ITMO

Laboratory task 2
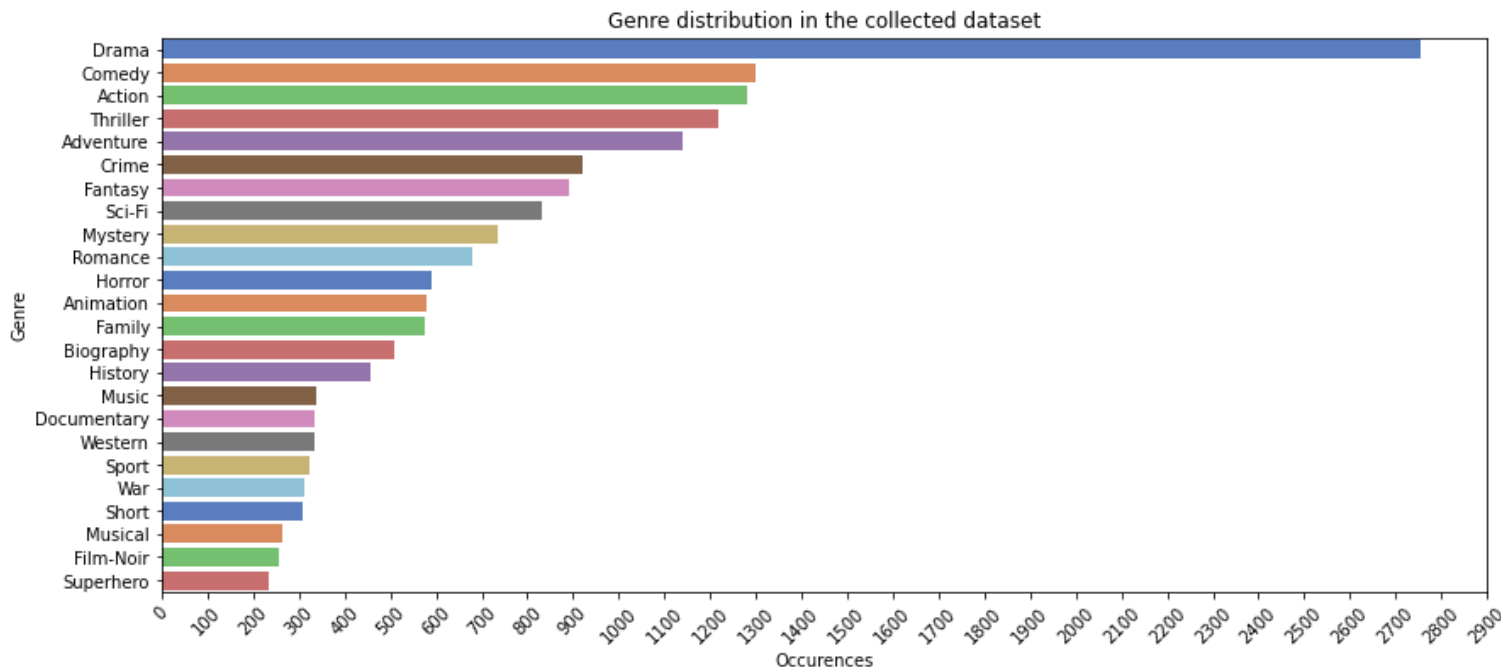
Shishkov V., Pinaev Z.

EDA

Data collected from intervals [1950, 1958], [2015, 2022].
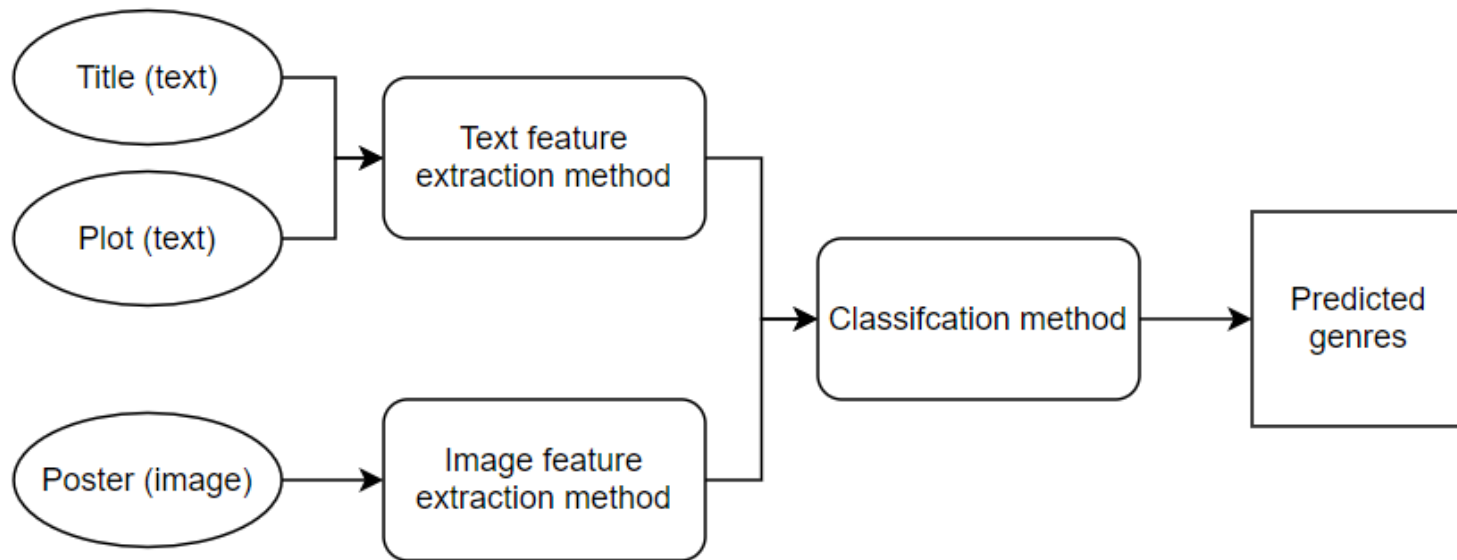
Most problem genre: "Superhero"

Total movies: 4702
Missing plot values: 1405

Genre distribution in the collected dataset

ML prediction pipeline

Used methods

ИТМО

Image methods:
- Resnet50
- img2vec
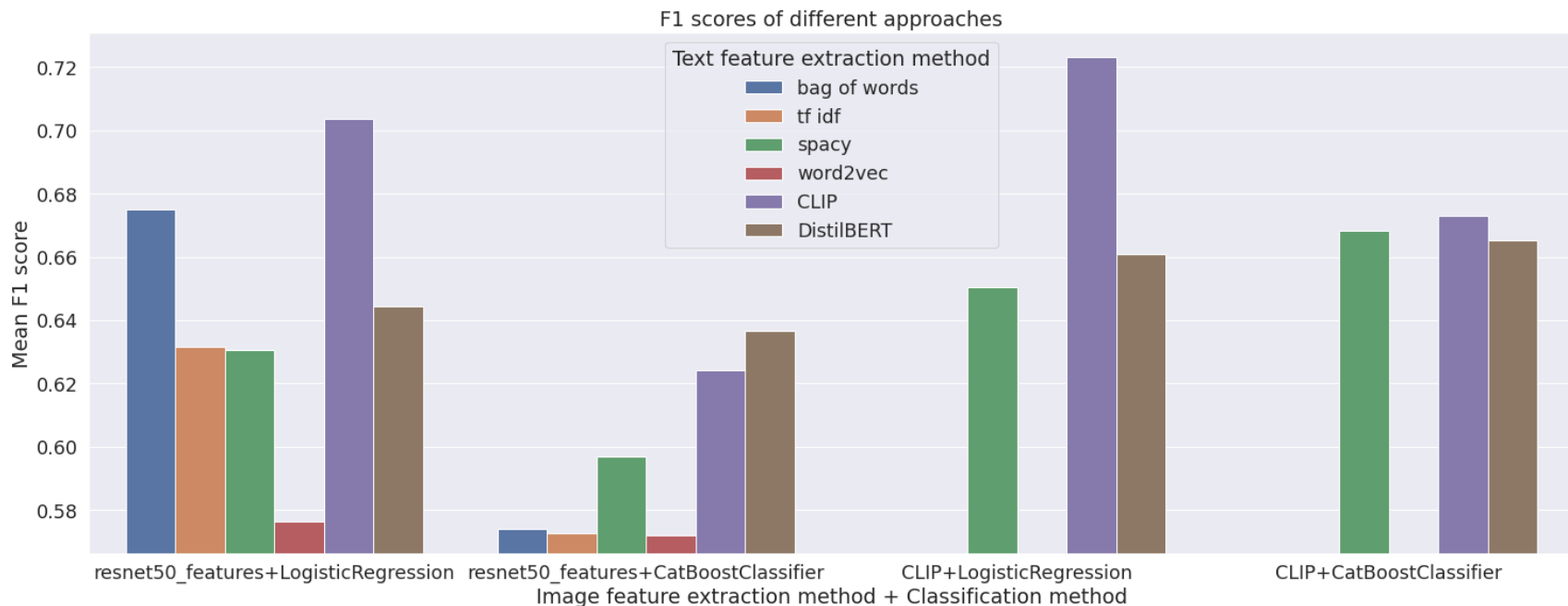- CLIP

Classification methods:
- Logistic Regression
- CatBoostClassifier
- Simple Neural Networks

Text methods:
- Bag of words
- TF – IDF
- spacy library (pretrained text transformers)
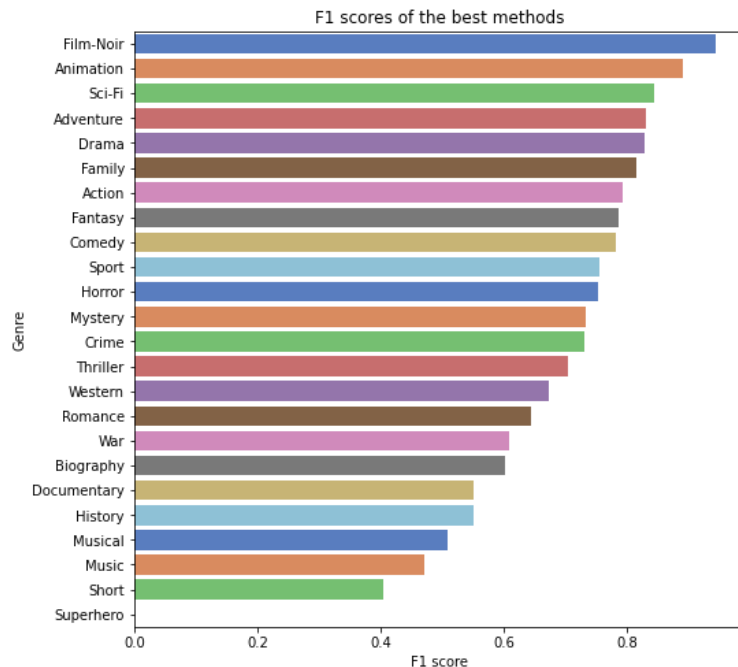- Gensim library (word2vec model)
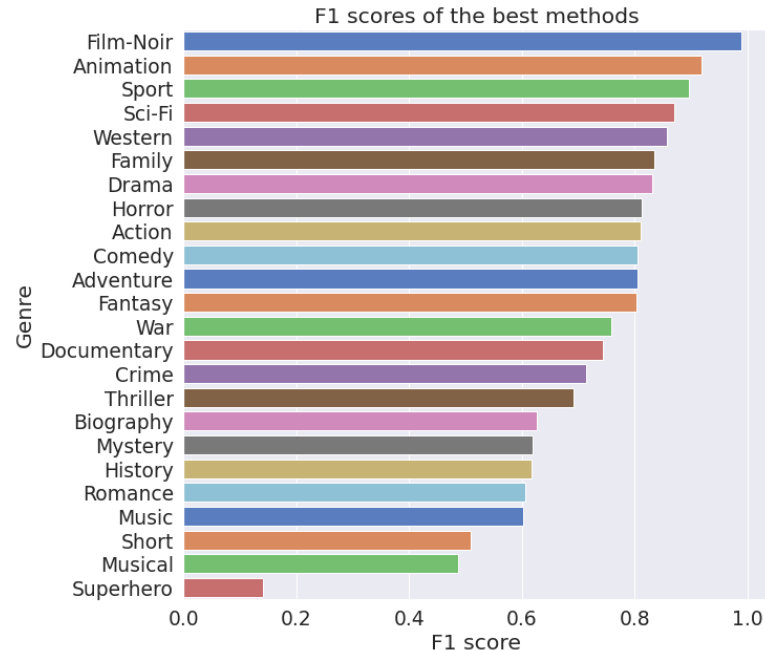- CLIP
- Distil BERT

Method comparison results

F1 scores of different approaches

Method comparison results

ResN + BoW + LR

CLIP + CLIP + LR

ИТМО

Title: 14 Peaks: Nothing Is Impossible

Poster

Plot: The documentary is about Project Possible, a plan by Nepali high altitude climber Nirmal Purja to climb all of the world's 14 highest peaks with an altitude greater than 8000m (called eight-thousanders) inside 7 months (i.e. from late spring to late summer, before the winter season begins).

Genre	Predicted probability	Ground truth
Adventure	0,82	1
Documentary	1,00	1
Sport	0,98	1

ITMO

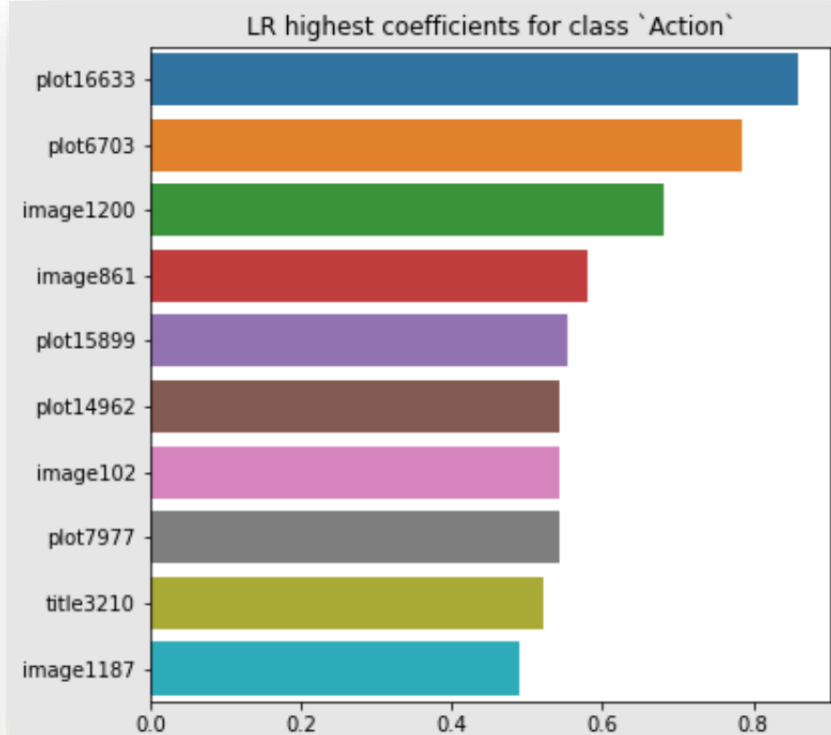**Title**: Venom: Let There Be Carnage

Poster

Plot: Eddie Brock struggles to adjust to his new life as the host of the alien symbiote Venom, which grants him super-human abilities in order to be a lethal vigilante. Brock attempts to reignite his career by interviewing serial killer Cletus Kasady, who becomes the host of the symbiote Carnage and escapes prison after a failed execution.
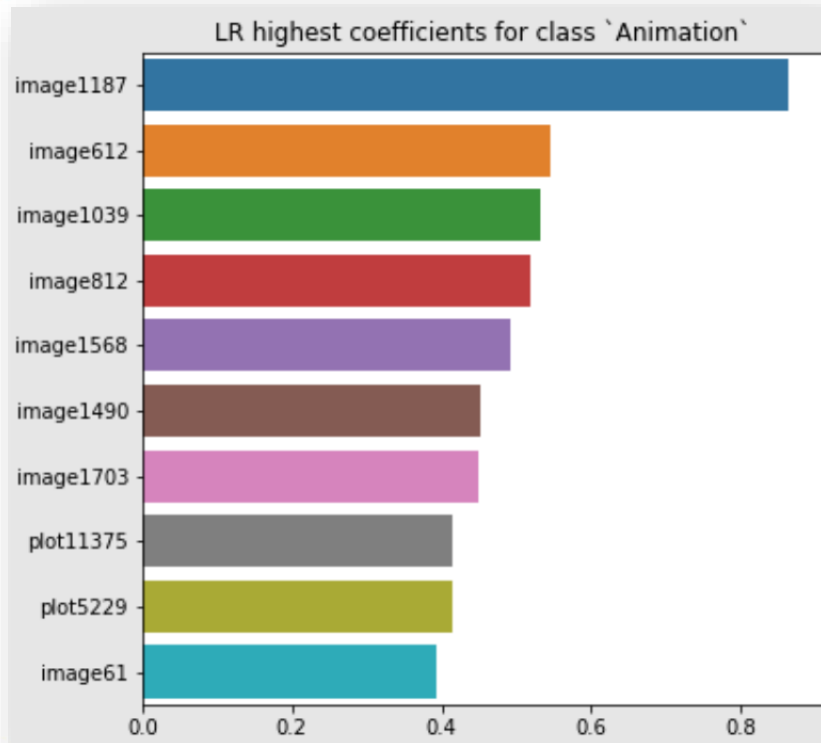

Genre	Predicted probability	Ground truth
Action	1,00	1
Adventure	0,97	1
Sci-Fi	0,99	1
Superhero	0,34	0
Thriller	0,96	1

8

Model analysis. ResN + BoW + LR

1. "super" in plot
2. "force" in plot
3. Image feature
4. Image feature
5. "soldier" in plot
6. "save" in plot
7. Image feature
8. "hero" in plot
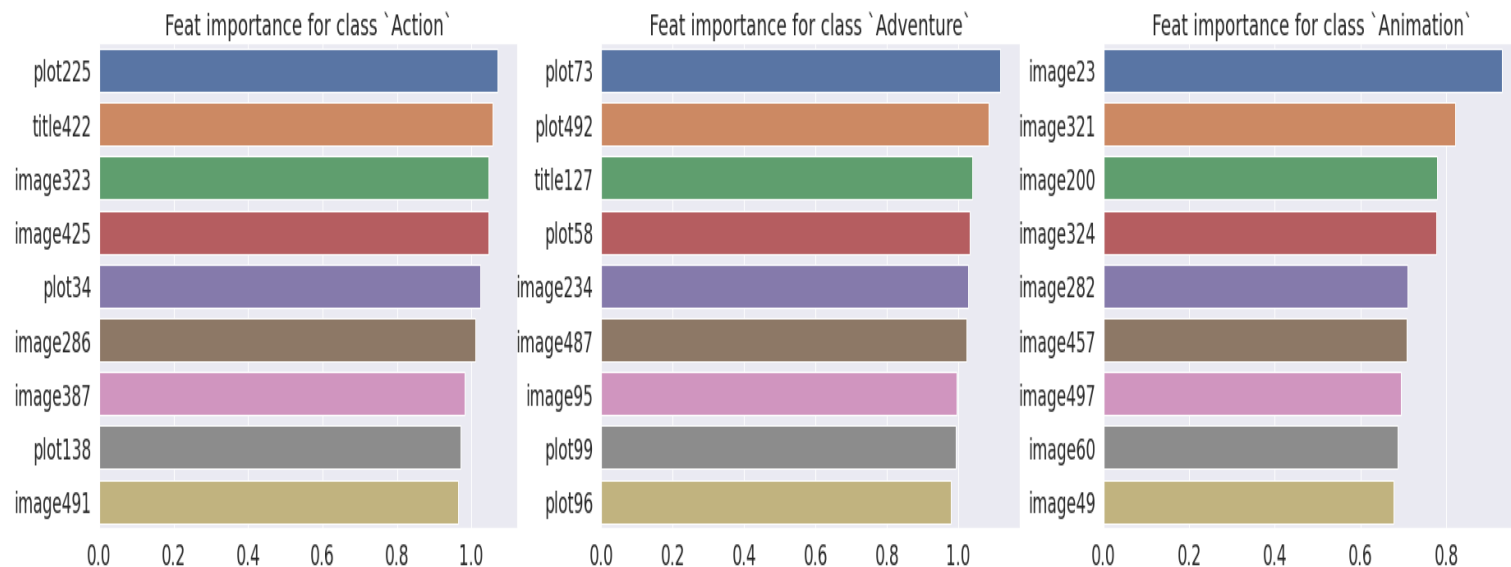9. "star" in title
10. Image feature

Model analysis. ResN + BoW + LR

LR highest coefficients for class `Animation`

1. Image feature
2. Image feature
3. Image feature
4. Image feature
5. Image feature
6. Image feature
7. Image feature
8. "monster" in plot
9. "dragon" in plot
10. Image feature

Model analysis. CLIP + CLIP + LR

Thank you for your attention!

iT'sMOre than a
UNIVERSITY