DAI Project Report on

Removing Bias from Movie Reviews

Submitted by:
Aarti Pol
(M20MA002)

Indian Institute Of Technology Jodhpur

Colab Link:
https://colab.research.google.com/drive/1Se8kfnvqmxLekp9xSJGDZpOg5WMQCmkU#scrollTo=2c5dd4a8-c759-46f4-a195-20cf5993a029

Abstract:

Movie reviews are an integral part of helping users decide whether to watch a movie or not. If a movie has raving reviews users are very much inclined to watch it either at home or in theatre. Similarly, bad reviews affect users' decisions in a negative manner. Thus it is important that the reviews are not biased towards a specific aspect of a movie like a genre, actor, actress, etc. But it's not easy to write an unbiased movie review because of our own personal biases. When we train ML models on movie reviews datasets, they also tend to pick up these implicit biases. For example,

Review1: Nice thriller. Overall it was a good story but the ending was not satisfying.
Review1': Nice comedy. Overall it was a good story but the ending was not satisfying.

Review2: The movie starred Amir Khan and Dipika Padukone. Overall it was a good story but the ending was not satisfying.
Review2': The movie starred Arjun Kapoor and Katrina Kaif. Overall it was a good story but the ending was not satisfying.

Here, the model should classify both sets of reviews in the same bucket (neutral/positive) with similar probabilities. However, if the genre Thriller or actors Amir Khan/Dipika Padukone are associated with highly rated movies then the model will be implicitly biased towards giving higher probability scores to review1 and review 2. This is a problem because we want the model to rate a movie based on the sentiments expressed towards the movie and not w.r.t other tangential information present in the reviews like genre or celebrity. So, we need to learn models which are not influenced by such information.

Problem Statement:
As discussed above, a model trained on movie review data can be biased towards some other information apart from the actual sentiment in the review. But we cannot ask users to not give such information in the reviews, this is something that we have to deal in the modeling part. Hence in this project, I will try to

1. Analyze how biased these models are w.r.t. celebrity name or genre
2. Come up with a method to remove this bias
3. Evaluate the model's performance after removing the bias

Proposed Solution:
I will try to solve this problem using the domain adaptation method described in the paper [5]. This method seemed to have worked well on removing gender/racial biases in various NLP tasks discussed above. So, the idea is to train a model in such a manner that it is able to learn the overall category/rating of review while not focusing on the protected information (genre, celebrity etc.).

Dataset you will be using. If not already present, how are you planning to create it?
There are few movie review datasets [6,7] but these do not contain additional information like an actor, actress, or genre information. So, the first task will be to build a dataset that contains <movie, genre, review, rating> tuple. I will try to look for some open source solution, if not I will use libraries like requests, BeatifulSoup to fetch and parse the data from IMDb.

Dataset Creation:

From my search, I couldn't find a movie review dataset which contains both movie info, reviews of movies, and their ratings. So, I have created my own movie review dataset from IMDB using the requests and BeautifulSoup and IMDbPY library.

- IMDbPY, we can get movie details and reviews as well but the library doesn't provide the rating for these reviews.
- Hence, I first fetched data for ~600 movies from IMDbPy by using functions get_top250_movies(), get_bottom100_movies() and get_top250_indian_movies().
- After that, I used the requests library to fetch the reviews page for these movies.
  - BeautifulSoup was used to parse the review page to get the review rating and content. For each movie, 20-25 reviews were fetched.

Following css selector rule was used to parse the data:

- reviews: elems = document.querySelectorAll(".review-container")
- rating: elems[0].querySelector(".ipl-ratings-bar").innerText
- review_text: elems[0].querySelector(".text").innerText

Sample movie + review data

movie	rating	genre_primary_mapped	review_title	review	review_rating	review_class_bin
Lakshya	7.8	action	Let your life not be sans any aim	Lakshya (aim) is the exemplary story of a care...	6/10	bad
Ben-Hur	8.1	adventure	Chariots of Fire	Ben-Hur is famous for its chariot race sequenc...	7/10	good
Gunday	3.3	action	Gunday (U/A) Hindi ----------- my Rating: ★★★...	New movie Reviews and lots more Hot news	8/10	good
Pan's Labyrinth	8.1	drama	Marvelous movie in which a bookish young stepd...	Set in the not so tranquil Spanish woodlands o...	8/10	good
Network	8.1	drama	Arch satire	NETWORK is one of those well-remembered '70s m...	6/10	bad

Data Analysis:

- Total movies: 591
- Total reviews: 7,500

Number of reviews under each rating:

Rating	Count
10/10	3357
8/10	2409
9/10	2206
7/10	1464
6/10	895
1/10	881
5/10	475
3/10	460
4/10	408
2/10	404

With respect to the review class, I used two strategies:

1. 3 class review system:
   `{"bad": rating <=5, "average": rating >=6 and rating <=7, "good": rating >=8}`

good	3000
bad	2500
average	2000

2. Binary review system:
   `{"bad": rating <=6, "good": rating >=7}`

good	4239
bad	3261

- *We will see below that it's difficult to learn reviews with the average class as the model confused them with either bad/good reviews constantly*

Count of each genre:

Action	2221
Comedy	1622
Drama	1484
Crime	739
Adventure	452
Biography	363
Animation	295
Horror	187
Mystery	96

```
Western        26
Thriller        9
Film-Noir       6
```

Some of the categories are quite similar in nature. So we used the following strategy to merge them.

```
genre_dict = {  'Action':'action','Comedy':'comedy',
                'Drama':'drama',    "Crime":'thriller/crime',
                'Adventure':'adventure',
                'Biography':'biography',
                'Animation':'animation',  'Horror':'horror',
                'Mystery':'thriller/crime',
                'Western':'western',
                'Thriller':'thriller/crime',
                'Film-Noir':'thriller/crime'
             }
```

Count of each genre after genres mapping:

```
action              2221
comedy              1622
drama               1484
thriller/crime       850
adventure            452
biography            363
animation            295
horror               187
western               26
```

Converting genres to ids:

```
genre2id = { 'action': 0,
             'comedy': 1,
             'drama': 2,
             'thriller/crime': 3,
             'adventure':4,
             'biography':5,
             'animation':6,
             'horror':7,
             'western':8
           }
```

- *A movie can belong to multiple genres, however for this task we are considering the primary genre of the movie.*

Genre/Rating Statistics:

	review_class_bin_int	
	count	mean
genre_primary_mapped		
action	2221	0.531292
adventure	452	0.475664
animation	295	0.610169
biography	363	0.732782
comedy	1622	0.417386
drama	1484	0.710916
horror	187	0.219251
thriller/crime	850	0.712941
western	26	0.730769

- From the above table, we can see that biography, drama, thriller/crime are on average rating highly while comedy and horror movies are rated low on average.
- This also matches with what we observe from movie awards - most Oscar-winning movies are drama.

Data Analysis:

Analyzing vocabulary size:
All tokens are a number of tokens or words in the vocabulary.

```
all tokens: 2074273
total unique tokens: 131489
```

Frequency of tokens:

```
tokens occuring >=2 times: 51923
tokens occuring >=3 times: 36043
tokens occuring >=5 times: 23665
tokens occuring >=10 times: 13432
```

Top 100 tokens in vocabulary:

[('the', 116105), ('and', 60672), ('a', 56859), ('of', 50576), ('to', 46342), ('is', 41160), ('in', 33208), ('it', 22365), ('that', 22187), ('this', 18744), ('i', 18599), ('as', 17874), ('with', 16727), ('for', 16187), ('but', 14654), ('was', 13920), ('his', 13910), ('film', 11883), ('he', 11672), ('on', 10813), ('movie', 10513), ('are', 10366), ('not', 10358), ('by', 9116), ('have', 8565), ('be', 8547), ('one', 8460), ('an', 8130), ('you', 8111), ('from', 7998), ('has', 7290), ('at', 7283), ('who', 7020), ('all', 6843), ('like', 6357), ('they', 6197), ("it's", 6076), ('so', 6075), ('just', 5465), ('about', 5436), ('or', 5212), ('some', 4952), ('her', 4938), ('there', 4800), ('very', 4701), ('when', 4641), ('which', 4569), ('if', 4413), ('good', 4411), ('their', 4393), ('more', 4382), ('out', 4378), ('what', 4192), ('even', 4156), ('also', 4077), ('story', 3986), ('its', 3679), ('up', 3634), ('no', 3607), ('best', 3540), ('can', 3527), ('really', 3442), ('she', 3433), ('would', 3303), ('much', 3286), ('only', 3251), ('than', 3218), ('will', 3196), ('we', 3118), ('been', 3090), ('into', 3083), ('how', 3029), ('most', 3015), ('had', 2997), ('first', 2904), ('my', 2890), ('after', 2879), ('him', 2871), ('see', 2845), ('were', 2840), ('other', 2768), ('get', 2731), ('time', 2704), ('make', 2675), ('great', 2674), ('because', 2662), ('-', 2654), ('many', 2619), ('too', 2605), ('being', 2570), ('well', 2480), ('do', 2424), ('made', 2404), ('where', 2352), ('me', 2321), ('could', 2289), ('films', 2282), ('watch', 2272), ('any', 2260), ('two', 2250)]

Bottom 100 tokens in vocabulary:

[('dumb....the', 1), ('magenta', 1), ('stupid!to', 1), ('cast....and', 1), ('freeze....he', 1), ('same...the', 1), ('given.in', 1), ('(1989).', 1), ('gimmickery.', 1), ('house...complete', 1), ('characters.overall,', 1), ('2...but', 1), ('good...at', 1), ('times.so', 1), ('special...and', 1), ('dumb...but', 1), ('dhar.', 1), ('film.plot:', 1), ('2016.story', 1), ('light-hearted.', 1), ("'bharat", 1), ("jai'", 1), ('ventures.', 1), ('sampat.', 1), ('bedi.', 1), ('great.music:', 1), ('sashwat', 1), ('sachdev', 1), ("'beh", 1), ("chala'.favorite", 1), ('half.verdict:', 1), ('finished.i', 1), ('right.way', 1), ('("2001")', 1), ("worse--he's", 1), ('spielberg!', 1), ('(165', 1), ('yammering', 1), ('cutouts', 1), ('deadened', 1), ('cry--like', 1), ('cry--tears', 1), ('"logic"', 1), ('praise?', 1), ('get.imagine', 1), ('imagination)', 1), ('fantasian', 1), ('one.god,', 1), ('you\'s"', 1), ('strong--particularly', 1), ('thing--that', 1), ('pensive', 1), ('closet-genius', 1), ('matt,', 1), ('dock.', 1), ('leaud).', 1), ("delinquent.that's", 1),

```
('entertainment.she', 1), ('biopic!', 1), ('embodiment,',
1), ('impersonation.', 1), ('world-cup-winning', 1),
('team.director', 1), ('parts.pandey', 1), ("jha's", 1),
('euphoric', 1), ('pitch.', 1),
('melodious.performance-wise:', 1), ('banerji,', 1),
('okay.on', 1), ('screen,"', 1), ('paper)', 1),
('college).', 1), ("'sexa'", 1), ('(varun', 1), ("'acid'",
1), ('(naveen', 1), ('polishetty),', 1), ('(tahir', 1),
('bhasin),', 1), ("'bevda'", 1), ('(saharsh', 1),
("'mummy'", 1), ('(tushar', 1), ('pandey).stress', 1),
('exams.', 1), ('whirlpool', 1), ('families.while', 1),
('lessons.as', 1), ('corridors,', 1), ('distance.the', 1),
("'three", 1), ('maya!),', 1), ('derek.', 1),
("relatable.'chhichhore'", 1), ('prospective.', 1),
("ryu's", 1), ('name..it', 1), ('cheesed', 1),
('movie...way', 1)]
```

- By looking at the top and bottom 100 tokens we can decide on a better pre-processing strategy.

Data Preprocessing:

Stopword removal:

We need to remove the stopwords from tokens in order to reduce max_sent_len as well as embedding matrix size. But we can't directly use default stopwords from the nltk library, as for review classification as we need words like 'not'. So, I used the following modified list of stopwords.

```
stop_words = set(['i', 'me', 'my', 'myself', 'we',
'our', 'ours', 'ourselves', 'you', 'your', 'yours',
'yourself', 'yourselves', 'he', 'him', 'his',
'himself', 'she', 'her', 'hers', 'herself', 'it',
"it's", 'its', 'itself', 'they', 'them', 'their',
'theirs', 'themselves', 'what', 'which', 'who',
'whom', 'this', 'that', 'these', 'those', 'am', 'is',
'are', 'was', 'were', 'be', 'a', 'an', 'the', 'and',
'd', 'll', 'm', 'o', 're', 've', 'y'])
```

```
len(stop_words): 55
```

Data cleaning:

- Replace "-" and "_" and ":" by space
- Remove extra spaces which have been added because of the removal of characters, or newline.
- Lowercase the review string
- Finally removed characters other than a-z, A-Z, and 0-9

- Split the review and keep only words that aren't in the stop_word list.

For example,

<u>review before preprocessing:</u>

```
Psychopath Risen
An origination story of a psychopath. The overall plot
was convincing but not satisfying enough.
Screenplay was gripping with unexpected twists &
turns.
The overall making of the movie could've been better
with more visual storytelling & the performance of
actors were also not so satisfactory but a meagre one.
Bgm was good but overall music could've been still way
better.
Probably, it was the hype by media which I feel was
too much.
But, nowadays negative content is getting more
traction than the postive ones, which definitely isn't
a good thing to the industry as a whole.
```

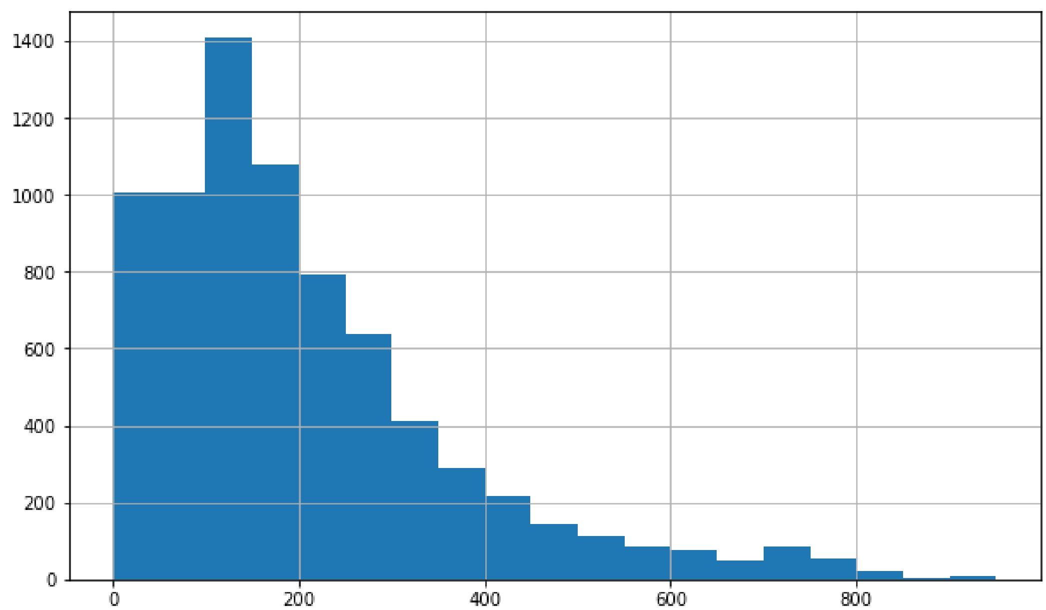
<u>review after preprocessing:</u>

```
psychopath risen origination story of psychopath
overall plot convincing but not satisfying enough
screenplay gripping with unexpected twists & turns
overall making of movie couldve been better with more
visual storytelling & performance of actors also not
so satisfactory but meagre one bgm good but overall
music couldve been still way better probably hype by
media feel too much but nowadays negative content
getting more traction than postive ones definitely
isnt good thing to industry as whole
```

Analyzing review length:

movie	rating	genre_primary_mapped	review_title	review	review_rating	review_class_bin	review_processed	num_tokens
Three Billboards Outside Ebbing, Missouri	8.1	comedy	Searching for ... aka the truth is out there	We know in real life that it's not easy to sol...	10/10	good	searching for aka truth out there know in real...	105
Radhe	2.4	action	No Less Than a Torment!	An official remake of Korean Thriller 'The Out...	2/10	bad	no less than torment official remake of korean...	380
Anbe Sivam	8.5	adventure	100% Magic!	If you want to know the intensity of tamil cin...	10/10	good	100 magic if want to know intensity of tamil c...	45

Number of tokens vs count of reviews:



We can see maximum reviews having tokens count below 400.

Review length analysis:

```
num of reviews having tokens >=2:  7500
num of reviews having tokens >=5:  7500
num of reviews having tokens >=10:  7433
num of reviews having tokens >=20:  7124
num of reviews having tokens >=30:  6858
num of reviews having tokens >=50:  6481
num of reviews having tokens >=100:  5479
num of reviews having tokens >=150:  4074
num of reviews having tokens >=200:  2998
num of reviews having tokens >=300:  1560
num of reviews having tokens >=400:  874
num of reviews having tokens >=500:  524
num of reviews having tokens >=1000:  17
```

Average length of review per class when considering 3 classes:

	num_tokens				
	min	max	median	mean	std
review_class					
average	8	1326	174	220.850251	180.639456
bad	6	1173	159	202.011245	162.816008
good	7	1212	160	204.124122	166.848278

Average length of review per class when considering 2 classes:

	num_tokens				
	min	max	median	mean	std
review_class_bin					
bad	6	1326	164	211.352651	176.822206
good	7	1212	164	205.206864	163.604748

Average length of review as per genre:

	num_tokens				
	min	max	median	mean	std
genre_primary_mapped					
action	6	1194	152	192.916254	164.718652
adventure	8	1283	188	233.311947	167.537433
animation	9	1013	150	204.610169	166.549294
biography	10	875	197	236.033058	164.679310
comedy	8	1212	157	204.860666	169.868020
drama	8	1326	172	222.545148	183.421608
horror	13	767	205	237.026738	163.637596
thriller/crime	8	1179	159	200.181176	159.965029
western	8	1046	190	241.500000	204.293759

train-test split: 80:20

tokenization:

```
Review tok: ['must', 'watch', 'extraordinary', 'movie',
'engages', 'by', 'last', 'moment', 'doesnt', 'let', 'miss',
'second', 'as', 'well', 'would', 'like', 'to', 'watch',
'more', 'movies', 'like']

Token idx: [150, 66, 1305, 10, 6790, 14, 191, 471, 118,
386, 736, 170, 5, 37, 45, 21, 3, 66, 31, 70, 21]
```

Word Embedding:

I have used Glove word embeddings for my dataset.

LSTM Model with bias:

Model for 3 class classification:

```
Model: "model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input (InputLayer)          [(None, 300)]             0

 embedding_1 (Embedding)     (None, 300, 50)           1750000

 lstm (Bidirectional)        (None, 128)               58880

 dense_1 (Dense)             (None, 64)                8256

 review_output (Dense)       (None, 3)                 195

=================================================================
Total params: 1,817,331
Trainable params: 67,331
Non-trainable params: 1,750,000
_________________________________________________________________
None
```

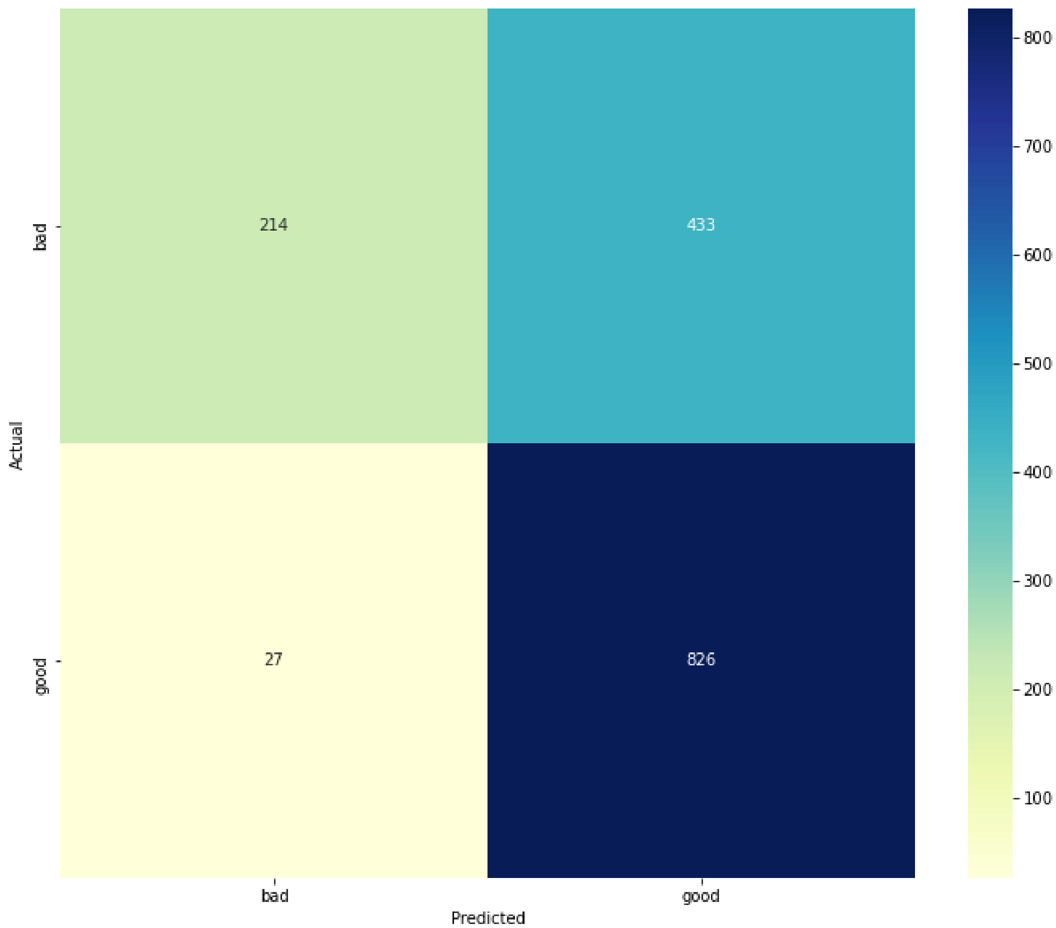
Model for 2 class classification:

```
Model: "model"
_________________________________________________________________
_
 Layer (type)                Output Shape              Param #
=================================================================
=
 input (InputLayer)          [(None, 300)]             0

 embedding (Embedding)       (None, 300, 50)           1750000

 lstm (Bidirectional)        (None, 128)               58880

 dense (Dense)               (None, 64)                8256

 review_output (Dense)       (None, 2)                 130

=================================================================
=
Total params: 1,817,266
Trainable params: 67,266
Non-trainable params: 1,750,000
_________________________________________________________________
_
None
```

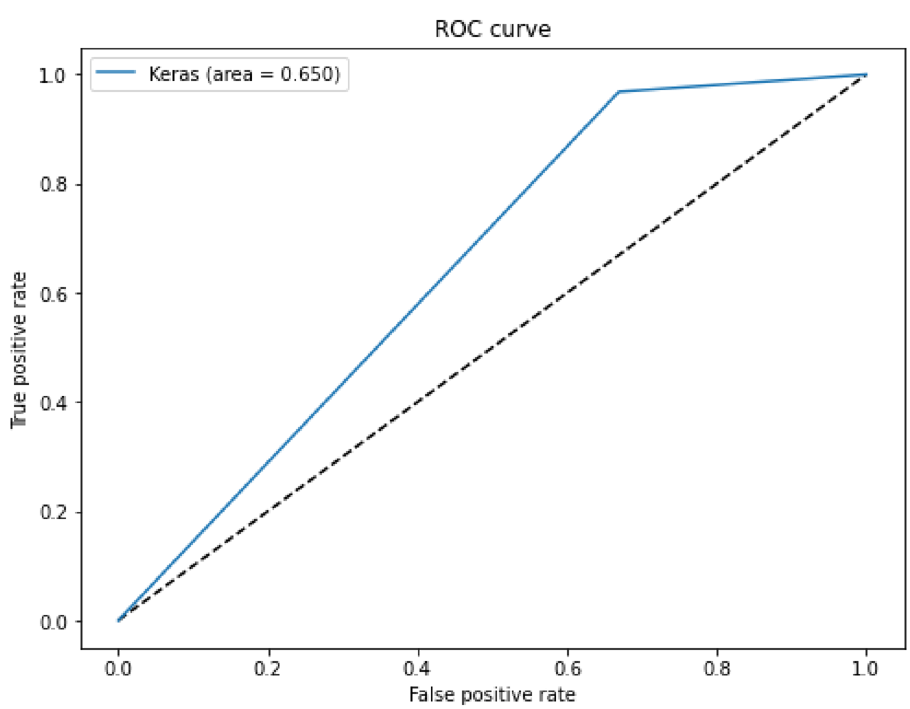
Classification report:

```
              precision    recall  f1-score   support

           0       0.89      0.33      0.48       647
           1       0.66      0.97      0.78       853

    accuracy                           0.69      1500
   macro avg       0.77      0.65      0.63      1500
weighted avg       0.76      0.69      0.65      1500
```

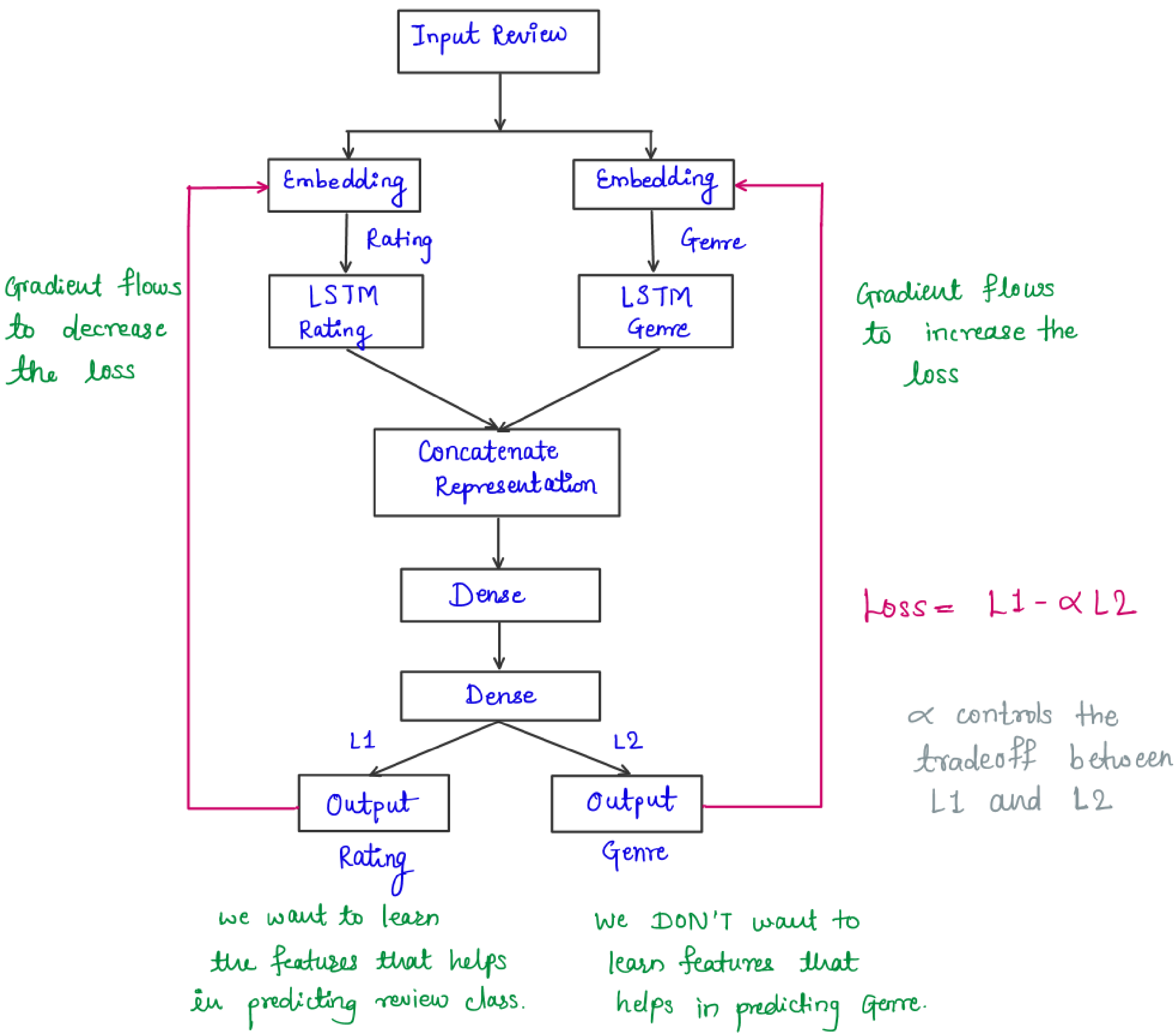
Confusion matrix:



ROC Curve:



Bias-free LSTM Model:



We are training the model on two objectives but we only want to learn one of them (review class) and not learn the other (genre). Both L1 and L2 are cross-entropy losses (sparse).

```
L1 = review rating loss
L2 = genre loss
alpha = hyperaparmeter to control trade-off between L1 and
L2
```

Model:

```
__________________________________________________________________
__________________________________
 Layer (type)                   Output Shape         Param #
Connected to
==================================================================
==================================
 input (InputLayer)               [(None, 300)]          0
[]

 embed_rating (Embedding)       (None, 300, 50)      1750000
['input[0][0]']

 embed_genre (Embedding)        (None, 300, 50)      1750000
['input[0][0]']

 lstm_rating (Bidirectional)    (None, 128)            58880
['embed_rating[0][0]']

 lstm_genre (Bidirectional)     (None, 128)            58880
['embed_genre[0][0]']

 concatenate (Concatenate)        (None, 256)            0
['lstm_rating[0][0]',

'lstm_genre[0][0]']

 dense (Dense)                    (None, 128)          32896
['concatenate[0][0]']

 dense_1 (Dense)                  (None, 64)            8256
['dense[0][0]']

 output_rating (Dense)            (None, 2)              130
['dense_1[0][0]']

 output_genre (Dense)             (None, 9)              585
['dense_1[0][0]']

==================================================================
==================================
Total params: 3,659,627
Trainable params: 159,627
Non-trainable params: 3,500,000
__________________________________________________________________
__________________________________
None
```

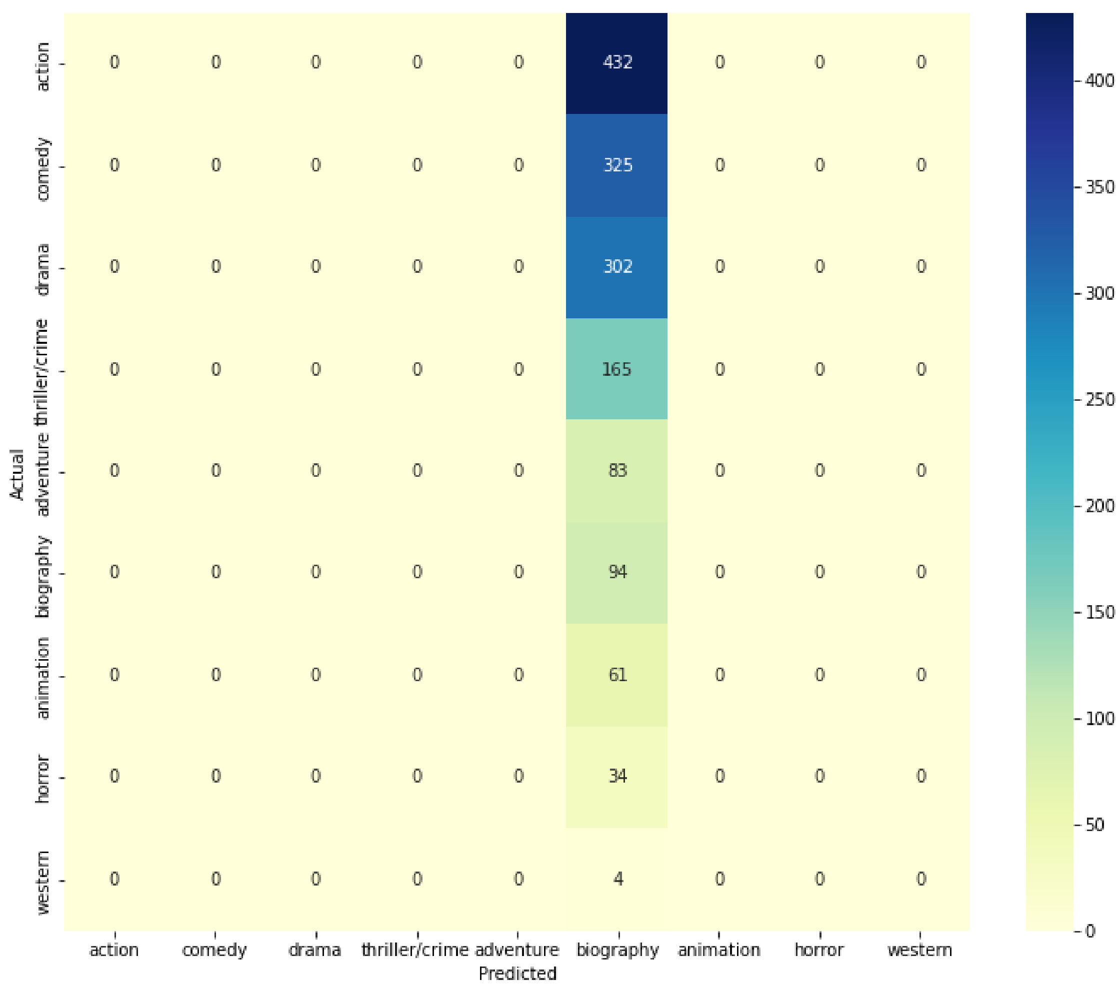
Output for Best value alpha and threshold:

I experimented with various values of alpha and thresholds and these are the best results obtained for:

```
alpha = -0.05
thres: 0.65
```

	precision	recall	f1-score	support
0	0.58	0.49	0.53	638
1	0.66	0.73	0.70	862
accuracy			**0.63**	1500
macro avg	0.62	0.61	0.61	1500
weighted avg	0.63	0.63	0.63	1500

Confusion matrix



Bias-free model weighted version:

Here I have used balanced class weights.

```
Number of samples in review rating(1/0)class:
     1: 3377
     0: 2623

Weights given:
     0: 1.1437285550895921
     1: 0.8883624518803672


Number of samples in 9 genre classes:
     0: 1789
     1: 1297
     2: 1182
     3: 685
     4: 369
     5: 269
     6: 234
     7: 153
     8: 22

Weights given:
     0: 0.3726476616359232
     1: 0.5140066820868672
     2: 0.5640157924421884
     3: 0.9732360097323601
     4: 1.8066847335140017
     5: 2.4783147459727384
     6: 2.849002849002849
     7: 4.357298474945534
     8: 30.303030303030305
```

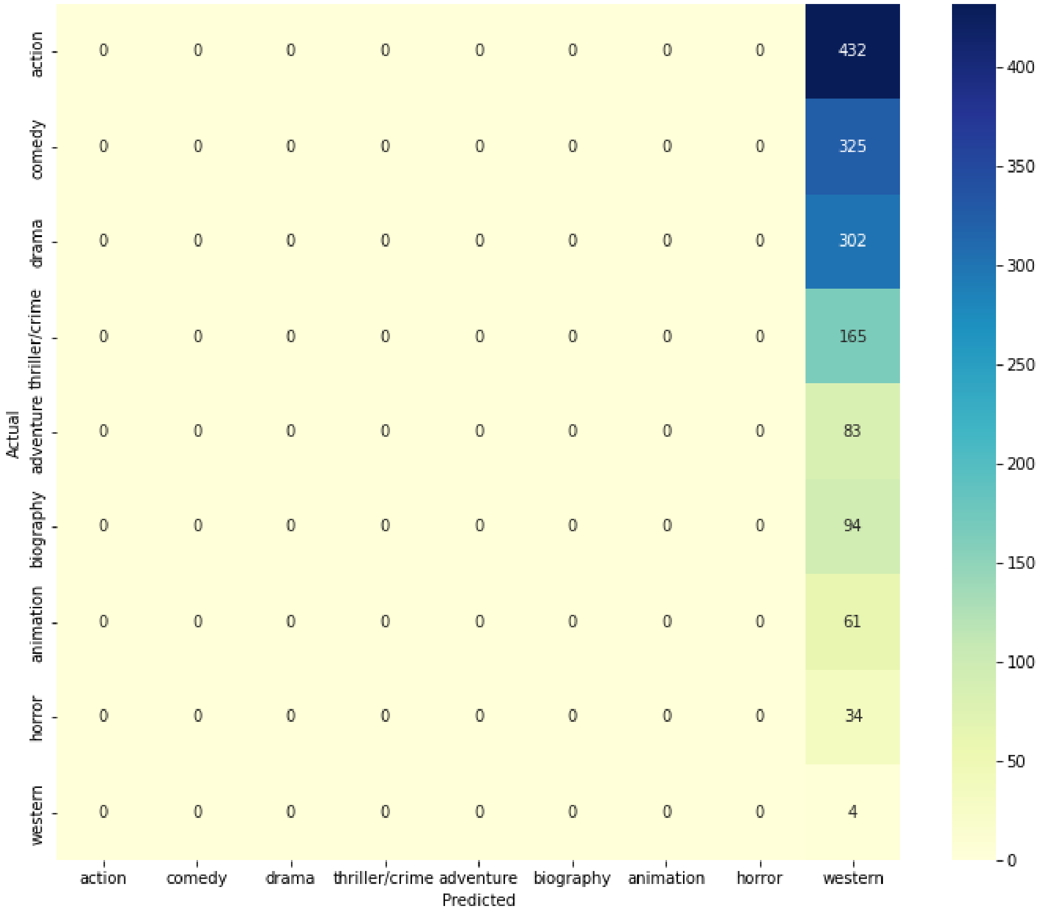
Output for Best value alpha and threshold:

I experimented with various values of alpha and thresholds and these are the best results obtained for:

Alpha = -0.05
thres: 0.82

	precision	recall	f1-score	support
0	0.59	0.56	**0.58**	638
1	0.69	0.71	**0.70**	862
accuracy			**0.65**	1500
macro avg	0.64	0.64	0.64	1500
weighted avg	0.65	0.65	0.65	1500

Confusion matrix:



Results:

As we have got best scores when we considered bias-free model with balanced class weights. The results comparison is made with the bias-free model weighted version and model with bias for binary classification that is we have 2 classes, good and bad for review ratings. Showing the classification report again for reference.

Model without Bias scores:

	precision	recall	f1-score	support
0	0.59	0.56	**0.58**	638
1	0.69	0.71	**0.70**	862
accuracy			**0.65**	1500
macro avg	0.64	0.64	0.64	1500
weighted avg	0.65	0.65	0.65	1500

Positive reviews comparison:

1. For review rating - 7/10

movie	rating	genre_primary_mapped	genre_label	review_title	review	review_processed	num_tokens	review_rating	review_class_bin	positive_prob	bias_free_positive_prob
Anjaam Pathiraa	7.8	thriller/crime	3	Deviated mold from Midhun.	I was surprised because I can't expect this ki...	deviated mold from midhun surprised because ca...	61	7/10	good	0.57	0.63
In the Name of the King: A Dungeon Siege Tale	3.9	action	0	One of Uwe Boll's Better Forays into Film-maki...	"Alone in the Dark" director Uwe Boll's PG-13 ...	one of uwe bolls better forays into film makin...	507	7/10	good	0.77	0.66
Earth	7.6	drama	2	Earth - did Deepa do it again?	1947 Earth is a story about the partition of I...	earth did deepa do again 1947 earth story abou...	318	7/10	good	0.44	0.67
Batman Begins	8.2	action	0	A vast improvement over the Batman films of th...	I enjoyed the film BATMAN from Tim Burton, tho...	vast improvement over batman films of 1990s en...	202	7/10	good	0.69	0.67
Queen	8.1	adventure	4	Finally, an Indian film that lives up to the hype	My experience with Indian cinema has been fair...	finally indian film lives up to hype experienc...	101	7/10	good	0.83	0.65

1. For review rating - 8/10

movie	rating	genre_primary_mapped	genre_label	review_title	review	review_processed	num_tokens	review_rating	review_class_bin	positive_prob	bias_free_positive_prob
The Elephant Man	8.1	biography	5	One of Us.	This has to be the only totally straight movie...	one of us has to only totally straight movie d...	743	8/10	good	0.66	0.65
The Kid	8.2	comedy	1	Sensitive and enjoyable film in which Chaplin ...	Wonderful picture mixes sentiment , drama , sl...	sensitive enjoyable film in chaplin meets stre...	335	8/10	good	0.66	0.65
hhichhore	8.1	comedy	1	One-Line Review: Chhichhore (8 Stars)	Nitesh Tiwari's Chhichhore (Loafers), although...	one line review chhichhore 8 stars nitesh tiwa...	53	8/10	good	0.74	0.66
Bhaag Milkha Bhaag	8.1	biography	5	Milkha Singh Has Won. So Has The Biographical ...	In less than ten minutes into the film, you ge...	milkha singh has won so has biographical film ...	237	8/10	good	0.85	0.67
Malik	7.9	action	0	Another best from Mahesh Narayanan, Fahadh Faazil	Superb movie from Mollywood. Made justice to t...	another best from mahesh narayanan fahadh faaz...	27	8/10	good	0.51	0.66

2. **For review rating - 9/10**

movie	rating	genre_primary_mapped	genre_label	review_title	review	review_processed	num_tokens	review_rating	review_class_bin	positive_prob	bias_free_positive_prob
Dangal	8.3	action	0	Dangal is unmissable.	Dangal review :In his 2001 Oscar nominated 'La...	dangal unmissable dangal review in 2001 oscar ...	405	9/10	good	0.66	0.62
Angoor	8.2	comedy	1	One of Bollywood's finest comedies	Bollywood comedies are normally associated wit...	one of bollywoods finest comedies bollywood co...	268	9/10	good	0.85	0.65
Kai Po Che	7.7	drama	2	Satire on the existing state of affairs in India	Three friends find their lives strangled as th...	satire on existing state of affairs in india t...	80	9/10	good	0.44	0.67
Talvar	8.1	thriller/crime	3	As Good as a Movie About a Real-Life Legal Cas...	Reviewed by: Dare Devil Kid (DDK)Rating: 4.3/5...	as good as movie about real life legal case ca...	353	9/10	good	0.78	0.66
Kirik Party	7.9	comedy	1	A treat for all college students	This is one of the best movies i watched in re...	treat for all college students one of best mov...	71	9/10	good	0.77	0.64

3. **For review rating - 10/10**

movie	rating	genre_primary_mapped	genre_label	review_title	review	review_processed	num_tokens	review_rating	review_class_bin	positive_prob	bias_free_positive_prob
Poove Unakkaga	7.9	comedy	1	Poovae unakaga	Vijay's first blockbuster movie and superb act...	poovae unakaga vijays first blockbuster movie ...	16	10/10	good	0.44	0.64
Mahanati	8.2	biography	5	Mahanati - The Greatest Biopic Ever Made In Th...	To be frank, I've no expectations over this fi...	mahanati greatest biopic ever made in history ...	118	10/10	good	0.69	0.64
Don	7.7	action	0	AB is the best!!!	This is another classic movie by Bachchan. The...	ab best another classic movie by bachchan rang...	88	10/10	good	0.87	0.64
Ghilli	7.7	action	0	Such an interesting movie espically Vijay's at...	Semmma semmma semmma movie to watch. Such an i...	such interesting movie espically vijays attitu...	28	10/10	good	0.62	0.66
Satya	8.2	action	0	One of the most intelligent movies ever made a...	SATYA stands out from the normal Bollywood rom...	one of most intelligent movies ever made about...	143	10/10	good	0.86	0.66

Negative reviews comparison:

1. **For review rating - 2/10**

movie	rating	genre_primary_mapped	genre_label	review_title	review	review_processed	num_tokens	review_rating	review_class_bin	negative_prob	bias_free_negative_prob
The Master of Disguise	3.6	adventure	4	Annoying and not funny at all	Pistachio Disguisey (Dana Carvey) is a waiter ...	annoying not funny at all pistachio disguisey ...	137	2/10	bad	0.39	0.35
Laxmii	3.4	action	0	What kind of movie this is !!!!	The story of this movie is so bad and acting i...	kind of movie story of movie so bad acting ave...	21	2/10	bad	0.30	0.33
The Music Room	7.9	drama	2	The sensible review	If you are not from India and find the cast sy...	sensible review if not from india find cast sy...	468	2/10	bad	0.31	0.35
Dabangg 3	3.7	action	0	Exaggerated with tons of overacting	Move on ...Nothing new to watch here. The nove...	exaggerated with tons of overacting move on no...	54	2/10	bad	0.27	0.37
The Love Guru	3.9	comedy	1	Mike Myers hits a new low!	The Love Guru is another stupid comedy written...	mike myers hits new low love guru another stup...	182	2/10	bad	0.25	0.35

2. For review rating - 3/10

movie	rating	genre_primary_mapped	genre_label	review_title	review	review_processed	num_tokens	review_rating	review_class_bin	negative_prob	bias_free_negative_prob
My Father and My Son	8.1	drama	2	An emotional drama that never hits hard	This is an emotional drama that, after the ope...	emotional drama never hits hard emotional dram...	149	3/10	bad	0.42	0.33
Dragon Wars: D-War	3.8	action	0	The monsters attacks are some of the best visu...	First off you don't see any dragons until the ...	monsters attacks some of best visuals in years...	313	3/10	bad	0.58	0.34
Date Movie	2.9	comedy	1	The best of the Friedberg-Seltzer spoof movies...	You may probably remember how much I dislike F...	best of friedberg seltzer spoof movies but not...	216	3/10	bad	0.28	0.33
Dragonball Evolution	2.6	action	0	Goki , Master Roshi and Bulma battle Lord Picc...	The story starts with Goku (Justin Chatwin of ...	goki master roshi bulma battle lord piccolo to...	284	3/10	bad	0.44	0.35
Daddy Day Camp	3.5	comedy	1	Daddy Day Camp	Always beware a sequel where none of the origi...	daddy day camp always beware sequel where none...	96	3/10	bad	0.22	0.35

Conclusion and Future Work:

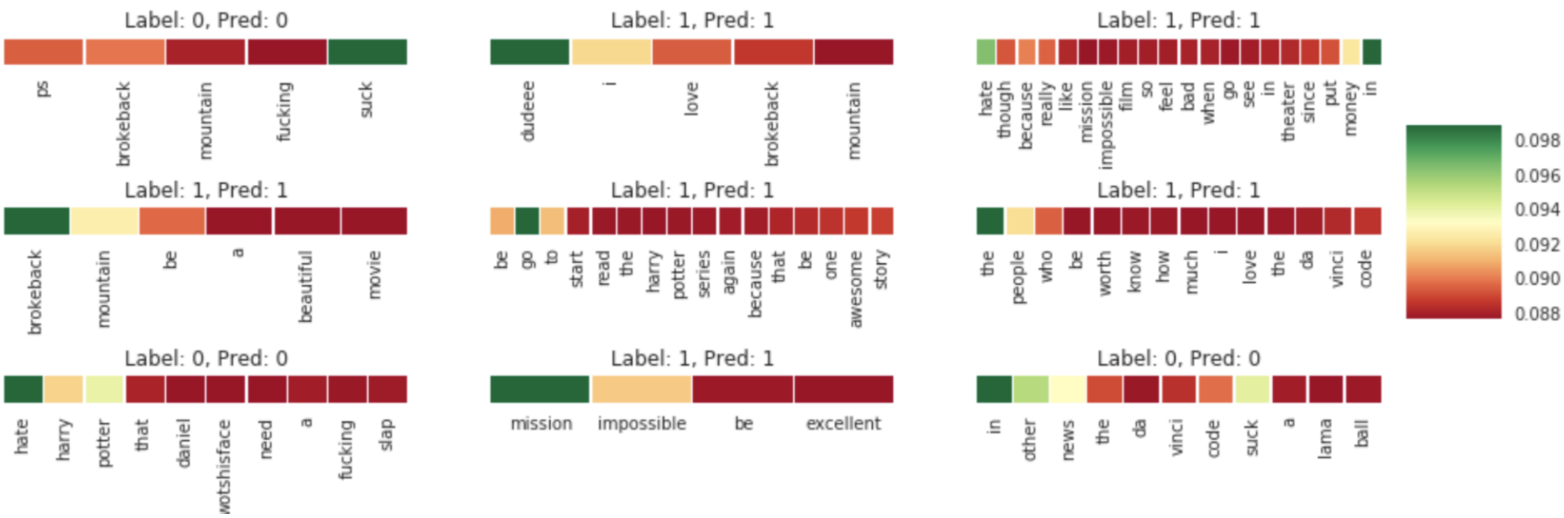
In the above project first, we built a dataset of (movie, genre, movie reviews, review rating). This dataset allowed us to learn the review ratings while removing the bias with respect to the genre of the movie. As we have seen from our analysis that some genres are always rated higher than others. So, its important to remove such biases.

Although we did this analysis on genre property, it could be extended to other attributes of movies as well, for example, celebrity, movie language, country, etc. For this, we will have to build datasets that contain such tuples. Similarly, we can also try to remove multiple biases at the same time.

Also, we are considering reviews from different users. It might happen that a user himself is biased towards one genre. For example, if a user generally gives a high rating to action and a low rating to comedy. If we want to remove this kind of bias then we need a dataset of reviews of the same user at the same rating in different genres and then we can remove user-level bias.

Explainability (Exam Question 1)

We can use an attention layer to show the features that were given more weightage while making a prediction. Attention layer is added on top of Bidirectional LSTM. While making predictions from the model, we can use the attention weights to explain the results.



References:

1. https://dl.acm.org/doi/abs/10.1145/3152494.3152520
2. http://ceur-ws.org/Vol-1253/paper1.pdf
3. https://aclanthology.org/P19-1159.pdf
4. https://aclanthology.org/2020.socialnlp-1.2.pdf
5. https://arxiv.org/pdf/1409.7495.pdf
6. https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews
7. https://ai.stanford.edu/~amaas/data/sentiment/