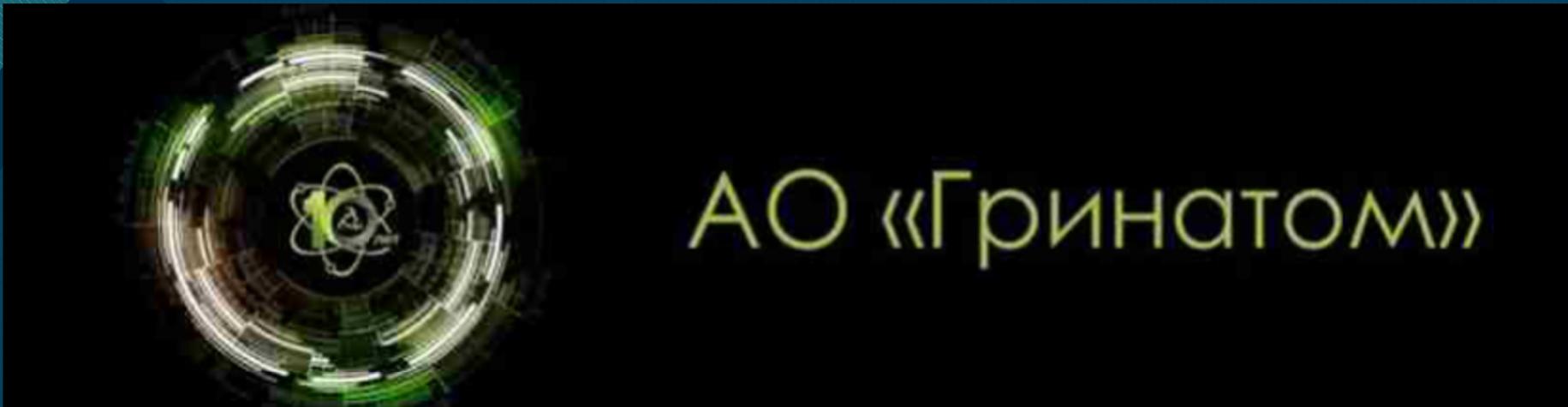


# Тестовое задание для junior-специалистов по направлению Data Science



# Отчет о работе с оценкой точности полученного результата на тестовой выборке

Задача: Разработать веб-сервис для оценки  
комментариев, рецензий, отзывов к  
фильмам.

Данные: открытый набор данных, который  
содержит в себе отзывы о фильмах, а также  
соответствующие им оценки рейтинга.

# Выполнено

- Произведена предобработка данных. Удалены стоп-слова из отзывов, произведены лемматизация, токинезация, векторизация учебной и тестовой выборки, стандартизация, отбор признаков.
- Сформированы датасеты для обучения.
- На сформированные датасетах обучены и сохранены для последующего создания прототипа приложения: векторизатор текста (CountVectorizer), стандартизатор(StandardScaler), система отбора признаков (SelectKBest).
- Произведено обучение моделей: GaussianNB, RandomForestClassifier, нейросети tensorflow в режиме многоклассовой классификации и бинарной классификации. Наилучшие результаты - за нейросетью.
- Создано приложение-прототип на фреймворке django, куда интегрированы ранее обученные модели.
- Код проекта выложен в публичный репозиторий github.
- Приложение развернуто на VPS по адресу <http://127.0.0.1:8000/>

# Технологический стек

- Python
- Pandas
- Sklearn
- Nltk
- tensorflow.

# Метрики обучения на тестовой выборке

782/782 [=====] - 1s 1ms/step				
	precision	recall	f1-score	support
0	0.87	0.85	0.86	12500
1	0.86	0.87	0.86	12500
accuracy			0.86	25000
macro avg	0.86	0.86	0.86	25000
weighted avg	0.86	0.86	0.86	25000

Для бинарной классификации позитивная или негативная направленность комментариев.

# Метрики обучения на тестовой выборке

	precision	recall	f1-score	support
1	0.51	0.81	0.63	5022
2	0.17	0.02	0.03	2302
3	0.25	0.16	0.19	2541
4	0.29	0.33	0.31	2635
7	0.29	0.14	0.19	2307
8	0.28	0.31	0.29	2850
9	0.13	0.00	0.01	2344
10	0.48	0.76	0.59	4999
accuracy			0.42	25000
macro avg	0.30	0.32	0.28	25000
weighted avg	0.34	0.42	0.35	25000

Рейтинг фильма на основании комментариев

# Ссылки

- Репозиторий гитхаб: <https://github.com/1kar1s/filmreviews> 
- Приложение развернуто по адрес: <http://127.0.0.1:8000/> 

## Контактная информация

Автор: Шуркин Александр Вячеславович  
Город Кстово.

Стажер - аналитик Fintech, Стажер-разработчик Fintech, Data Scientist.

 89873994104

 keenji1388@gmail.com