

Classifying Movie Genres from Descriptions

Nikita Baklazhenko, Miguel Conner, David Vallmanya, Dominik Wielath

Outline

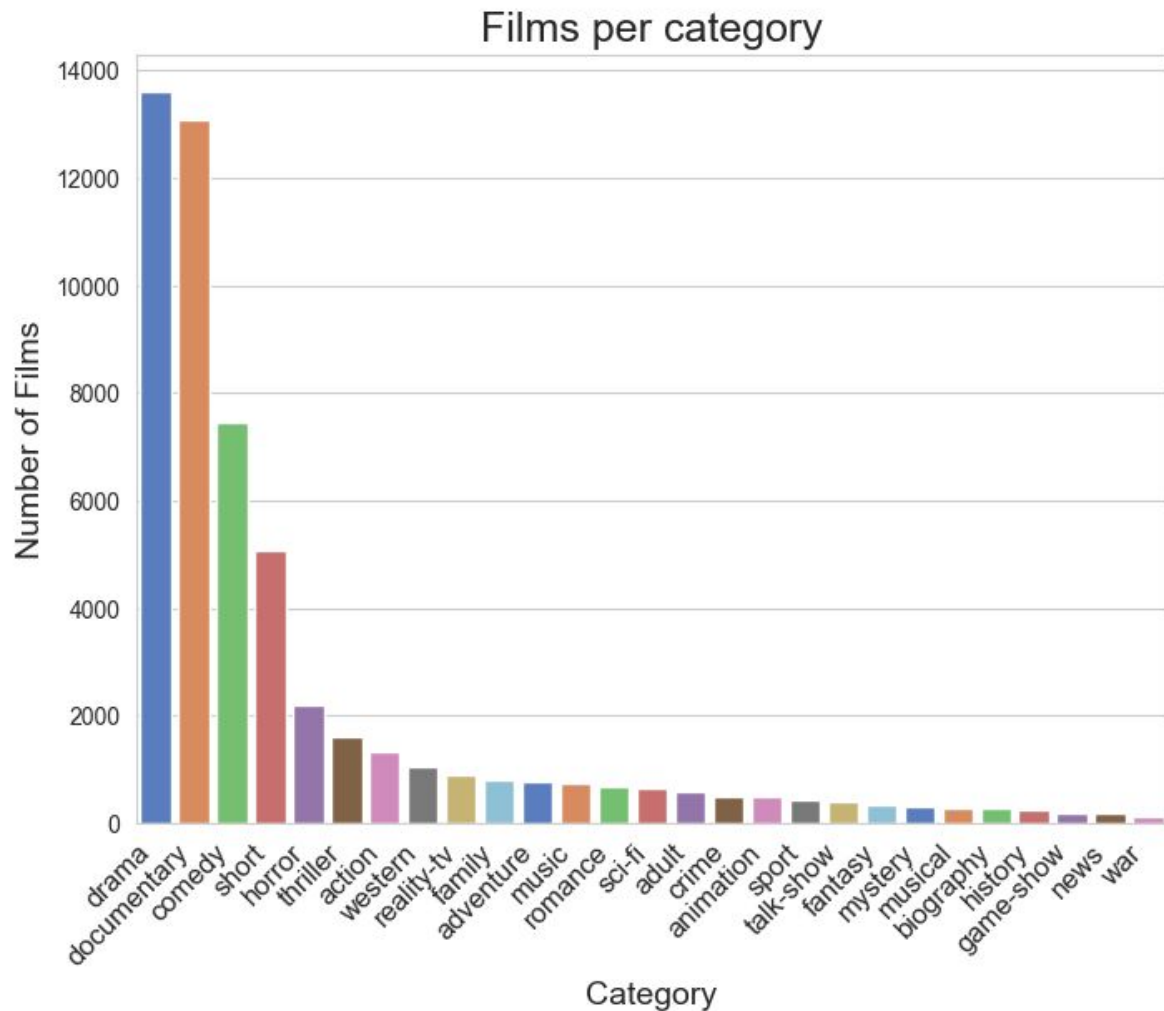
1. Dataset and Error Metrics
2. Baseline Model
3. RNN Model
4. Transformer Models
5. Final Comparison of Results and Biases

1. Dataset

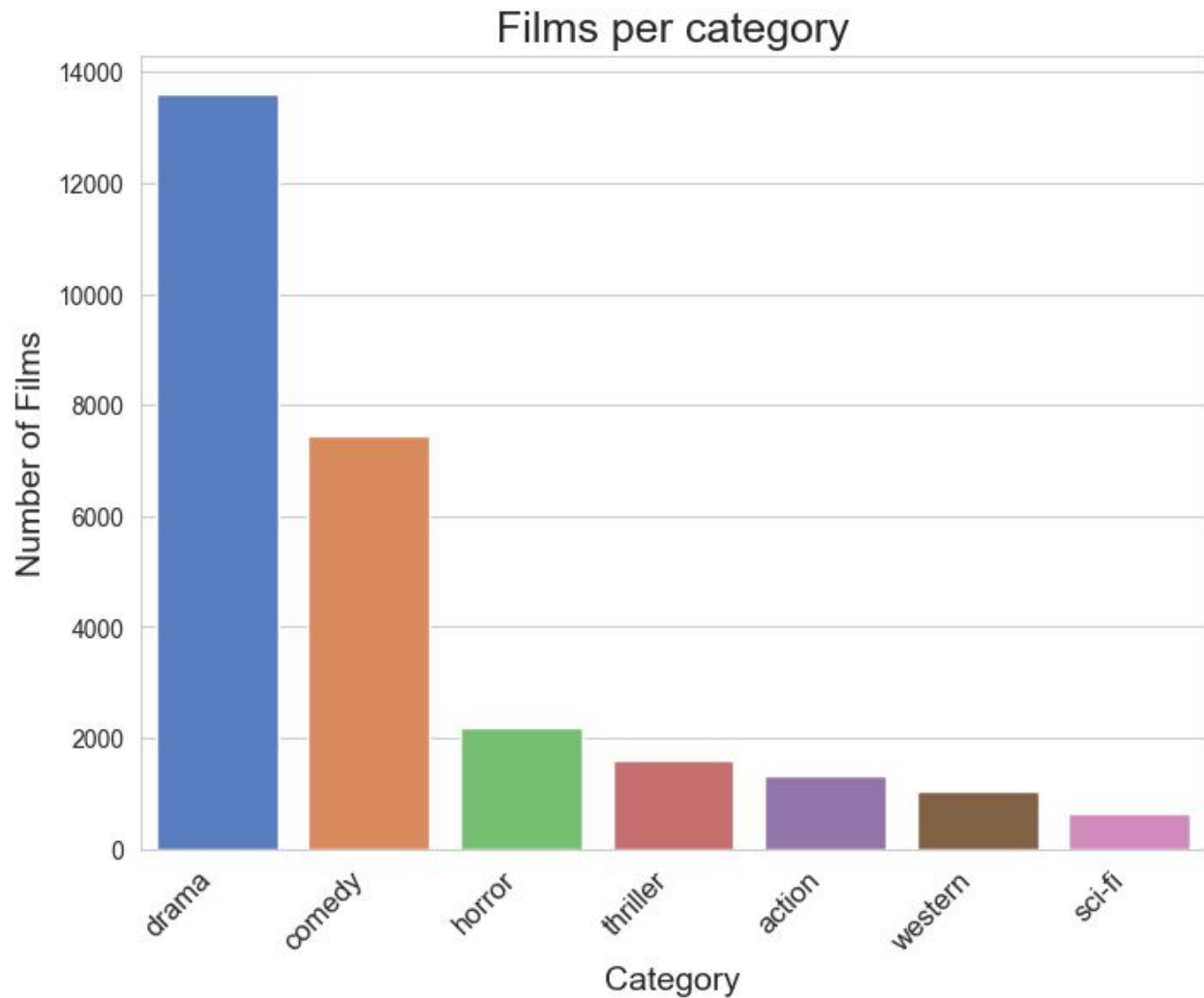
- Train set: 27,849 movies
- Test set: 27,844 movies

	title	genre	description
0	Edgar's Lunch (1998)	thriller	L.R. Brane loves his life - his car, his apar...
1	La guerra de papá (1977)	comedy	Spain, March 1964: Quico is a very naughty ch...
2	Off the Beaten Track (2010)	documentary	One year in the life of Albin and his family ...
3	Meu Amigo Hindu (2015)	drama	His father has died, he hasn't spoken with hi...
4	Er nu zhai (1955)	drama	Before he was known internationally as a mart...

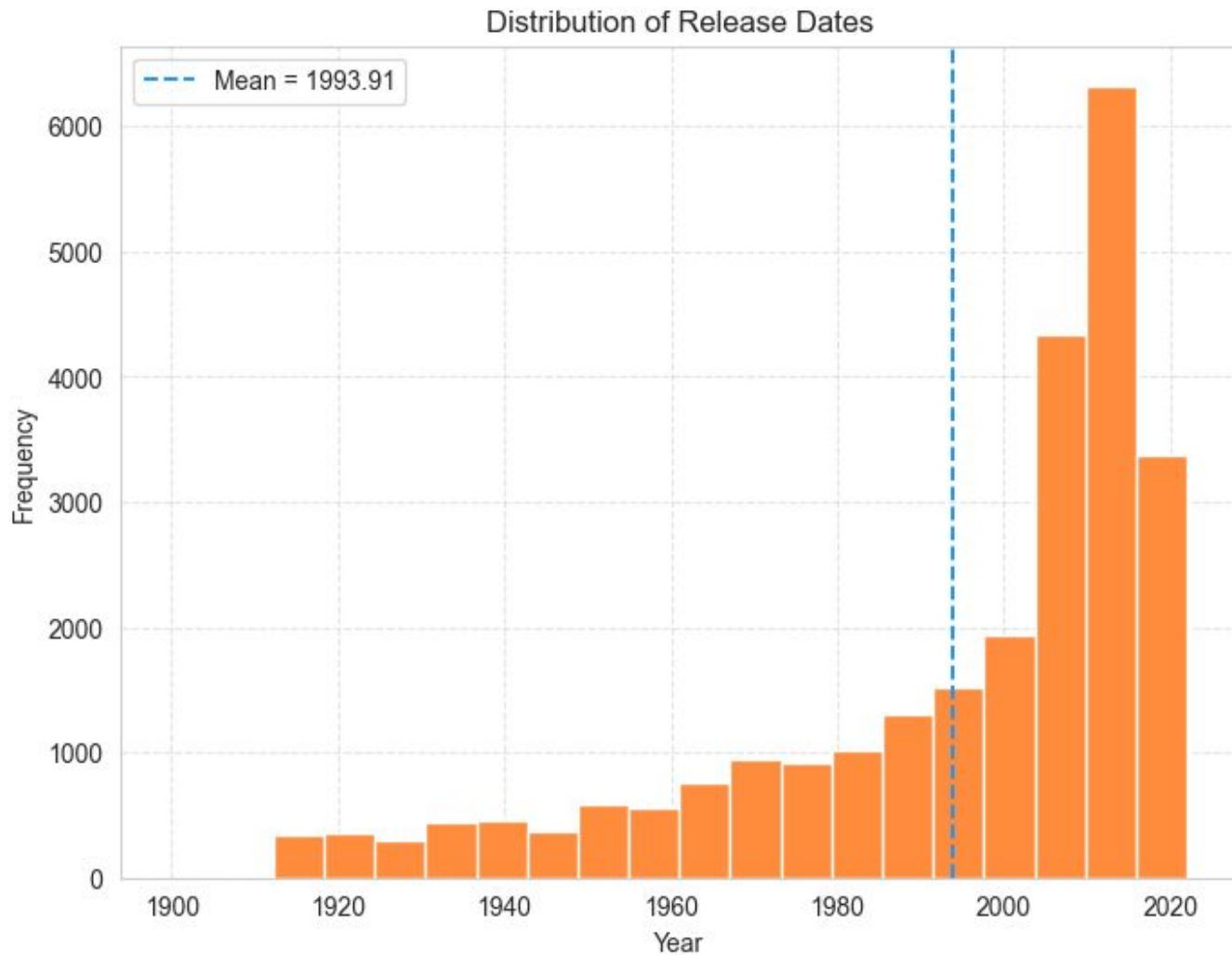
1. Dataset



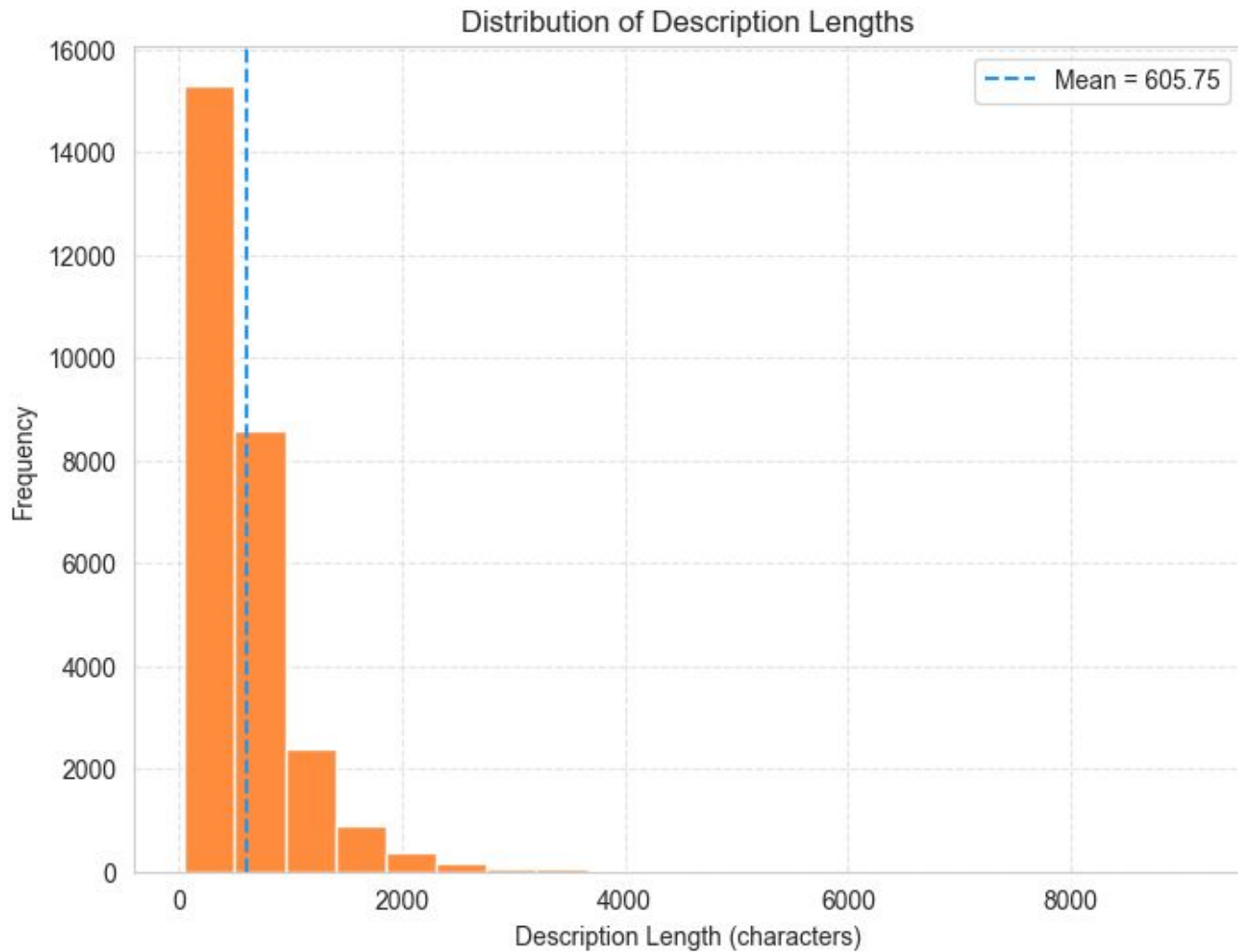
1. Dataset



1. Dataset



1. Dataset



1. Dataset

Preprocessing:

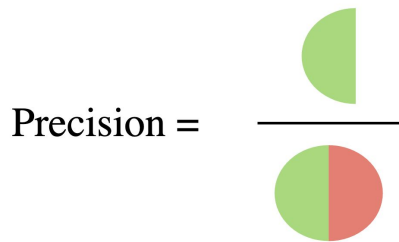
- Remove stopwords
- Lemmatize

For BERT we used **no** preprocessing

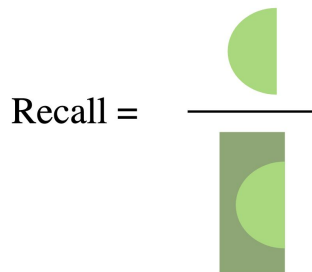
1. Error Metrics

We use mainly a **weighted F-1 score**.

How many retrieved items are relevant?

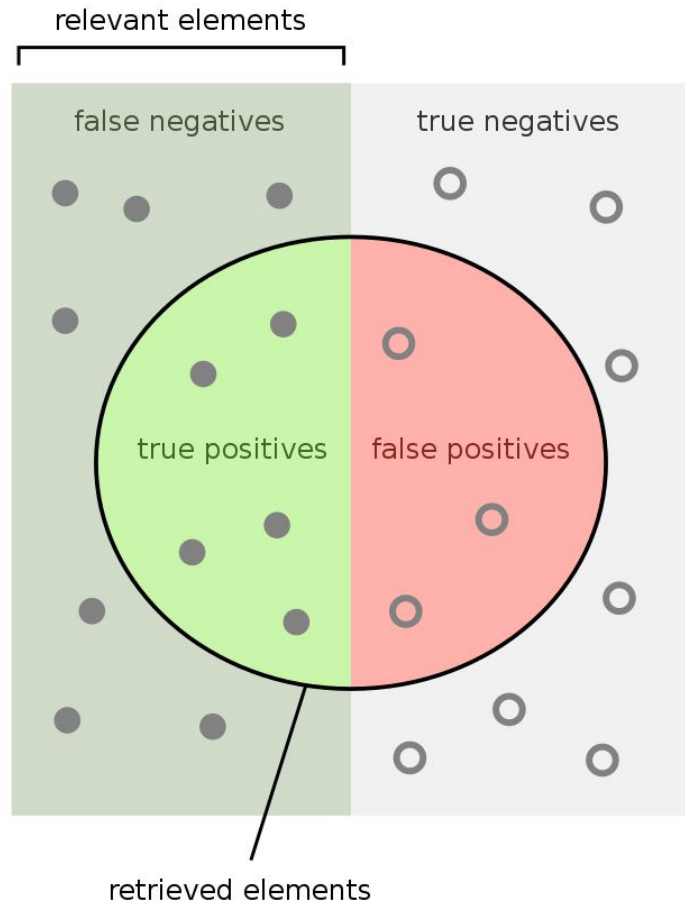


How many relevant items are retrieved?



F-1 score is the harmonic mean of recall and precision.

A weighted F-1 score takes into account the **size** of each category



2. Baseline Models

a) Random Assignment (Prior Probabilities)

Strategy: Predict movies randomly based on overall size of each category.

Weighted F-1 Score: **0.327**

Recall: **0.327**

2. Baseline Models

b) Term Frequency

Strategy: Predict movies based on keywords that have been weighted by frequency. If we don't know, guess based on size of each category.

- Share of data which was labeled based on dictionary: **0.521**
- Share of data randomly labeled using prior distribution: **0.479**

Weighted F-1 Score: **0.410**

Recall: **0.393**

2. Baseline Models

c) TF-IDF

Strategy: Predict movies based on keywords that have been weighted based on frequency **relative to the length of the document**. If we don't know, guess based on size of each category.

- Share of data which was labeled based on dictionary: **0.303**
- Share of data randomly labeled using prior distribution: **0.697**

Weighted F-1 Score: **0.408**

Recall: **0.401**

2. Baseline Models

TF vs. TFIDF

	TF	TFIDF
% of data labeled based on dictionary	52.1%	30.3%
% of data randomly labeled using prior distribution	47.9%	69.7%

2. Baseline Models

Model	Weighted F-1	Recall	Precision
Random Assignment	0.327	0.327	0.328
Term Frequency	0.410	0.393	0.439
TFIDF	0.408	0.401	0.421

3. RNN Models

a) Gated Recurrent Unit Neural Network (GRU)

Strategy: An RNN with a gating mechanism that allows it to “forget” things.

Weighted F-1 Score: **0.633** / Recall: **0.636**

Computation time: ~205s per epoch (16 epochs)

b) Long Short-Term Memory Neural Network (LSTM)

Strategy: Like a GRU but has an output gate.

Weighted F-1 Score: **0.642** / Recall: **0.647**

Computation time: ~180s per epoch (14 epochs)

4. LSTM Model - **Examples** (1)

True label: comedy

TFIDF pred: drama

LSTM pred: comedy

Title: "Low Strung" (2017)

Description: Shervin and Victoria experience a number of mini-adventures smattered with elements of the surreal, often blurring the lines themselves using mind-altering drugs. This lends a fresh, satirical, and often perplexing way for the viewer to vicariously experience tinder dates, pet ownership, Xboxs, exercise, sex, technology, and social media. All against the backdrop of Chicago, a city that shares characteristics with the main characters: inviting, yet frightening, beautiful, yet raw, teeming with the unexpected. An audacious pair of friends - Victoria and Shervin - explore chaotic endeavors through a surreal lens in Chicago. Fresh, satirical, and often perplexing, Low Strung gives viewers a chance to vicariously experience sex, drugs, and montages.

4. LSTM Model - **Examples** (2)

True label: comedy

TFIDF pred: drama

LSTM pred: comedy

Title: Married People Only! (2005)

Description: A romantic comedy about a single man seeking a married woman and a single woman seeking a married man. Somehow, somehow, the two are matched up. Needless to say that neither are ready for a commitment. Come on, we all know now that Married people are already in that magical relationship, so they don't need another one. Right? Keep your sunglasses near, being discreet is a must! Don't ask, don't tell is the law of this land, at least until one of the secret lovers slips up and falls in love. BUMMER! This match made in the Lost In Love section of Misguided Souls News paper is only another crazy tale from the mind of Khari Ajene. It's only a movie, I promise!

4. Transformers (encoder)

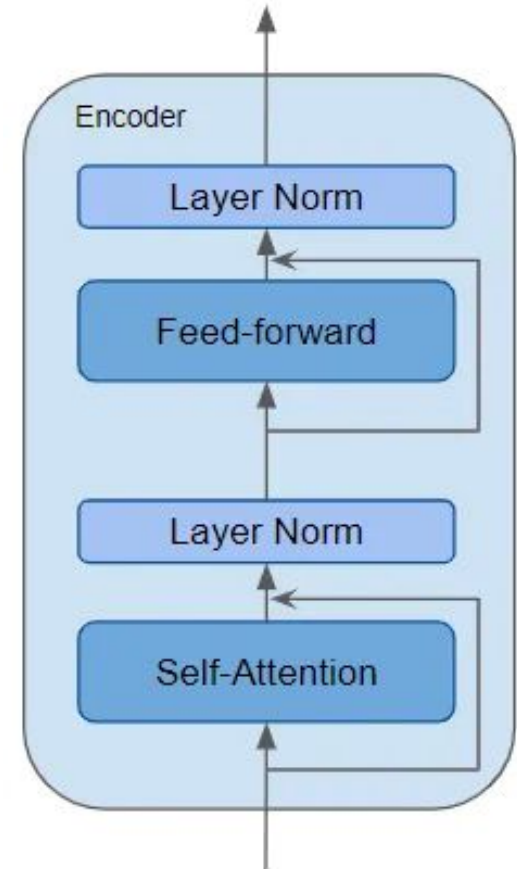
a) Simple Transformer with Multi Headed Attention

A variation of a transformer encoder with:

- Self-attention layer that computes the relationship between different words in the sequence (64 hidden units)
- Pooling layer
- "Relu" layer (32 hidden units)
- Softmax to 7 classes

Weighted F-1 Score: **0.659** / Recall: **0.672**

Computation time: ~3800s per epoch (8 epochs)



4. Transformers (encoder) - **Examples** (1)

True label: drama

TFIDF pred: sci-fi

Transformer pred: drama

Title: Sides (2007)

Description: "Sides" The Boston-based Feature Length drama opens with a panorama of the diverse Boston architecture at night, reflecting a rippling mirror image on the Charles River. A view of Memorial Drive moves to Storrow, gliding around the harbor to Back Bay and finally resting at the colorful lights of the North End waterfront, where the sky is slowly illuminating with the glow of dawn. Al (Aly) Gianetti, a traditional Italian man approaching middle age, lies sleeping with his beautiful live-in girlfriend Gianna. As they softly awaken, their deep loving rest will soon be replaced with the frantic routine of Al's chosen career path. Al, an intense, bohemian actor with a brash, sarcastic sense of humor, is rushing out the door to attend a casting call where he will meet Cinthia (Cin). Meanwhile, we parallel the hurried preparations of Cin leaving her home for the same casting call. A pretty, well-organized, 30-something formally trained stage actress, Cinthia juggles her baby in one arm and her script in the other while bickering with her husband in this typical morning scene. Al introduces himself to Cinthia in the casting company waiting

4. Transformers (encoder) - **Examples** (2)

True label: drama

TFIDF pred: horror

Transformer pred: drama

Title: Pearl Diver (2004)

Description: One night changed them forever. Twenty years later, a night will change them again. Pearl Diver is the story of two sisters, haunted by the twenty-year old murder of their mother and what happens when a farming accident rips away the layers of secrecy and buried trauma surrounding that night. They learn that sometimes the secrets from the past can bring hope for the future.

True label: drama

LSTM pred: comedy

Transformer pred: drama

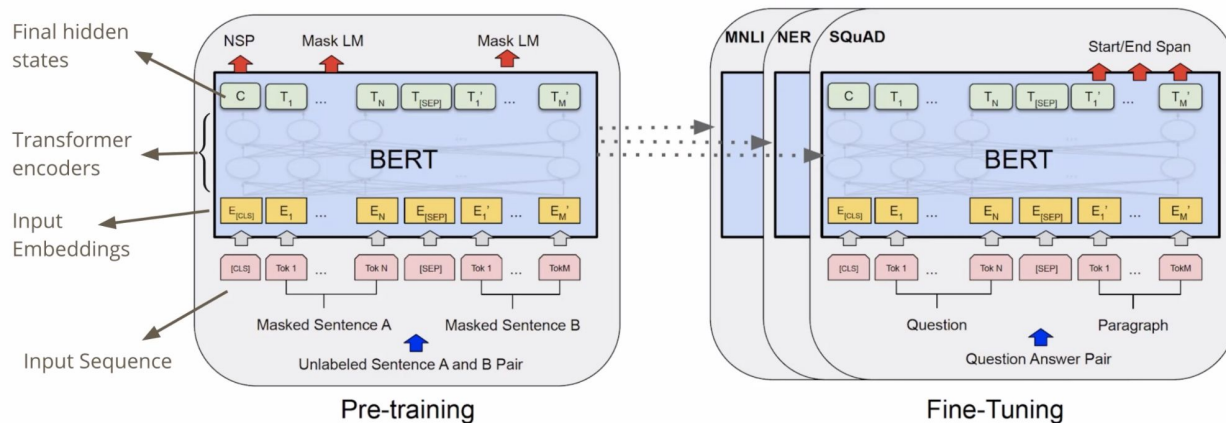
Title: Fuga de cerebros (1998)

Description: A teenage Argentine street gang are targeted for revenge by the overweight street cop whom they wrongfully accused of police brutality. Gang leader Fideo was born in the U.S. and speaks both English and Spanish. He dreams of returning to the US to become a recording star.

4. Transformers (BERT)

BERT = Transformer architecture, which consists of an encoder and a decoder. In BERT, only the encoder part of the Transformer is used

Pre-trained on a large corpus of text using two unsupervised tasks: Masked Language Model (MLM) and Next Sentence Prediction (NSP).



For specific tasks
requires fine
tuning!

Bert->DropOut->
Linear->Relu

4. Transformers (BERT)

Bert-base uncased (given model was used)	BERT-Large	Smaller/different Bert variations
12 layers in the encoder, 768 hidden units, and 110 million parameters	This is a more complex model with 24 layers (i.e., 24 Transformer blocks) in the encoder, 16 attention heads, and 340 million parameters.	DistilBERT, ALBERT, Roberta, BERT-Multilingual and others

Very computationally heavy to even tune the model, even when fully utilize RTX3060

For 1 epoch, 2000 observations: Weighted F-1 Score: **0.43** / Recall: **0.46** / Macro F1 **0.183** (4 min)

For 1 epoch, 25 000 observation: Weighted F-1 Score: **0.653** / Recall: **0.693** / Macro F1: **0.365** (45 mins)

For 4 epoch 25 000 observations: Weighted F-1 Score: **0.746** / Recall: **0.75** / Macro F1: **0.66** (3,5 hours)

!More learning + more data = can achieve even better results

4. Transformers (BERT) - Examples (1)

True label: comedy

LSTM pred: sci-fi

BERT pred: comedy

Title: Flyboy baksa sodong (1959)

Description: When he learns that famous rocket scientist is apparently missing, a pilot nicknamed Flyboy decides to pass himself off as the scientist to collect the awards being bestowed on him. Unfortunately, Flyboy is anything but a rocket scientist...

True label: drama

Encoder pred: thriller

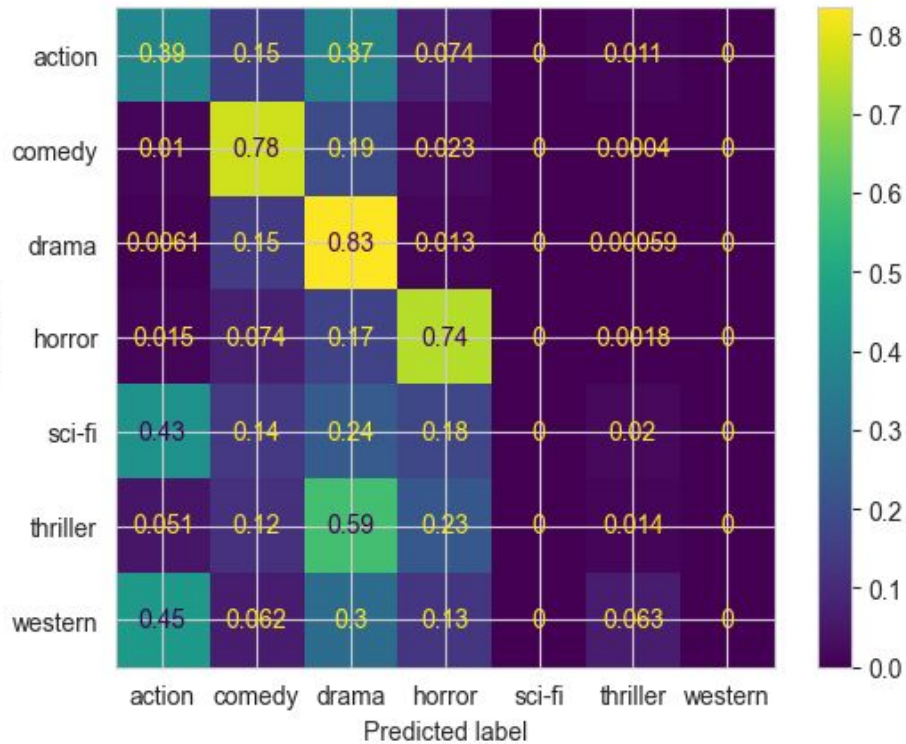
BERT pred: drama

Title: Doch (2012)

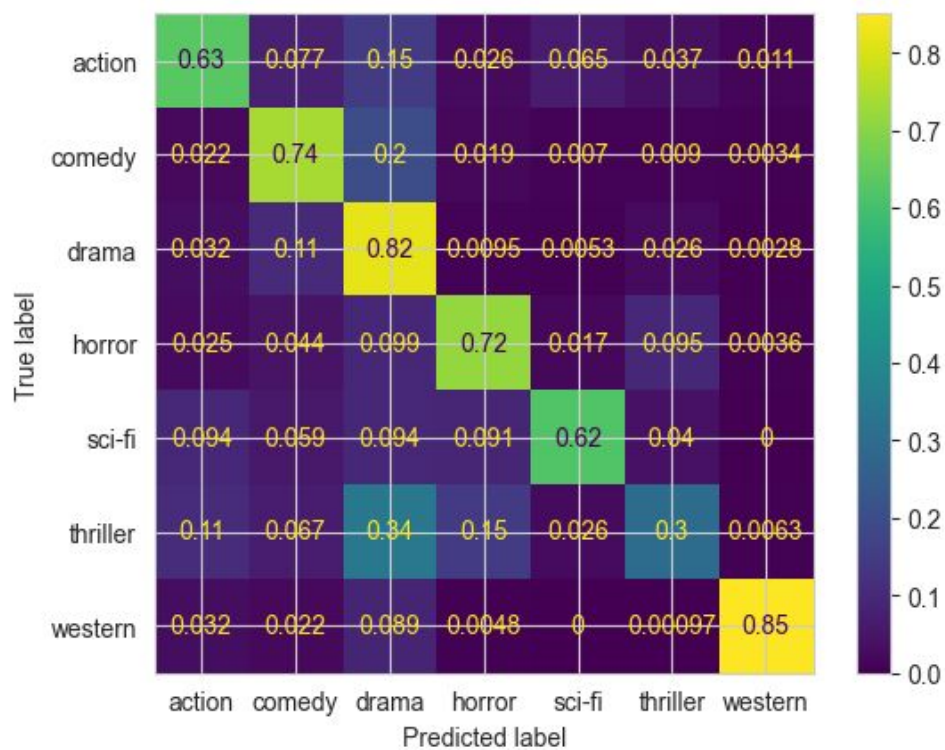
Description: Inna lives with her dad and little brother in a small town. Her life changes when a new girl, Masha, comes to her school. They become friends. But Masha is murdered by a maniac serial killer, who kills teenage girls in the town. Inna goes to church for the funeral service. There she meets the priest's son Ilya, whose sister was also murdered.

5. Comparison between Models and Biases

BERT 1 epoch



BERT 4 epoch

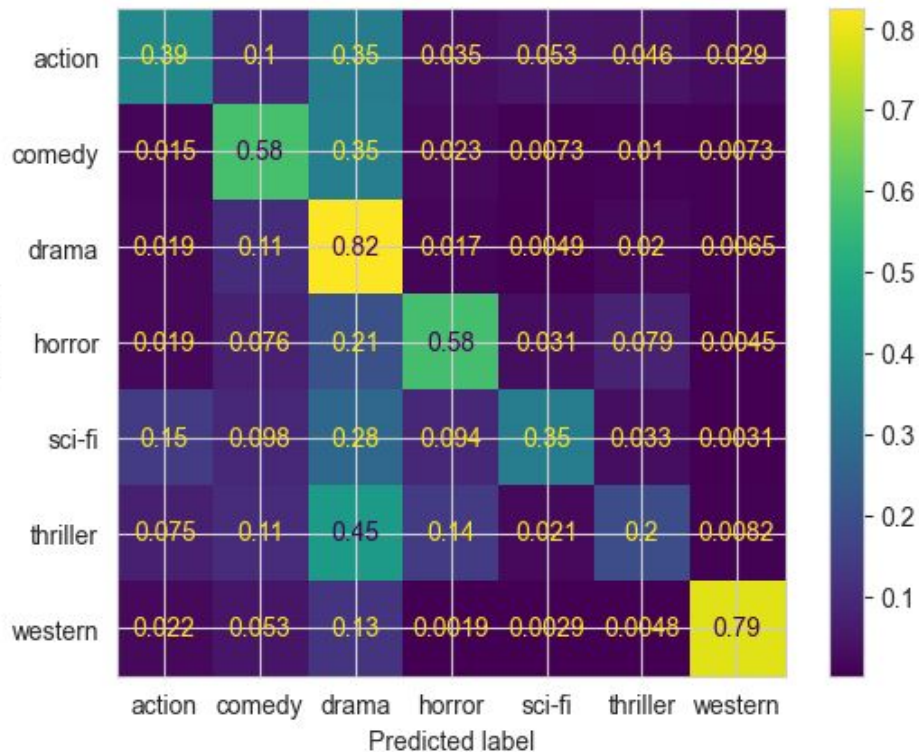


5. Biases: BERT 1 Epoch confusion matrix

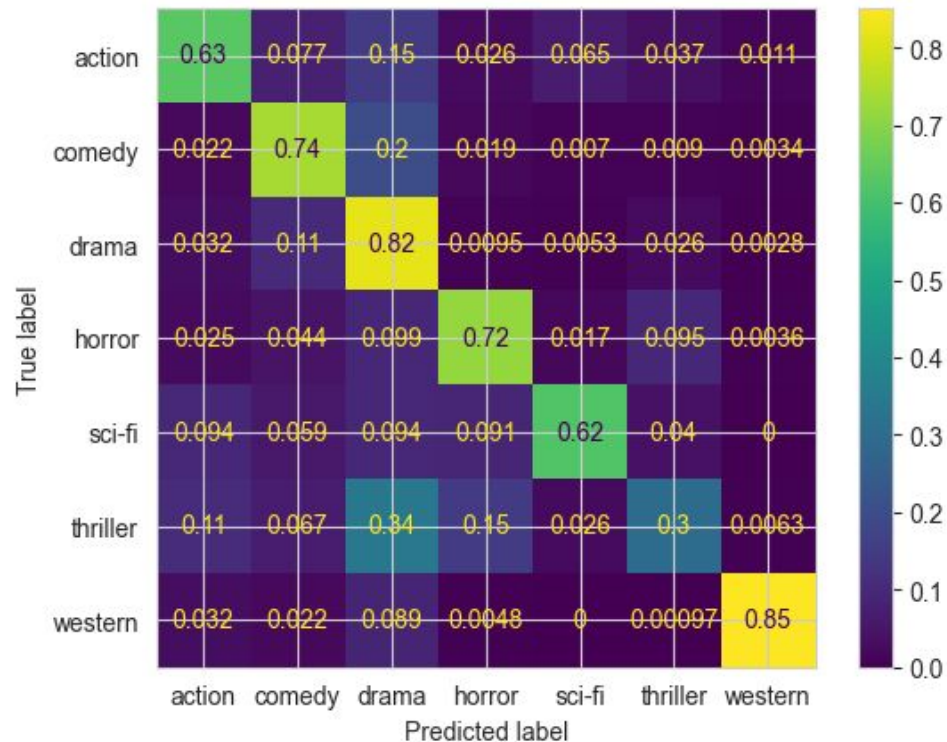
	Precision	Recall	F1-score	Support
Action	0,34	0,38	0,36	1314
Comedy	0,68	0,77	0,73	7446
Drama	0,75	0,88	0,79	13612
Horror	0,61	0,74	0,67	2204
Sci-fi	0,00	0,00	0,00	646
Thriller	0,18	0,01	0,03	1590
Western	0,00	0,00	0,00	1032
Accuracy			0,69	27844
Macro Avg	0,37	0,39	0,37	27844
Weighted Avg	0,62	0,69	0,65	27844

5. Comparison between Models and Biases

Simple Transformer



BERT 4 epoch



5. Comparison between Models and Biases

Macro F-1 weights each category **equally**.

Model	Weighted F-1	“Macro” F-1	Computation Time
TF (Baseline)	0.440	0.289	~ 1 min
LSTM (RNN)	0.640	0.547	~ 45 min
Simple Transformer	0.659	0.576	~ 8 hrs
BERT	0.746	0.663	~ 96 hrs *

* ~ 3.5 hours using GPU

5. Comparison between Models and Biases

Bias: Year

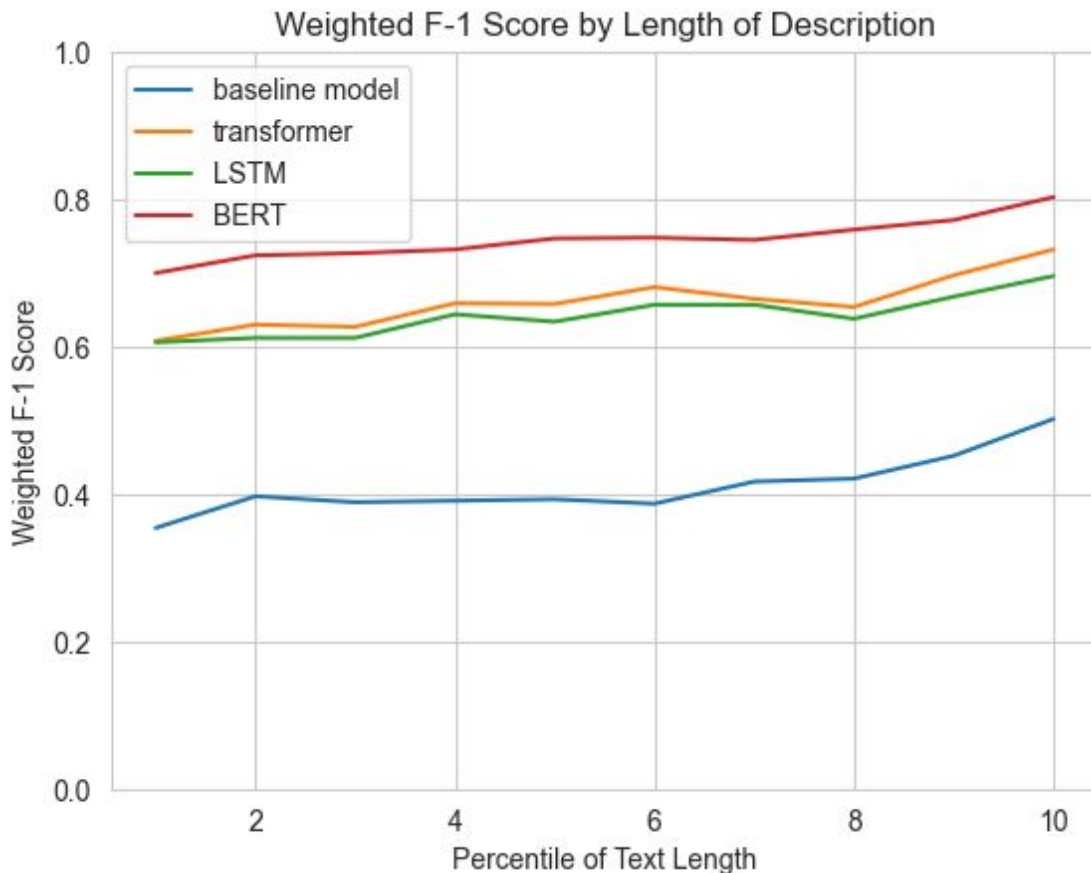
Split movies into quartiles based on **year of release** and compute weighted F-1 score for each quartile.

	TF (baseline)	LSTM (RNN)	Simple Transformer	BERT
pre-1982	0.449	0.687	0.703	0.783
1982 - 2005	0.421	0.638	0.662	0.75
2005 - 2013	0.420	0.652	0.671	0.756
2013 - 2022	0.399	0.630	0.645	0.725

5. Comparison

Bias: Length of Description

Split test set into 10 quantiles based on **character length of description** and calculate weighted F-1 score for each quantile.



5. Comparison between Models and Biases

Macro F-1 weights each category **equally**.

Model	Weighted F-1	"Macro" F-1	Computation Time
TF (Baseline)	0.440	0.289	~ 1 min
LSTM (RNN)	0.640	0.547	~ 45 min
Simple Transformer	0.659	0.576	~500 min (8 hrs)
BERT	0.746	0.663	~ 210 min * (3.5 hours)

* using GPU