3 loại node sau:

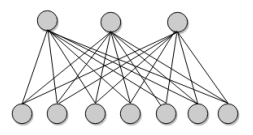
* Input nodes (Node đầu vào): Chứa dữ liệu đầu vào từ bên ngoài và đưa trực tiếp vào các Hidden nodes.
* Hidden nodes (Node ẩn): Nó không chứa kết nối trực tiếp đến dữ liệu từ bên ngoài. Nó thực hiện tính toán các dữ liệu nhận được từ các input nodes, thực hiện tính toán và đưa ra các output nodes. Tập hợp các node ẩn trong mạng tạo thành tầng ẩn. Một mạng nơ ron truyền thẳng có thể có hoặc không có tầng ẩn.
* Output nodes (Node đầu ra): Có nhiệm vụ tính toán và đưa dữ liệu từ trong mạng ra bên ngoài.

#### 5.1.2.2. Mạng nơ ron tích chập

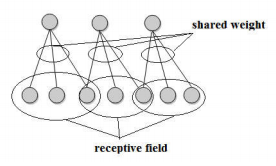
Mạng nơ ron tích chập là một mạng neuron nhân tạo với các toán tử tích chập. Nó có khả năng học một lượng lớn các dữ liệu trong khoảng thời gian ngắn hơn nhiều so với mạng nơ ron thông thường. Lý do là nó sử dụng ít trọng số hơn trong khi độ chính xác chỉ kém hơn một phần nhỏ so với kiến trúc truyền thống. Mô hình này sử dụng khá tốt trong bài toán phân loại ảnh.

Trong mạng nơ ron truyền thống, các node ở các tầng phía sau sẽ liên kết với toàn bộ các node ở layer phía dưới thông qua một tập các trọng số. Với mỗi nơ ron khác nhau, chúng ta cần một tập trọng số hoàn toàn độc lập để liên kết với các nơ ron ở tầng trước đó.

Điểm khác biệt của mạng nơ ron tích chập so với mạng nơ ron truyền thống đó là trong liên kết giữa 2 tầng liên tiếp nhau việc các node ở các tầng phía sau chỉ liên kết với một bộ phận các node ở tầng phía trước đó gọi là receptive field thông qua một tập các trọng số. Hơn nữa tập trọng số này là như nhau đối với mỗi nơ ron ở tầng sau. Do đó số lượng tham số cần huấn luyện ít hơn trong khi vẫn giữ được lượng thông tin cần thiết.



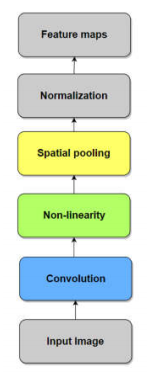
Hình 12. Kết nối giữa các tầng trong mạng nơ ron truyền thống



Hình 13. Kết nối giữa các tầng trong mạng nơ ron tích chập

Một mạng nơ ron tích chập thường được thực hiện thông qua các bước sau:

* Convolutional layer.
* Pooling layer.
* Non-linearity layer.
* Fully-connected layer.

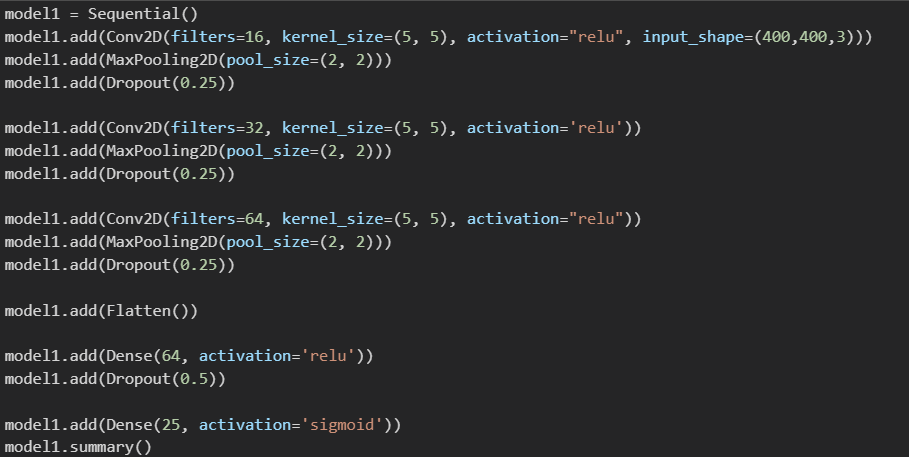


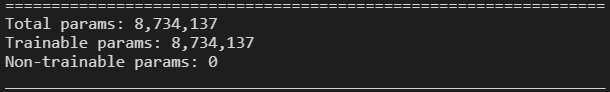
Hình 14. Các bước cơ bản trong mạng neuron tích chập

Trên đây là các khối cơ bản trong một mạng neuron tích chập, mạng neuron tích chập được cấu thành bằng việc xếp trồng nhiều các lớp thành một cấu trúc chặt chẽ.

## 5.2. Xây đựng Model

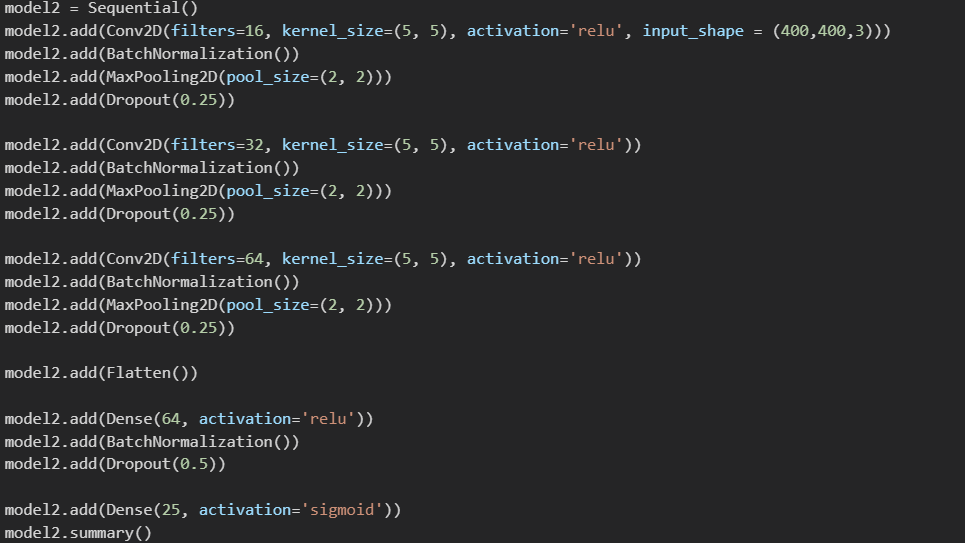
Ở đây nhóm sẽ sử dụng 3 model với cấu trúc như bên dưới:

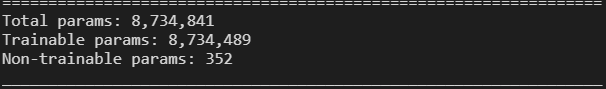




Hình 15. Model 1

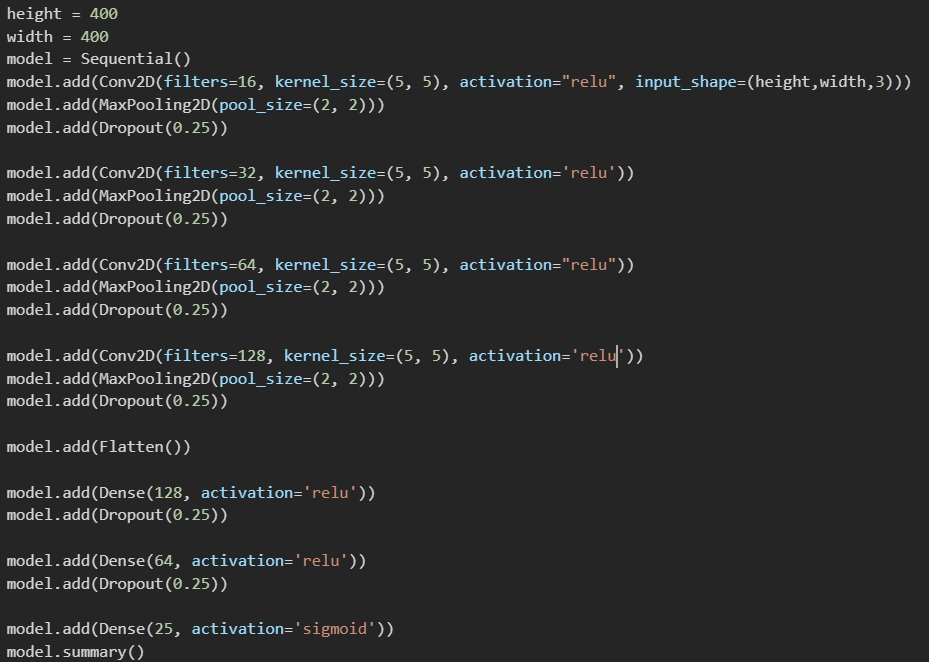
Model 2 được thêm các lớp BatchNormalization để giảm độ ảnh hưởng giữa các lớp với nhau.

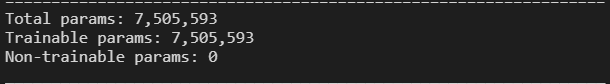




Hình 16. Model 2

Model 3 có thêm 2 lớp ẩn giúp giảm trọng số.





Hình 17. Model 3

Vì kích thước của poster không cố định nên đầu vào sẽ được resize về kích thước (400x400).

## 5.3. UnderFitting & OverFitting

Đối với model1 và model2, mỗi model sẽ được huấn luyện trên tập dữ liệu trên 50 lần. Validation\_data được dùng chính là test\_data.



Bên dưới là độ thị hàm mất mát của quá trình huấn luyện:



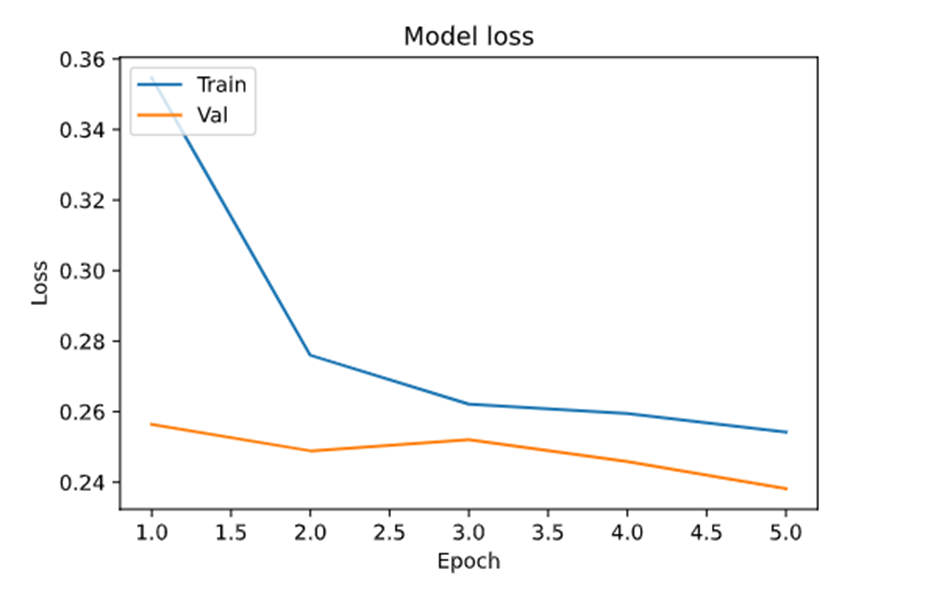


Hình 18. Đồ thị hàm mất mát Model 1&2

Vì ở 2 model trước đều bắt đầu overfitting ở lần huấn luyện thứ 5 model 3 nhóm sẽ chỉ huấn luyện 5 lần nhưng với validation\_set được lấy từ train\_set.



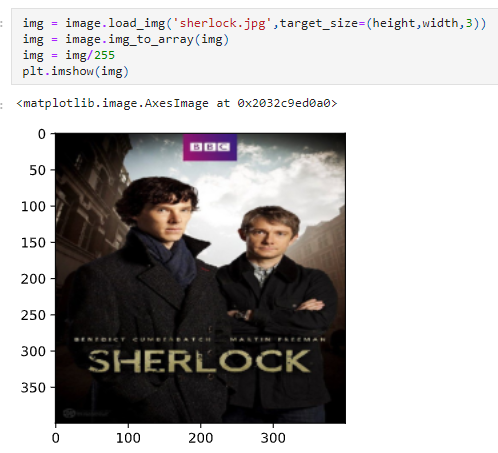
Sau khi huấn luyện model3 cho kết quả tốt hơn so với 2 model trước:



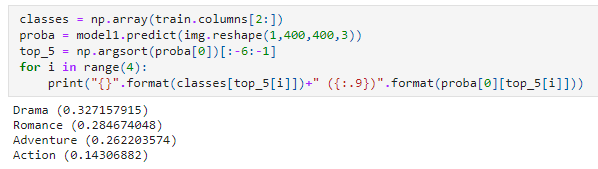
Hình 19. Đồ thị hàm mất mát Model 3

## 5.4. Kiểm tra kết quả

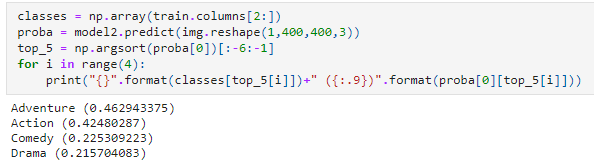
Nhóm sẽ dùng cả 3 model để dự đoán thể loại của poster phim sherlock và chọn 4 thể loại có khả năng cao nhất:



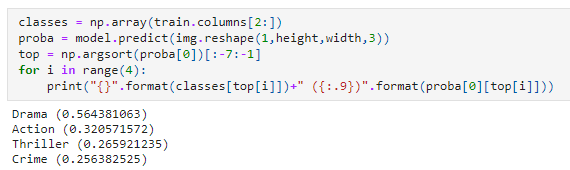
Hình 20. Poster dự đoán



Hình 21. Kết quả 4 thể loại có khả năng cao nhất của Model 1



Hình 22. Kết quả 4 thể loại có khả năng cao nhất của Model 2



Hình 23. Kết quả 4 thể loại có khả năng cao nhất của Model 3

Theo như kết quả thì model thứ 3 cho dự đoán sát với thực tế nhất trong cả 3 model.

# Chương 6: TỔNG KẾT

## 6.1. Hạn chế

Vì dataset bị mất cân bằng nên khuynh hướng dự đoán bị kéo về các class có nhiều trọng số hơn.

Vì thiếu bộ nhớ nên nhóm chỉ có thể sử dụng 1 phần của dataset trong quá trình huấn luyện.

Chưa thể phân chia train\_set và test\_set theo tỉ lệ đồng đều cho từng thể loại.

## 6.2. Hướng phát triển

Dùng Early Stopping để ngăn model overfitting.

Sử dụng Kerras.Sequence để tạo một datagenarator giúp load ảnh trong lúc huấn luyện nhằm sử dụng toàn bộ tập dữ liệu cho việc huấn luyện.

# TÀI LIỆU THAM KHẢO

1. Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media.
2. Wei-Ta Chu and Hung-Jui Guo, “Movie Genre Classification based on Poster Images with Deep Neural Networks,” Proceedings of International Workshop on Multimodal Understanding of Social, Affective and Subjective Attributes, pp. 39-45, 2017. (in conjunction with ACM Multimedia 2017).
3. Rachel Wolff. Multi-Label Classification: Overview & How to Build A Model.

Truy cập tại: https://monkeylearn.com/blog/multi-label-classification/

1. (PDF) A multimodal approach for multi-label movie genre classification (researchgate.net).
2. Jason Brownlee. (2017). Gentle Introduction to the Adam Optimization Algorithm for Deep Learning.

Truy cập tại: https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/

1. Aditya Mishra. (2018). Metrics to Evaluate your Machine Learning Algorithm.

Truy cập tại: https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234

1. How to use sklearn train test split to stratify data for multi label classification.

Truy cập tại: https://datascience.stackexchange.com/questions/45174/how-to-use-sklearn-train-test-split-to-stratify-data-for-multi-label-classificat