ITMO

Laboratory task 2

Shishkov V., Pinaev Z.
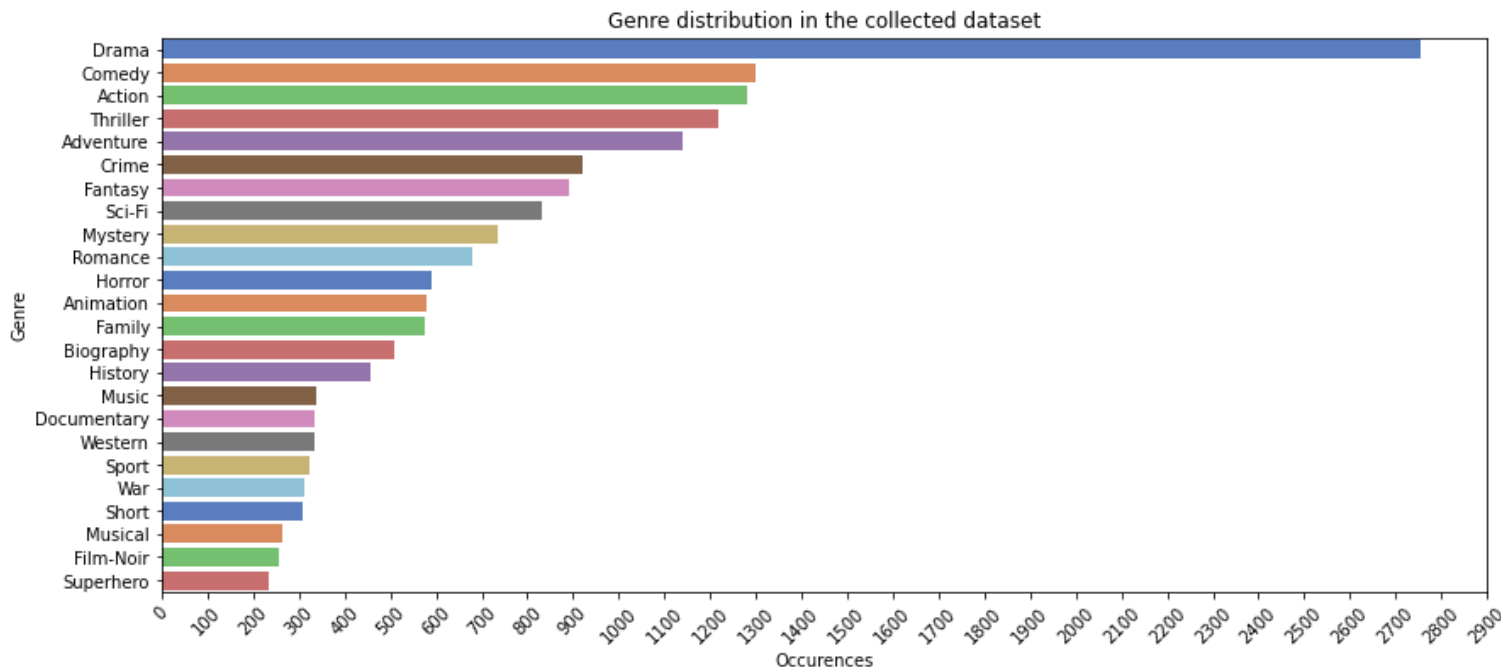
Collected data exploration

ИТМО

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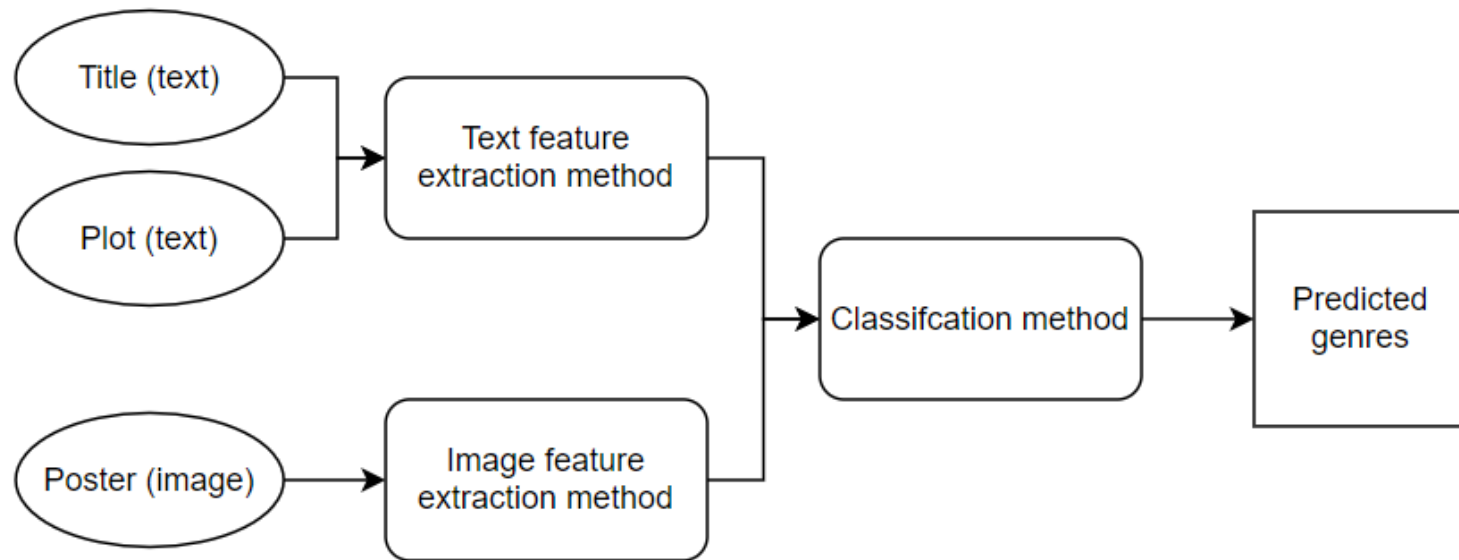
Most problem genre: "Superhero"

Total movies: 4702
Missing plot values: 1405

Genre distribution in the collected dataset

ML prediction pipeline

Used methods

Image methods:
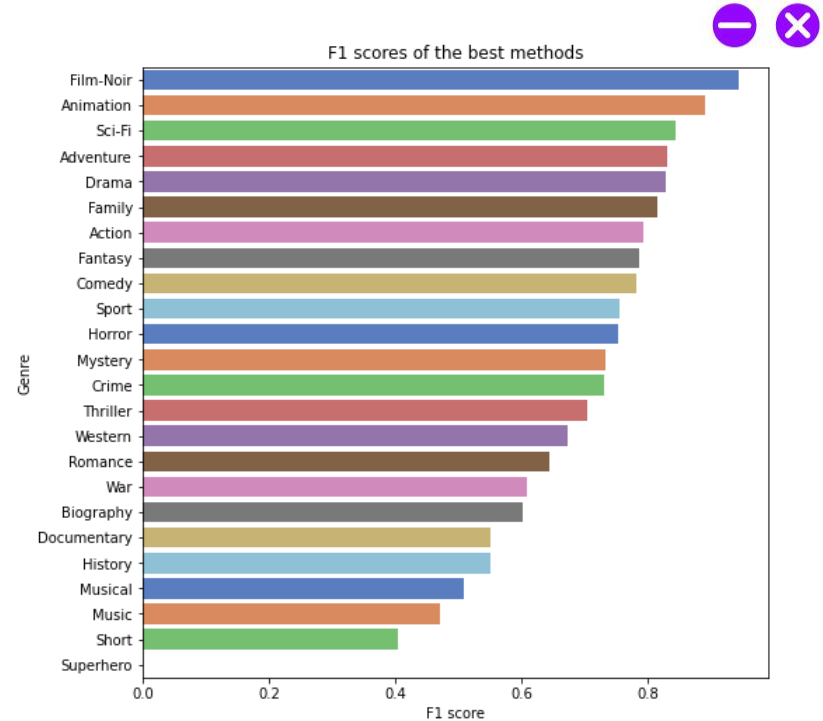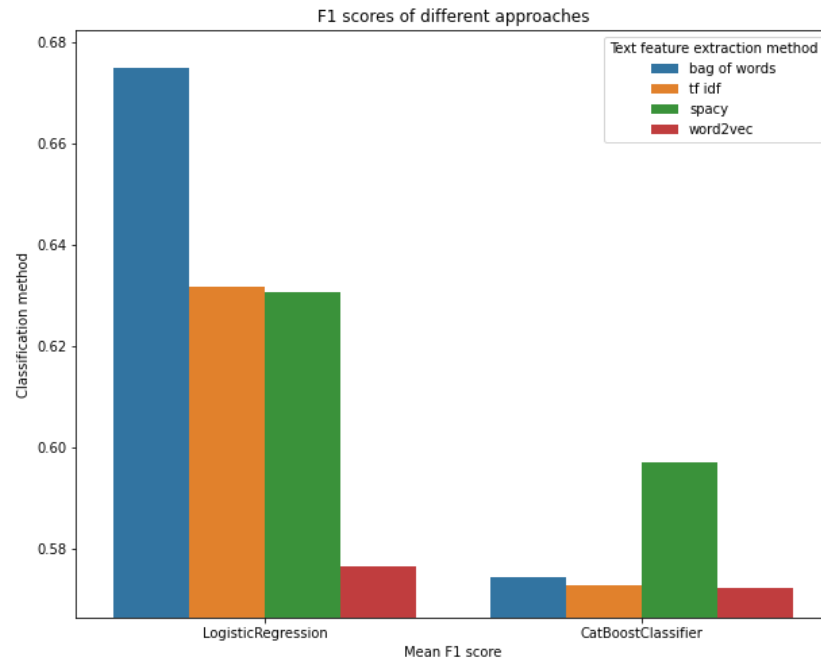
- Resnet50
- img2vec
- CLIP

Text methods:

- Bag of words
- TF – IDF
- spacy library (pretrained text transformers)
- Gensim library (word2vec model)

Classification methods:

- Logistic Regression
- CatBoostClassifier
- Simple Neural Networks

Method comparison results

Title: `14 Peaks: Nothing Is Impossible`

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).`

Poster

	genres	predicted probability	ground truth
0	Action	0.00	0
1	Adventure	0.87	1
2	Animation	0.00	0
3	Biography	0.02	0
4	Comedy	0.00	0
5	Crime	0.00	0
6	Documentary	0.95	1
7	Drama	0.09	0
8	Family	0.00	0
9	Fantasy	0.00	0
10	Film-Noir	0.00	0
11	History	0.05	0
12	Horror	0.02	0
13	Music	0.01	0
14	Musical	0.00	0
15	Mystery	0.02	0
16	Romance	0.01	0
17	Sci-Fi	0.00	0
18	Short	0.00	0
19	Sport	0.92	1
20	Superhero	0.00	0
21	Thriller	0.08	0
22	War	0.00	0
23	Western	0.01	0

Prediction examples (2)

Title: Venom: Let There Be Carnage

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.
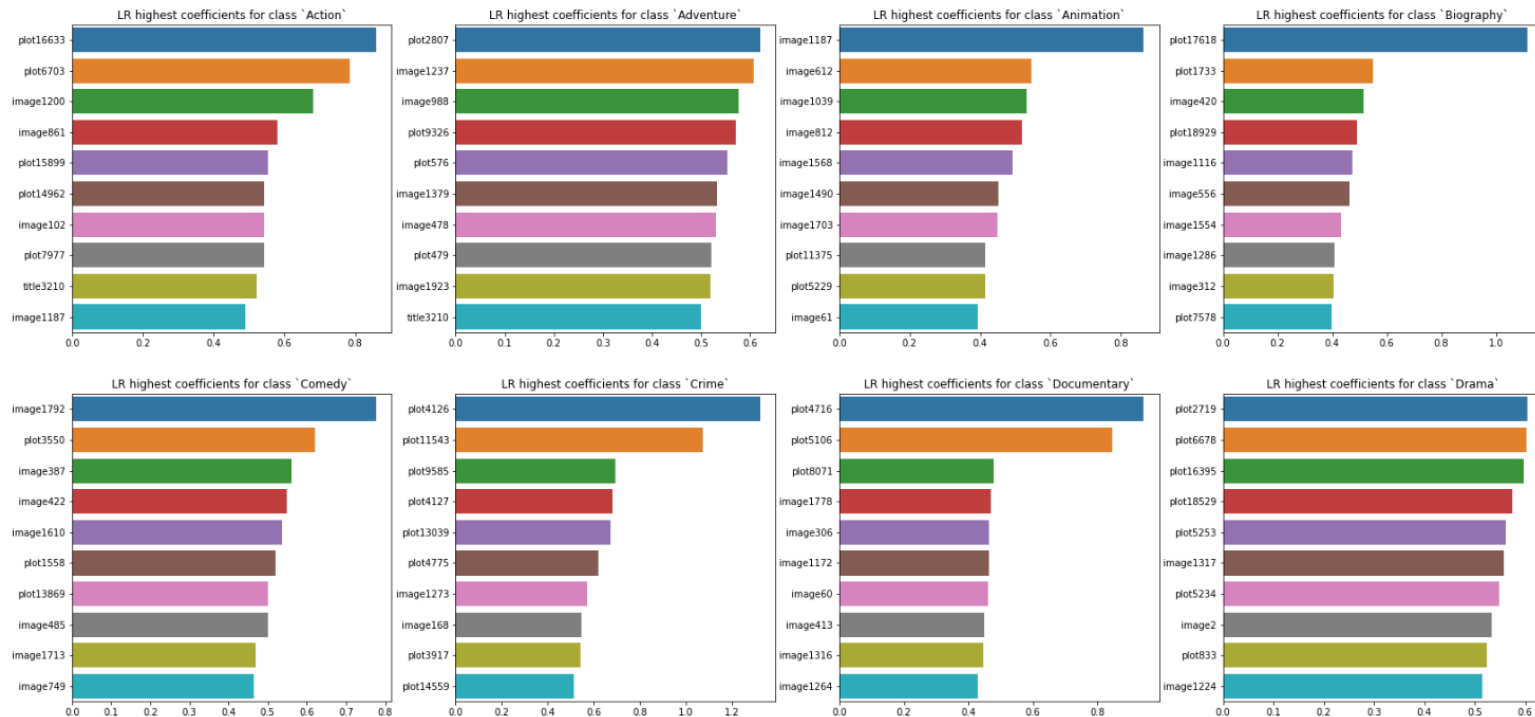
Poster

	genres	predicted probability	ground truth
0	Action	0.96	1
1	Adventure	0.96	1
2	Animation	0.00	0
3	Biography	0.00	0
4	Comedy	0.02	0
5	Crime	0.04	0
6	Documentary	0.00	0
7	Drama	0.08	0
8	Family	0.01	0
9	Fantasy	0.03	0
10	Film-Noir	0.00	0
11	History	0.00	0
12	Horror	0.05	0
13	Music	0.00	0
14	Musical	0.00	0
15	Mystery	0.03	0
16	Romance	0.01	0
17	Sci-Fi	0.99	1
18	Short	0.00	0
19	Sport	0.00	0
20	Superhero	0.34	0
21	Thriller	0.97	1
22	War	0.00	0
23	Western	0.00	0

Model analysis

Thank you for your attention!

iT's MOre than a UNIVERSITY