

Applied Machine Learning - Final Presentation

Predicting Opening-Weekend Box Office Sales Based on Movie Trailers

Institute for Data Processing (LDV)

Aug 7th, 2019

Akram Louati
Andre Karinda
Jan Tóth
Oussema Dhaouadi
Patrick Willner
Rachid Ellouze
Roland Bode
Syrine Chelly

Motivation



The New York Times



SUBSCRIBE NOW | LOG IN

'Hobbs & Shaw' Leads at the Box Office

The "Fast and Furious" spinoff brought in about \$60.8 million during its opening weekend — and significantly more overseas.

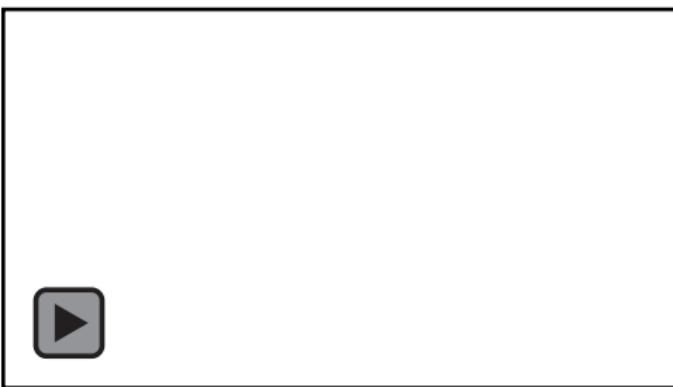


Jason Statham, left, and Dwayne Johnson in "Fast & Furious Presents: Hobbs & Shaw." The movie debuted to \$60.8 million in domestic ticket sales this weekend.
Daniel Smith/Universal Pictures

Source: <https://www.nytimes.com/2019/08/04/arts/hobbs-shaw-leads-at-the-box-office.html>

Motivation

Experiment: let's predict the revenue based on trailers



2006



2015



Motivation

Experiment: let's predict the revenue based on trailers



UNIVERSAL

A COMCAST COMPANY

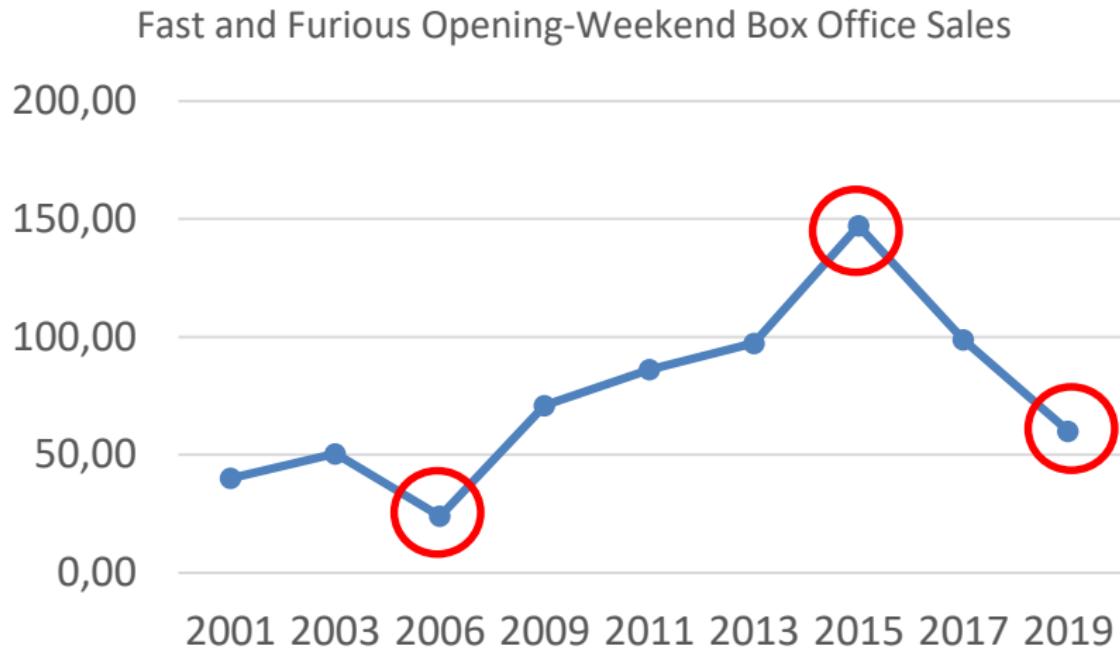
2015



2019



Motivation



Source: <https://www.boxofficemojo.com/franchises/chart/?view=main&id=fastandthefurious.htm&sort=date&order=DESC&p=.htm>

Overview

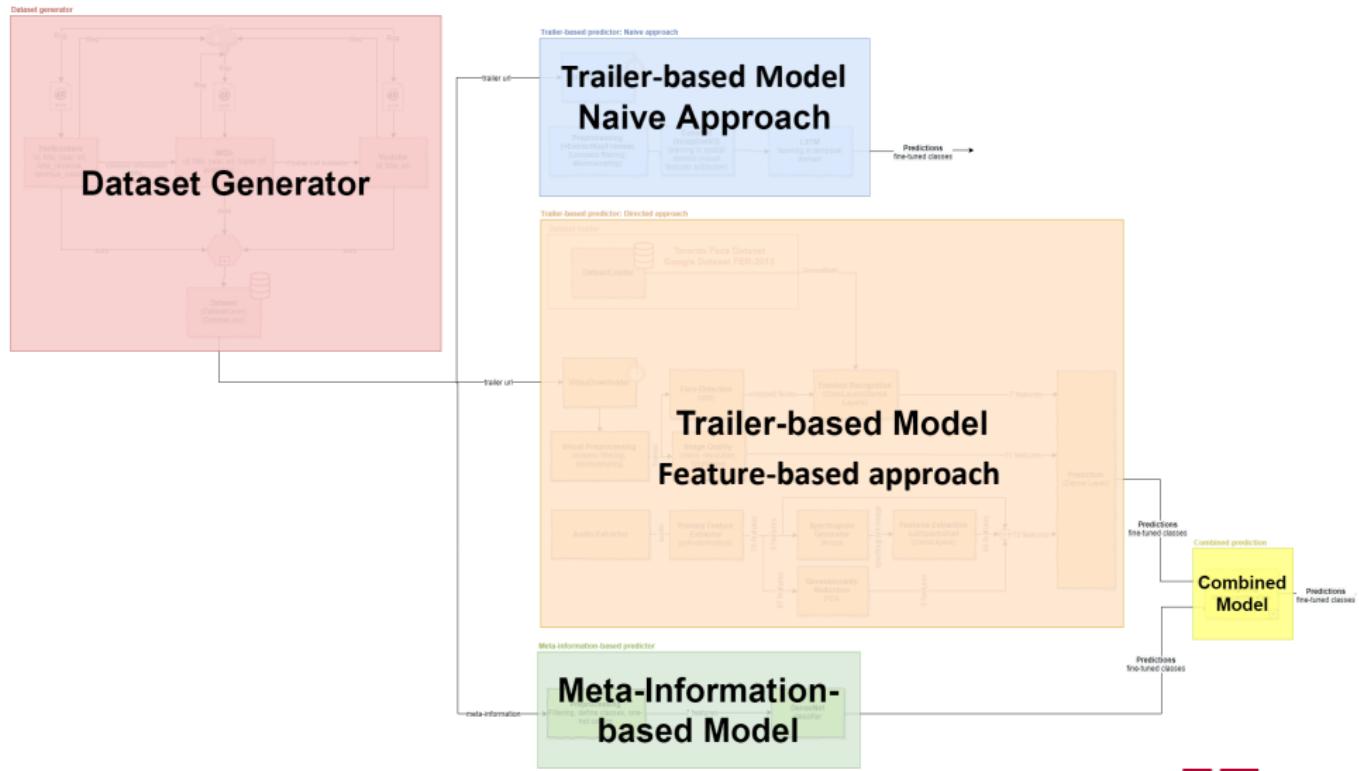


Table of contents

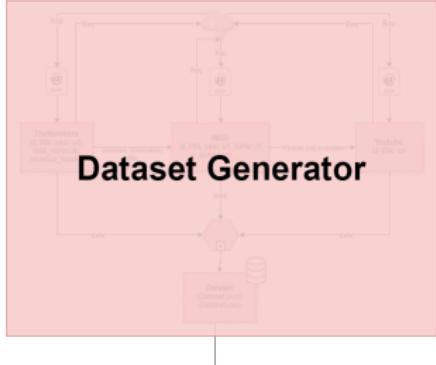
- I. Dataset Generation
- II. Meta-Information-based Model
- III. Trailer-based Model
- IV. Combined Model
- V. Frontend
- VI. Conclusion and Discussion

Table of contents

- I. Dataset Generation
- II. Meta-Information-based Model
- III. Trailer-based Model
- IV. Combined Model
- V. Frontend
- VI. Conclusion and Discussion

Dataset Generation

Dataset generator



Dataset Generator

Trailer Lf

Trailer-based Model Naive Approach

Trailer-based Predictor: Naive approach

Process: **Face detection** (haar classifier), **Feature extraction** (haar classifier), **Dimensionality reduction** (PCA), **Classification** (SVM)Predictions
fine-tuned classes

Trailer-based Model Feature-based approach

Meta-information based predictor

Meta-Information- based Model

meta-information

or

or

or

Process: **Face detection** (haar classifier), **Feature extraction** (haar classifier), **Dimensionality reduction** (PCA), **Classification** (SVM)Predictions
fine-tuned classes

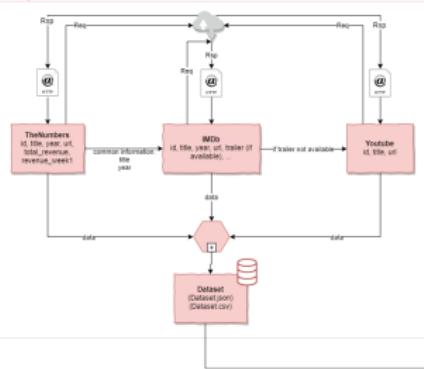
Combined Model

Predictions
fine-tuned classes

or

Dataset Generation

Dataset generator



Trailer-based predictor: Naive approach

**Trailer-based Model
Naive Approach**

Trailer-based predictor: Directed approach

**Trailer-based Model
Feature-based approach**

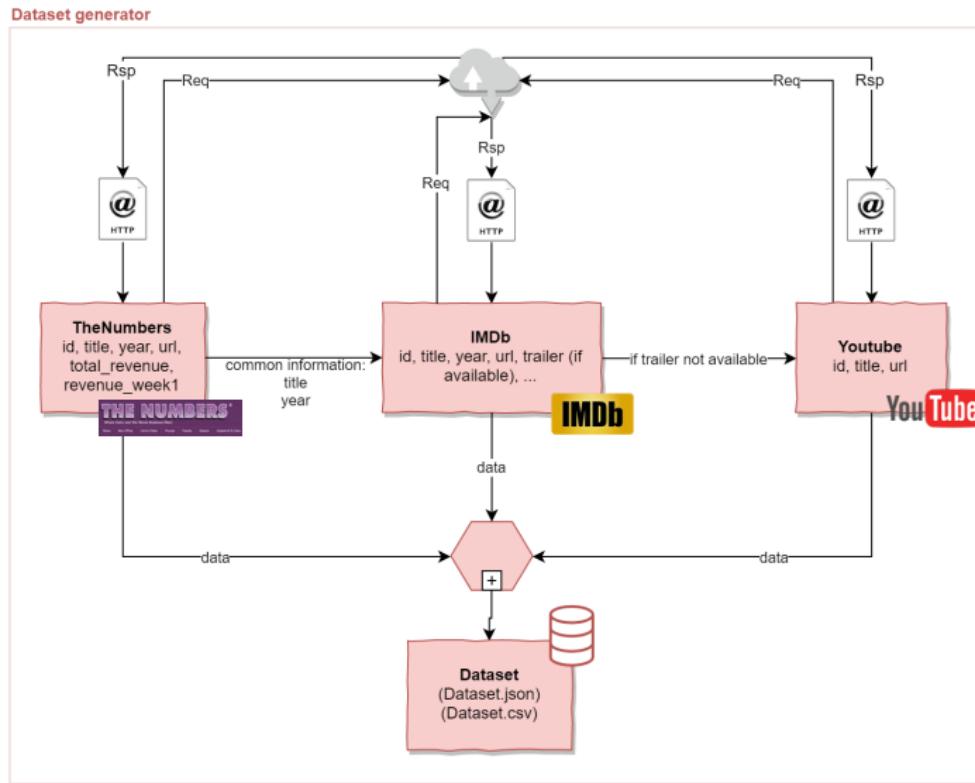
Meta-information based predictor

Meta-Information-based Model

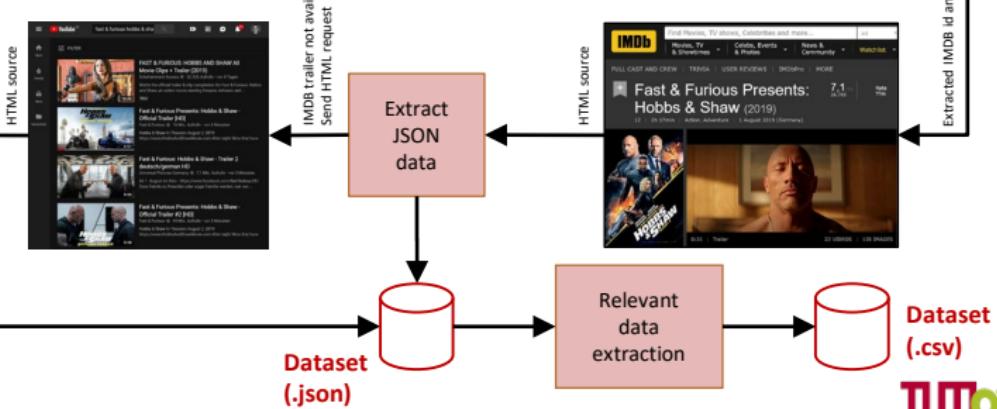
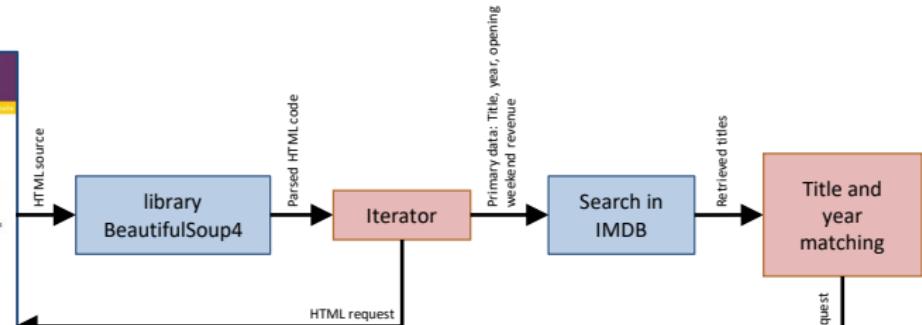
Predictions
fine-tuned classes

Combined Model

Dataset Generation



Dataset Generation



Dataset Generation

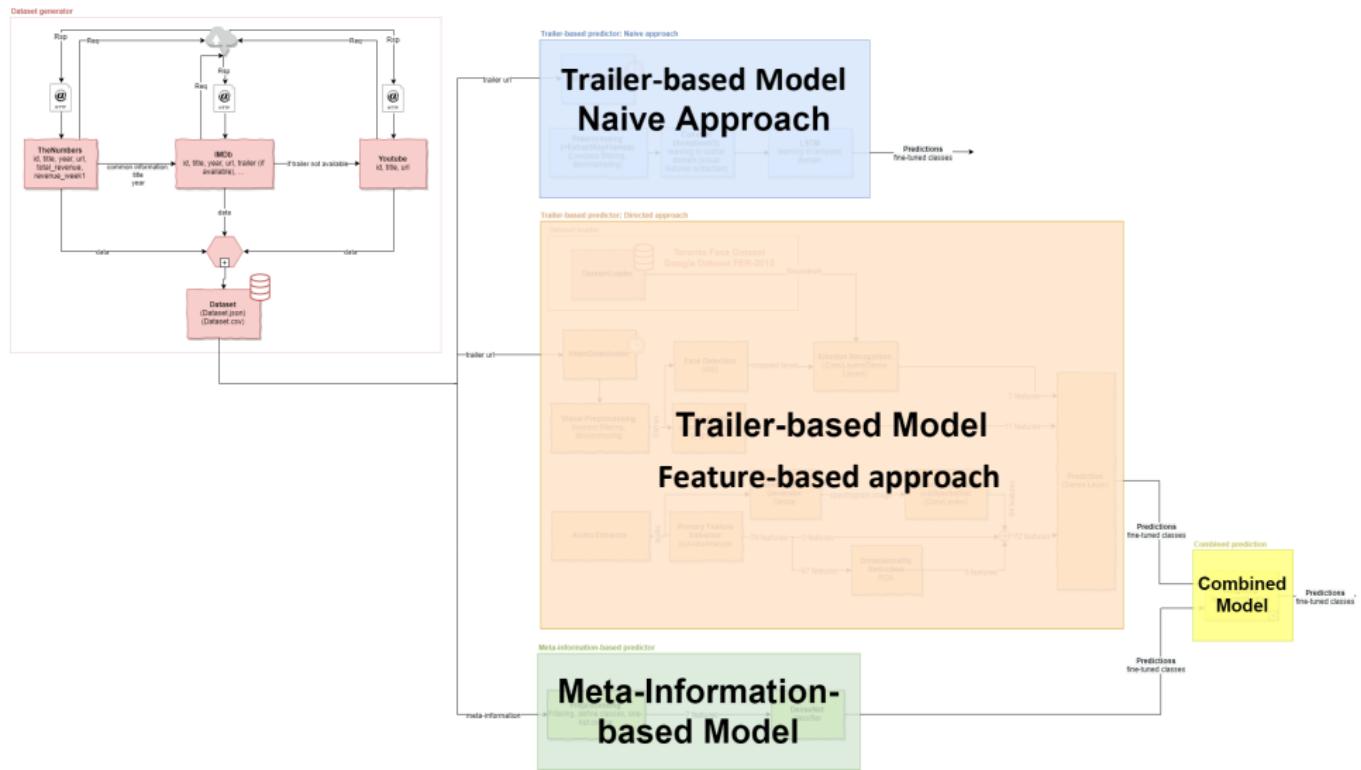
Problems	Solutions
Titles in TheNumbers and IMDb do not match.	Use a correlation metric by counting the cardinality of the intersection set.
Years in TheNumbers and IMDb do not match.	Set an interval of confidence ± 3 years.
Extracting the german version of the trailer.	Set the geographical place as US in the header when sending the HTML request.
Trailer URL extraction is also based on search machines, which are not 100% accurate.	Some False Positives could be filtered using thresholds (e.g. duration < 10min).

→ In some cases, manual correction.

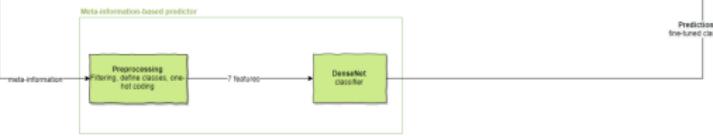
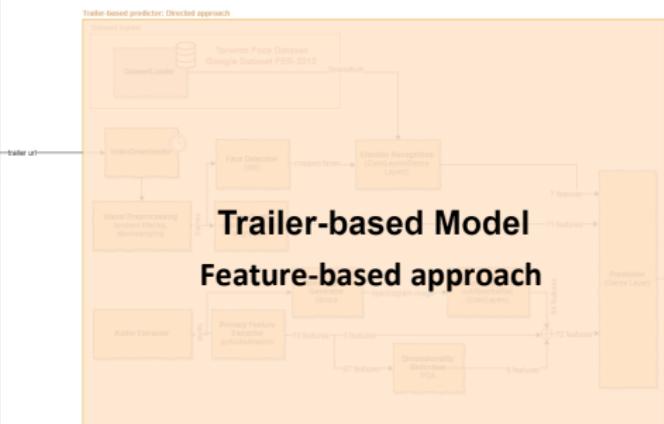
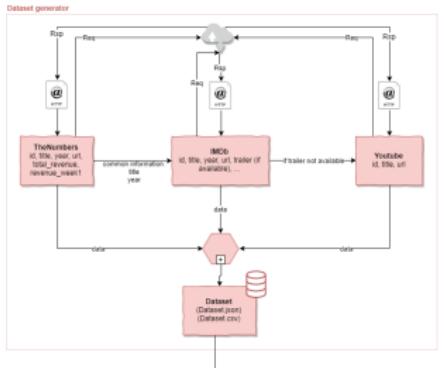
Table of contents

- I. Dataset Generation
- II. Meta-Information-based Model
- III. Trailer-based Model
- IV. Combined Model
- V. Frontend
- VI. Conclusion and Discussion

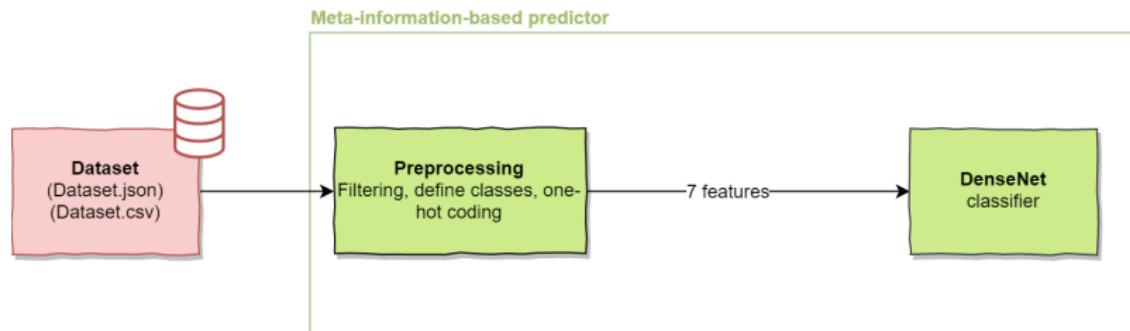
Meta-Information-based Model



Meta-Information-based Model



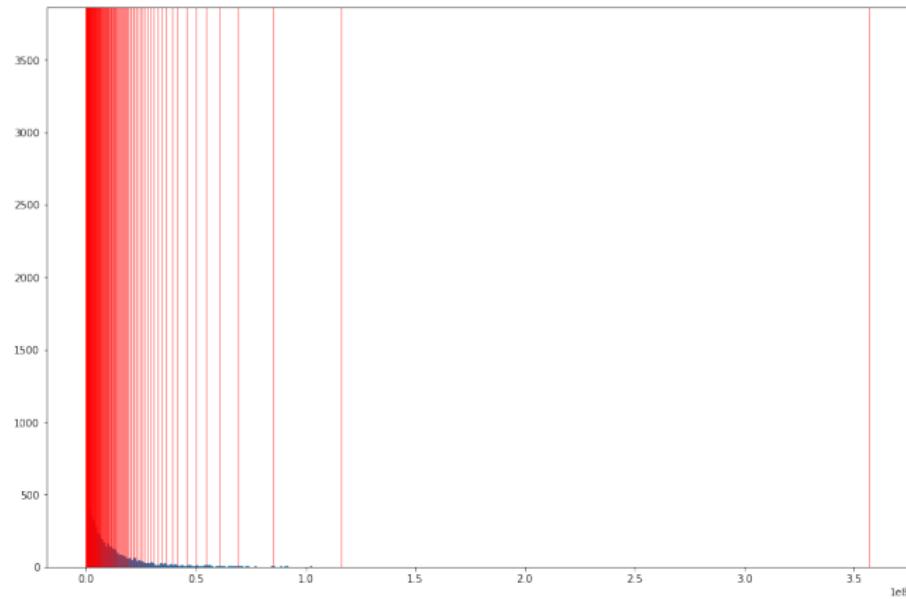
Meta-Information-based Model



- Get feature from our dataset for each movie.
- Filter the movies by country and select only English movies, i.e. from US or England.
- Model input: $X = [\text{'actors'}, \text{'creators'}, \text{'directors'}, \text{'genres'}, \text{contentRating'}, \text{'year'}, \text{'duration'}]$
- Module: Classifier.

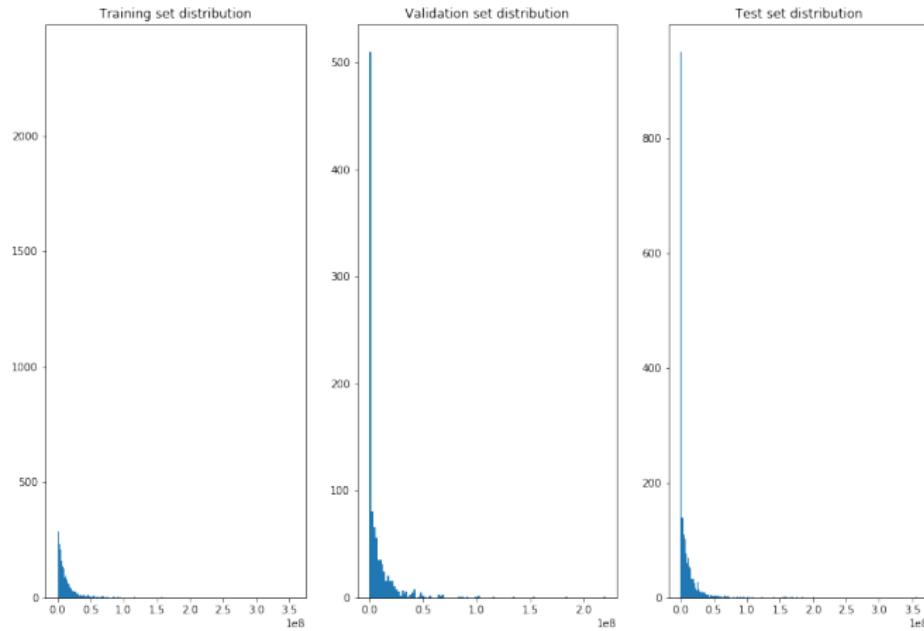
Meta-Information-based Model – Labeling

- 200 classes were defined by splitting the probability distribution into equally probable regions using quantile function.



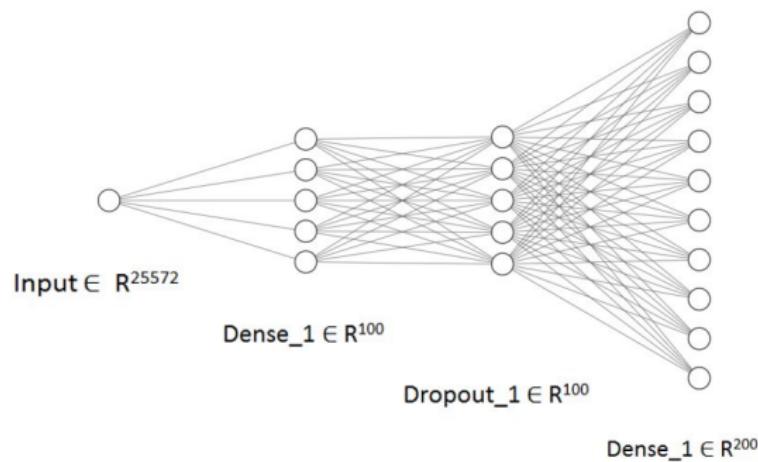
Meta-Information-based Model – Dataset Splitting

- Split the dataset after shuffling while conserving the shape of the probability of the original data.



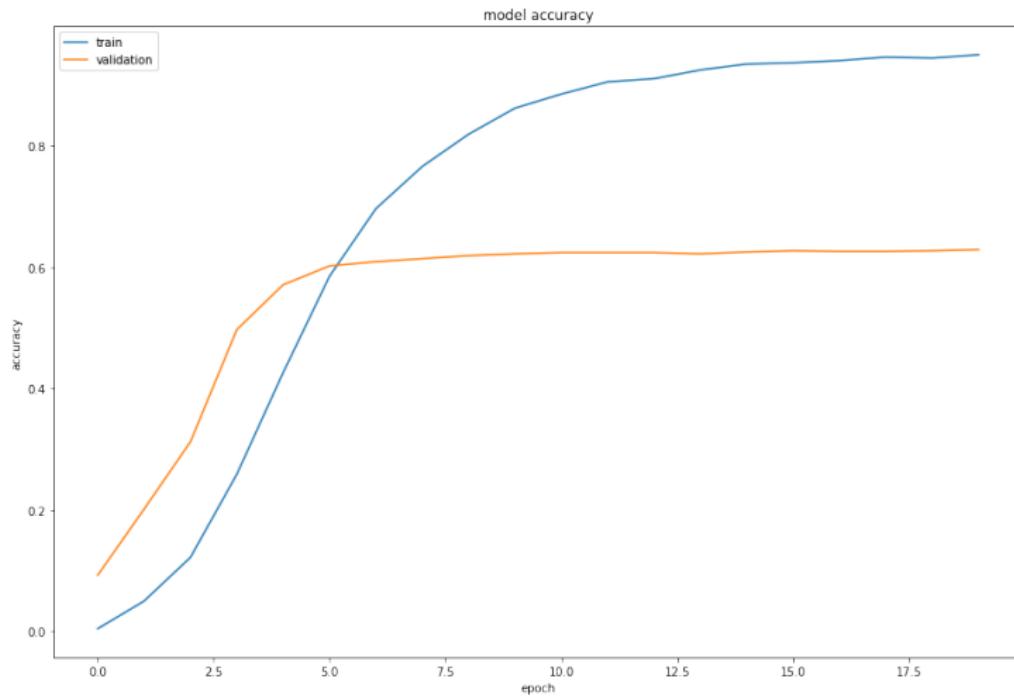
Meta-Information-based Model – Training Model

- Sequential Model
- 3 Layers
- Input dimension: Dimension from extracted one-hot encoded features
- Output Dimension: one-hot classes (200 classes)
- Total number of parameters: 2,577,500



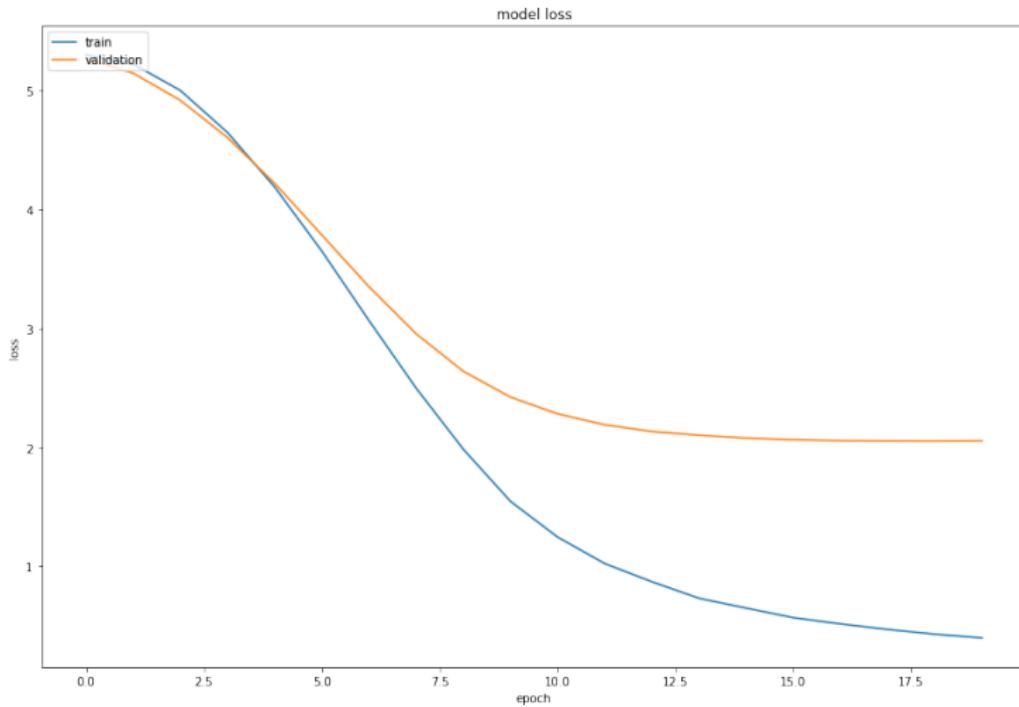
Meta-Information-based Model – Evaluation

- Accuracy



Meta-Information-based Model – Evaluation

- Loss: Categorical cross entropy loss



Meta-Information-based Model – Evaluation

- Testing

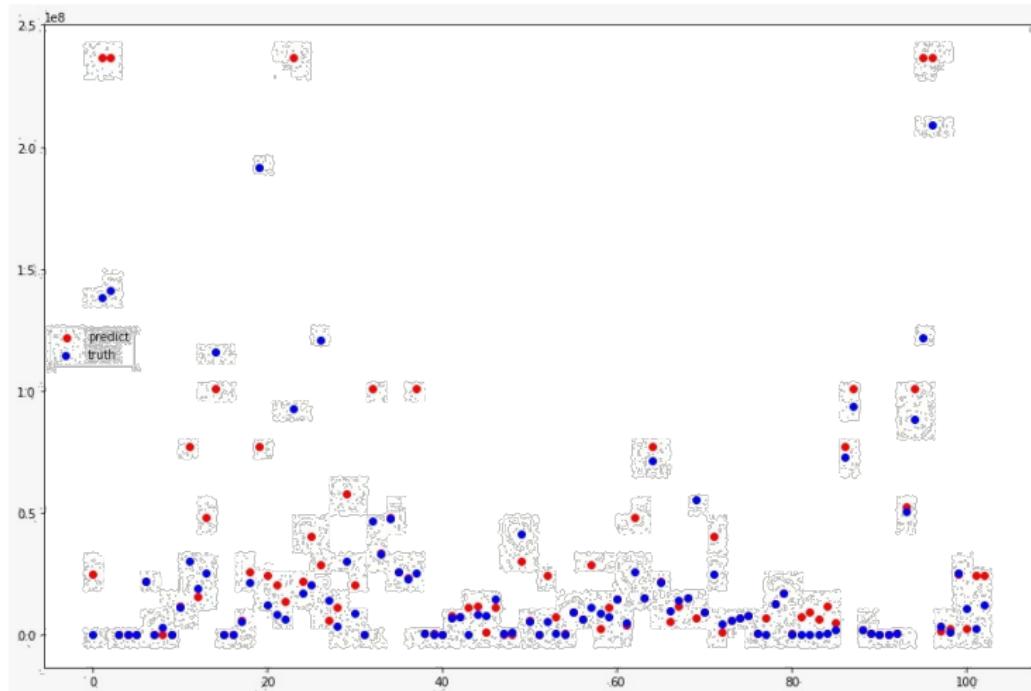
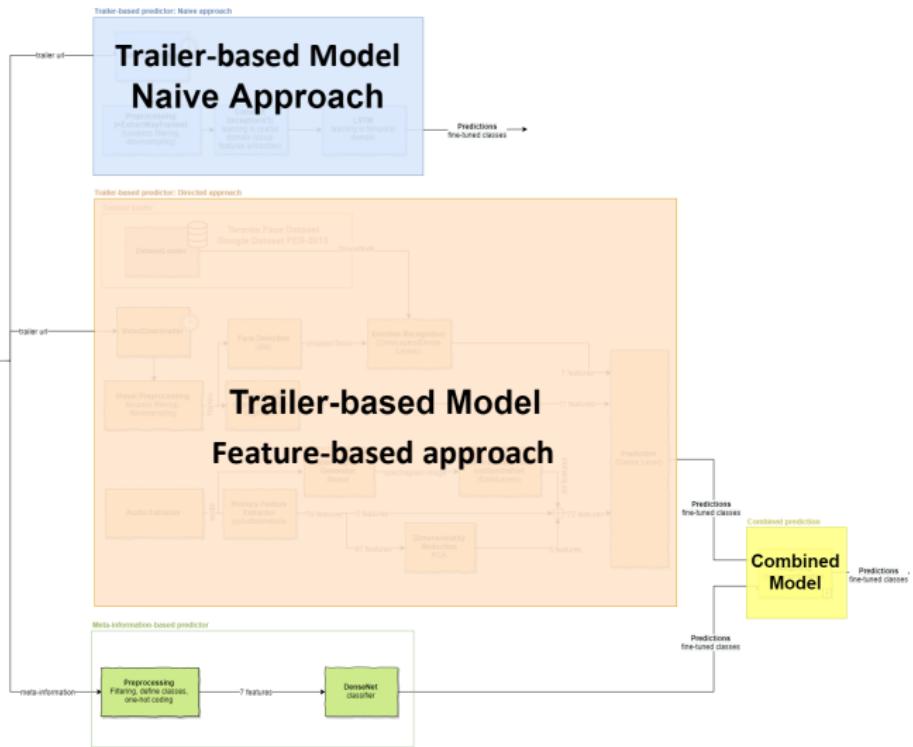
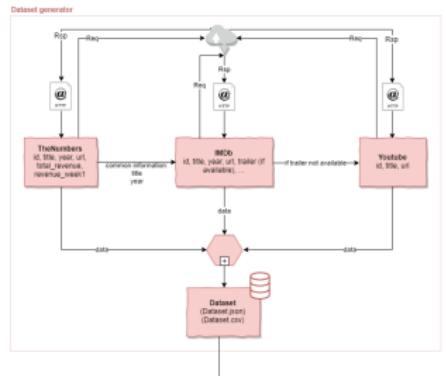


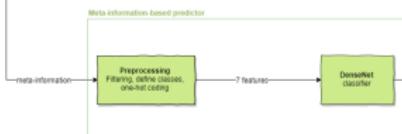
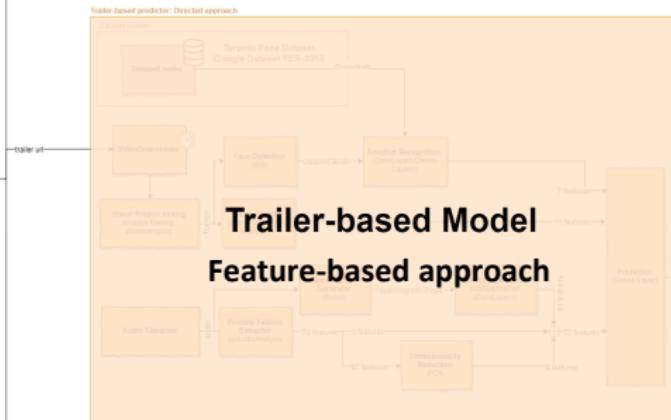
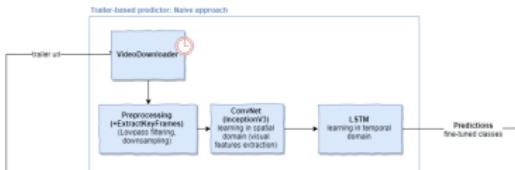
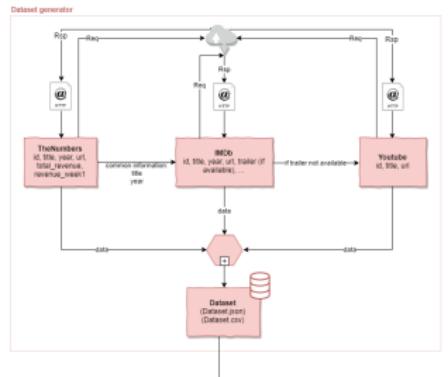
Table of contents

- I. Dataset Generation
- II. Meta-Information-based Model
- III. Trailer-based Model
- IV. Combined Model
- V. Frontend
- VI. Conclusion and Discussion

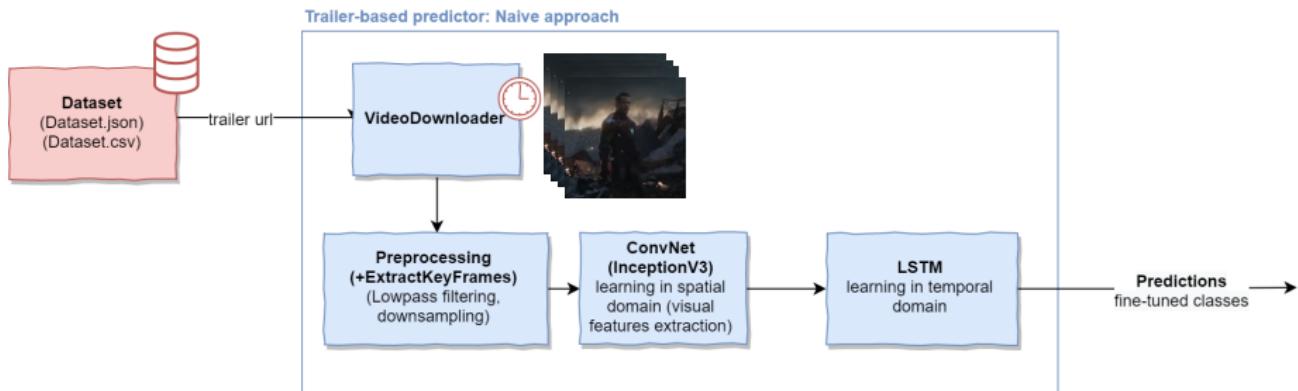
Trailer-based-Model



Trailer-based-Model – Naive Approach

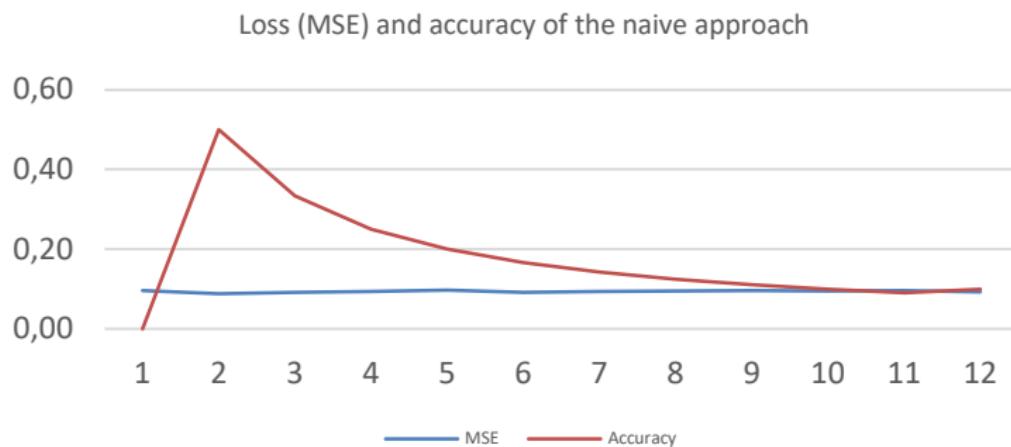


Trailer-based-Model – Naive Approach



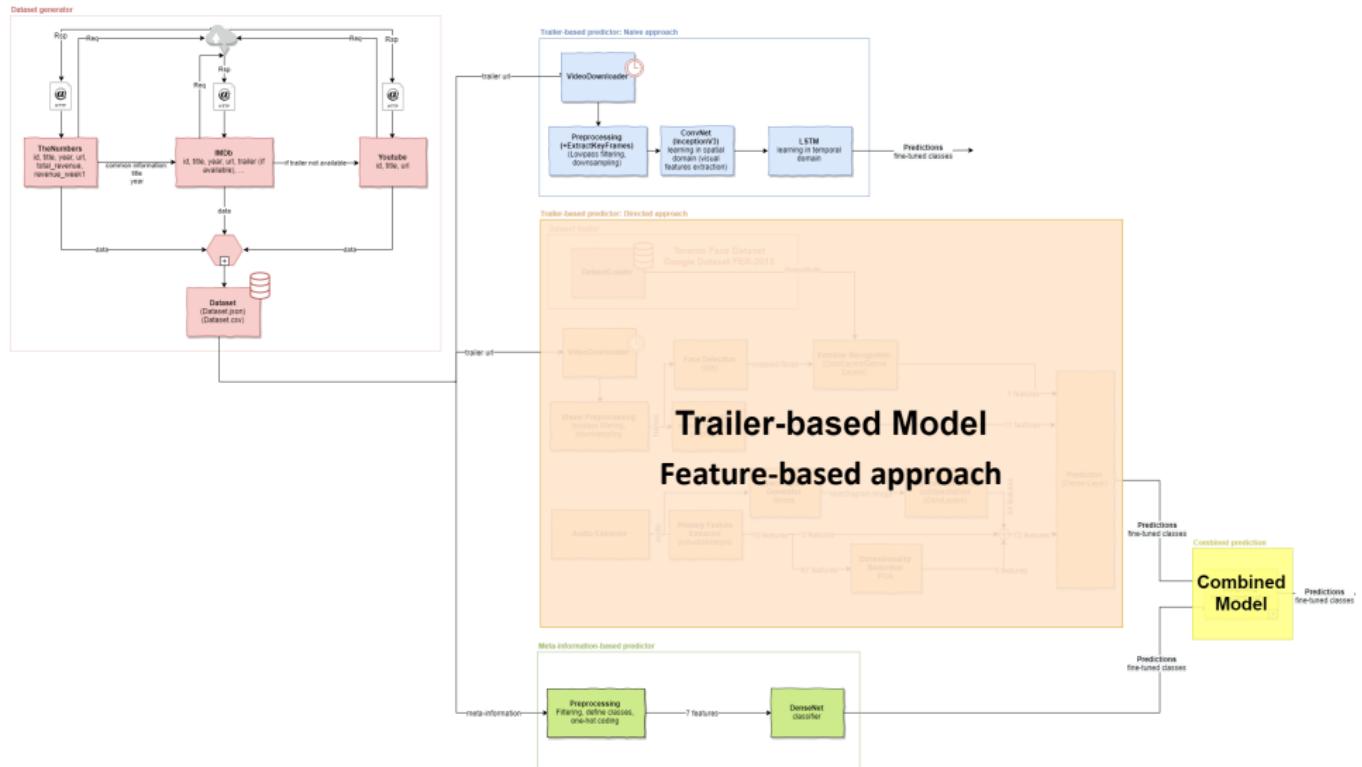
- Architecture: Input Layer → TimeDistributedDense Layer connected to InceptionV3 → LSTM → Dense Layer
- Unguided Learning
- Total number of parameters: 1,115,914

Trailer-based-Model – Naive Approach

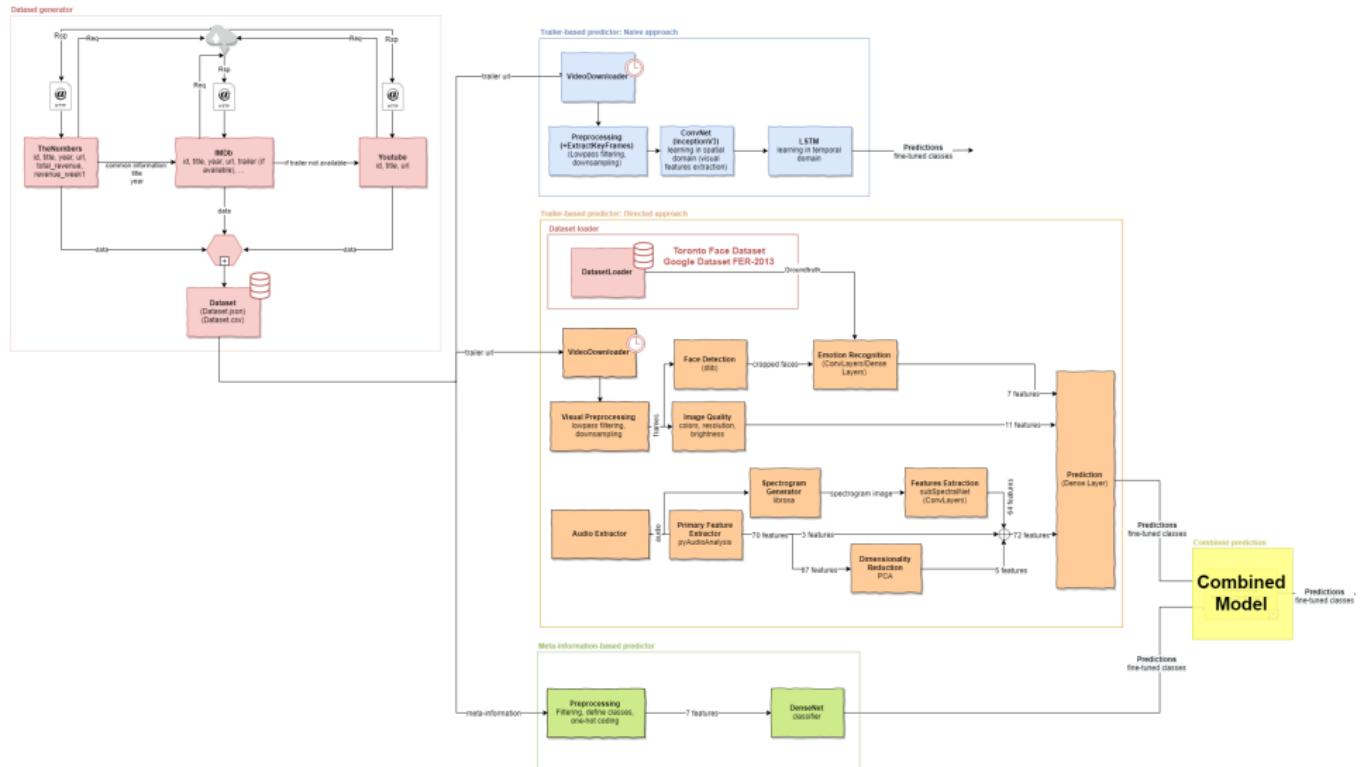


Problems	Solution
Poor performance.	Guided learning based on feature extraction preprocessing.
Unable to learn relevant features	

Trailer-based-Model

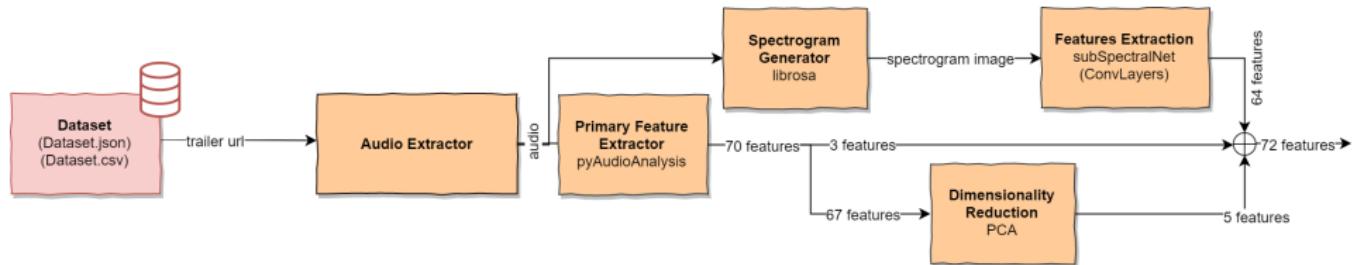


Trailer-based-Model – Feature-based approach



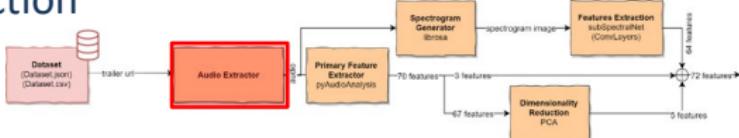
Trailer-based-Model – Feature-based approach

- Audio Analyzer



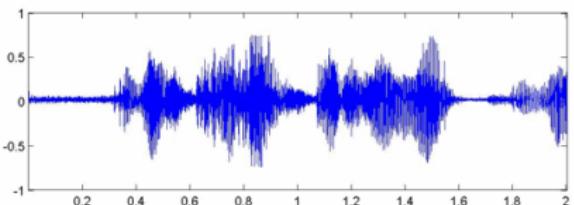
Trailer-based-Model – Feature-based approach

- Audio Analyzer - Audio Extraction



Trailer

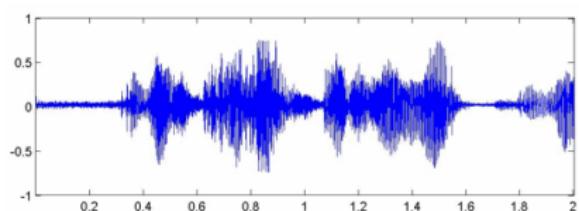
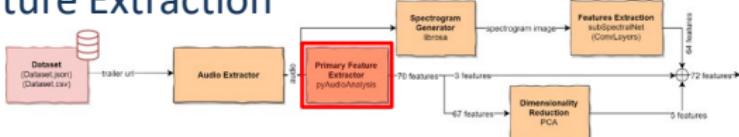
FFmpeg



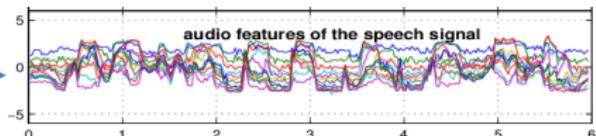
Raw audio signal

Trailer-based-Model – Feature-based approach

- Audio Analyzer – Primary Feature Extraction



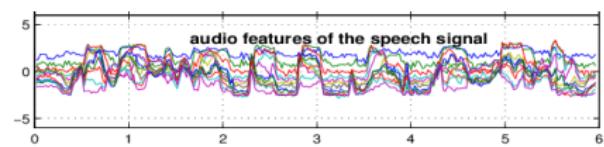
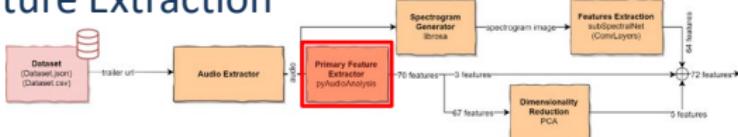
Raw audio signal



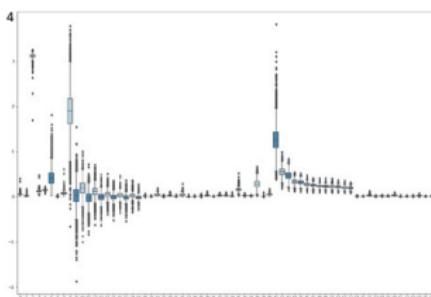
70 features

Trailer-based-Model – Feature-based approach

- Audio Analyzer – Primary Feature Extraction



70 features

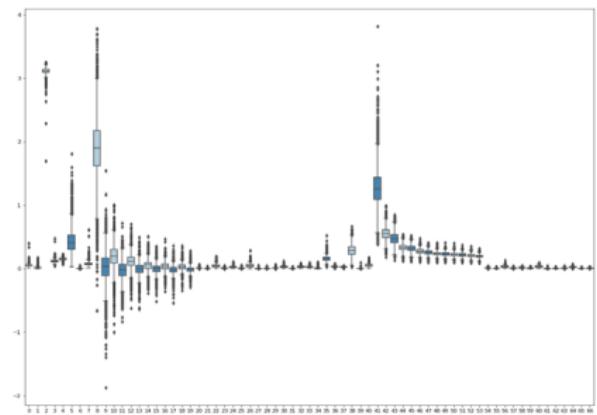
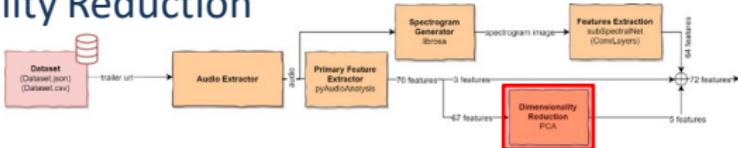


67 dense features

3 significant features
(beat, its confidence, mfcc mean)

Trailer-based-Model – Feature-based approach

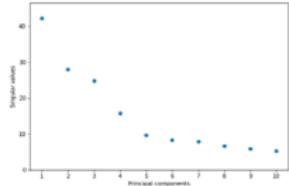
- Audio Analyzer – Dimensionality Reduction



scikit-learn

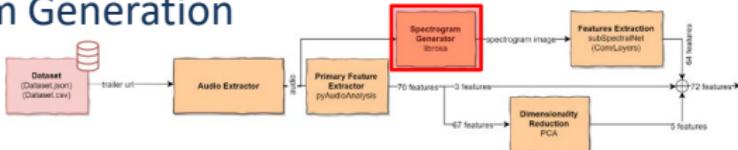
PCA

5 features

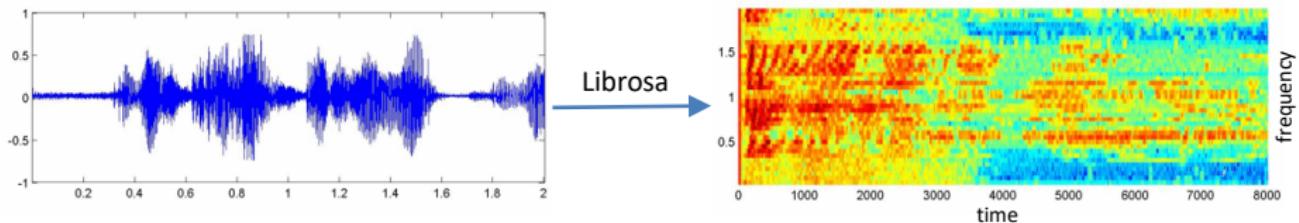


Trailer-based-Model – Feature-based approach

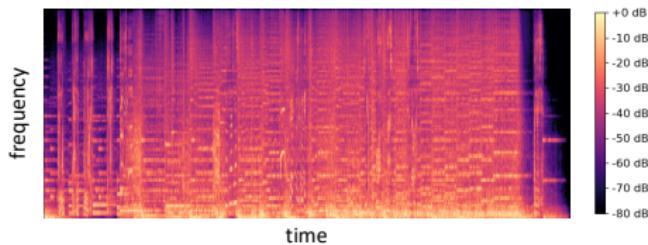
- Audio Analyzer – Spectrogram Generation



Spectrogram: Representation of the power spectrum of a sound.

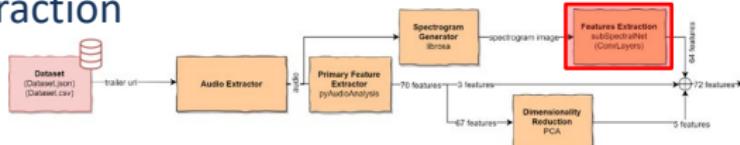


e.g. MEL-Spectrogram of Avengers Endgame

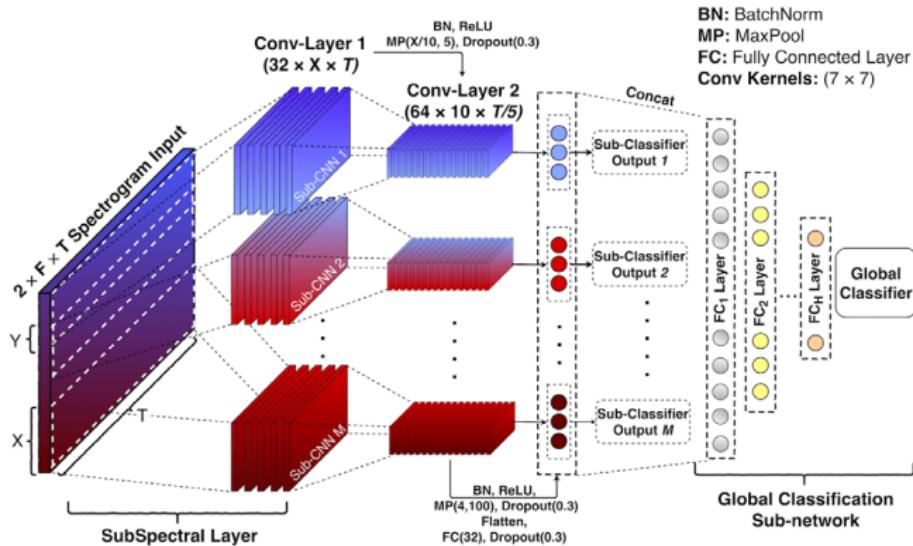


Trailer-based-Model – Feature-based approach

- Audio Analyzer – Feature extraction



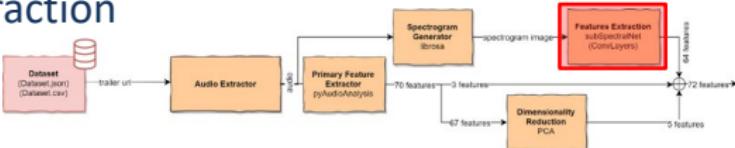
Architecture: SubSpectral Net* (R Phaye et. al.)



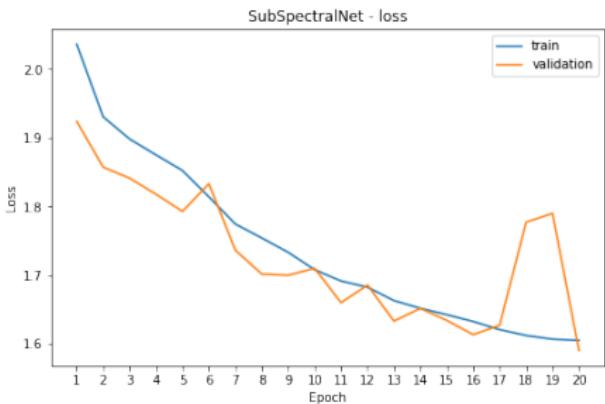
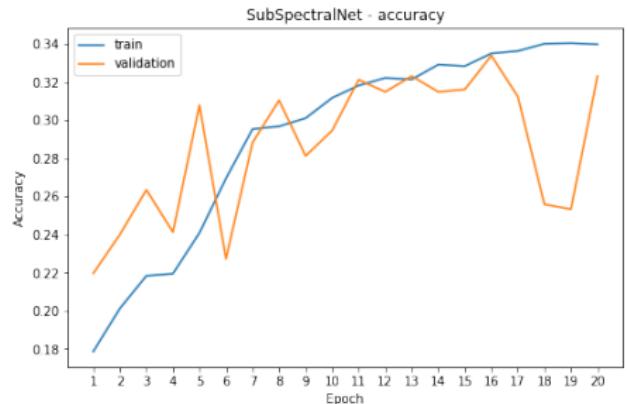
*SUBSPECTRALNET – USING SUB-SPECTROGRAM BASED CONVOLUTIONAL NEURAL NETWORKS FOR ACOUSTIC SCENE CLASSIFICATION (2019)

Trailer-based-Model – Feature-based approach

- Audio Analyzer – Feature extraction



Training and validation



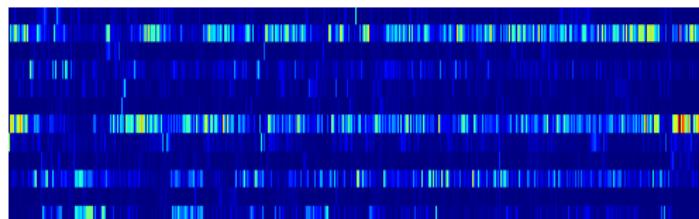
- Training the network on 8000 spectrograms of our trailers with 6 target classes.
- Best accuracy: 30%.

Trailer-based-Model – Feature-based approach

- Audio Analyzer – Attempt

Use chromograms in addition to the spectrograms:

- Try to retrain ResNet50, VGG19 and VGG16.

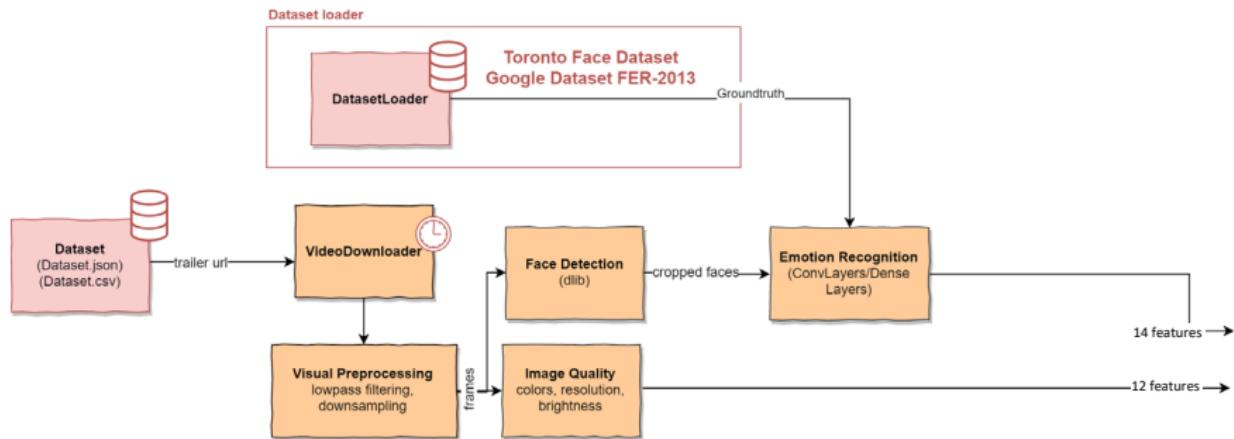


Chromogram of Avengers Endgame

Problem	Possible Solution
1/6 accuracy (deterministic classification of 1 class out of 6 classes)	Try different data normalization

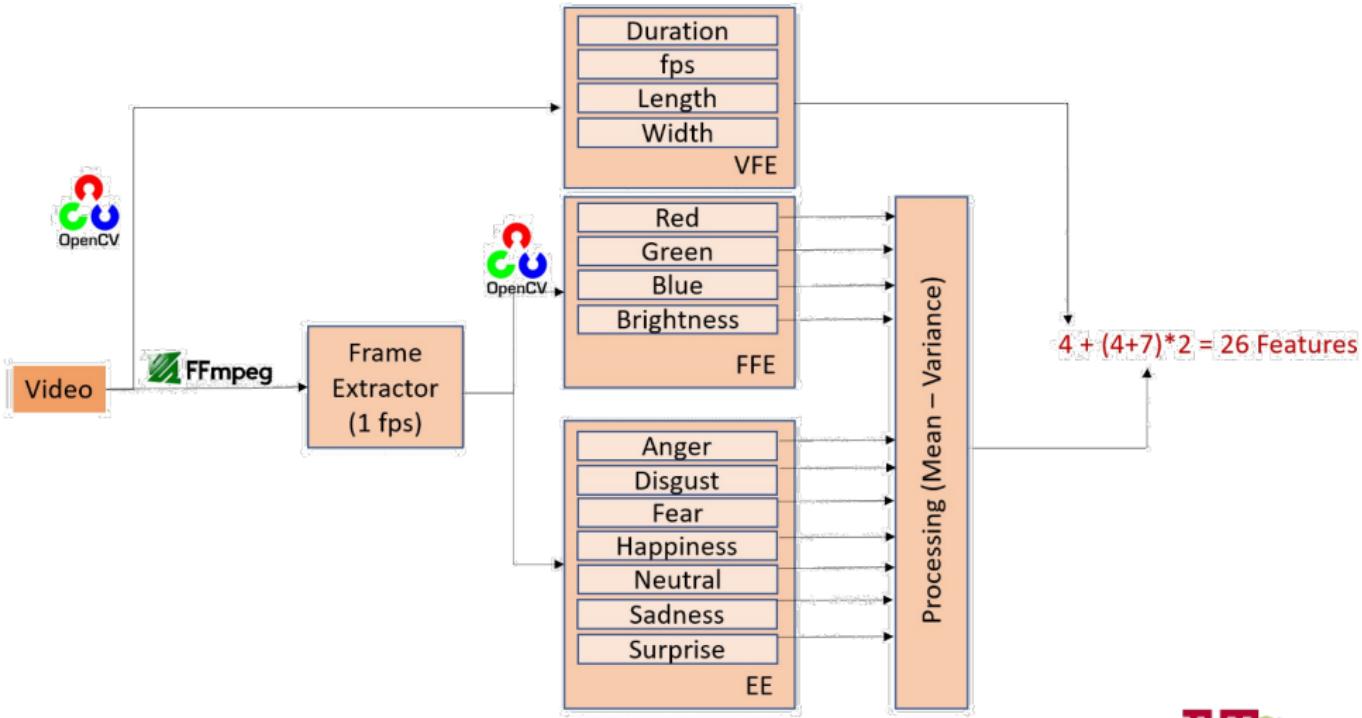
Trailer-based-Model – Feature-based approach

- Visual Analyzer



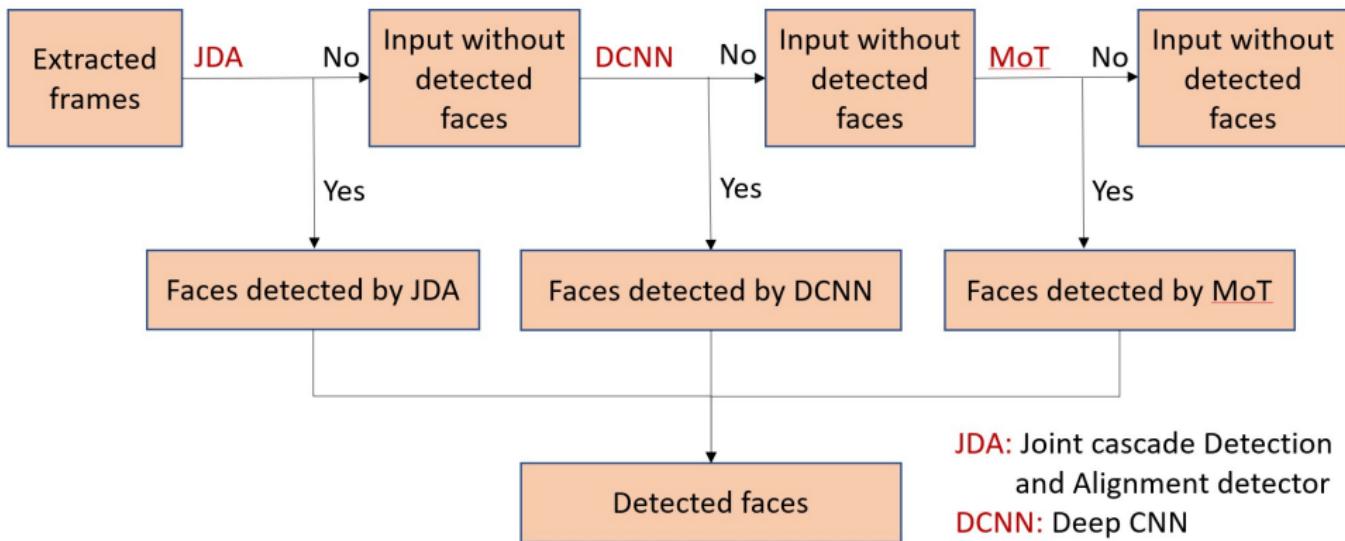
Trailer-based-Model – Feature-based approach

- Visual Analyzer – Feature extraction



Trailer-based-Model – Feature-based approach

- Visual Analyzer – Face Detector



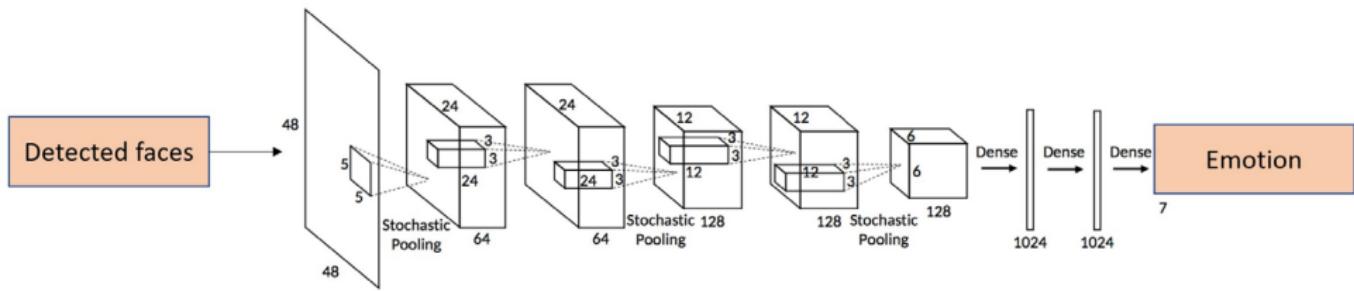
JDA: Joint cascade Detection
and Alignment detector

DCNN: Deep CNN

MoT: Mixtures of Trees

Trailer-based-Model – Feature-based approach

- Visual Analyzer – Emotion Predictor



Source: Yu, Z., & Zhang, C. (2015, November). Image based static facial expression recognition with multiple deep network learning. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (pp. 435-442). ACM.

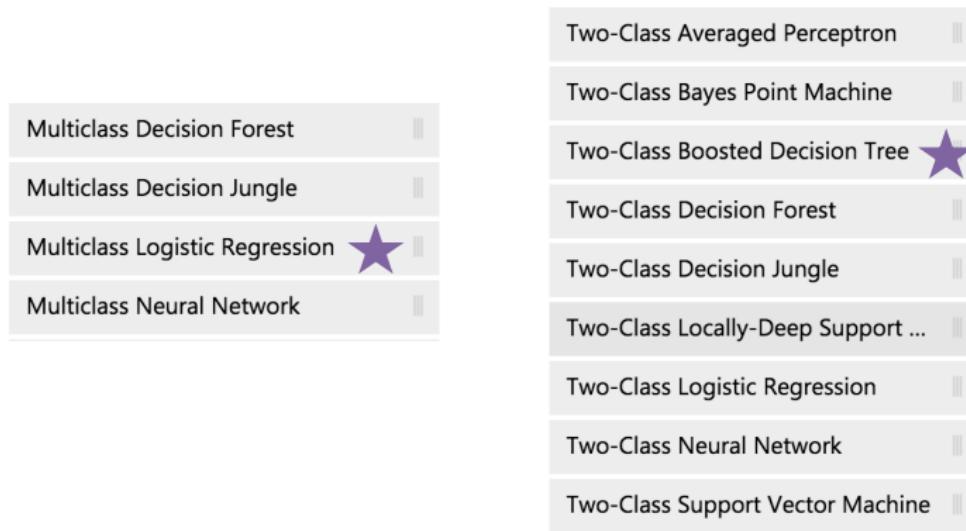
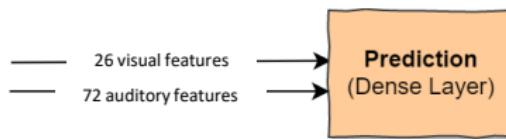
Trailer-based-Model – Feature-based approach

- Visual Analyzer – Performance

Problem	Solution
Computational performance: 4min/video → $4 \times 7900 = 31600\text{min} = 22$ days	<ul style="list-style-type: none">• Use only one extractor (down to 11 days)
Emotion Extraction is the most computationally expensive	<ul style="list-style-type: none">• Extract faces at a 0.1 fps rate (down to 5 days)

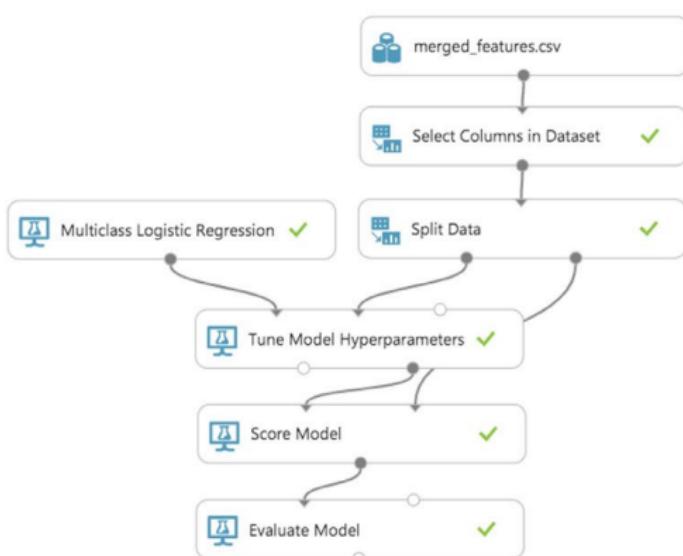
Trailer-based-Model – Feature-based approach

- Model - Selection



Trailer-based-Model – Feature-based approach

- Model - Flow Diagram



Microsoft Azure Machine Learning Studio

Specs:

B1S 1 vCPU 1 GiB RAM

Training data: 3525 trailer *98 feature
Training set: 80% Training data
Test set: 20% Training data
Computation time: 5 min
Model Complexity: $O(nm)$

Trailer-based-Model – Feature-based approach

- Model - Evaluation

Metrics

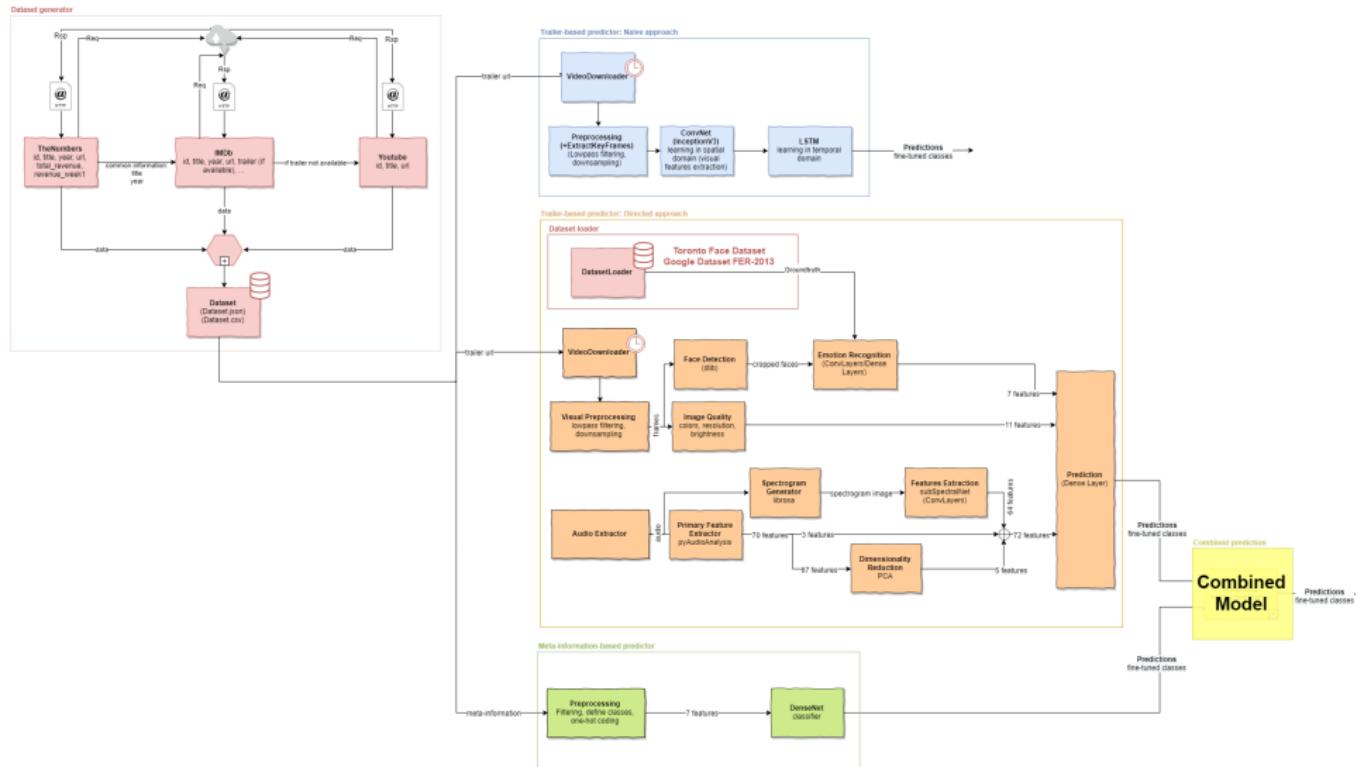
Overall accuracy	0.30922
Average accuracy	0.861844
Micro-averaged precision	0.30922
Macro-averaged precision	0.31057
Micro-averaged recall	0.30922
Macro-averaged recall	0.288811

		Predicted Class									
		3358.05	14589.95	47481.65	169284.85	1004789.6	2885210.4	5755499.85	10046085...	17734825...	18998828...
Actual Class	3358.05	50.7%	16.4%	4.5%	6.0%		7.5%			14.9%	
	14589.95	29.9%	26.0%		7.8%	9.1%	2.6%	5.2%	1.3%		2.6% 15.6%
	47481.65	18.6%	18.6%	18.6%	15.3%	1.7%	10.2%				3.4% 13.6%
	169284.85	15.6%	6.3%	7.8%	17.2%		14.1%	9.4%	6.3%		23.4%
	1004789.6	9.1%	6.8%	10.2%	12.5%	9.1%	26.1%	9.1%			2.3% 14.8%
	2885210.4	5.3%		3.5%	8.8%	1.8%	42.1%	10.5%	3.5%	8.8%	15.8%
	5755499.85	6.3%		7.8%	1.6%	6.3%	28.1%	25.0%	1.6%	3.1%	20.3%
	10046085...	2.6%	1.3%	3.9%	7.8%	1.3%	7.8%	10.4%	5.2%	6.5%	53.2%
	17734825...		2.0%	2.0%			4.1%	2.0%		14.3% 75.5%	
	18998828...	1.9%		1.0%	1.0%		1.9%	1.0%		12.6% 80.6%	

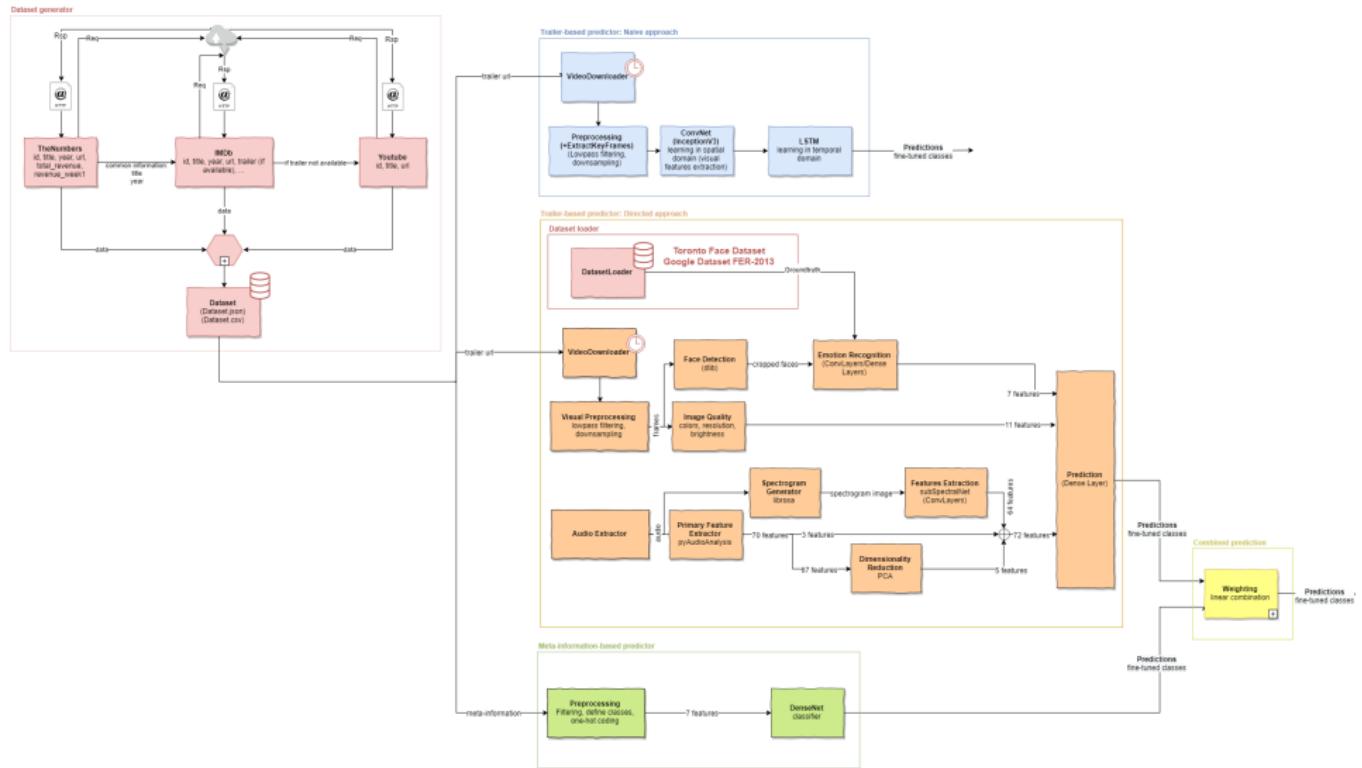
Table of contents

- I. Dataset Generation
- II. Meta-Information-based Model
- III. Trailer-based Model
- IV. Combined Model
- V. Frontend
- VI. Conclusion and Discussion

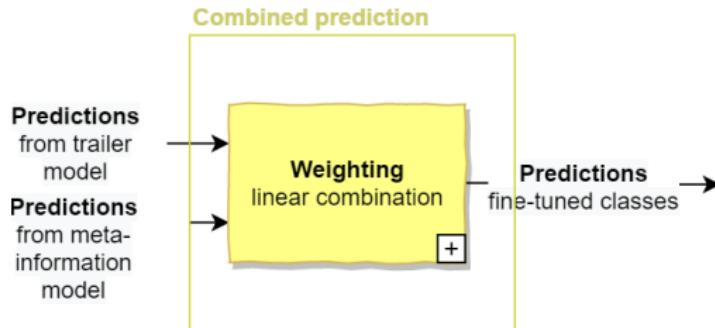
Combined Model



Combined Model



Combined Model



- Simple linear combination.
- Weight was learned by solving a least square error problem, whereby 20 data samples were randomly selected from the training dataset.
- Result: $\alpha_{\text{trailer-based-pred}} \approx 0.1$ $\alpha_{\text{meta-info-based-pred}} \approx 0.9$

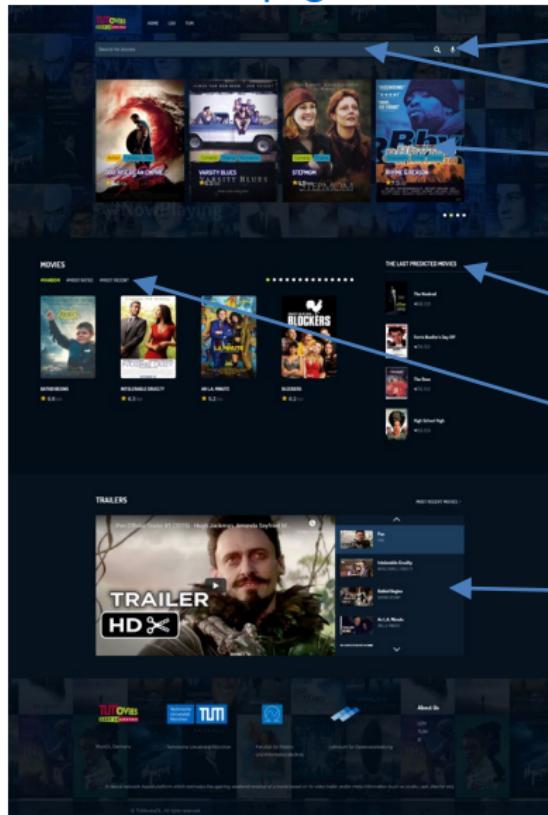
Table of contents

- I. Dataset Generation
- II. Meta-Information-based Model
- III. Trailer-based Model
- IV. Combined Model
- V. Frontend
- VI. Conclusion and Discussion

Frontend – Demo

Live Demo

Frontend – Homepage



Speech recognition

Search field (Title, IMDB id or Youtube URL/id)

Randomly selected movies

Last predicted movies

Random/most rated/ most recent movies

Random trailers

Frontend – Search page

The screenshot shows the TUMavies search interface. At the top, there is a navigation bar with the TUMavies logo, 'HOME', 'LDV', and 'TUM'. Below it is a search bar with the placeholder 'Search for movies' and a magnifying glass icon. The main content area has a dark background with movie posters. The title 'RESULTS FOR "FAST AND FURIOUS"' is displayed above a breadcrumb navigation 'HOME > SEARCH'. Below this, a subtitle 'Results from the IMDB database' is shown. The search results are organized in two rows of five movie posters each. The first row includes: 'THE FAST AND THE FURIOUS', 'FAST & FURIOUS', 'FAST AND FURIOUS', 'FAST & FURIOUS 9', 'THE FATE OF THE FURIOUS', and 'FURIOUS 7'. The second row includes: 'THE FAST AND THE FURIOUS', 'FAST & FURIOUS PRESENTS: HOBBS & SHAW', 'FAST AND FURIOUS', 'FURIOUS 6', '2 FAST 2 FURIOUS', and 'THE FAST & FURIOUS'. At the bottom of the page, there is a footer with links to 'TUMavies', 'Technische Universität München', 'TUM', 'Munich, Germany', 'Technische Universität München', 'Faculty of Electrical and Information Engineering', 'Lehrstuhl für Datenverarbeitung', 'About Us', 'HOME', 'LDV', 'TUM', and 'E1'. A small note at the bottom states: 'A neural-network based platform which estimates the opening weekend revenue of a movie based on its video trailer and/or meta-information (such as studio, cast, director etc.)'. A copyright notice at the very bottom reads: '© TUMaviesDL. All rights reserved.'

Results from IMDb online
database

Frontend – Movie page

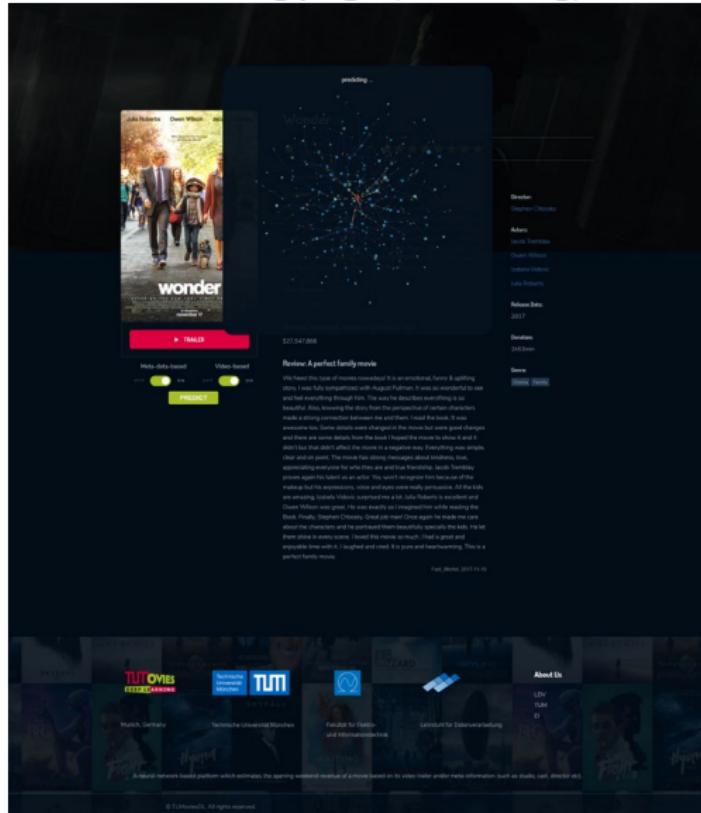
Randomly selected movies

Some metadata

Trailer player

Prediction mode (Metadata-based
and/or trailer-based)

Frontend – Loading page (Predicting)



Loading (10 sec for meta-data based prediction / 3 ~ 5min for trailed-based prediction)

Frontend – Results

The screenshot shows a movie search interface for 'Wonder'. At the top, there's a search bar with placeholder text 'Search for movies...'. Below it, a movie poster for 'Wonder' is displayed, followed by its title and a 5-star rating. To the right of the poster, there's a sidebar with navigation links like 'Home', 'About', 'Contact', 'Logout', and a dropdown menu for 'Machine Learning'. The main content area contains several sections: 'Storyline' (a brief plot summary), 'Trailer Statistics' (including 'Total Revenue' of \$123,000,000 and 'Opening Weekend Revenue (Domestic total)' of \$2,171,000), 'Emotions' (a pie chart showing emotion distribution: Anger 1%, Surprise 1%, Fear 1%, Joy 10%, Sadness 20%, and Neutral 50%), 'Color Features' (a table showing Red, Green, Blue, and Brightness values), 'Video Parameters' (Frame Rate, Height, Width, Duration), and 'Audio Features' (MP3, Bass, Bass confidence). At the bottom, there's a detailed 'Synopsis & Plot Summary' section with a large amount of text about the movie's plot.

Extracted emotions

Other trailer features (not all
features are shown)

Frontend – Results

The screenshot shows the TUTOVIES Frontend interface. At the top, there's a navigation bar with the TUTOVIES logo, 'HOME', 'UDV', and 'TUM'. Below it, a YouTube video player displays the trailer for 'FREAKONOMICS TRAILER HD OFFICIAL - YOUTUBE'. The video has 210,000 views and 47,481 likes. Below the video, the page displays various prediction results:

- Emotions:**

Emotion	Percentage
Anger	23.8%
Disgust	2.0%
Fear	26.8%
Happy	6.0%
Sad	26.0%
Surprise	2.3%
Neutral	13.0%
- Color features:**

Color	Value
Red	84.24
Green	84.52
Blue	85.48
Brightness	103.10
- Video parameters:**

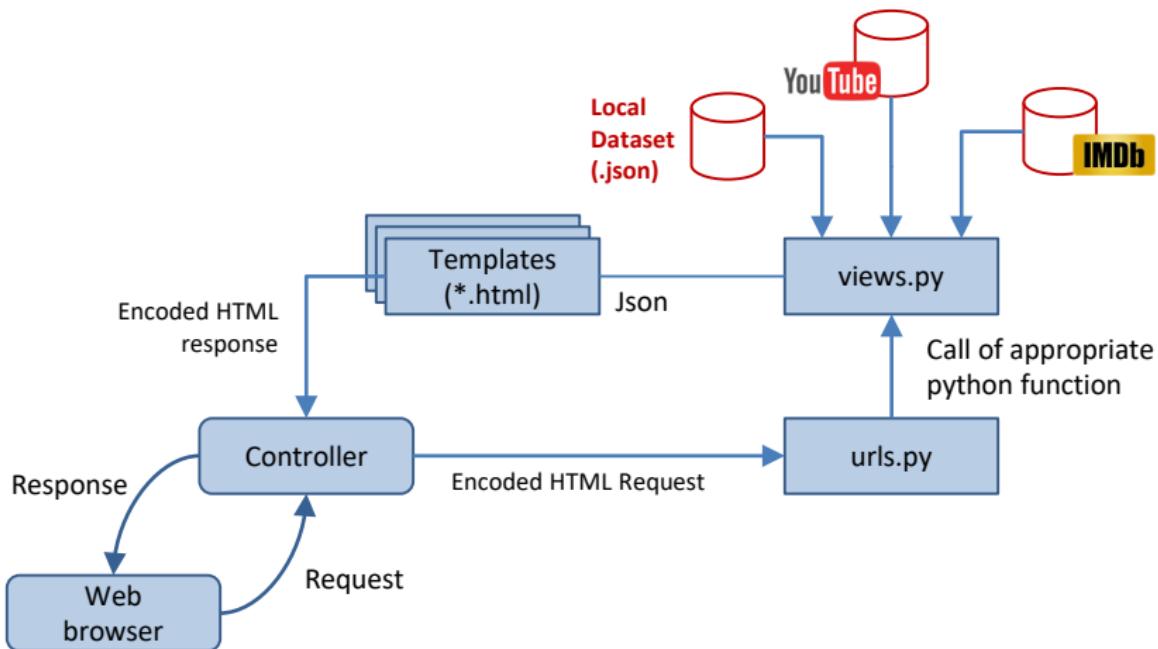
Parameter	Value
Frame rate	24.0
Height	220.0
Width	1380.0
Duration	155.65
- Audio features:**

Feature	Value
MFCC	28.43
Beat	150.00
Beat confidence	0.29

At the bottom of the page, there's a footer with links to 'About Us', 'UDV', 'TUM', and 'TUTORIALS'. It also includes the TUM logo and a note about being a neural-network based platform.

Movie page, when youtube url was provided and the results of prediction

Frontend – Framework **django**



- Programming languages: HTML, CSS, Python, Javascript, JQuery.

Table of contents

- I. Dataset Generation
- II. Meta-Information-based Model
- III. Trailer-based Model
- IV. Combined Model
- V. Frontend
- VI. Conclusion and Discussion

Conclusion and Discussion

- A machine learning algorithm requires 2 basic pillars: Data Generation and Model Development (architecture building, training, validation and testing).
- Solving a regression required a large dataset → Turning it into a classification problem may solve the problem, since classification is easier to learn.
- Slowing down the visual and auditory feature extraction is necessary, since downloading trailer from the Youtube may cause an IP suspension.
- Predicting the revenue is a hard task; Neural Networks solves this problem.
- The meta-information-based model outperforms the trailed-based model.
- Combining both models improves the accuracy.

Resources and References

- Resources:



Google
Colab



Microsoft Azure
ML Studio



GitHub



Google Drive

- References:

Yu, Z., & Zhang, C. (2015, November). Image based static facial expression recognition with multiple deep network learning. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (pp. 435-442). ACM.

SUBSPECTRALNET – USING SUB-SPECTROGRAM BASED CONVOLUTIONAL NEURAL NETWORKS FOR ACOUSTIC SCENE CLASSIFICATION (2019)

Thank you for
your attention!