

# Movie Genre Classification Based on Plot

Dishank Aggarwal, 21Q05R003

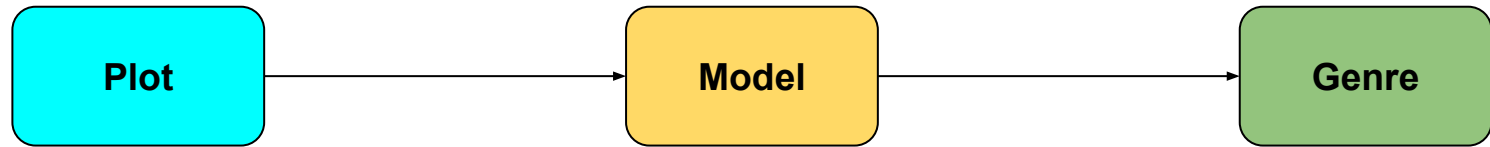
Kishan Maharaj, 22M2122

Rupasai Rangaraju, 213059006

08/11/2022

# Problem Statement

Given a multi-genre plot with a genre set  $\mathbb{L} = \{l1, l2, l3, \dots, lK\}$ , where  $K$  is the total number of genres predict a genre subset  $y_i = \{y_{i1}, y_{i2}, \dots, y_{in}\}$  where  $n$  is the size of target genre.



$$x_i = \{Tok_i1, Tok_i2, \dots, Tok_iL\}$$

$$y_i = \{y_{i1}, y_{i2}, \dots, y_{in}\}$$

# Related Works

- Naiyin Liu, Qianlong Wang & Jiangtao Ren. **Label-Embedding Bi-directional Attentive Model for Multi-label Text Classification**. Neural Processing Letters volume 53, pages 375–389 (2021)
- Ali Mert Ertugrul, Pinar Karagoz. **Movie Genre Classification from Plot Summaries using Bidirectional LSTM**. 2018 12th IEEE International Conference on Semantic Computing

# Dataset

- Around 35,000 detailed movie plot description
- Columns
  - Release Year
  - Title
  - Origin/ Ethnicity
  - Director
  - Cast
  - Genre (multilabel)
  - Wiki Page
  - Plot

[https://www.kaggle.com/code/aminejallouli/genre-classification-based-on-wiki-movies-plots/data?select=wiki\\_movie\\_plots\\_deduped.csv](https://www.kaggle.com/code/aminejallouli/genre-classification-based-on-wiki-movies-plots/data?select=wiki_movie_plots_deduped.csv)

# Dataset Cleaning

- Total Genres Initially: 2265
- Inconsistency
  - Different Separators : “/”, “ ”, “-” etc.
  - Synonyms: Romance and Comedy - > “rom com” || “romantic comedy”
  - Solution: Regex and Manual Annotation
- Redundant Values and Sparsity
  - Examples : “dark slasher” , “imax”, cast names etc.
  - Solution: Frequency analysis
- Spelling mistakes (Genres)
  - Examples: “drama” - > “dram”
  - Manual Annotation

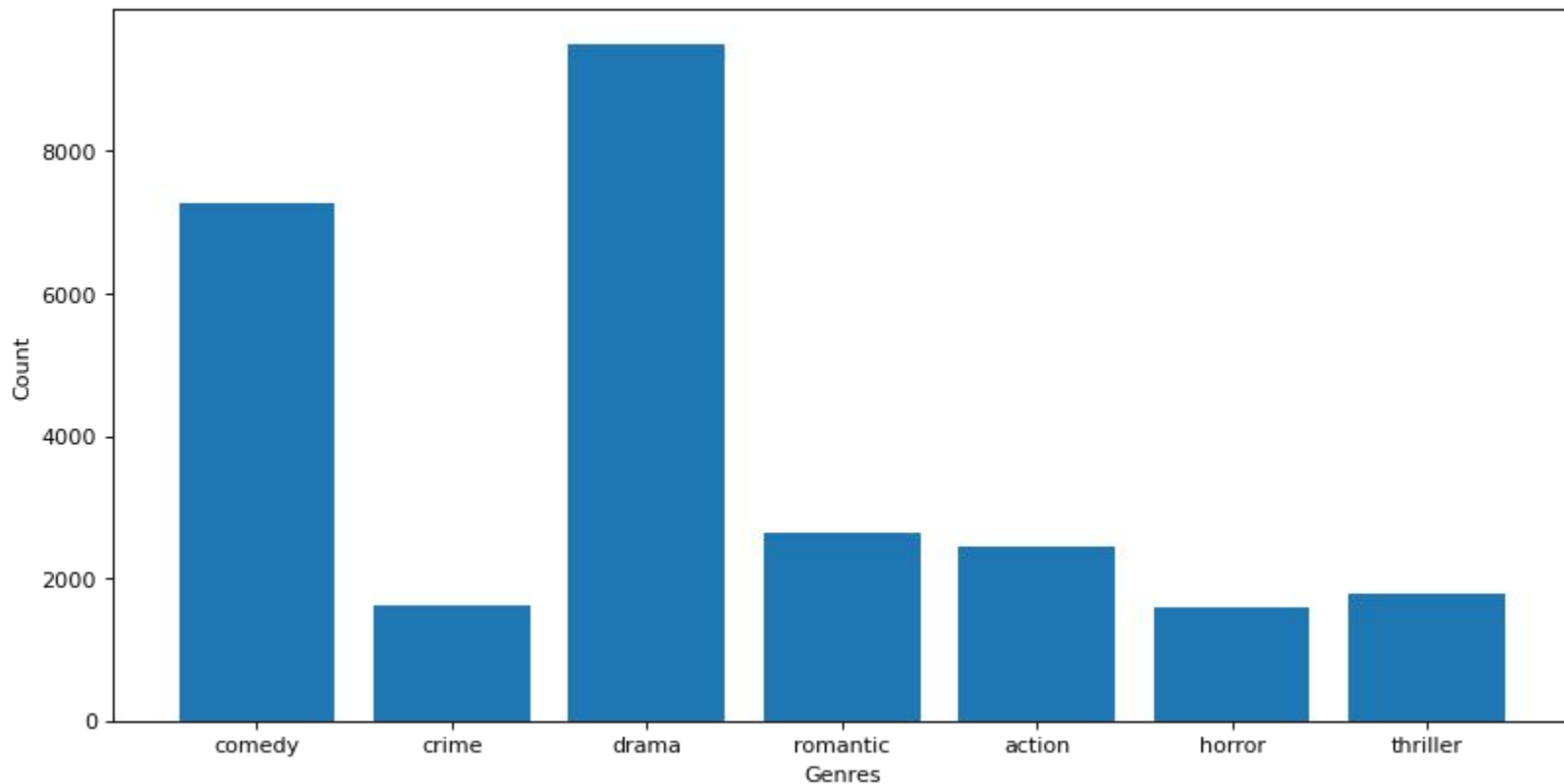
# Dataset (Initially preprocessed)

	dramedy	biblical	miniseries	eiga	bhoomika chawla	sports (volleyball)	environmental film	samurai	human rights	radhika joshi	...	ballard	exploitation	noir	in	concert film	psycho	i
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
34881	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
34882	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
34883	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
34884	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
34885	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	

# Dataset (Cleaned)

	Plot	comedy	crime	drama	romantic	action	horror	thriller
0	The film is about a family who move to the sub...	1	0	0	0	0	0	0
1	The film features a train traveling through th...	0	1	0	0	1	0	0
2	Before heading out to a baseball game at a nea...	1	0	0	0	0	0	0
3	The plot is that of a black woman going to the...	1	0	0	0	0	0	0
4	On a beautiful summer day a father and mother ...	0	0	1	0	0	0	0
...	...	...	...	...	...	...	...	...
22164	Zeynep lost her job at weaving factory, and he...	0	0	1	0	0	0	0
22165	Two musicians, Salih and Gürkan, described the...	1	0	0	0	0	0	0
22166	Zafer, a sailor living with his mother Döndü i...	1	0	0	0	0	0	0
22167	The film centres around a young woman named Am...	1	0	0	1	0	0	0
22168	The writer Orhan Şahin returns to İstanbul aft...	0	0	0	1	0	0	0

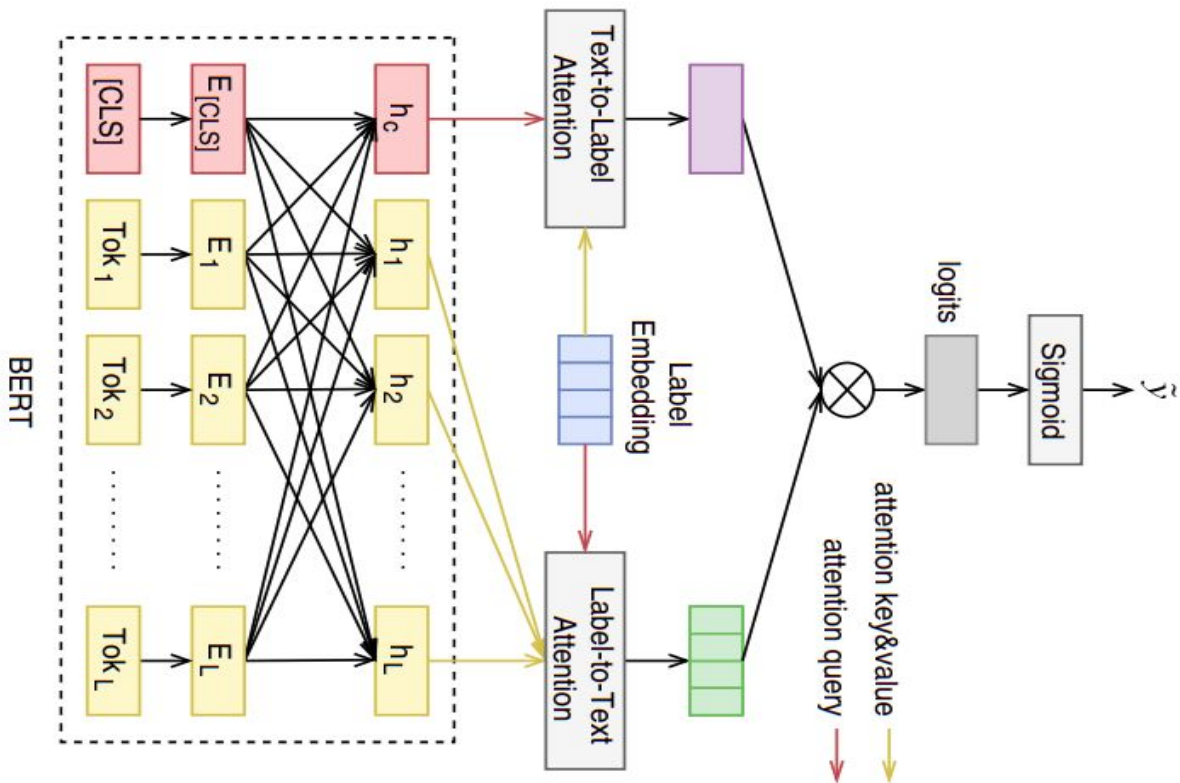
# Dataset Distribution (cleaned)



- Final Genres : 7 (22169)



# Architecture



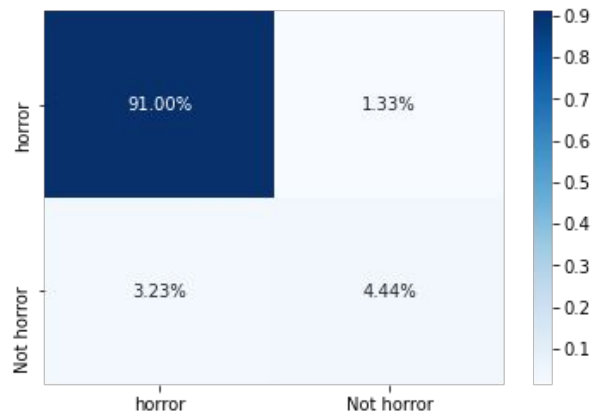
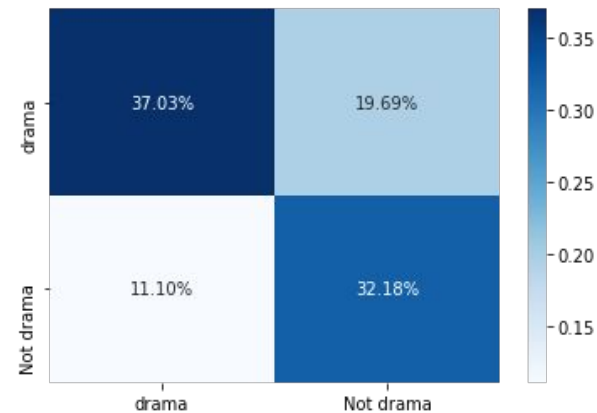
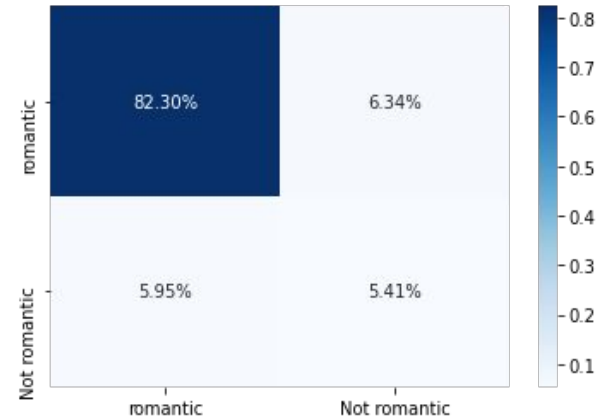
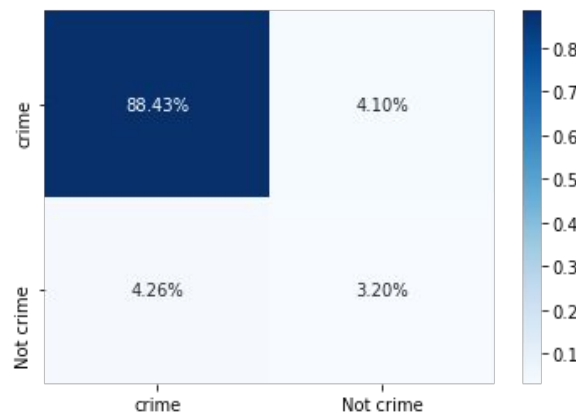
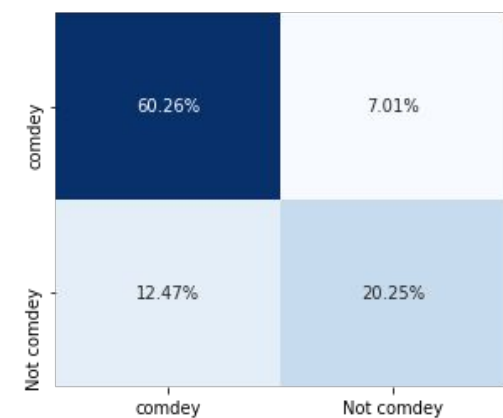
# Results

Genre	Precision (Macro)	Recall (Macro)	F1 (Macro)
Comedy	0.78	0.75	0.76
Crime	0.696	0.692	0.694
Drama	0.694	0.698	0.691
Romantic	0.696	0.70	0.699
Action	0.78	0.73	0.75
Horror	0.86	0.78	0.81
Thriller	0.65	0.65	0.65

# Results Overall

Metric	Values
Hamming Loss	0.135
Accuracy	0.8646497841355758
Precision	0.7388193858424178
Recall	0.7170113446357045
F1	0.7253216378032165

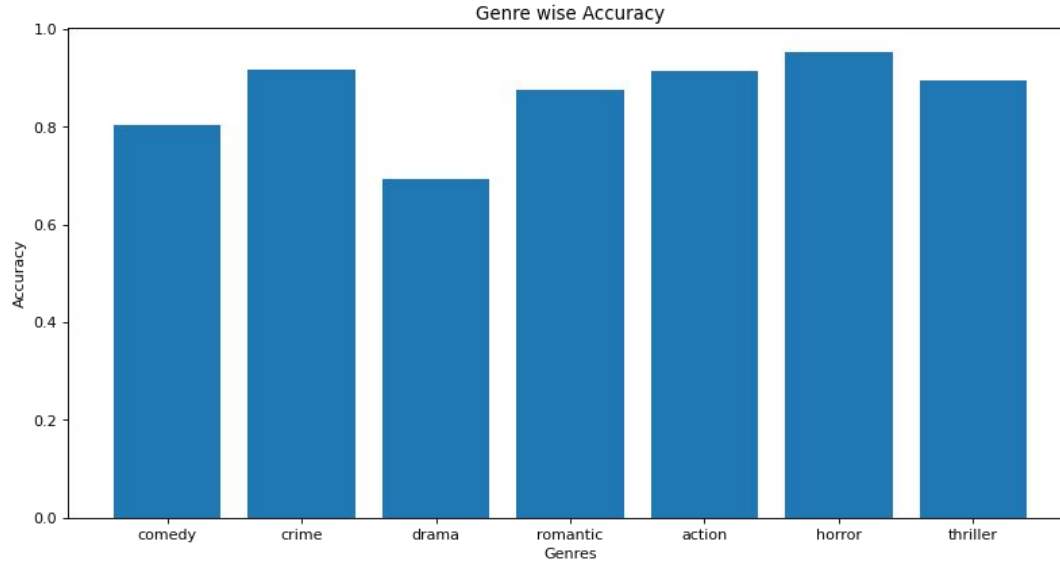
# Confusion Matrix



# Confusion Matrix

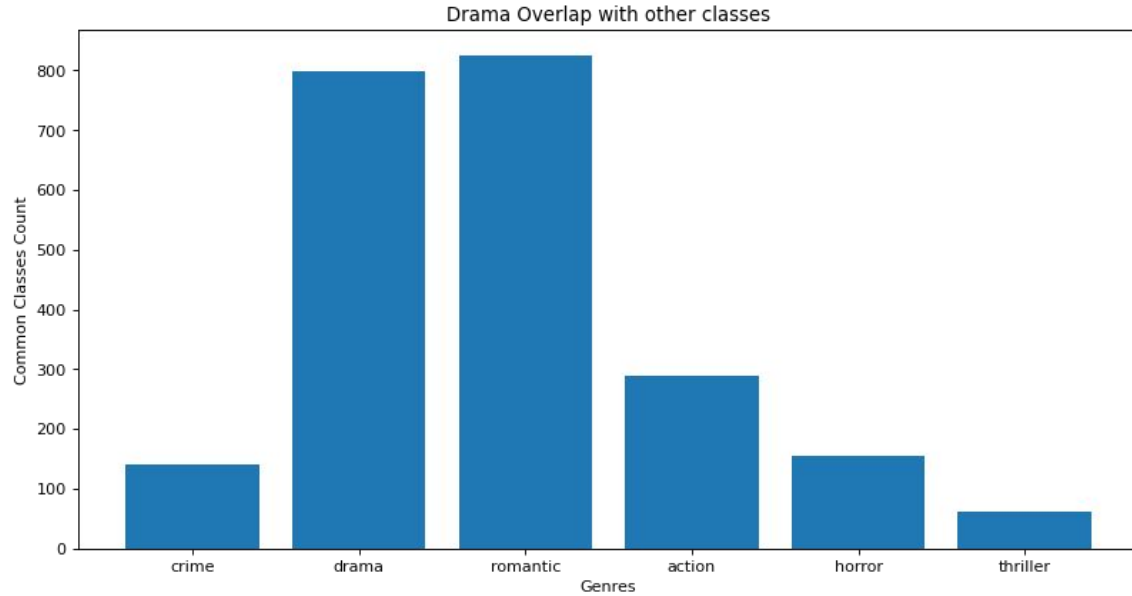


# Analysis



- Classes with lesser overlap with other classes performed better
- Classes with lower accuracy had lesser overlap

# Analysis : Comedy



- 28% of Comedy plots overlaps with other classes

# Analysis : Comedy Examples

**True:** ["comedy", "Romantic"]

**Predicted:** ["Drama", "Romantic"]

Ralph Prescott (Marmont) is a New York divorce lawyer tired of his clientele. Woodbury (Pallette), who runs a ladies hosiery business across the hall, suggests that they get away from the city and camp in Mantrap, Canada. Bachelor Joe Easter (Torrence) runs a dry-goods store in Mantrap. Joe, wanting female company, goes to Minneapolis. In a barbershop there, backwoods Joe meets flirtatious manicurist Alverna (Bow), who agrees to meet Joe for dinner...

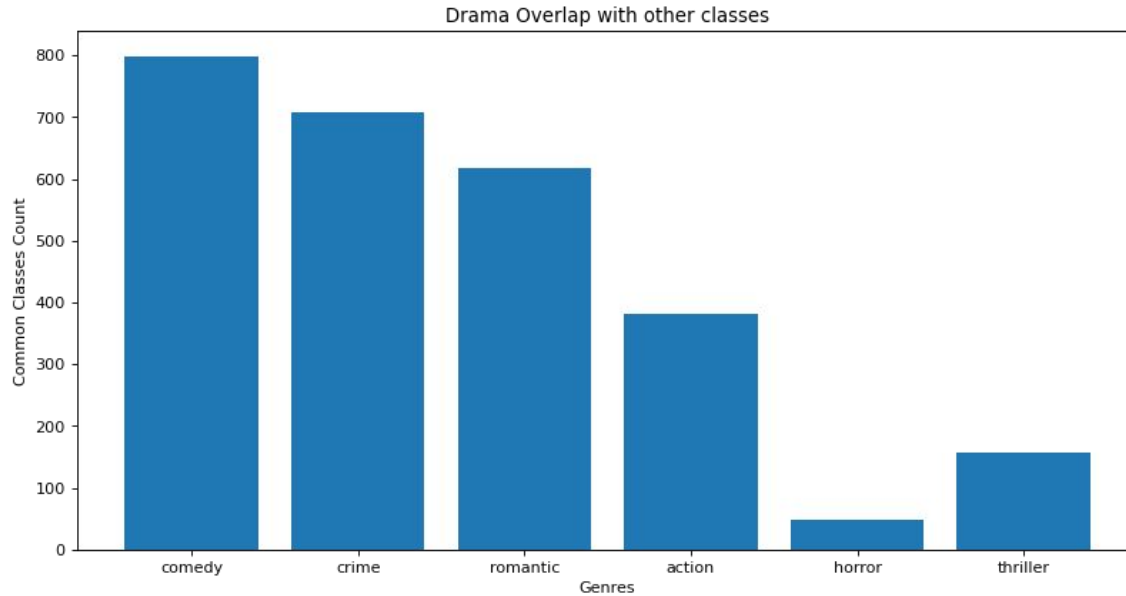
**True:** ["comedy", "Romantic"]

**Predicted:** ["Drama"]

Ann Barton (Laura La Plante), a girl from a once-wealthy family, must make a living by clerking in a cigar store. There she meets and falls in love with James McDonald (Pat O'Malley), a cigar salesman. She is then adopted by Margaret Wyndham (Hedda Hopper), her rich and aristocratic aunt, who disapproves of James due to his crude manners...



# Analysis : Drama



- 25% of Drama plots overlaps with other classes

# Analysis : Drama Examples

**True:** ["Drama", "Romantic"]

**Predicted:** ["Romantic"]

The Queen of Sardalia is in a bad marriage with the brutal King Constantine II. She decides to get away from her normal life for a period and goes on vacation to Switzerland. There, she meets Paul Verdayne. They have an affair, which lasts for three weeks.

**True:** ["Drama", "Action"]

**Predicted:** ["Action"]

Antar is sent by Selima, head of the Ottoman Empire, to prevent Pasha Hammam from attempting to overthrow the emperor.

Selima blames Hammam and his assassin Kasseim for the death of her father.

Kasseim's wife, Rosanna, falls in love with Antar, but he wants Selima for himself.

# Demo

## JAMES BOND

<b>PLOT</b>	<b>OUTPUT</b> 0.0s
MI6 operative James Bond gains his licence to kill and promotion to 00 agent sta	['action']
<b>Clear</b>	<b>Submit</b>
<b>Flag</b>	

## LA LA Land

<b>PLOT</b>	<b>OUTPUT</b> 0.0s
While stuck in Los Angeles traffic, aspiring actress Amelia "Mia" Dolan has a mon	['drama', 'romantic']
<b>Clear</b>	<b>Submit</b>
<b>Flag</b>	

## American Horror Story

<b>PLOT</b>	<b>OUTPUT</b> 0.0s
An anthology series centering on different characters and locations, including a h	['horror']
<b>Clear</b>	<b>Submit</b>
<b>Flag</b>	

# Demo

## Mission Impossible 6

PLOT

Two years after Solomon Lane's capture, the remnants of his organization The Sy

Clear

Submit

OUTPUT

0.0s

['action']

Flag

## Kal ho na ho

PLOT

Naina Catherine Kapur is a pessimistic, uptight MBA student who lives in New Yoi

Clear

Submit

OUTPUT

0.0s

['drama', 'romantic']

Flag

## Dear Zindagi

PLOT

Based in Mumbai and Goa, Kaira is a promising cinematographer who dreams of

Clear

Submit

OUTPUT

0.0s

['drama']

Flag

**Thank You !!!**