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

TUMOVIES DEEP LEARNING

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

Source: Youtube

Motivation

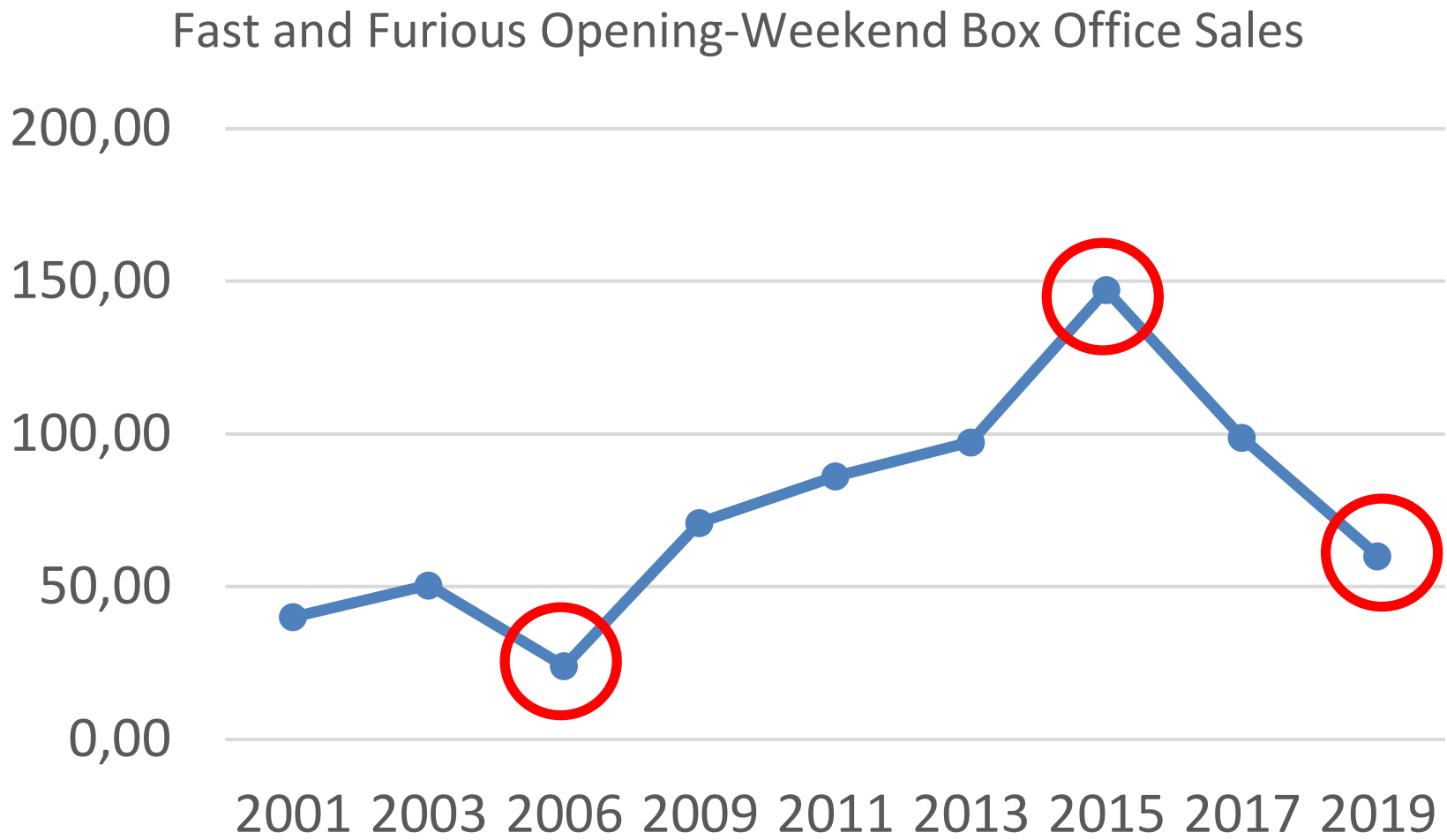
Experiment: let's predict the revenue based on trailers

2015

2019

Source: Youtube

Motivation



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

Overview

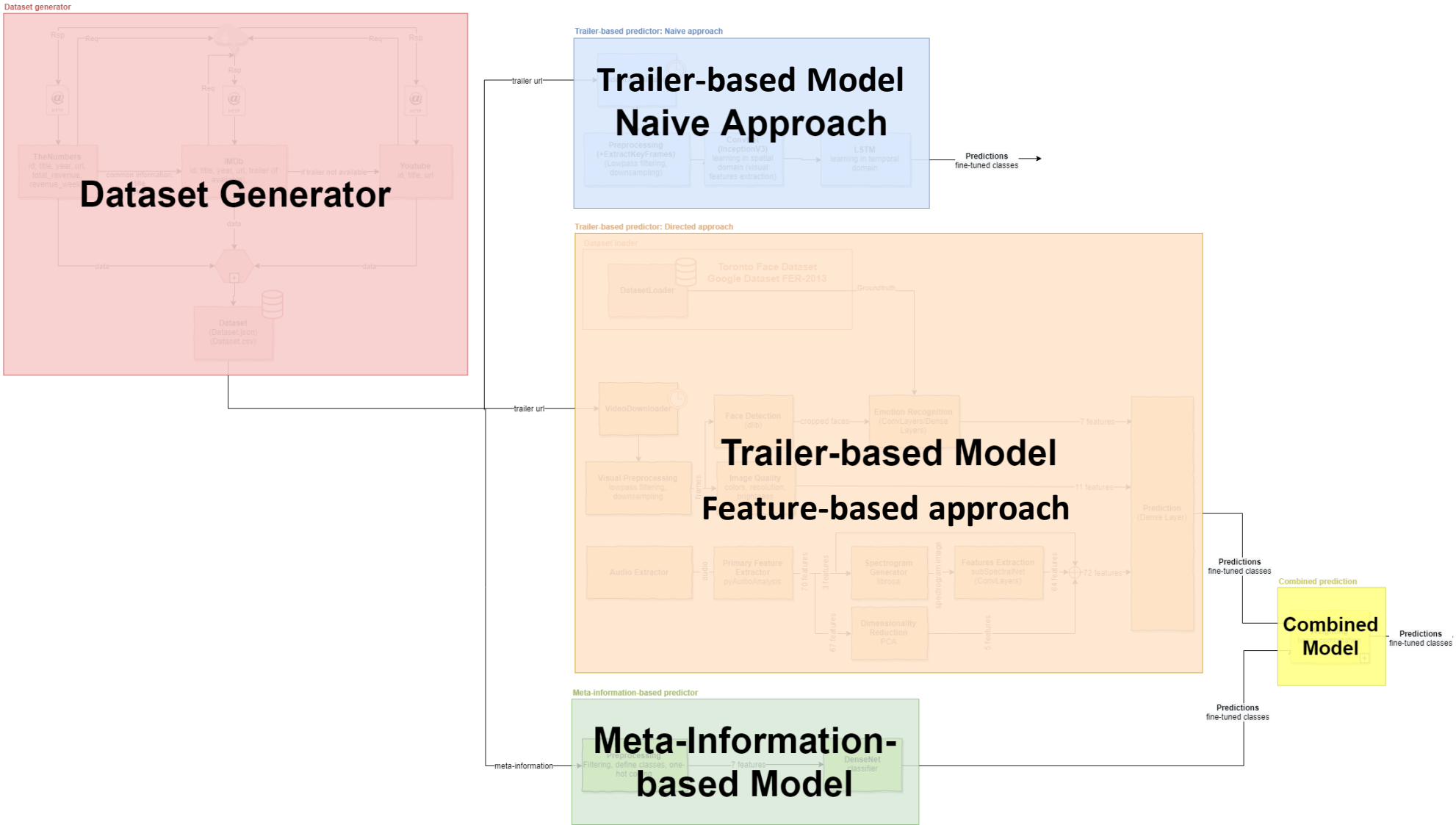
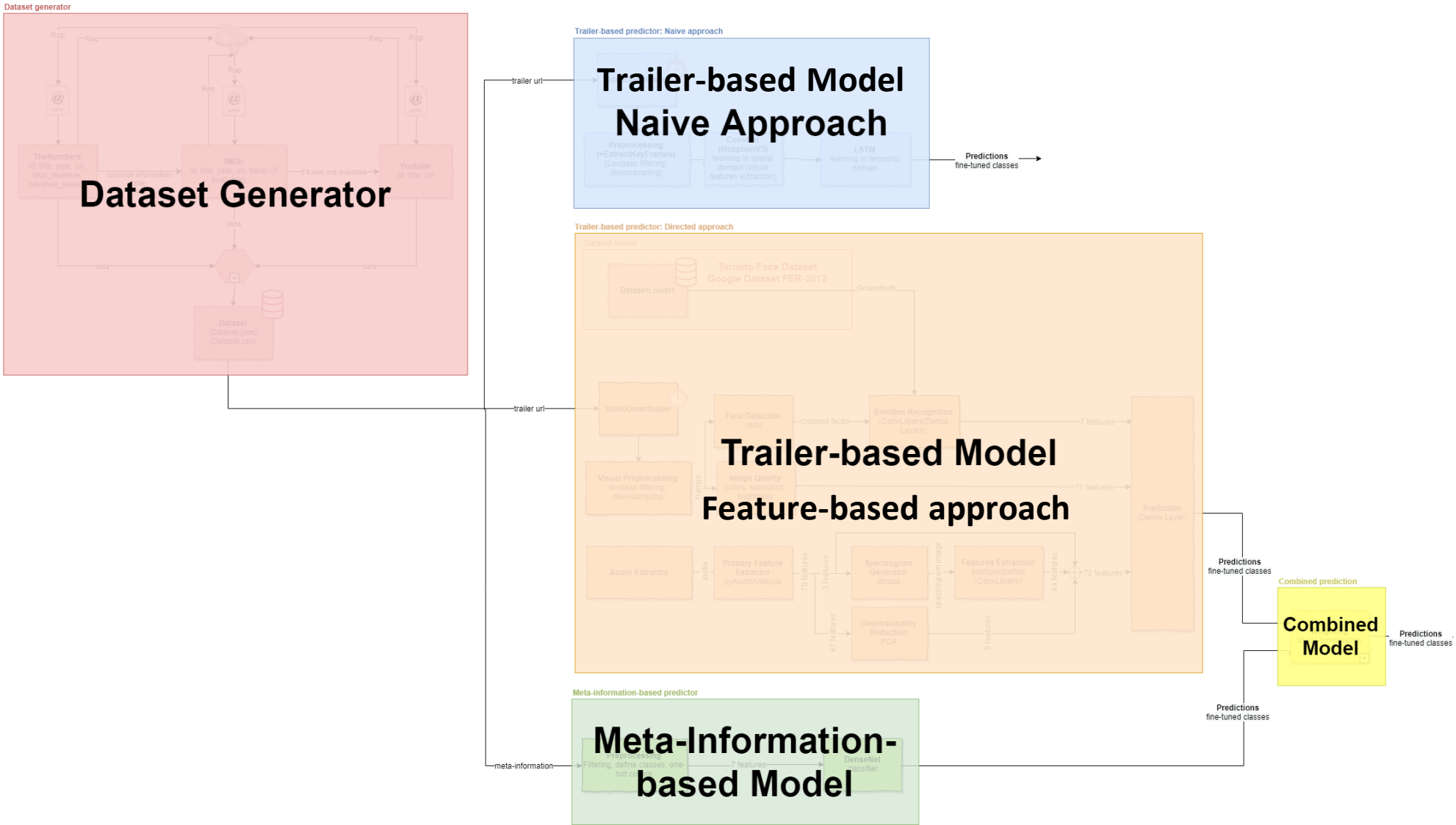

Dataset generator
Dataset Generator
trailer url
Trailer-based predictor: Naive approach
Trailer-based Model
Naive Approach
Predictions
fine-tuned classes
Trailer-based predictor: Directed approach
Trailer-based Model
Feature-based approach
trailer url
Predictions
fine-tuned classes
Combined prediction
Combined
Model
Predictions
fine-tuned classes
Meta-information-based predictor
Meta-Information-
based Model
meta-information
Predictions
fine-tuned classes

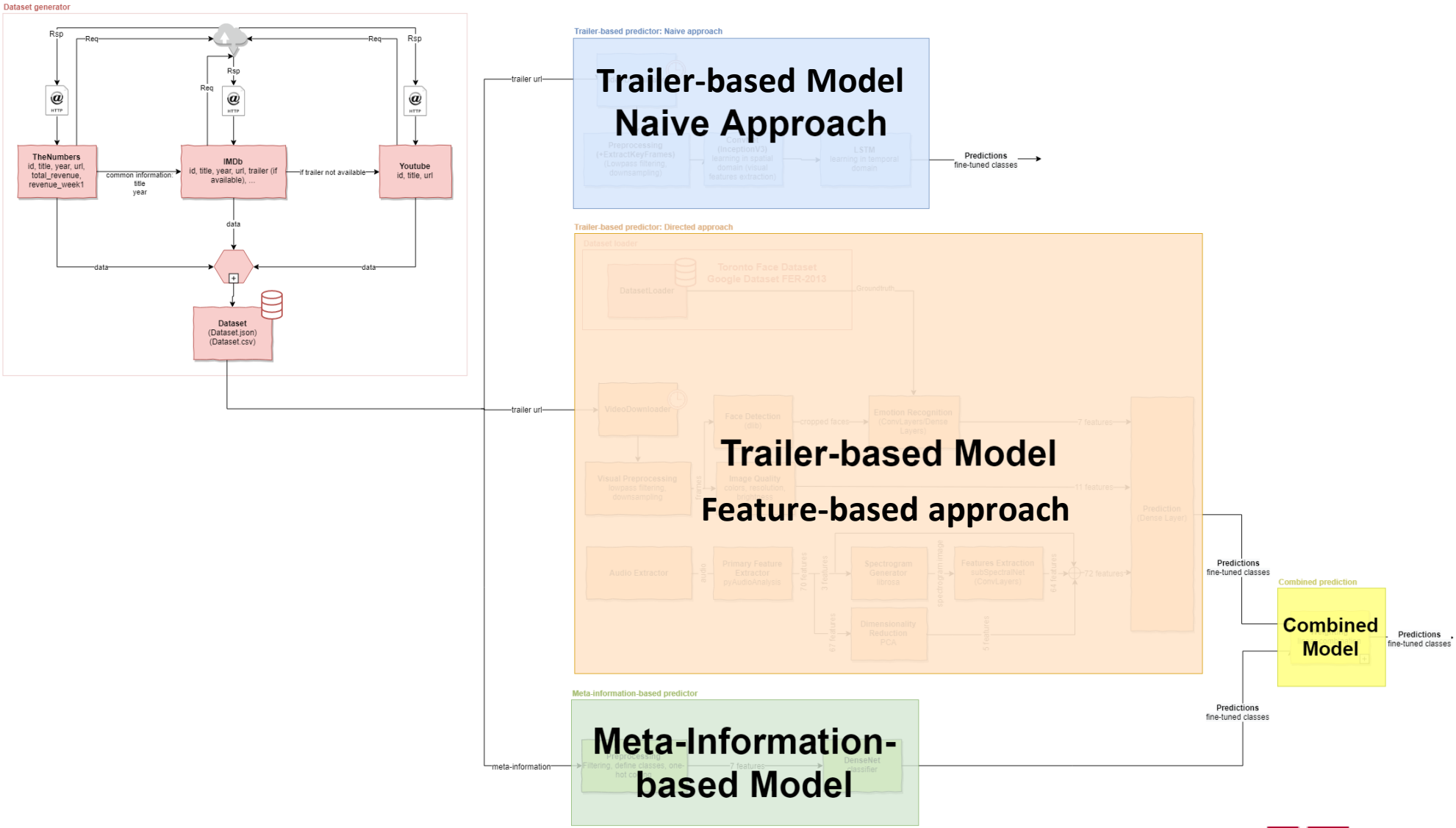
Table of contents

Table of contents

Dataset Generation

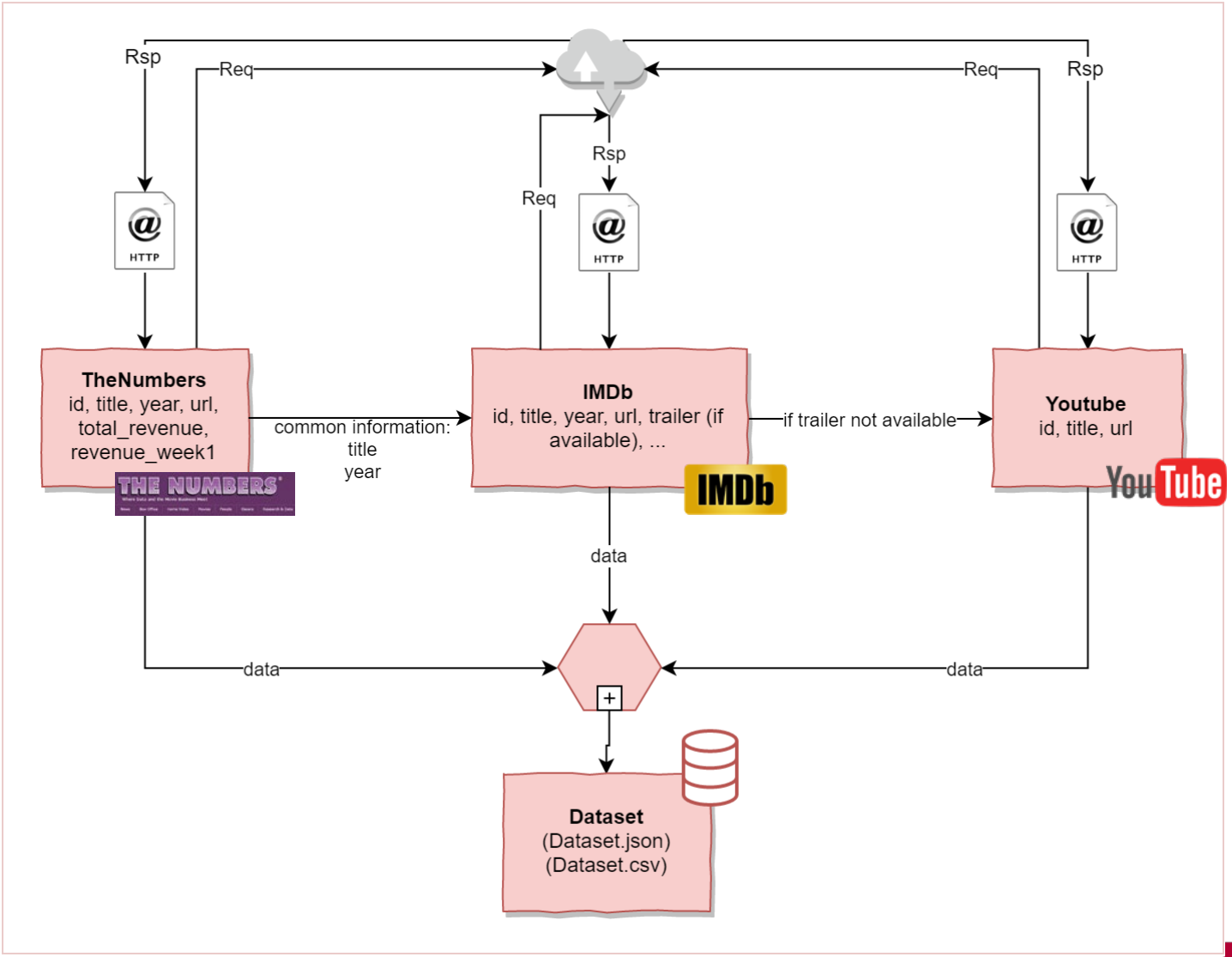


Dataset Generation

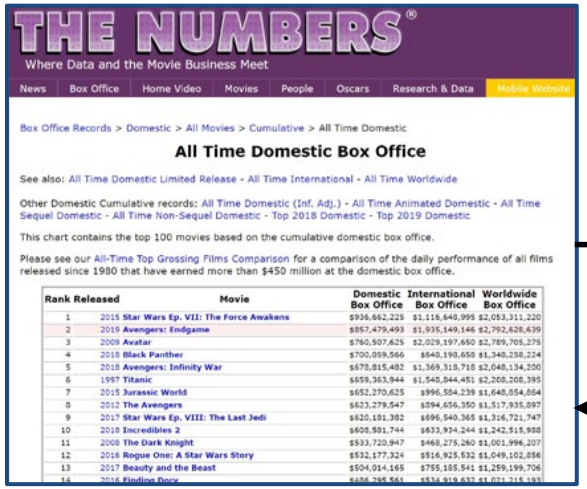


Dataset Generation

Dataset generator



Dataset Generation



HTML source

library BeautifulSoup4

Parsed HTML code

Iterator

HTML request

Primary data: Title, year, opening weekend revenue

Search in IMDB

Retrieved titles

Title and year matching

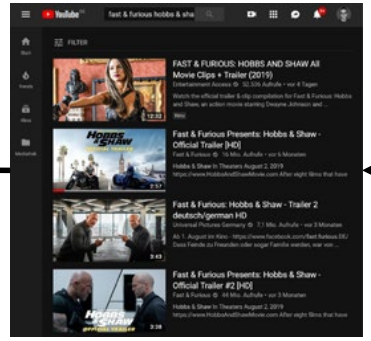
Extracted IMDB id and send HTML request

HTML source

Extract JSON data

IMDB trailer not available ? Send HTML request

HTML source



Extract URL of most relevant video

Dataset (.json)

Relevant data extraction

Dataset (.csv)

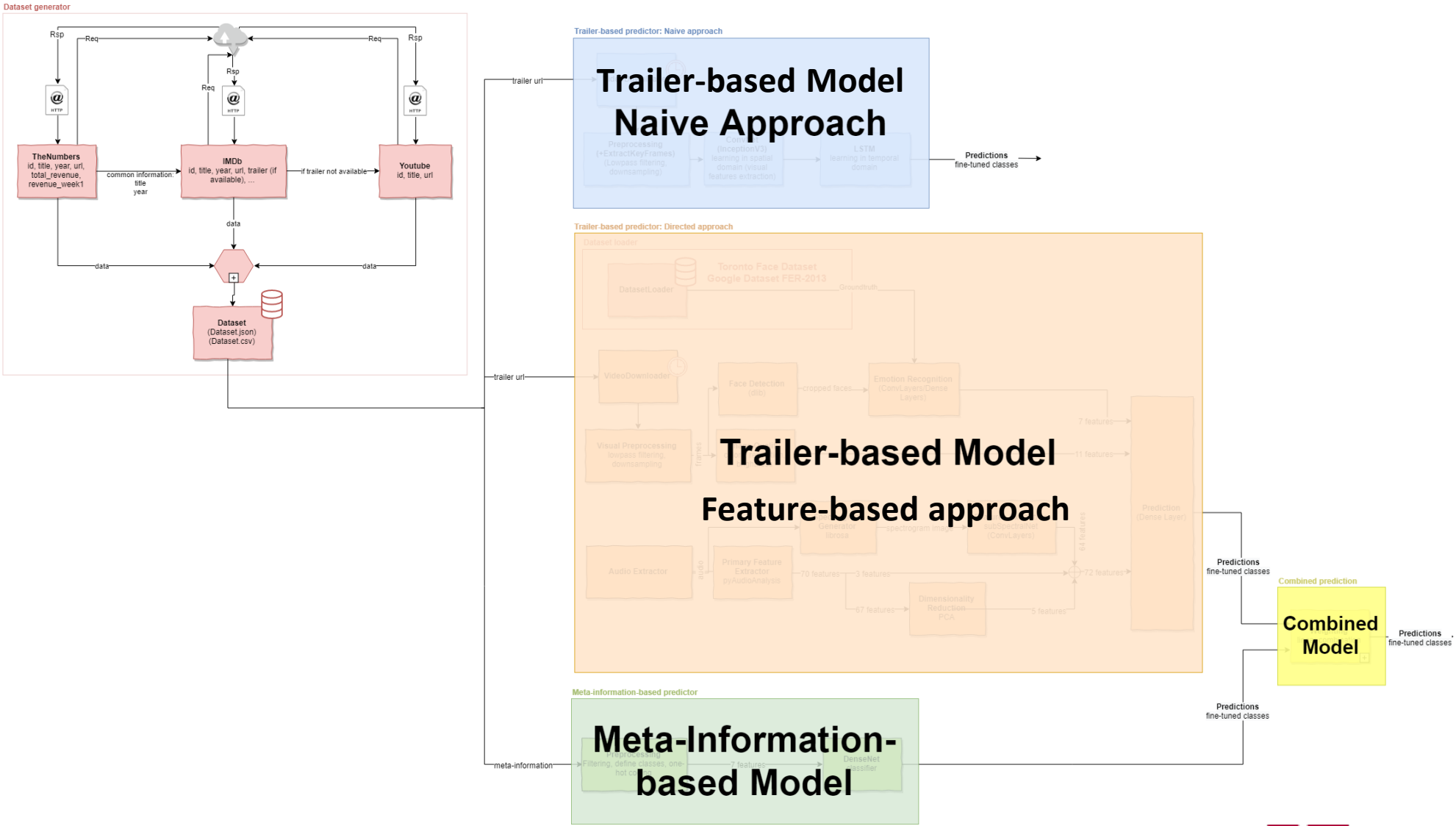
Dataset Generation

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 ±3years.
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).

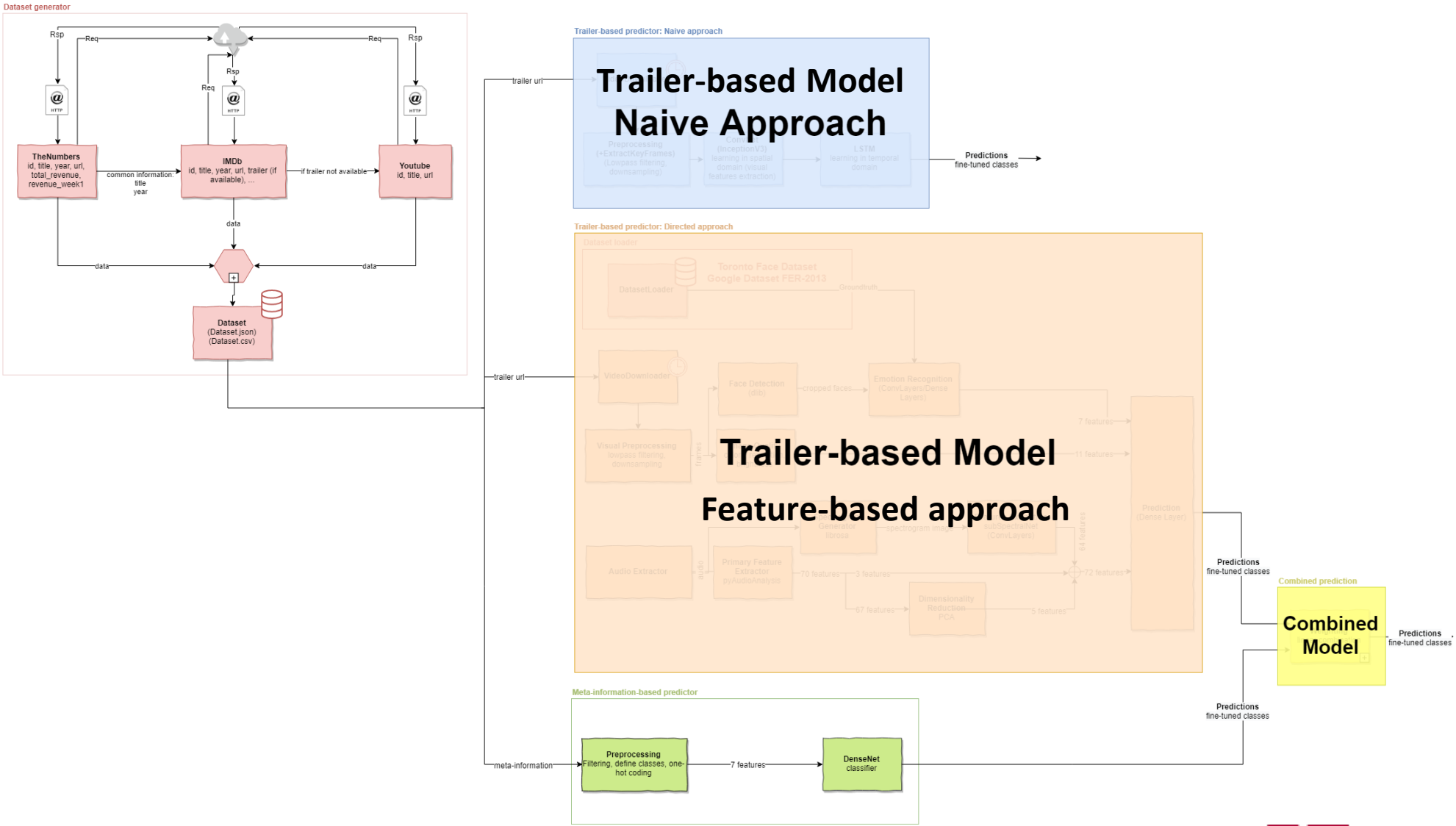
→ In some cases, manual correction.

Table of contents

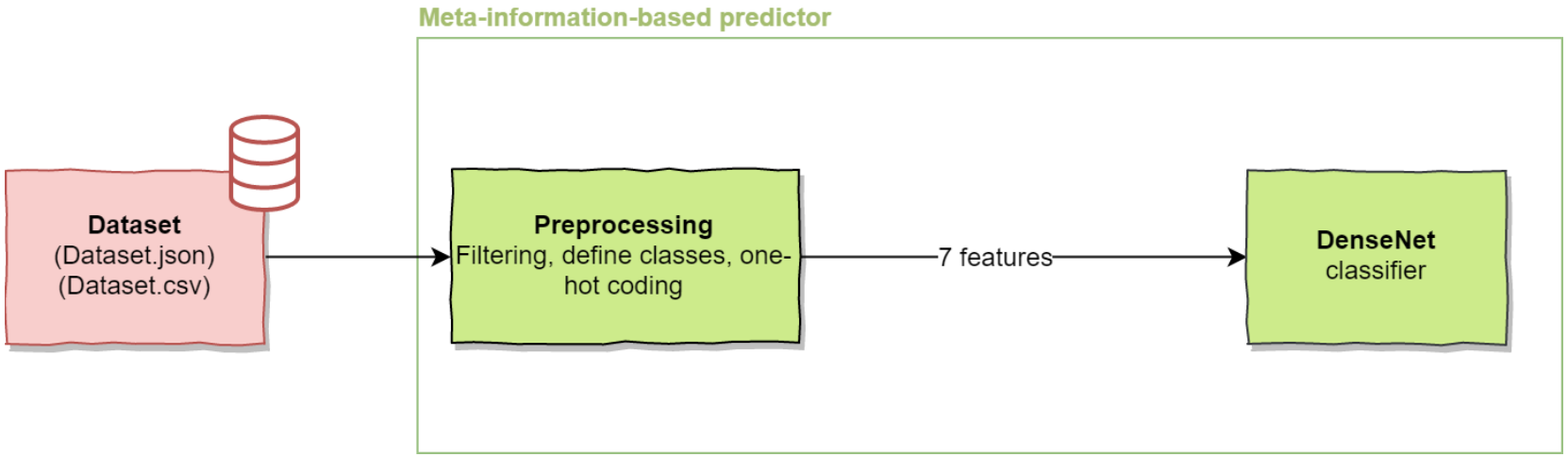
Meta-Information-based Model



Meta-Information-based Model



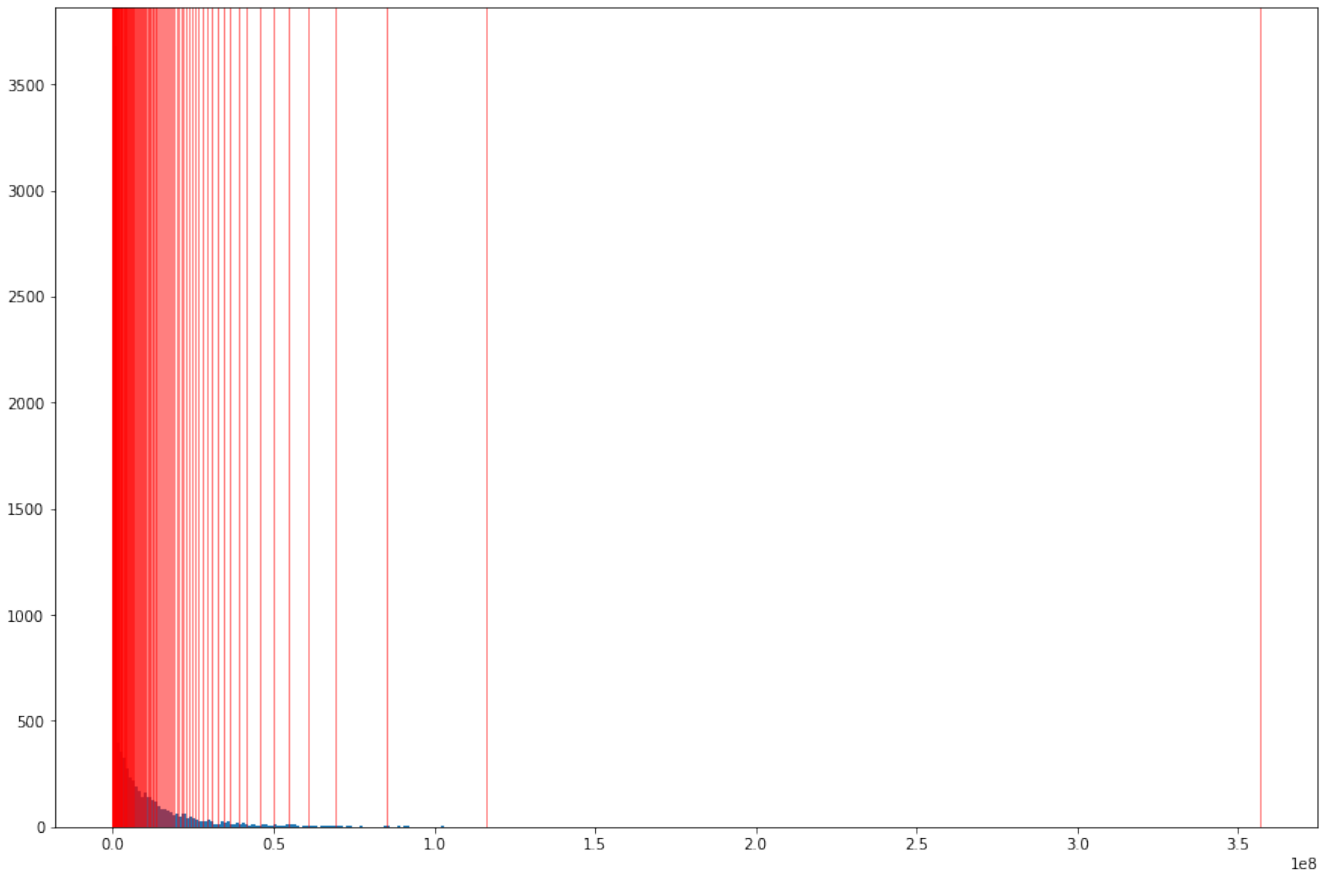
Meta-Information-based Model



- 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 = [‘actors’, ‘creators’, ‘directors’, ‘genres’, contentRating’, ‘year’, ‘duration’]
- Module: Classifier.

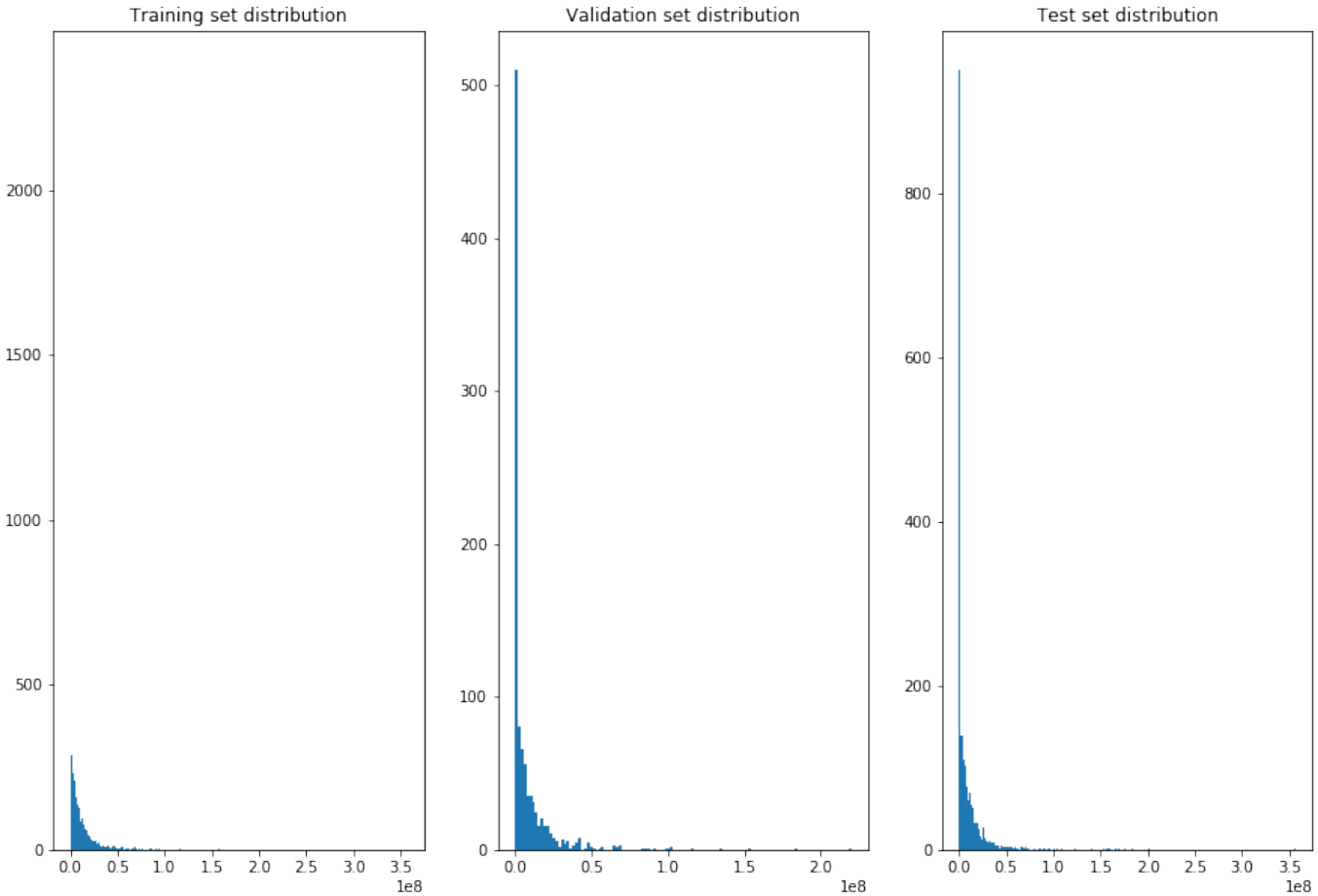
Meta-Information-based Model – Labeling

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



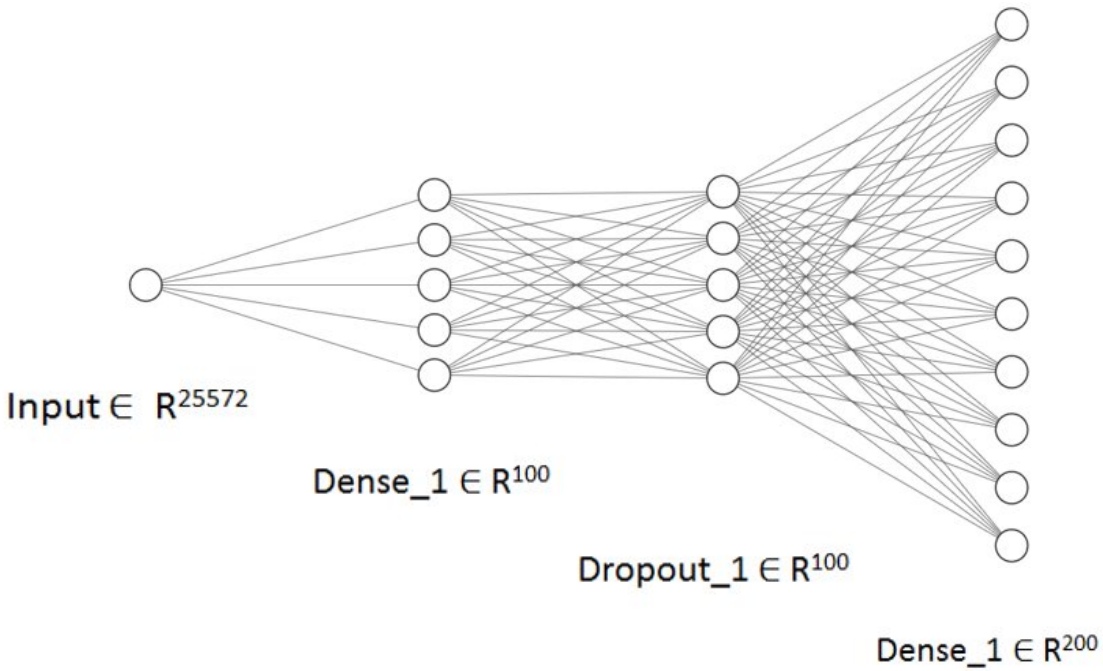
Meta-Information-based Model – Dataset Splitting

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



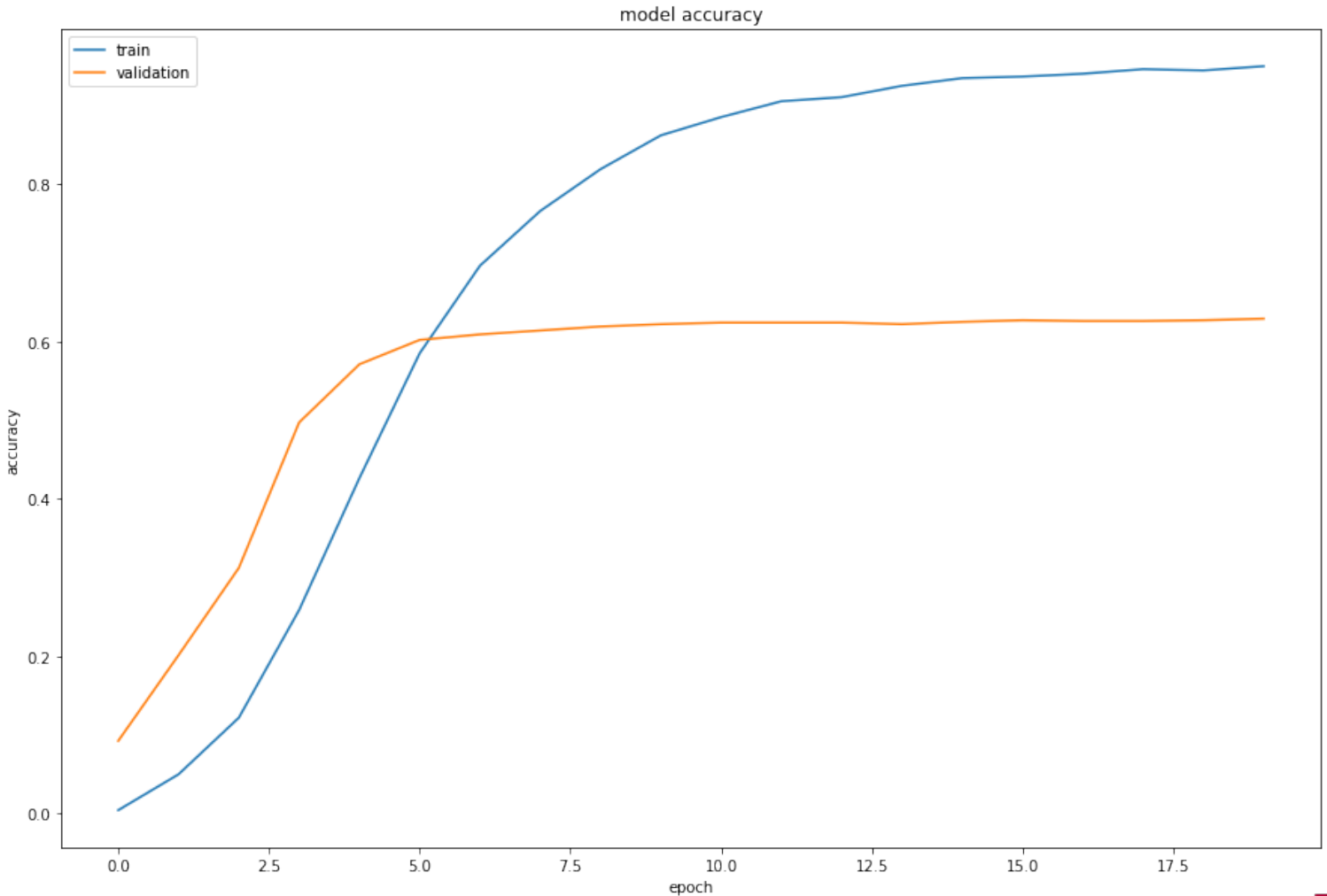
Meta-Information-based Model – Training Model

- 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



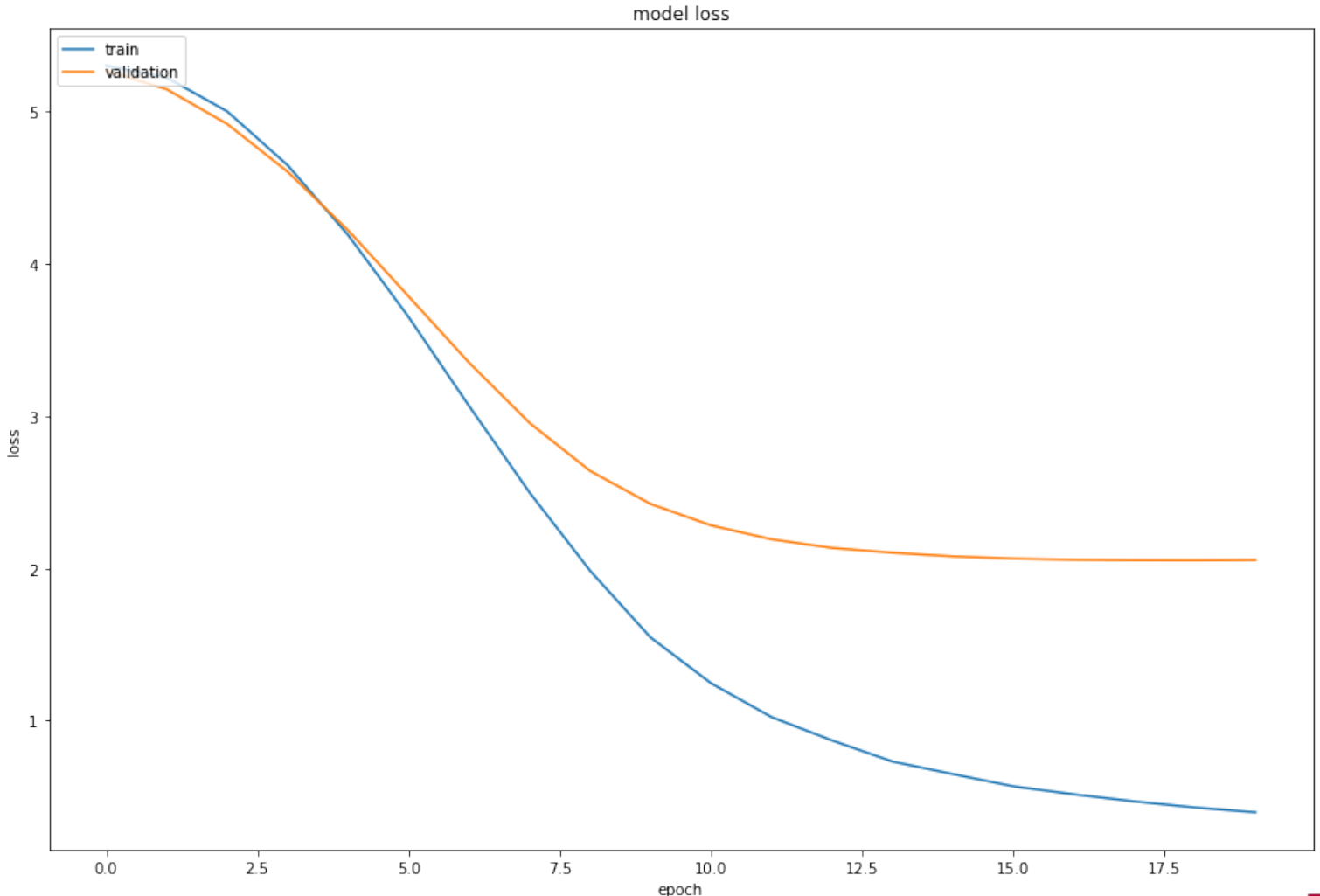
Meta-Information-based Model – Evaluation

- Accuracy



Meta-Information-based Model – Evaluation

- Loss: Categorical cross entropy loss



Meta-Information-based Model – Evaluation

- Testing

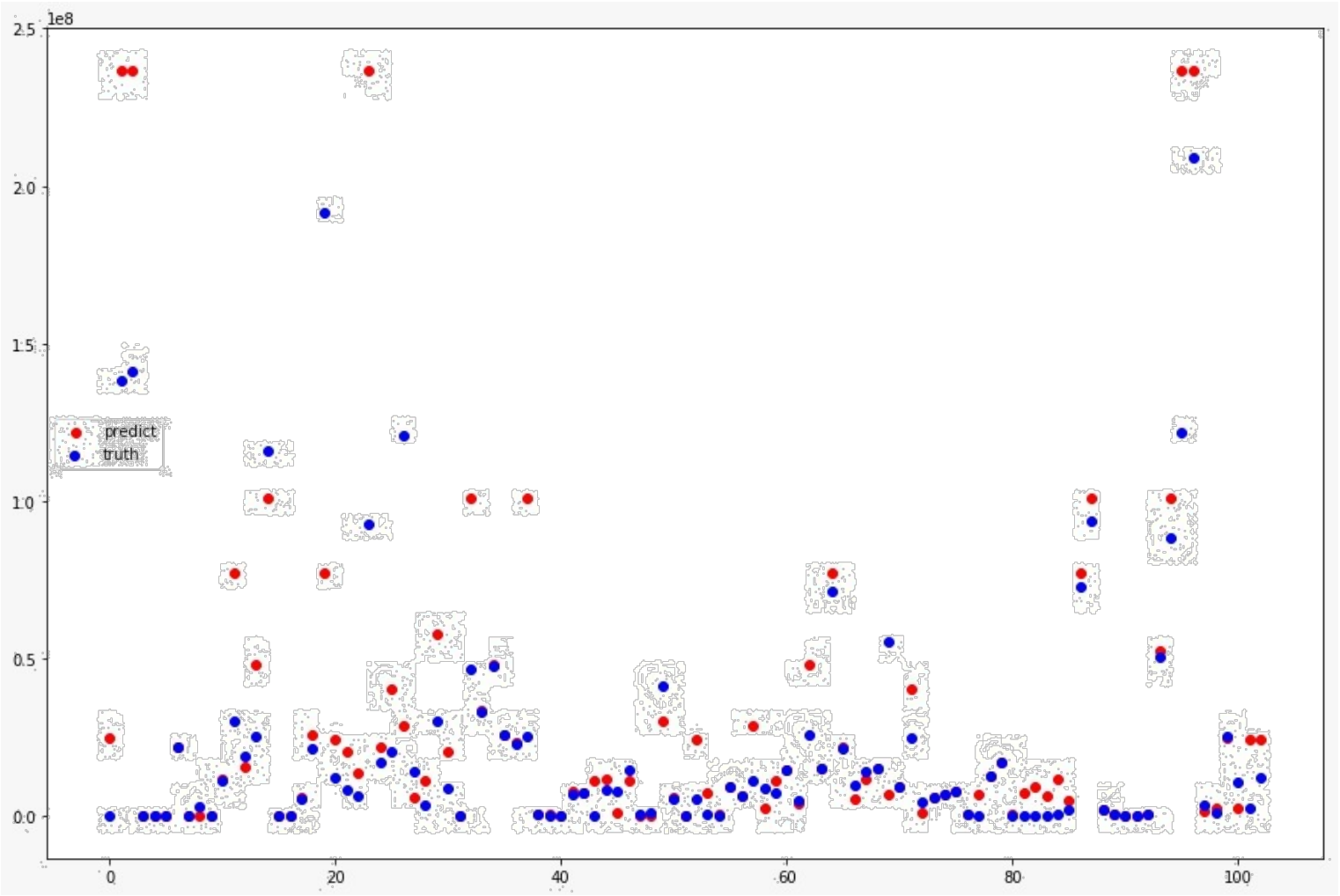
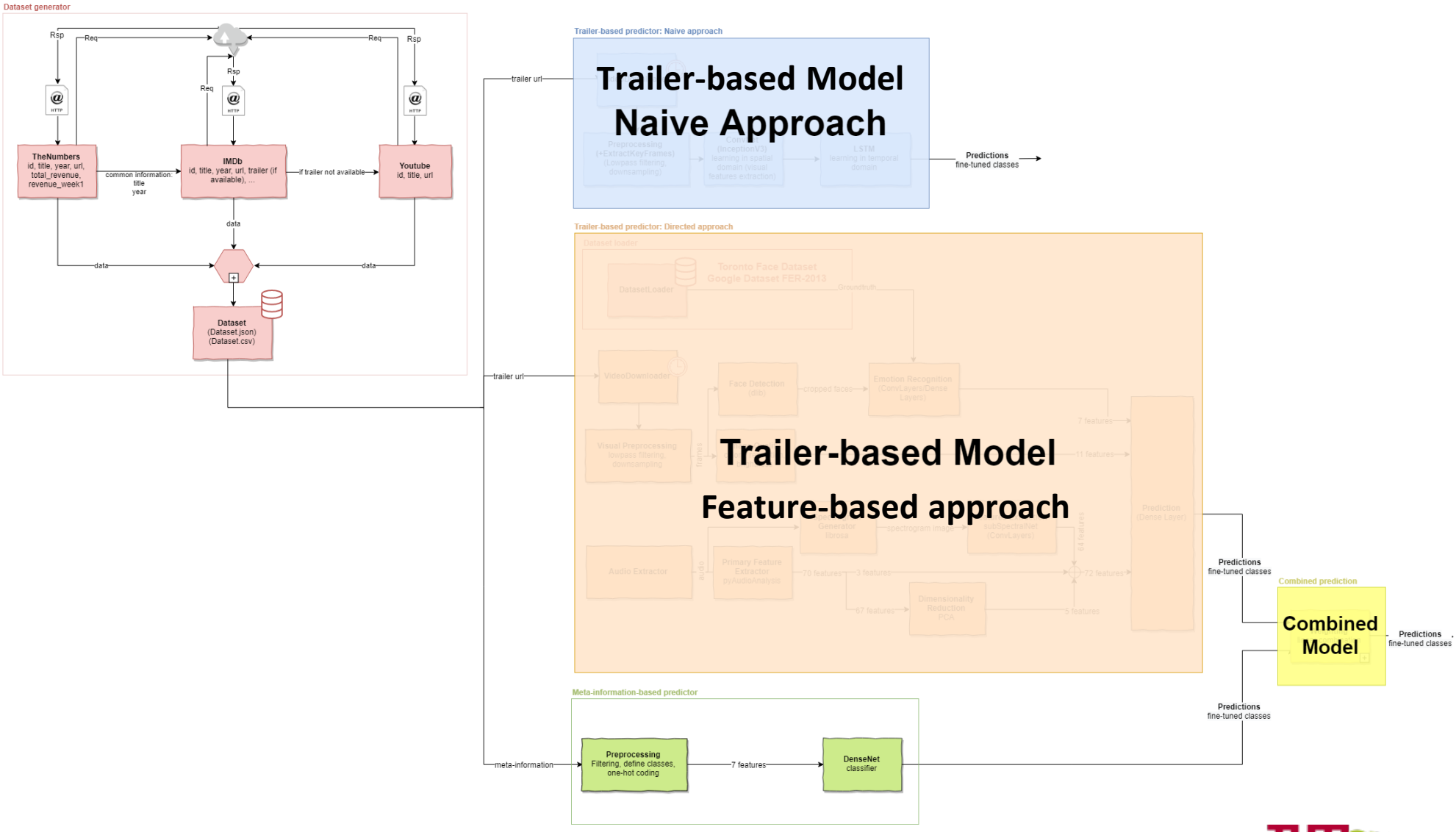
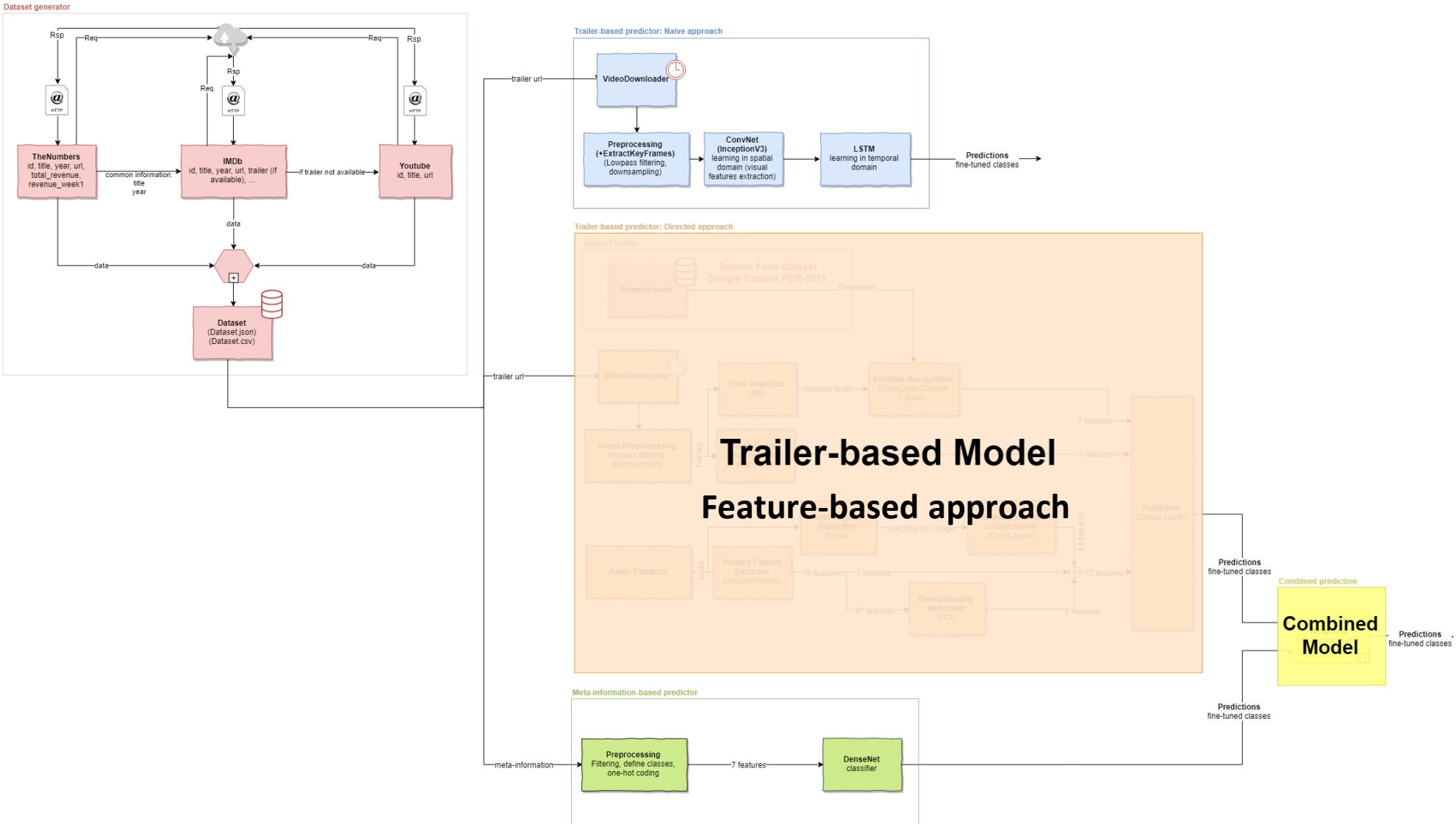


Table of contents

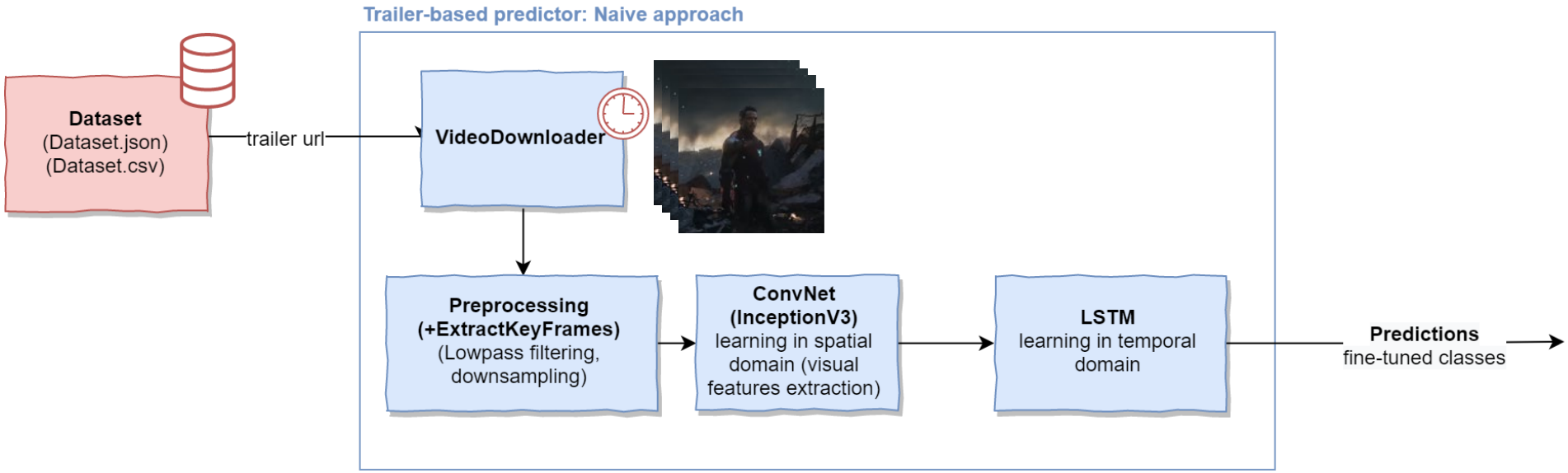
Trailer-based-Model



Trailer-based-Model – Naive Approach

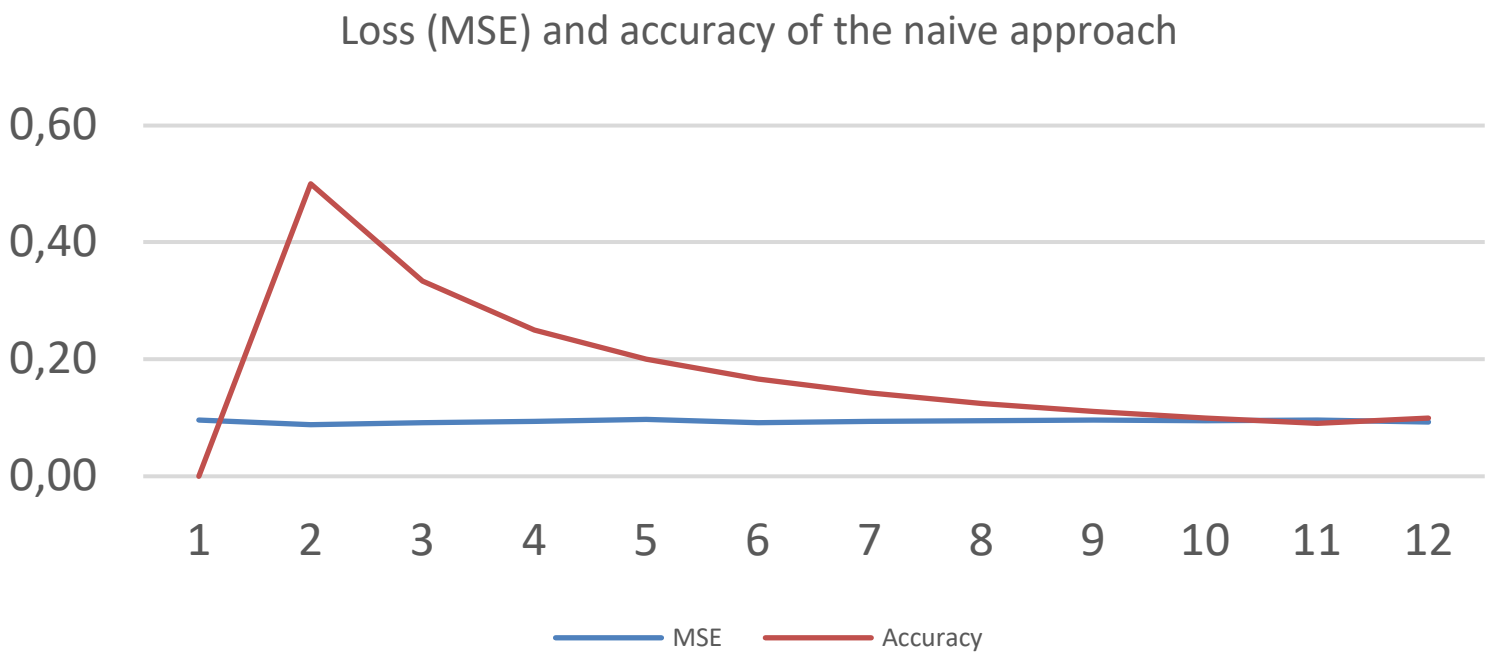


Trailer-based-Model – Naive Approach



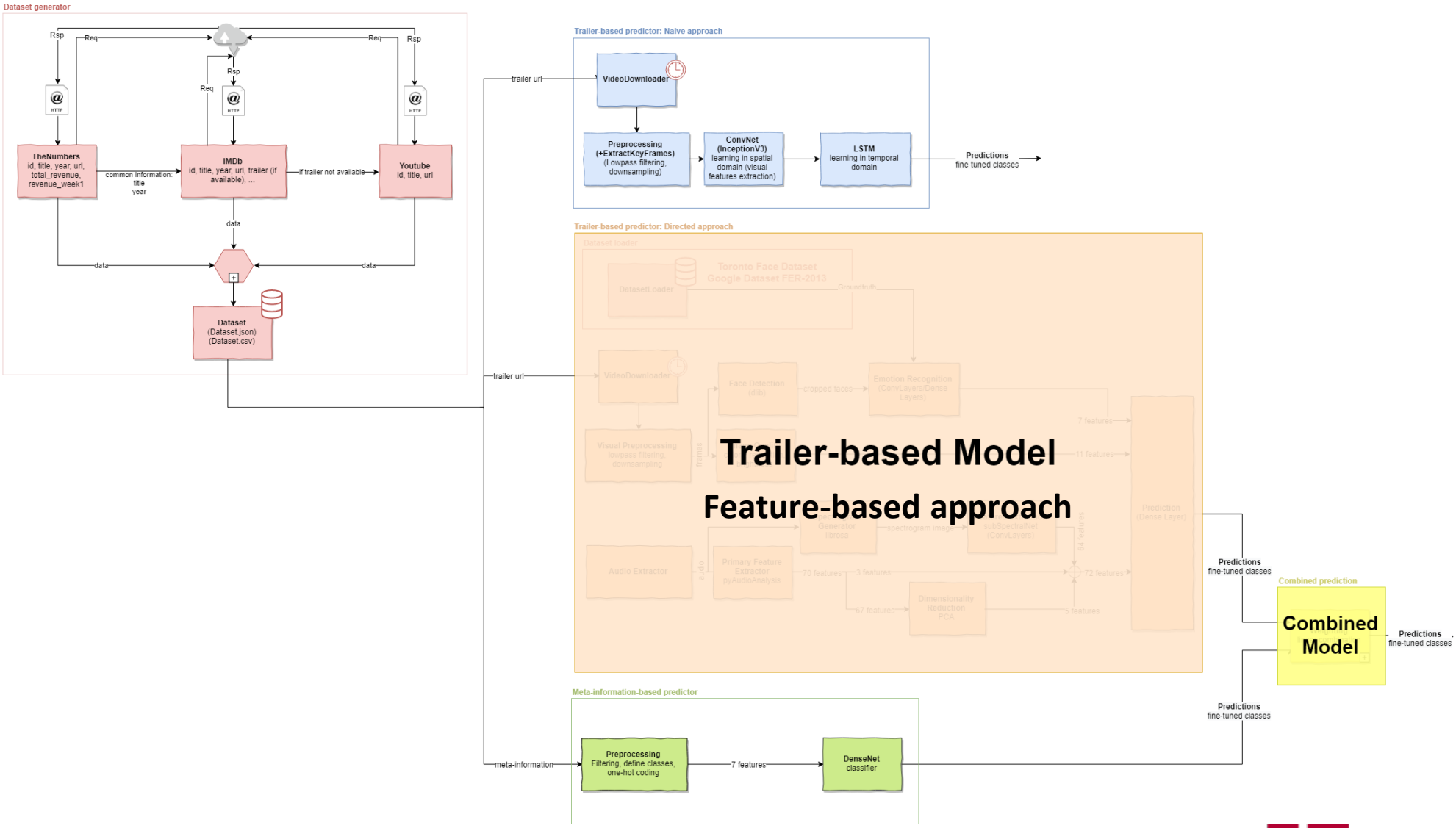
- Architecture: Input Layer → TimeDistributedDense Layer connected to InceptionV3 → LSTM → Dense Layer
- Unguided Learning
- Total number of parameters: 1,115,914

Trailer-based-Model – Naive Approach

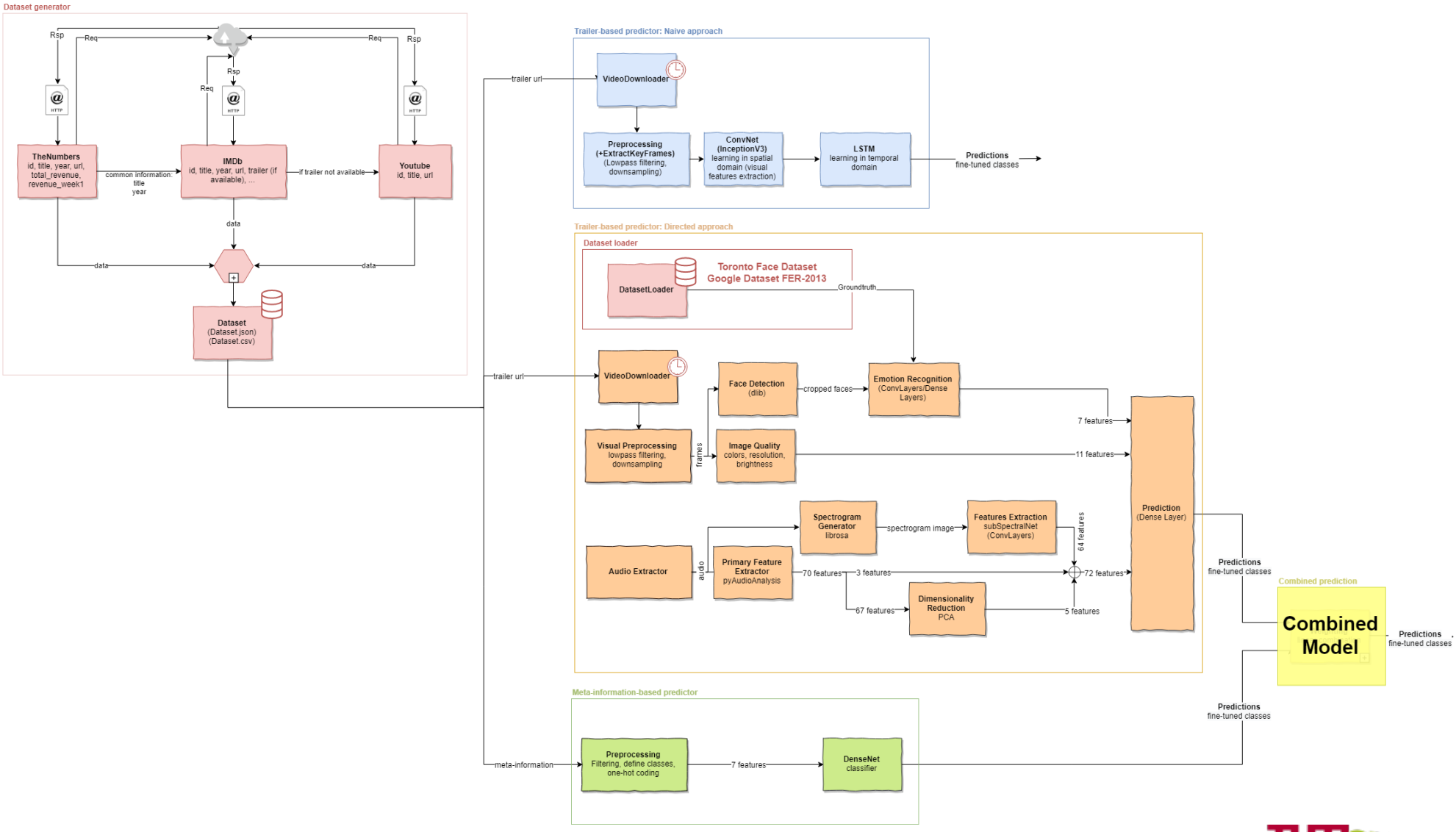


Problems	Solution
Poor performance.	Guided learning based on feature extraction preprocessing.
Unable to learn relevant features	

Trailer-based-Model

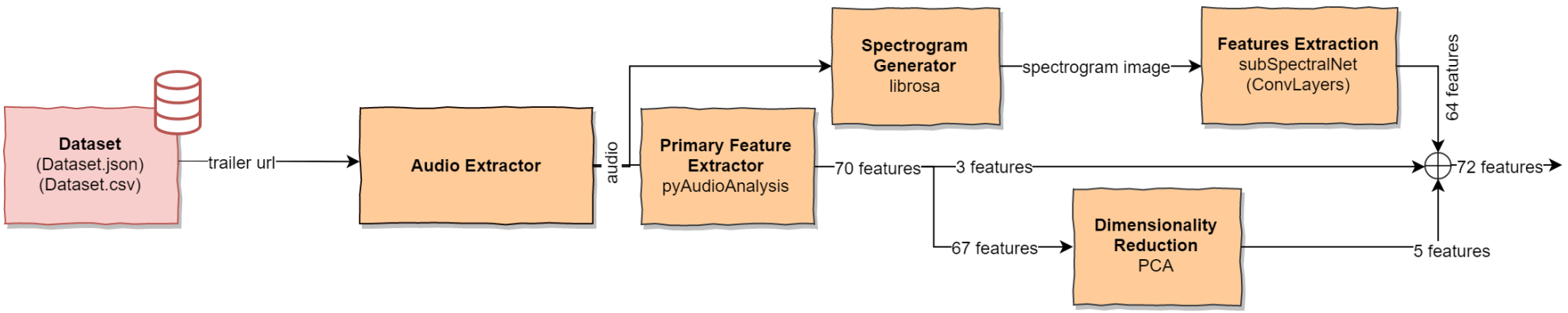


Trailer-based-Model – Feature-based approach



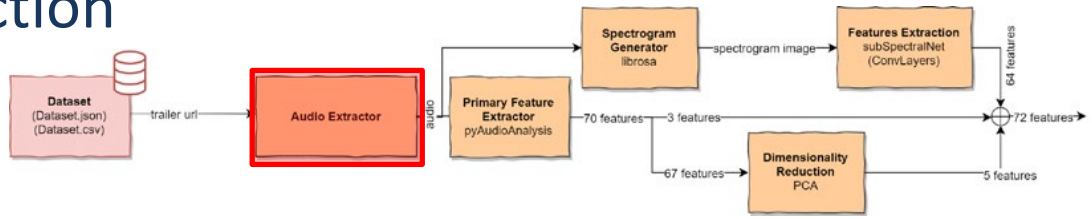
Trailer-based-Model – Feature-based approach

- Audio Analyzer

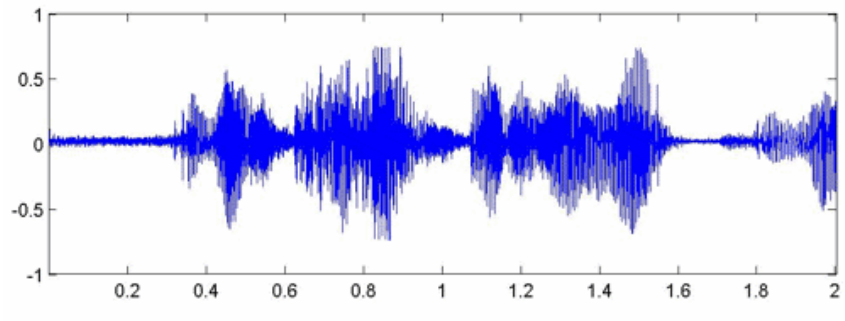


Trailer-based-Model – Feature-based approach

- Audio Analyzer - Audio Extraction



FFmpeg

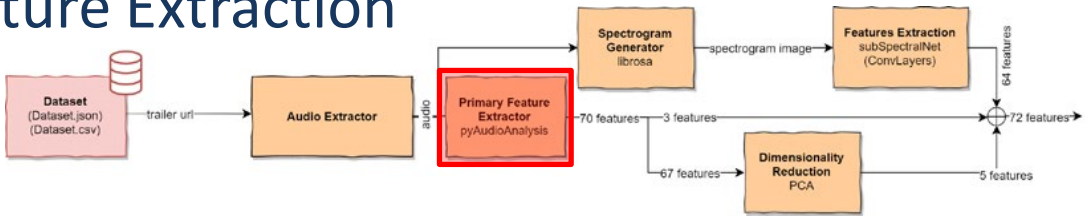


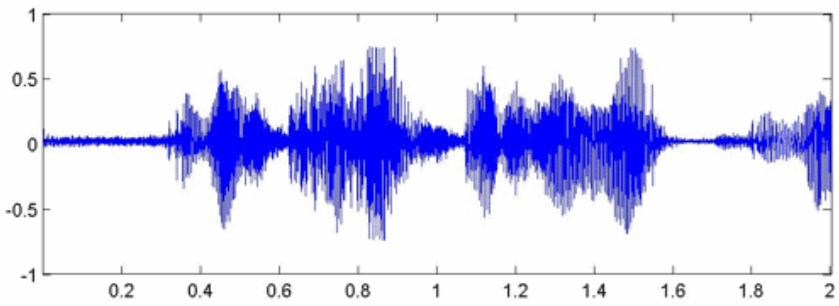
Trailer

Raw audio signal

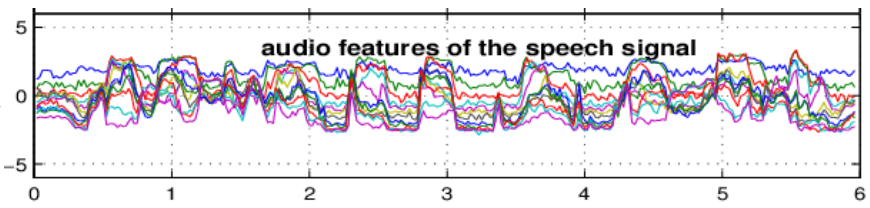
Trailer-based-Model – Feature-based approach

- Audio Analyzer – Primary Feature Extraction





pyAudio Analysis

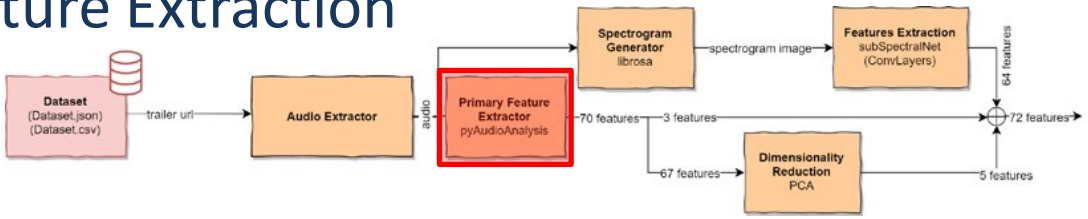


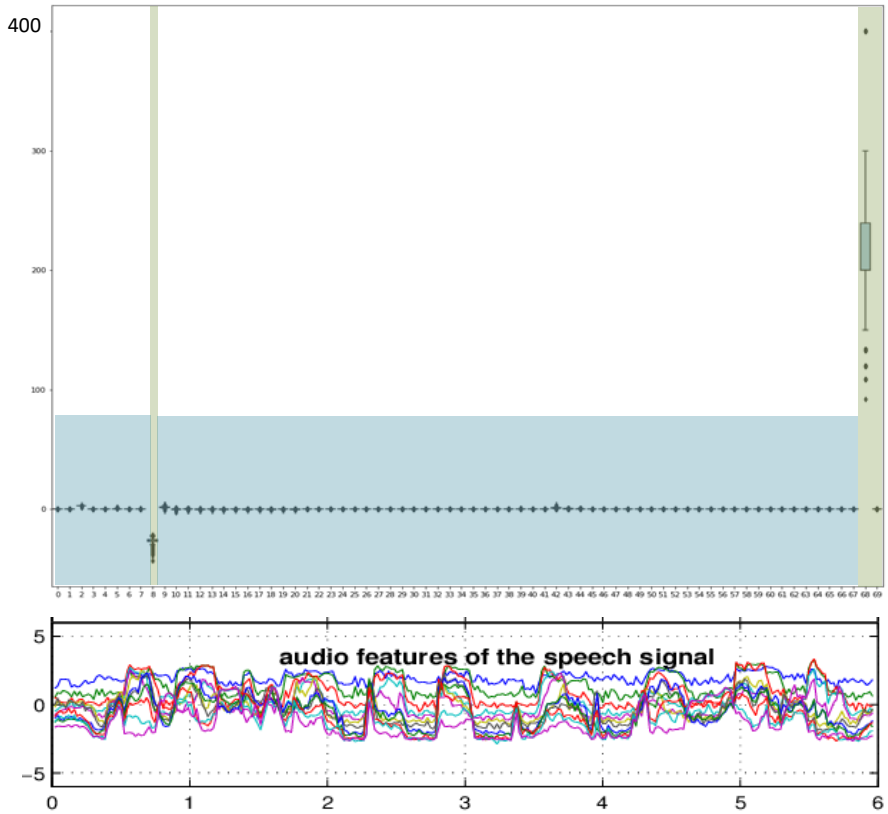
Raw audio signal

70 features

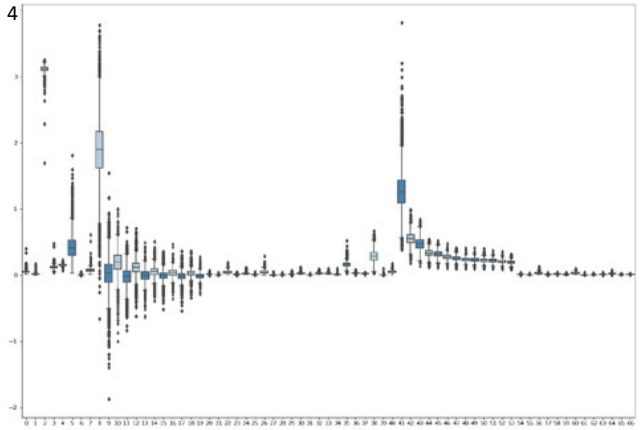
Trailer-based-Model – Feature-based approach

- Audio Analyzer – Primary Feature Extraction





70 features



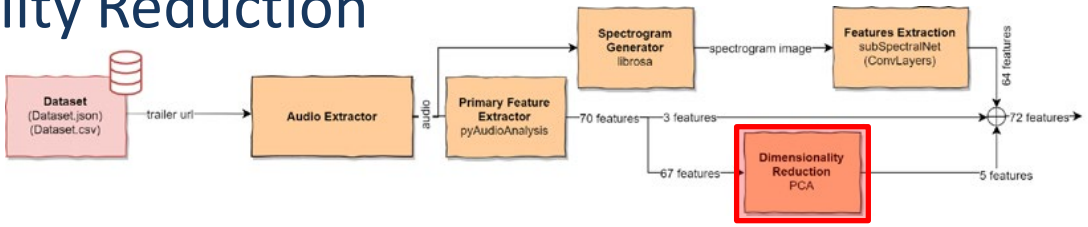
67 dense features

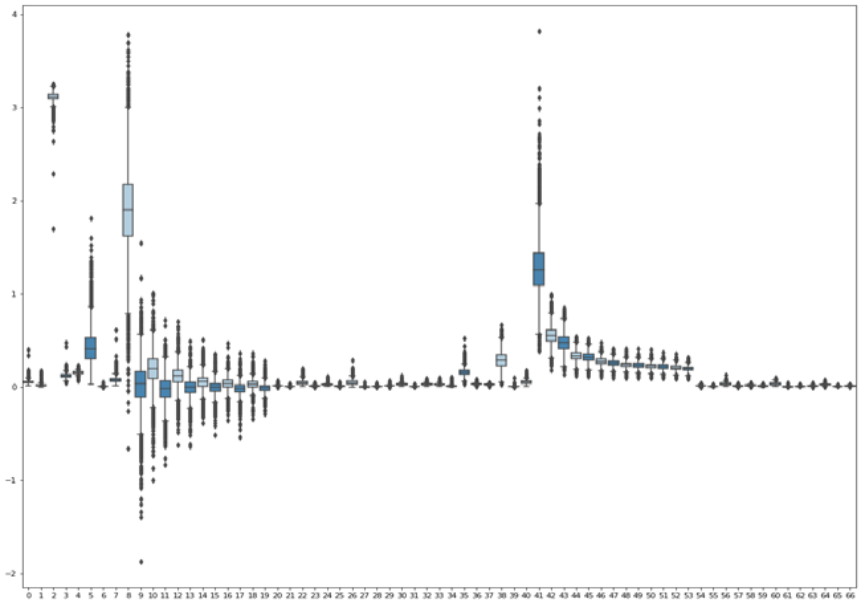
3 significant features

(beat, its confidence, mfcc mean)

Trailer-based-Model – Feature-based approach

- Audio Analyzer – Dimensionality Reduction

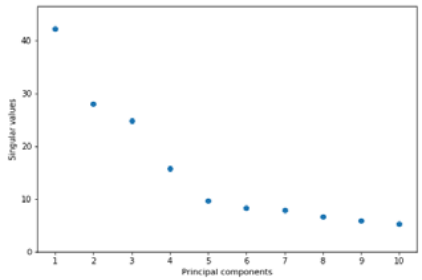




scikit learn

PCA

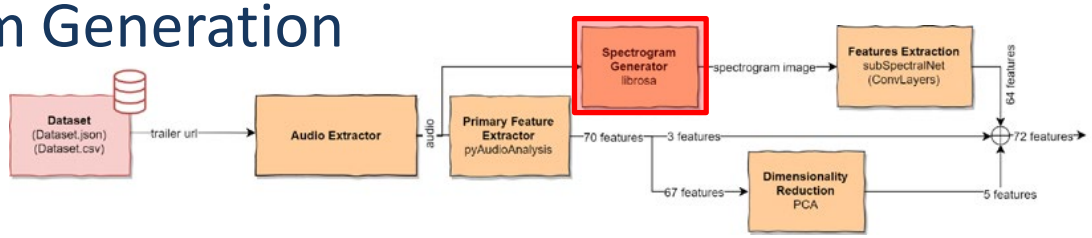
5 features



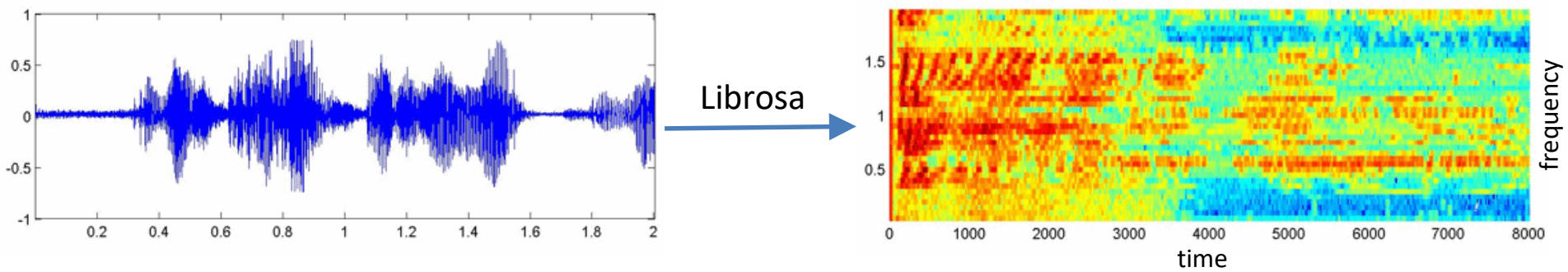
TUMOVIES DEEP LEARNING

Trailer-based-Model – Feature-based approach

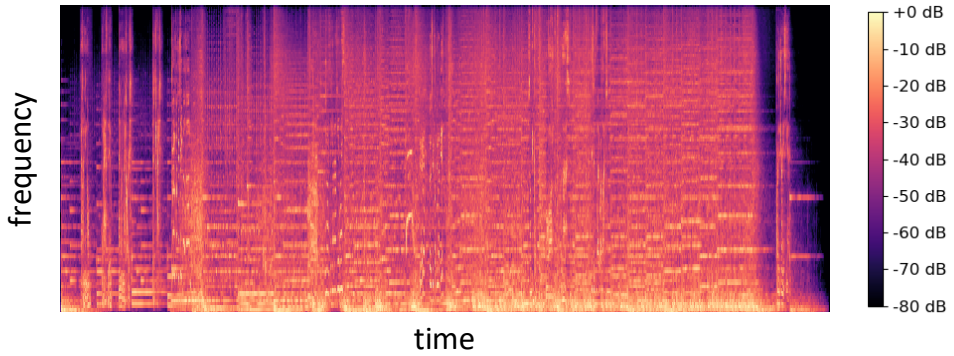
- Audio Analyzer – Spectrogram Generation



Spectogram: Representation of the power spectrum of a sound.

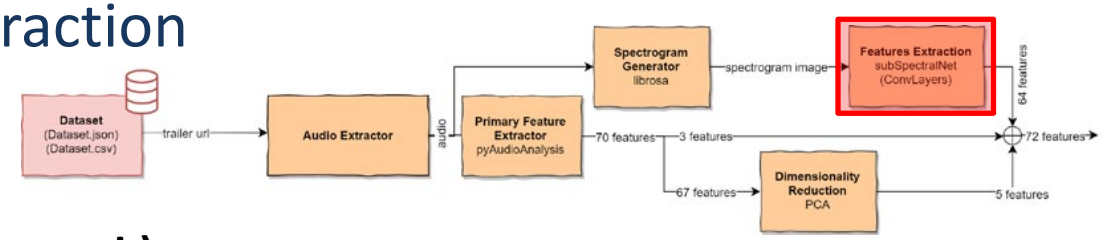


e.g. MEL-Spectogram of Avengers Endgame

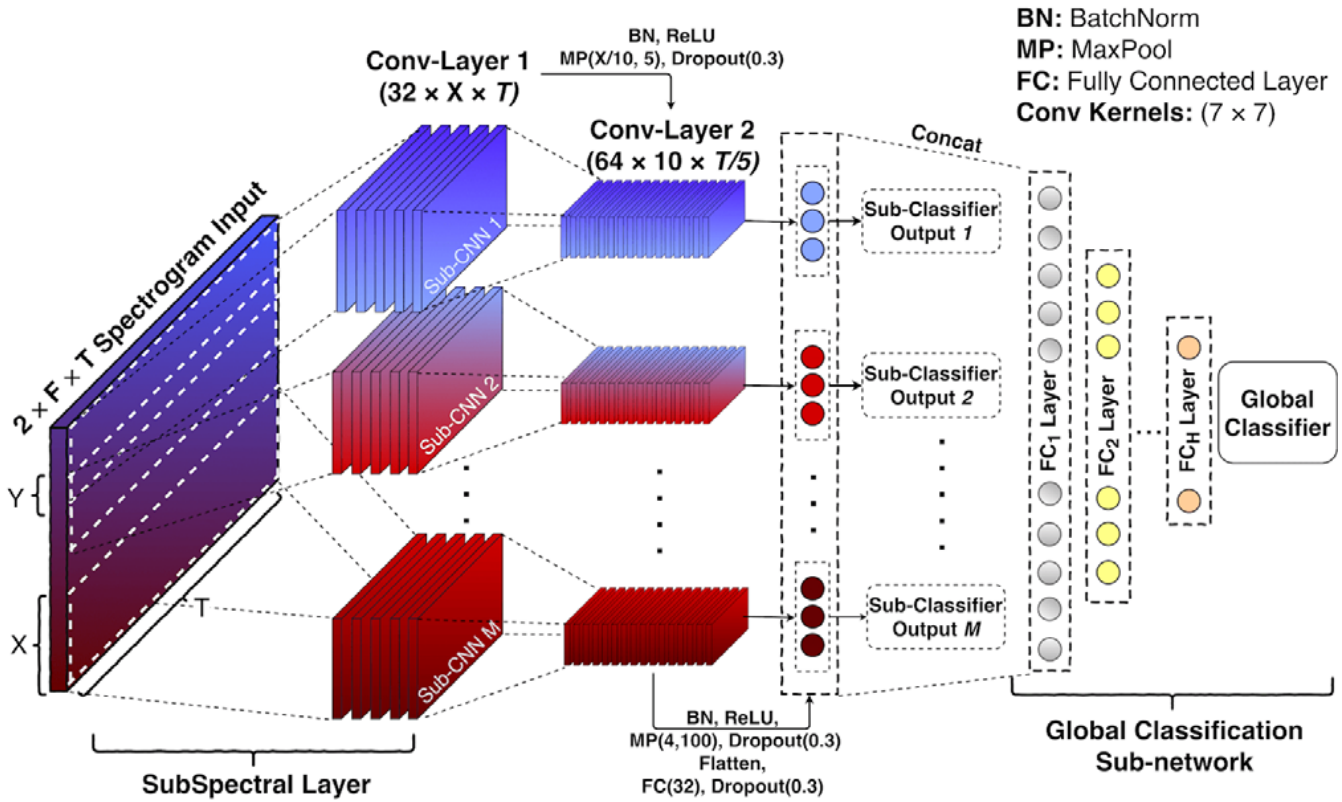


Trailer-based-Model – Feature-based approach

- Audio Analyzer – Feature extraction



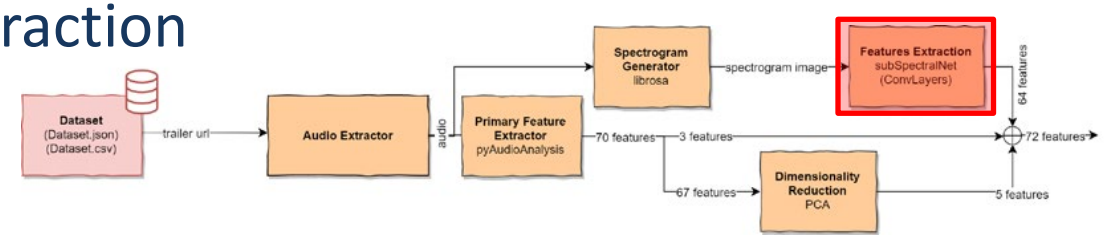
Architecture: SubSpectral Net* (R Phaye et. al.)



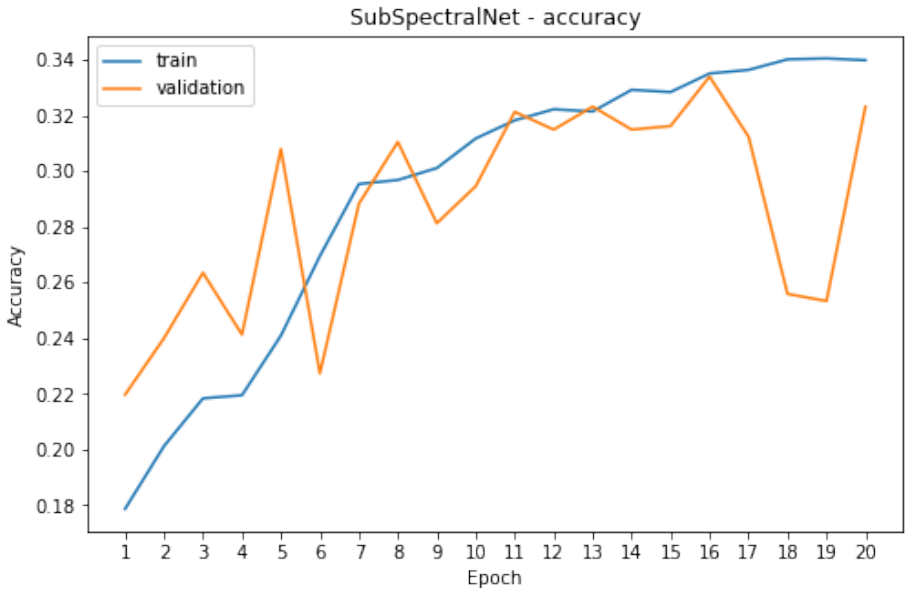
*SUBSPECTRALNET – USING SUB-SPECTROGRAM BASED CONVOLUTIONAL NEURAL NETWORKS FOR ACOUSTIC SCENE CLASSIFICATION (2019)

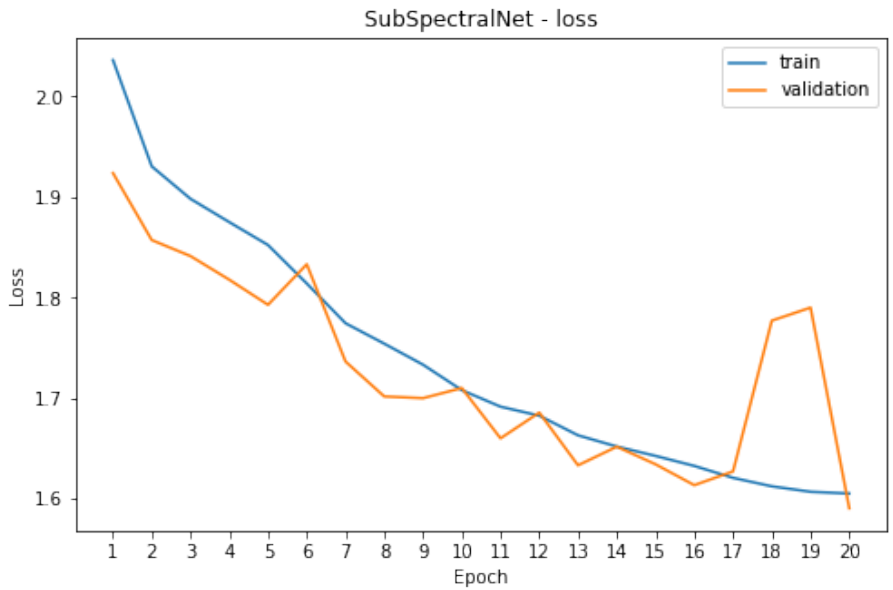
Trailer-based-Model – Feature-based approach

- Audio Analyzer – Feature extraction



Training and validation





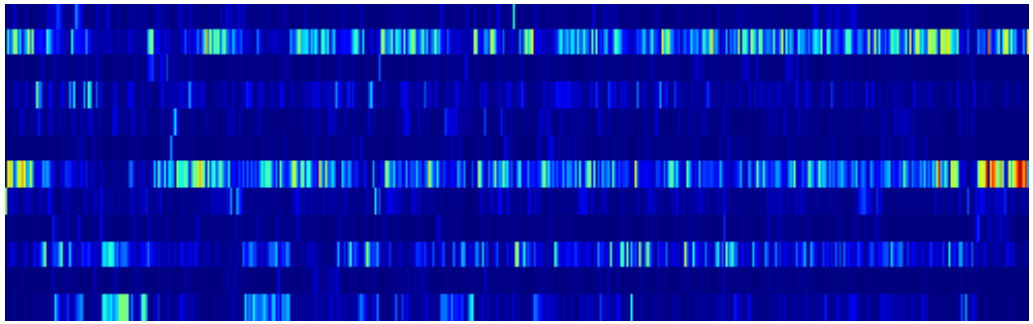
- Training the network on 8000 spectrograms of our trailers with 6 target classes.
- Best accuracy: 30%.

Trailer-based-Model – Feature-based approach

- Audio Analyzer – Attempt

Use chromograms in addition to the spectrograms:

- Try to retrain ResNet50, VCG19 and VGG16.

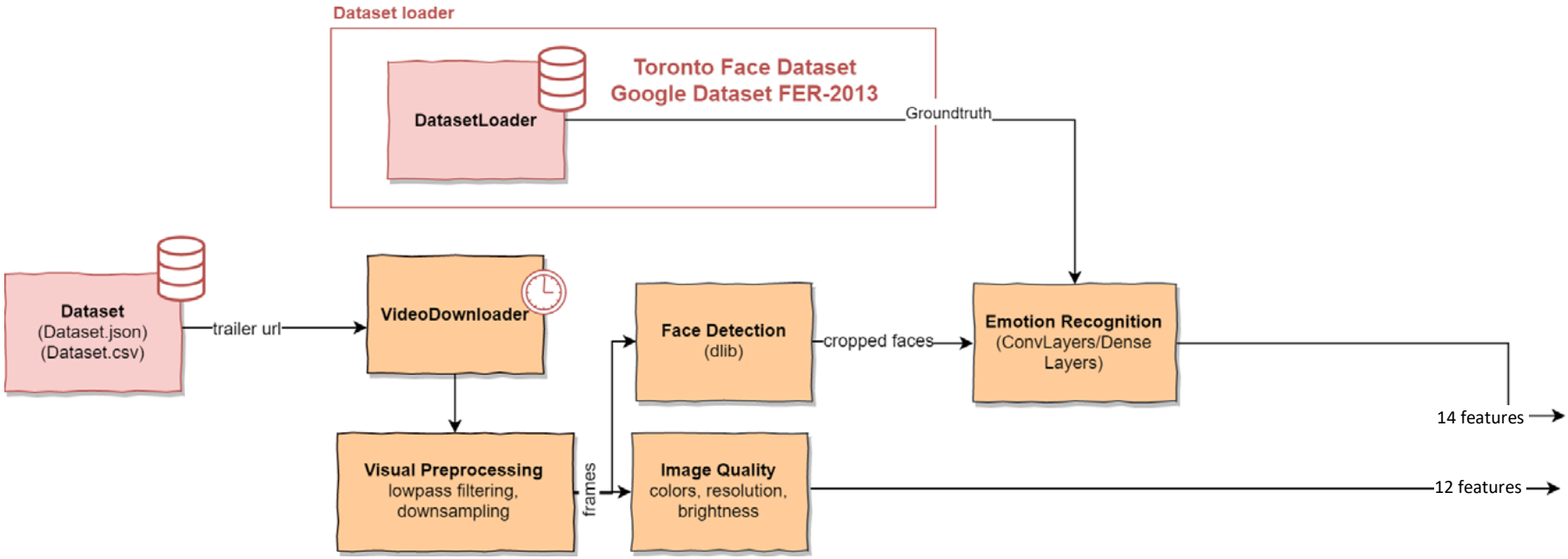


Chromogram of Avengers Endgame

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

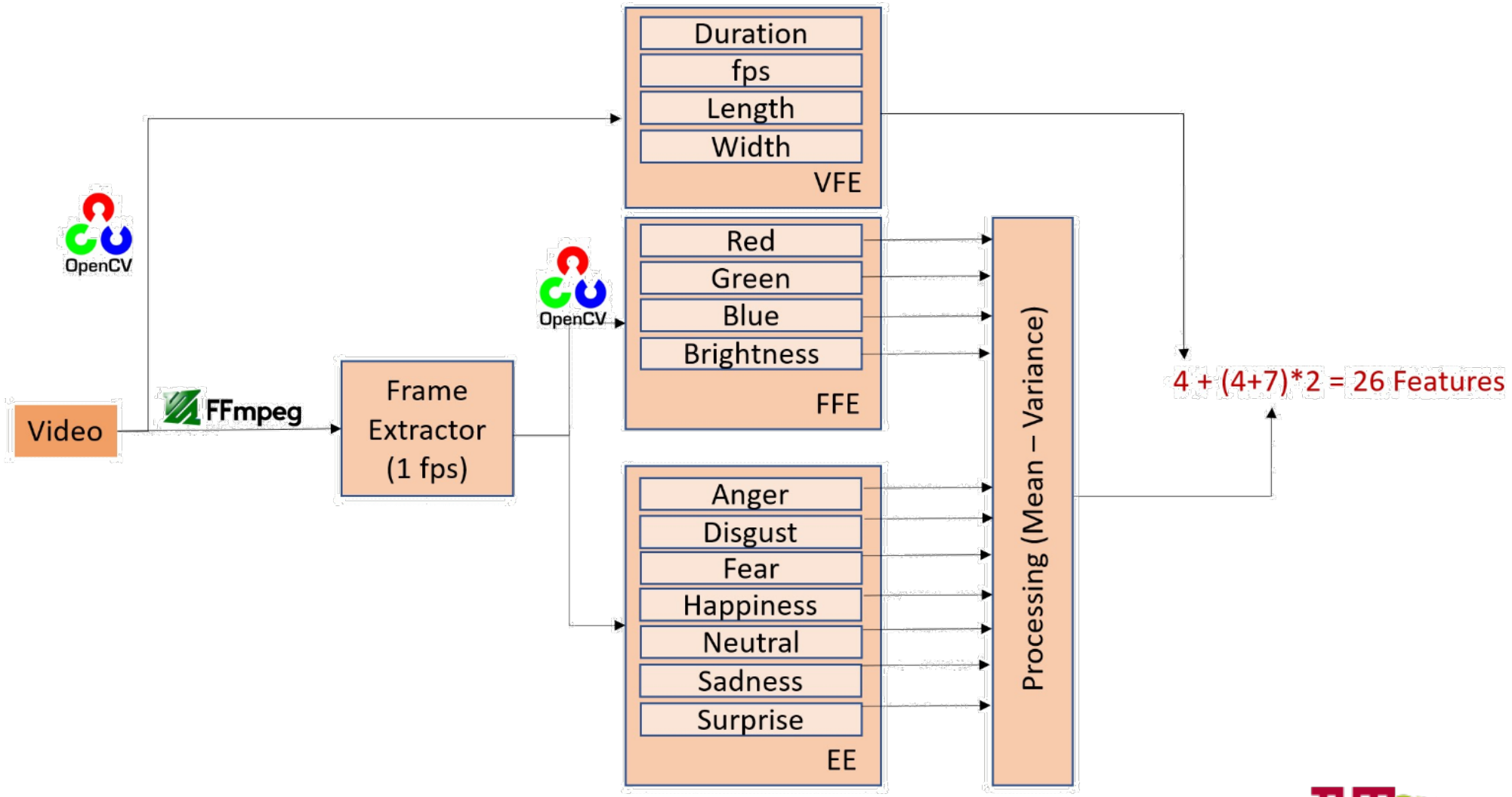
Trailer-based-Model – Feature-based approach

- Visual Analyzer



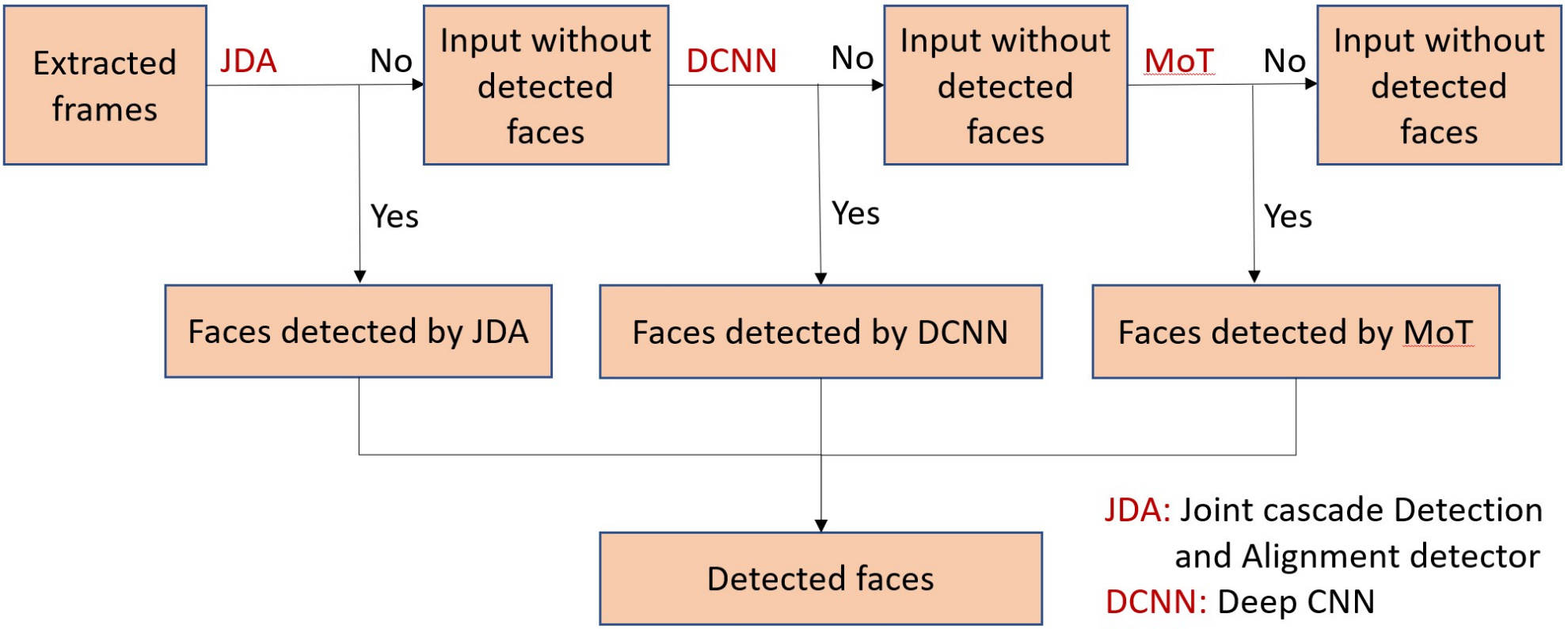
Trailer-based-Model – Feature-based approach

- Visual Analyzer – Feature extraction



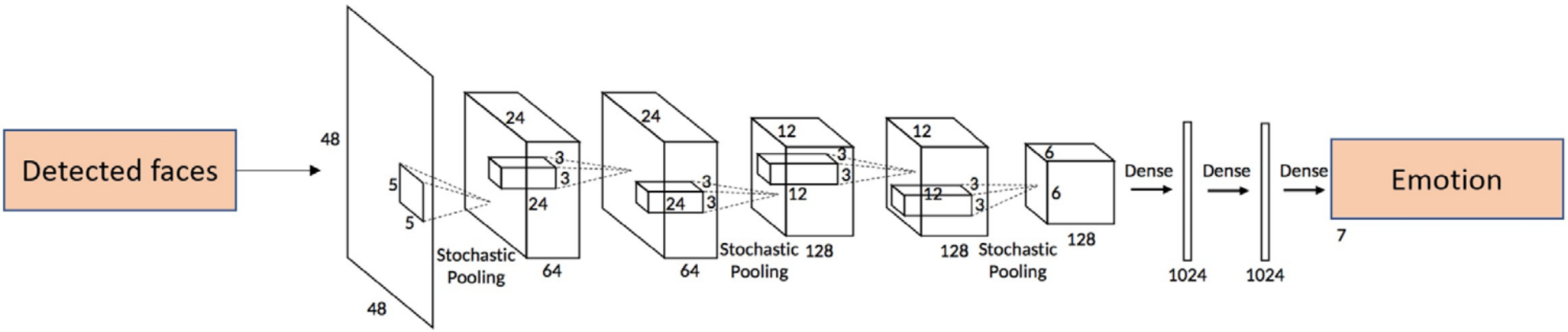
Trailer-based-Model – Feature-based approach

- Visual Analyzer – Face Detector



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.

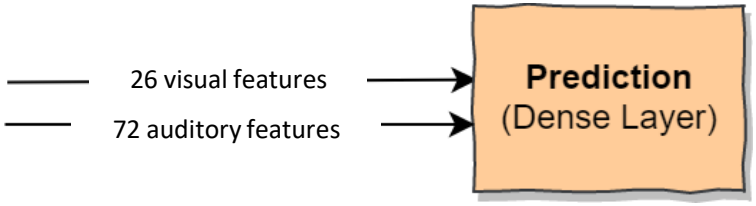
Trailer-based-Model – Feature-based approach

- Visual Analyzer – Performance

Problem	Solution
Computational performance: 4min/video→ 4*7900 = 31600min = 22 days Emotion Extraction is the most computationally expensive	• Use only one extractor (down to 11 days) • Extract faces at a 0.1 fps rate (down to 5 days)

Trailer-based-Model – Feature-based approach

- Model - Selection



Multiclass Decision Forest

Multiclass Decision Jungle

Multiclass Logistic Regression ★

Multiclass Neural Network

Two-Class Averaged Perceptron

Two-Class Bayes Point Machine

Two-Class Boosted Decision Tree ★

Two-Class Decision Forest

Two-Class Decision Jungle

Two-Class Locally-Deep Support ...

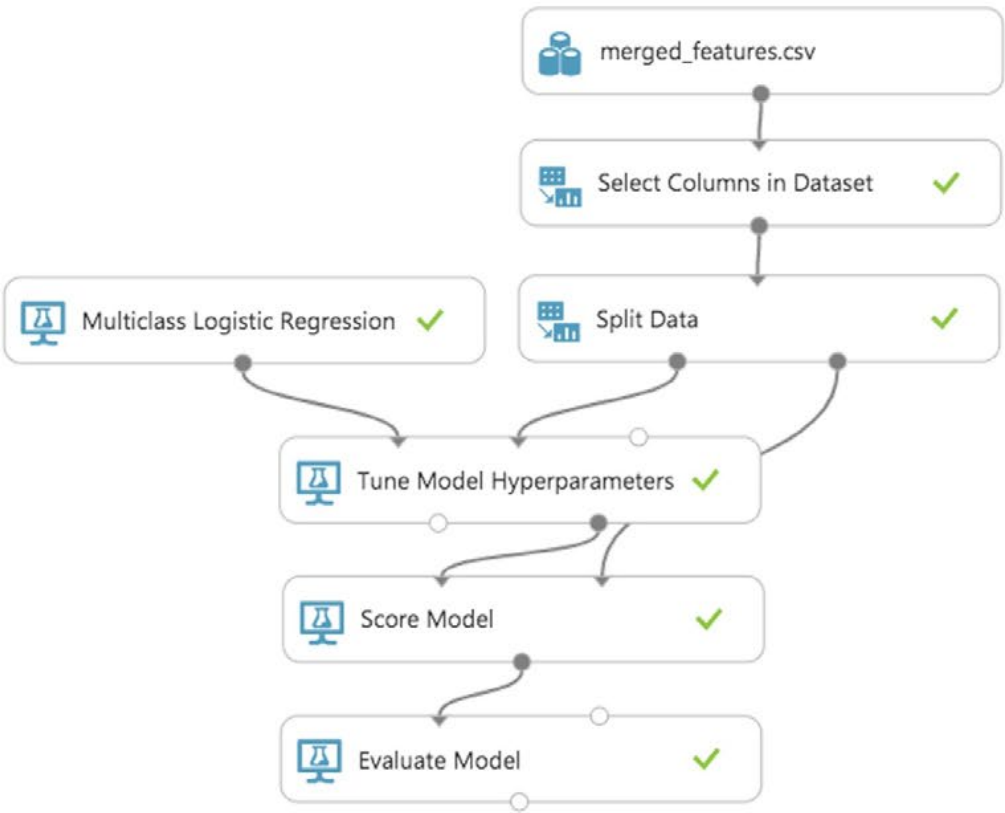
Two-Class Logistic Regression

Two-Class Neural Network

Two-Class Support Vector Machine

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

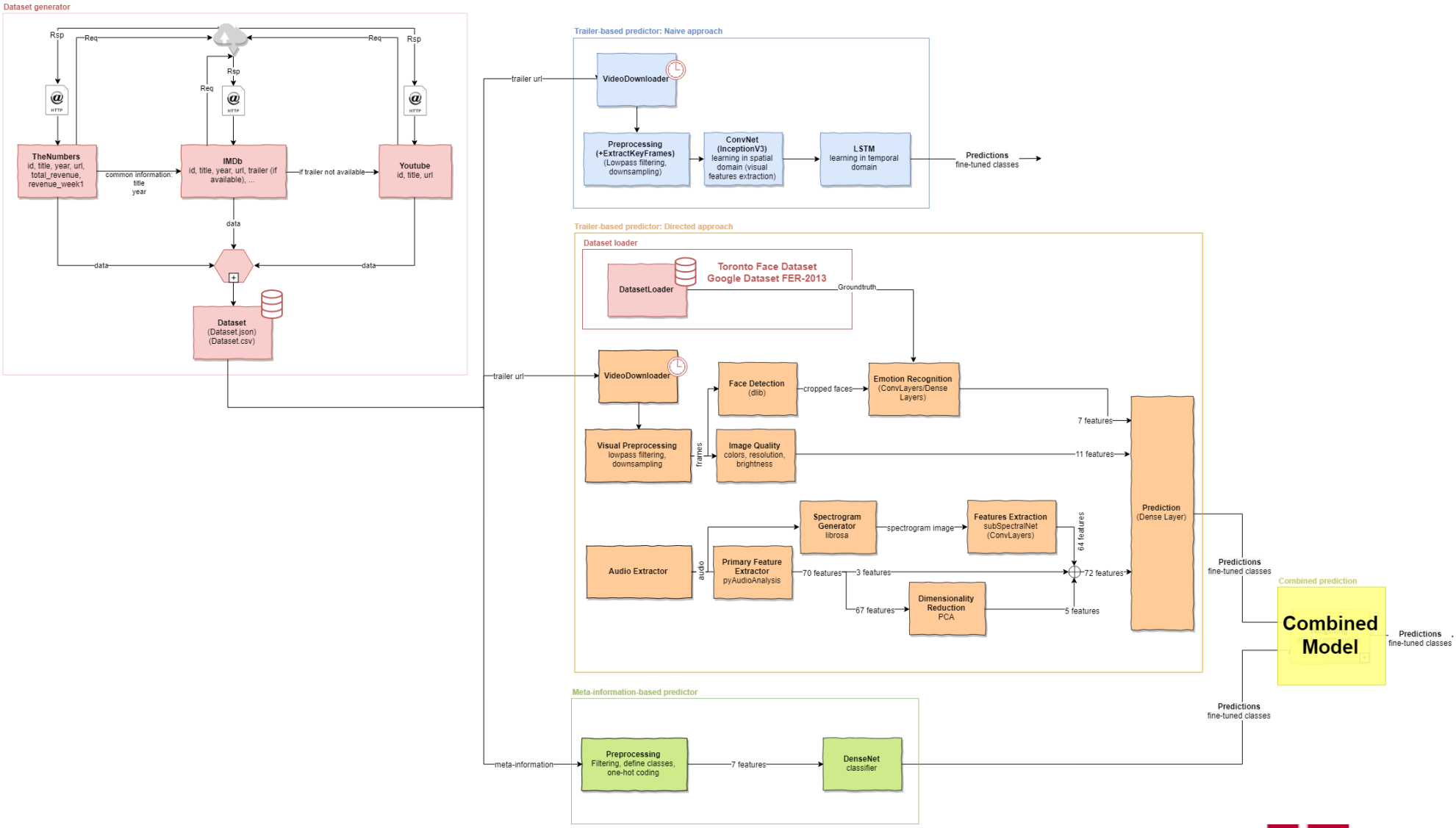
- Model - Evaluation

Metrics

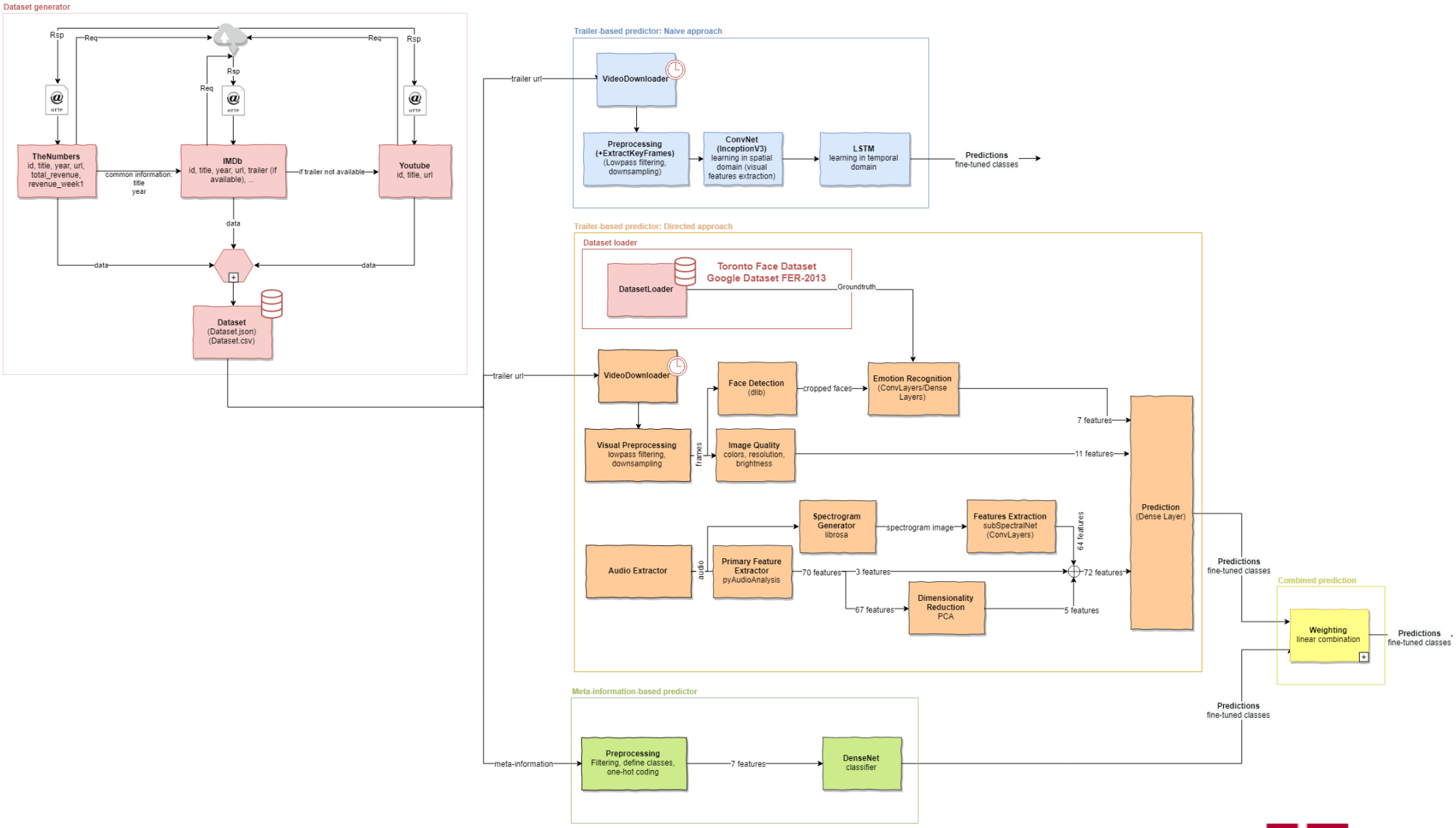
Overall accuracy	0.30922
Average accuracy	0.861844
Micro-averaged precision	0.30922
Macro-averaged precision	0.31057
Micro-averaged recall	0.30922
Macro-averaged recall	0.288811

Table of contents

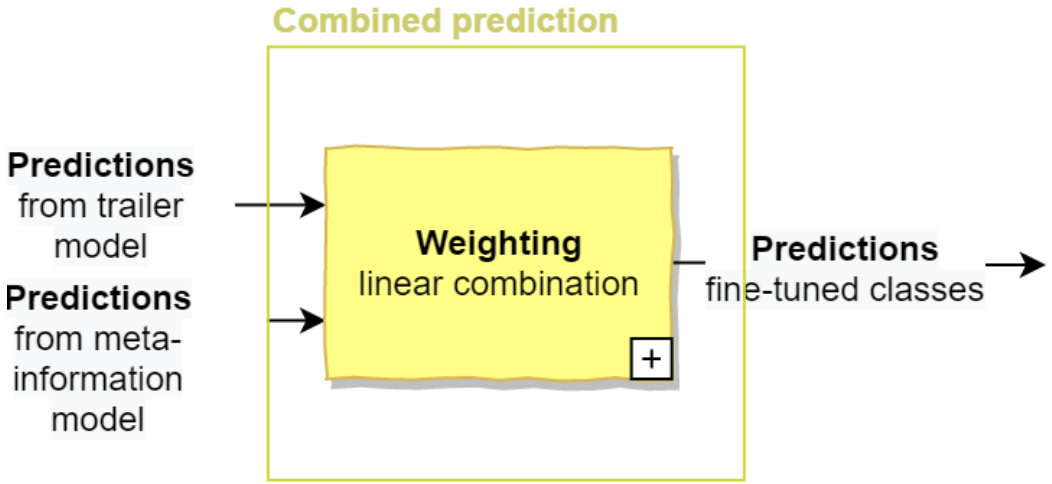
Combined Model



Combined Model



Combined Model



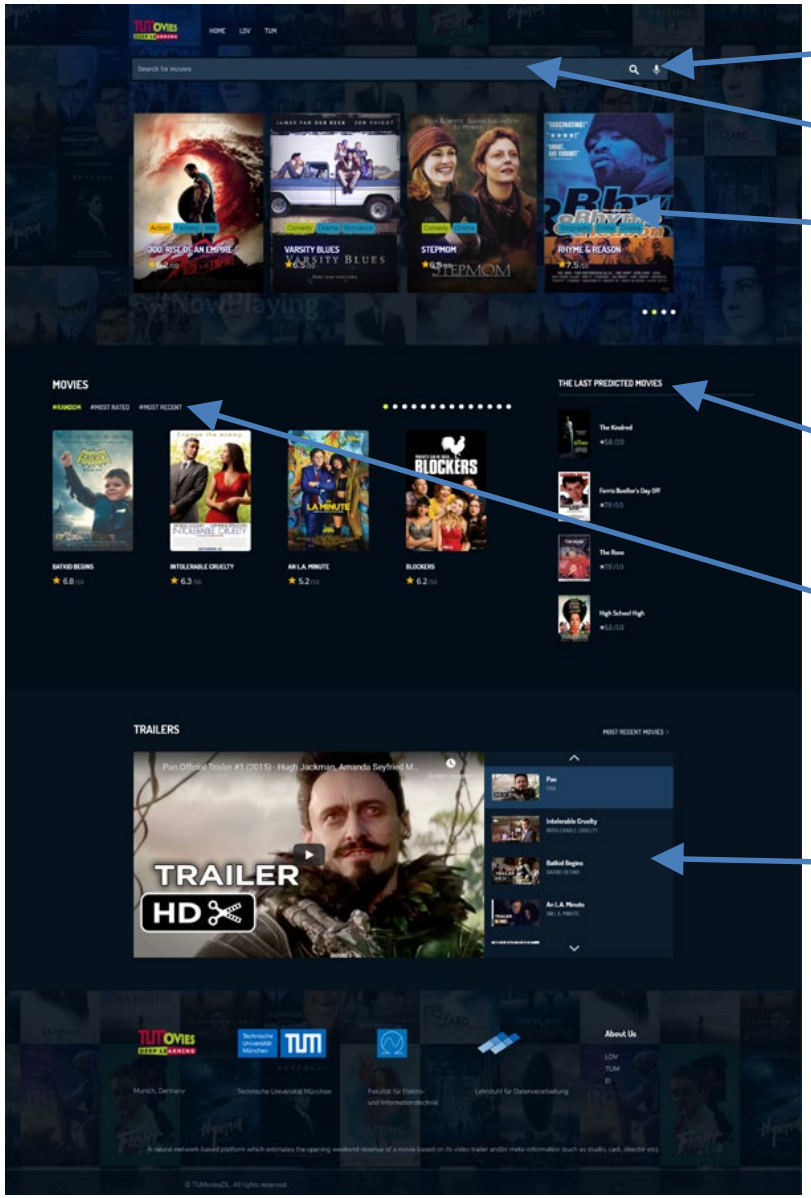
- Simple linear combination.
- Weight was learned by solving a least square error problem, whereby 20 data samples were randomly selected from the training dataset.
- Result: $\alpha_{trailer-based-pred} \approx 0.1$ $\alpha_{meta-info-based-pred} \approx 0.9$

Table of contents

Frontend – Demo

Live Demo

Frontend – Homepage



Speech recognition

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

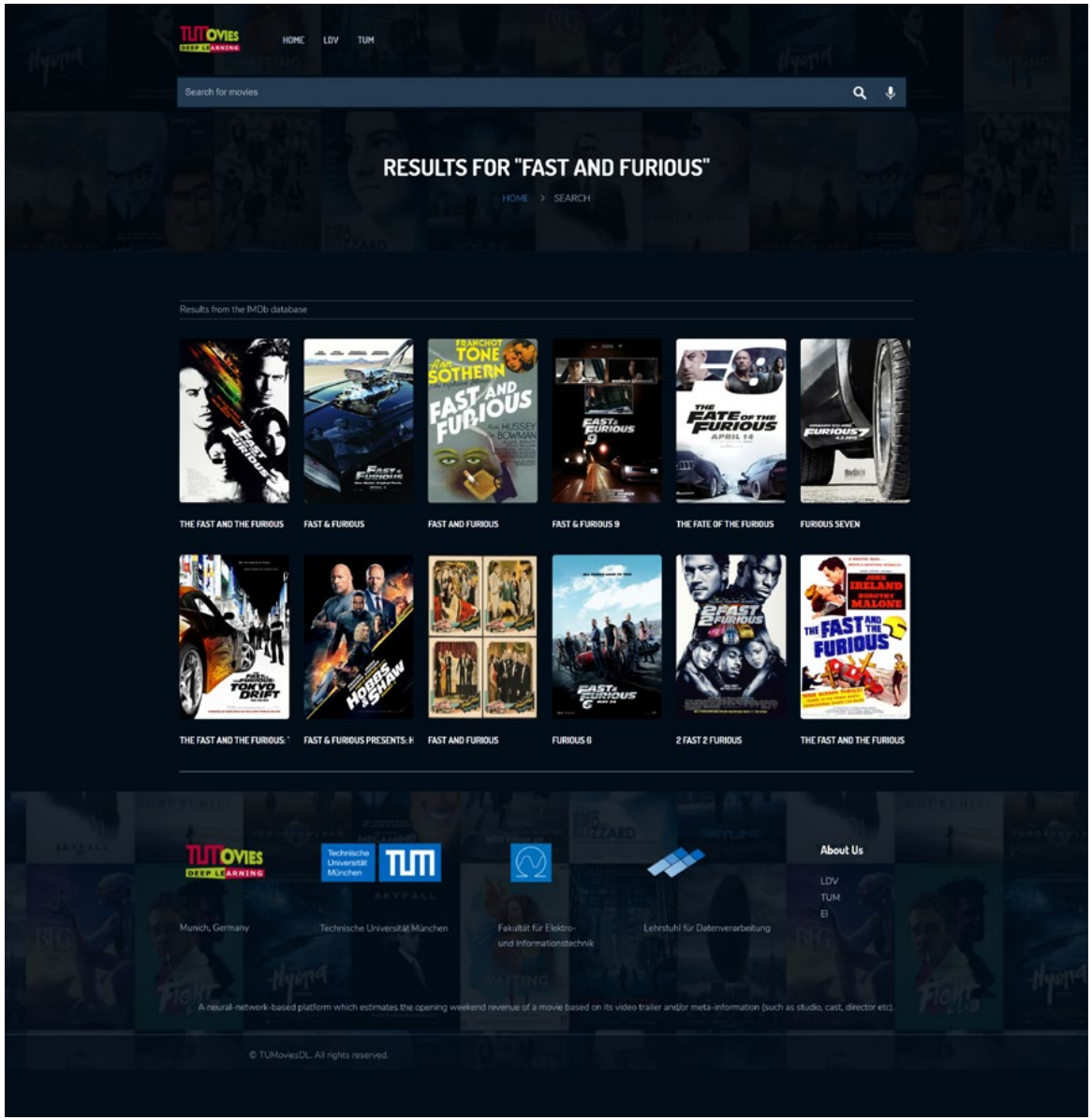
Randomly selected movies

Last predicted movies

Random/most rated/ most recent movies

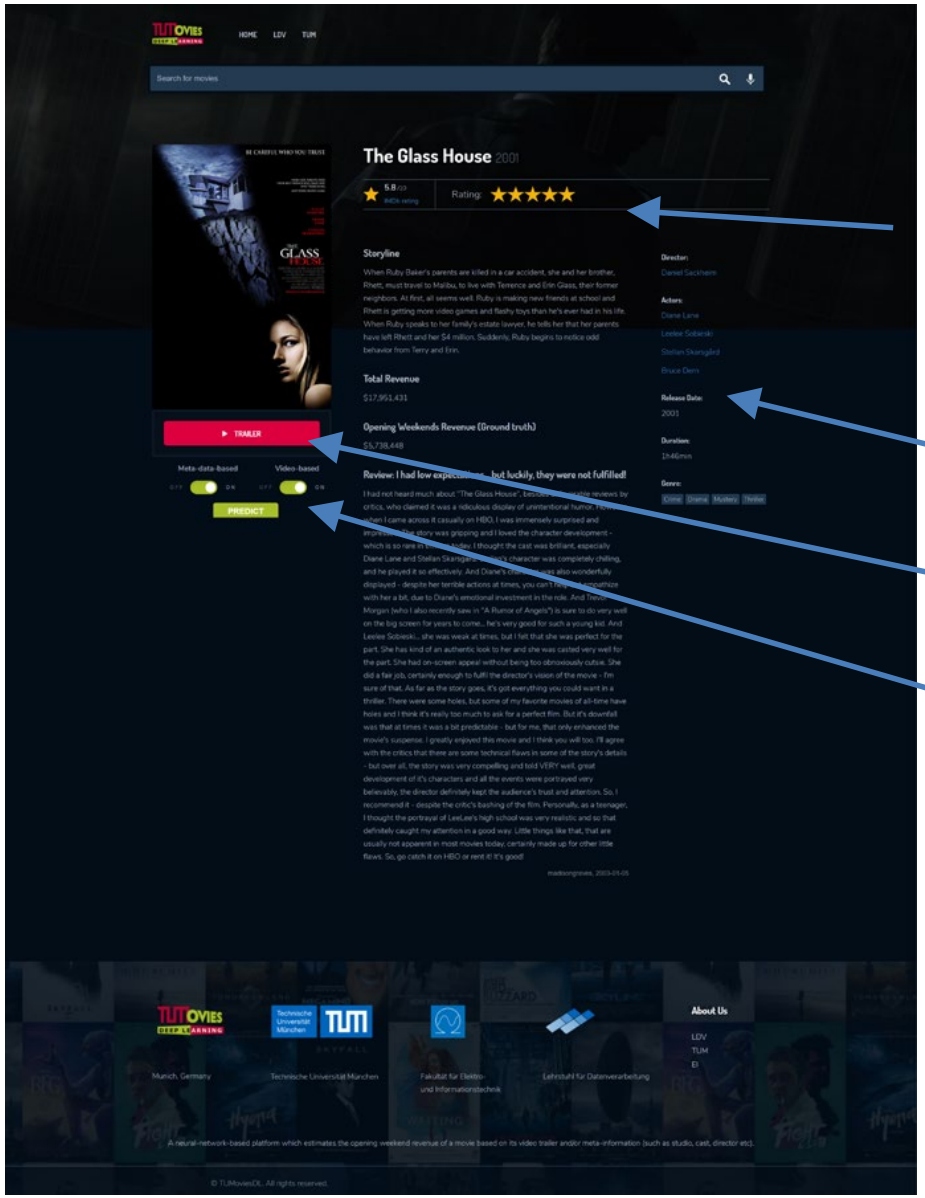
Random trailers

Frontend – Search page



Results from IDMb online database

Frontend – Movie page



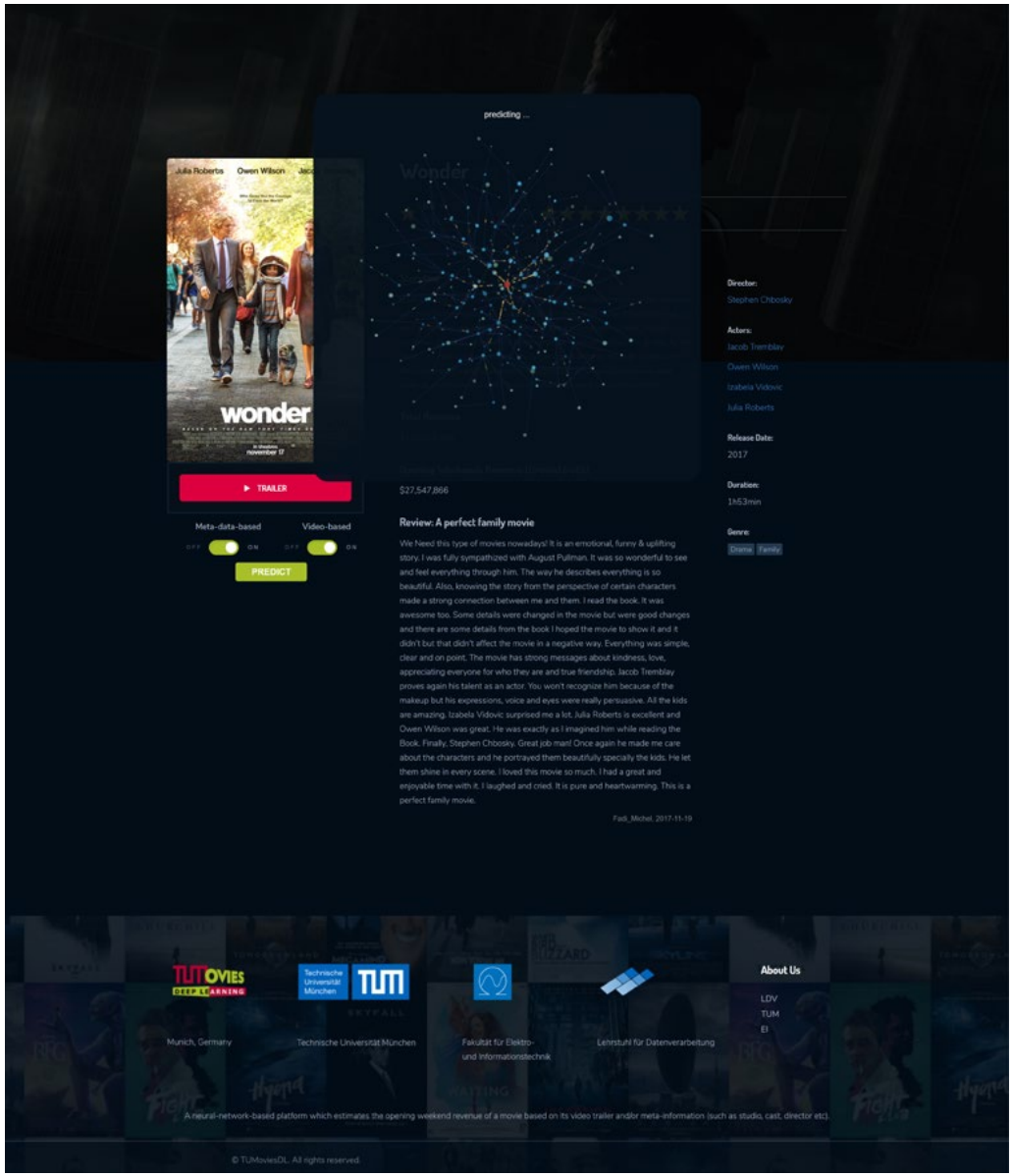
Randomly selected movies

Some metadata

Trailer player

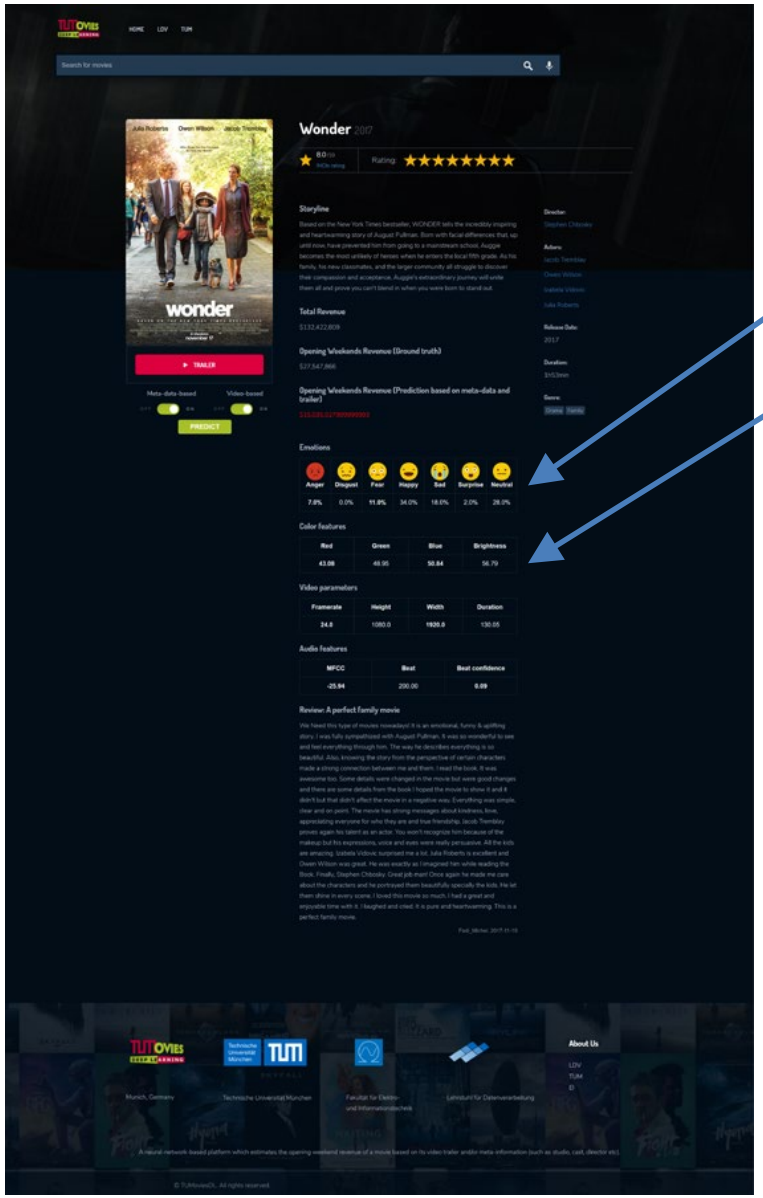
Prediction mode (Metadata-based and/or trailer-based)

Frontend – Loading page (Predicting)



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

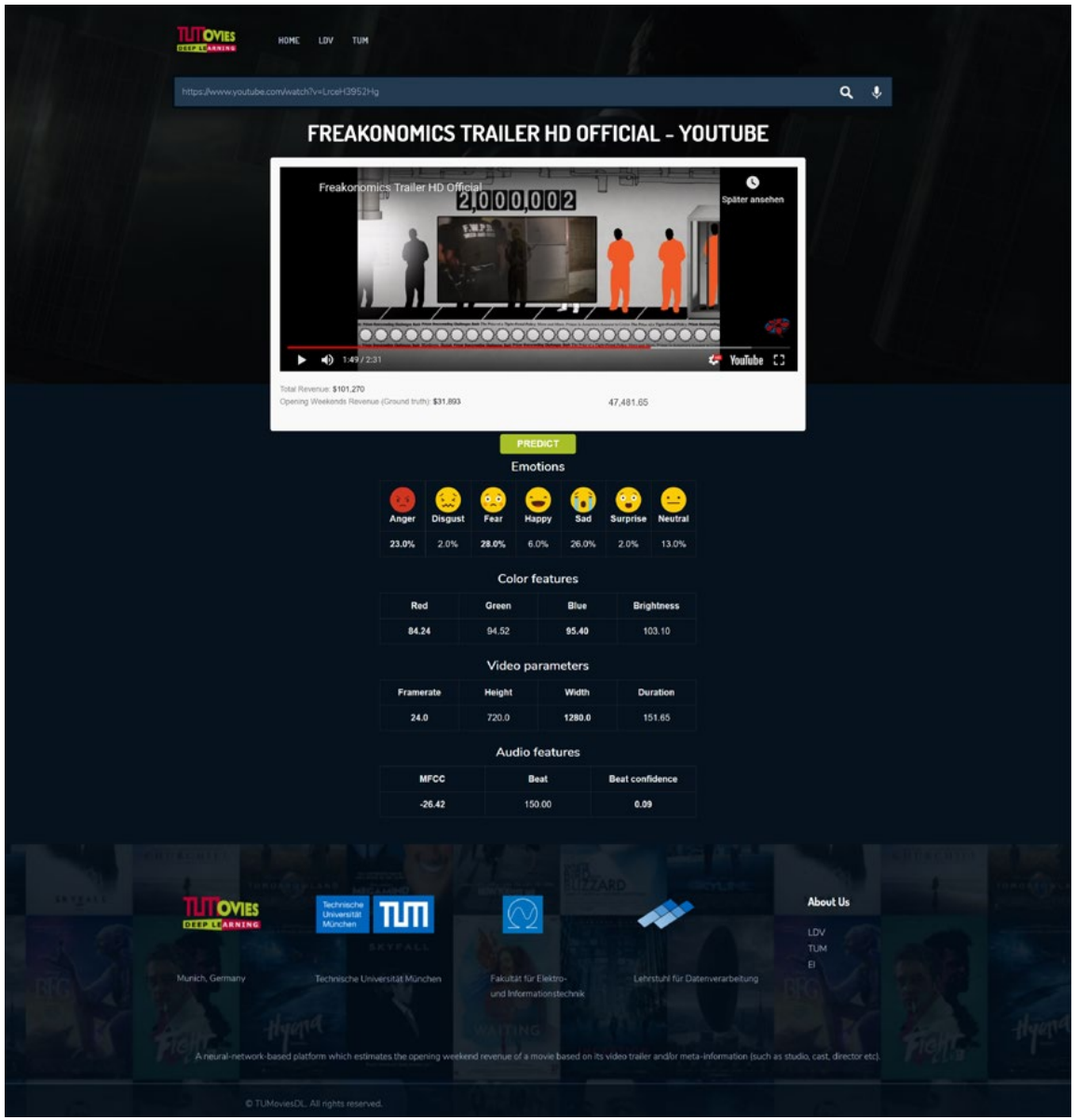
Frontend – Results



Extracted emotions

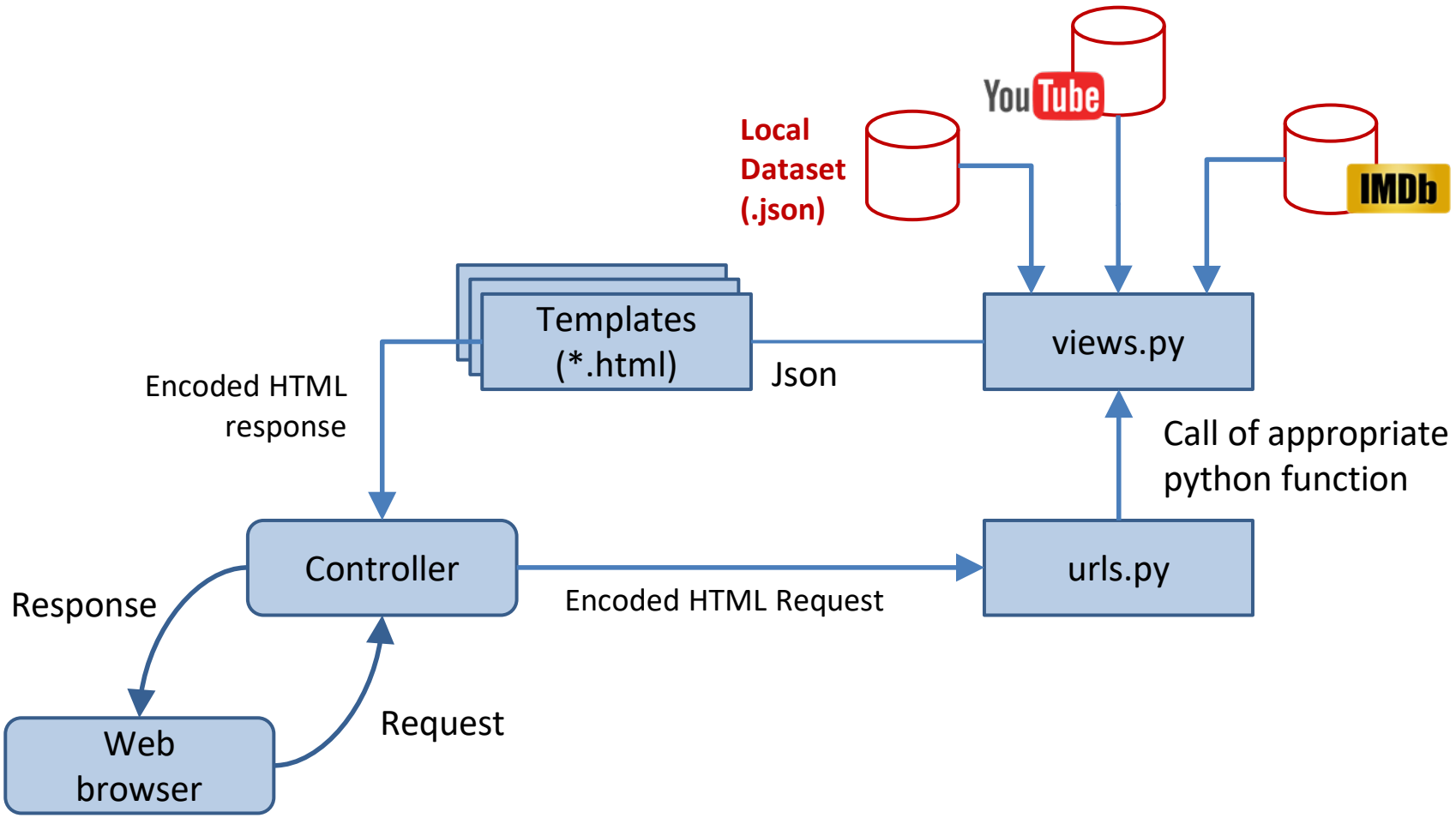
Other trailer features (not all features are shown)

Frontend – Results



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

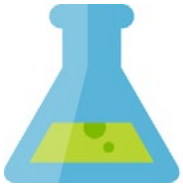
Conclusion and Discussion

- 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!