

Задача предложена компанией Тинькофф

Задача 5: Поиск "токсичных" комментариев

Название команды: Трансформеры

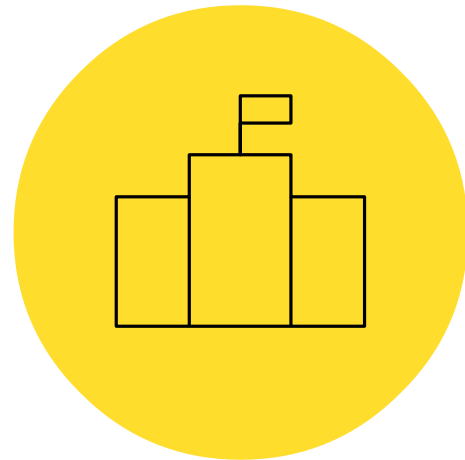
Ладанова Светлана

Холькин Сергей

Шакиров Ренат

Описание задачи:

- ➔ Научиться определять тональность комментариев и постов с использованием state-of-the-art подходов для обработки текстов.
- ➔ Токсичные комментарии являются большой проблемой для онлайн платформ, включая русскоязычные.
- ➔ По данным опросов 73% онлайн пользователей сталкивались с оскорблениями в свой адрес



Технический обзор задачи

Датасет

- Тинькофф Пульс

Метрика

- F1 score
- AUC
- GMC-AUC

Мат Модель

- Deep Learning
- Classic ML

Решение задачи: Данные

- “Чистка” ненужных данных
- Удаление стоп-слов
- Удаление ненужных символов
- Векторизация данных





Решение задачи: Модель



CatBoost



RNN



Transformers (BERT, ELMo)



Решение задачи: Модель

Обработка литературы

Table 1: Binary classification of toxic comments in Russian

System	<i>P</i>	<i>R</i>	<i>F₁</i>
<i>MNB</i>	87.01%	81.22%	83.21%
<i>BiLSTM</i>	86.56%	86.65%	86.59%
<i>M – BERT_{BASE} – Toxic</i>	91.19%	91.10%	91.15%
<i>ruBert – Toxic</i>	91.91%	92.51%	92.20%
<i>M – USE_{CNN} – Toxic</i>	89.69%	90.14%	89.91%
<i>M – USE_{Trans} – Toxic</i>	90.85%	91.92%	91.35%



Решение задачи: Модель

Результат работы CatBoost

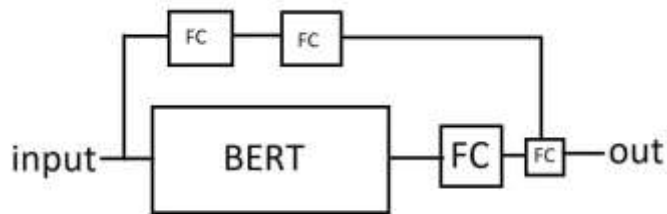


Yandex
CatBoost

	precision	recall	f1-score	support
0	0.99	0.97	0.98	27446
1	0.09	0.34	0.15	255
accuracy			0.96	27701
macro avg	0.54	0.65	0.56	27701
weighted avg	0.99	0.96	0.97	27701

Решение задачи: Модель

Неудачный эксперимент с ruBert



Evaluating...

	precision	recall	f1-score	support
0	0.99	0.93	0.96	6855
1	0.24	0.67	0.36	228
accuracy			0.92	7083
macro avg	0.62	0.80	0.66	7083
weighted avg	0.96	0.92	0.94	7083



Итоговая модель



	precision	recall	f1-score	support
0	0.99	0.97	0.98	6855
1	0.39	0.68	0.50	228
accuracy			0.96	7083
macro avg	0.69	0.82	0.74	7083
weighted avg	0.97	0.96	0.96	7083



Задача предложена компанией Тинькофф

Спасибо за внимание

Название команды: Трансформеры

Ладанова Светлана https://t.me/sveta_lad,
<https://github.com/Ladanovasv>

Холькин Сергей <https://t.me/kholkins>,
https://github.com/SKholkin/toxic_bert

Шакиров Ренат <https://t.me/Varfalamei13>,
https://github.com/Varfalamei/Toxic_detection_bert