



Выпускная квалификационная работа на тему:

Определение эмоций персонажей в тексте посредством анализа
диалогов с помощью машинного обучения

Студент: группа М8О-406Б-19,
Гаврилов Максим Сергеевич

Дипломный руководитель: к.ф.-м.н.,
Лемтюжникова Дарья Владимировна

Москва, 2023 г.



Определение эмоций вне художественного текста

Применение предобученного на примерах из художественного текста детектора эмоций для решения задач в области конфликтологии.

Определение эмоций в художественном тексте

Использование эмоций, как признаков при классификации текста

Определение эмоций персонажей, как часть автоматического анализа текста

Применения автоматического анализа текстов:

- Обработка большого количества текстов
- Анализ тенденций в литературе различных эпох
- Разметка художественного текста для обучения искусственных нейронных сетей

Данная работа выполнена в рамках проекта по автоматическому анализу текстов, разрабатываемого по заказу сообщества Руниверс.

Цель и задачи



Цель: Получить модель, способную определять эмотивную тональность речи персонажей в диалоге на основе модели эмоций Келлермана-Плутчика.

Задачи:

- Создать алгоритм извлечения диалога из текста, способный устанавливать соответствия персонаж-реплика.
- Создать анализатор эмотивной тональности по Келлерману-Плутчику.
- Собрать из реализованных частей модель, способную отслеживать эмоции персонажей в ходе повествования.

Постановка задачи



Дано:

- **Текст на русском языке.**

Надо получить:

- **Список действующих лиц и список реплик.**
- **Графики, отображающие изменение эмоционального состояния персонажей в процессе повествования.**

Архитектура программного комплекса



- **Извлечение диалогов**

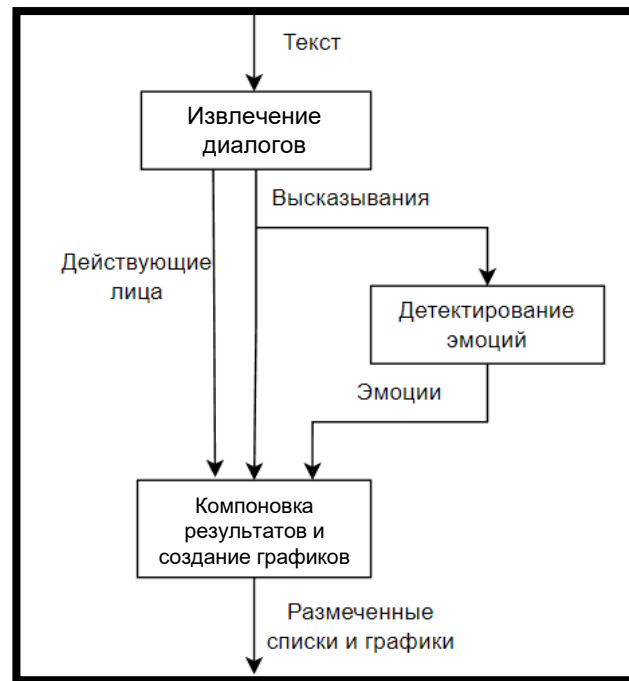
Вход — текст

Выход — список действующих лиц и их высказываний.

- **Детектирование эмоций**

Вход — предложение

Выход — эмоции, выраженные в нем.



Стек технологий





Набор данных cedr *

	Удовольствие	Печаль	Страх	Удивление	Злость	Нет эмоций
Количество предложений	1569	1417	589	607	411	3043

Набор данных на основе 10 первых глав романа «Мастер и Маргарита»

	Радость	Восхище ние	Злость	Грусть	Удивлен ие	Ожидани е	Отвраще ние	Страх	Нет эмоций
Количество предложений	83	45	114	46	73	88	46	70	308

* <https://huggingface.co/datasets/cedr>

Извлечение диалога из художественного текста



На вход подается текст,
разбитый по предложениям.

**Задача: получить список
действующих лиц и их
высказываний.**



Извлечение диалога из художественного текста



Используемый метод	Точность выделения слов, указывающих на действующих лиц	
	Набор данных А	Набор данных Б
Подлежащее по синтаксическому парсеру	0.715	0.767
Имя собственное	0.482	0.630
Имя с подчиненным глаголом.	0.919	0.848
Любое имя	0.88	0.832
Композиция методов	0.974	0.921

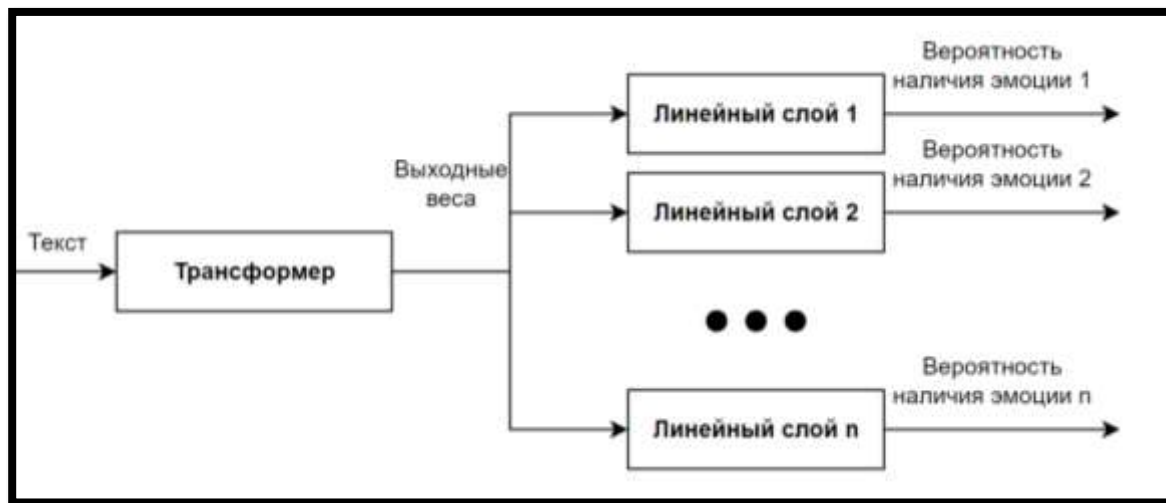
	Точность определения действующих лиц и их высказываний	
	Набор данных А	Набор данных Б
Точность	0.847	0.775

Определение эмоций



На вход поступает предложение.

Задача – определить, какие эмоции в нем представлены.



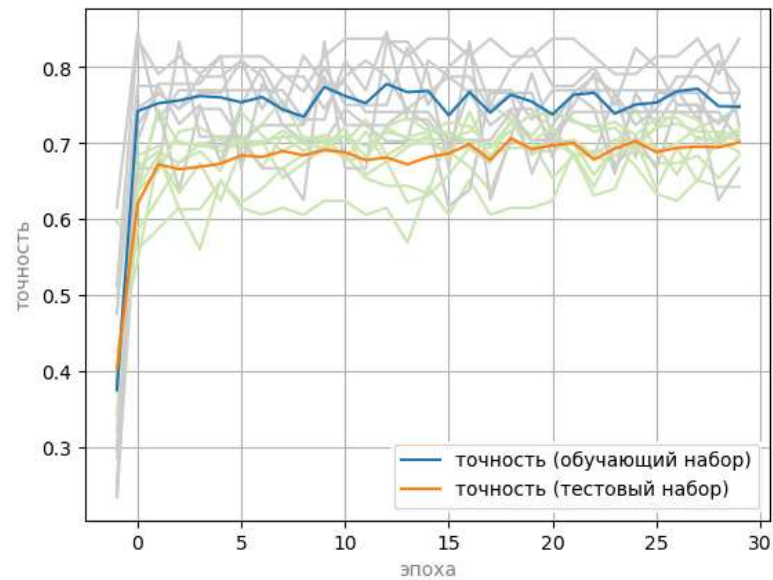
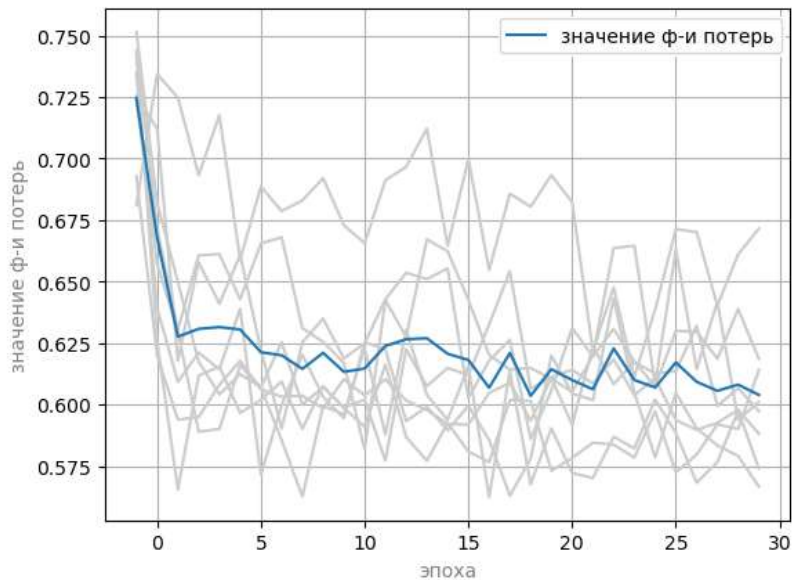
Определение эмоций



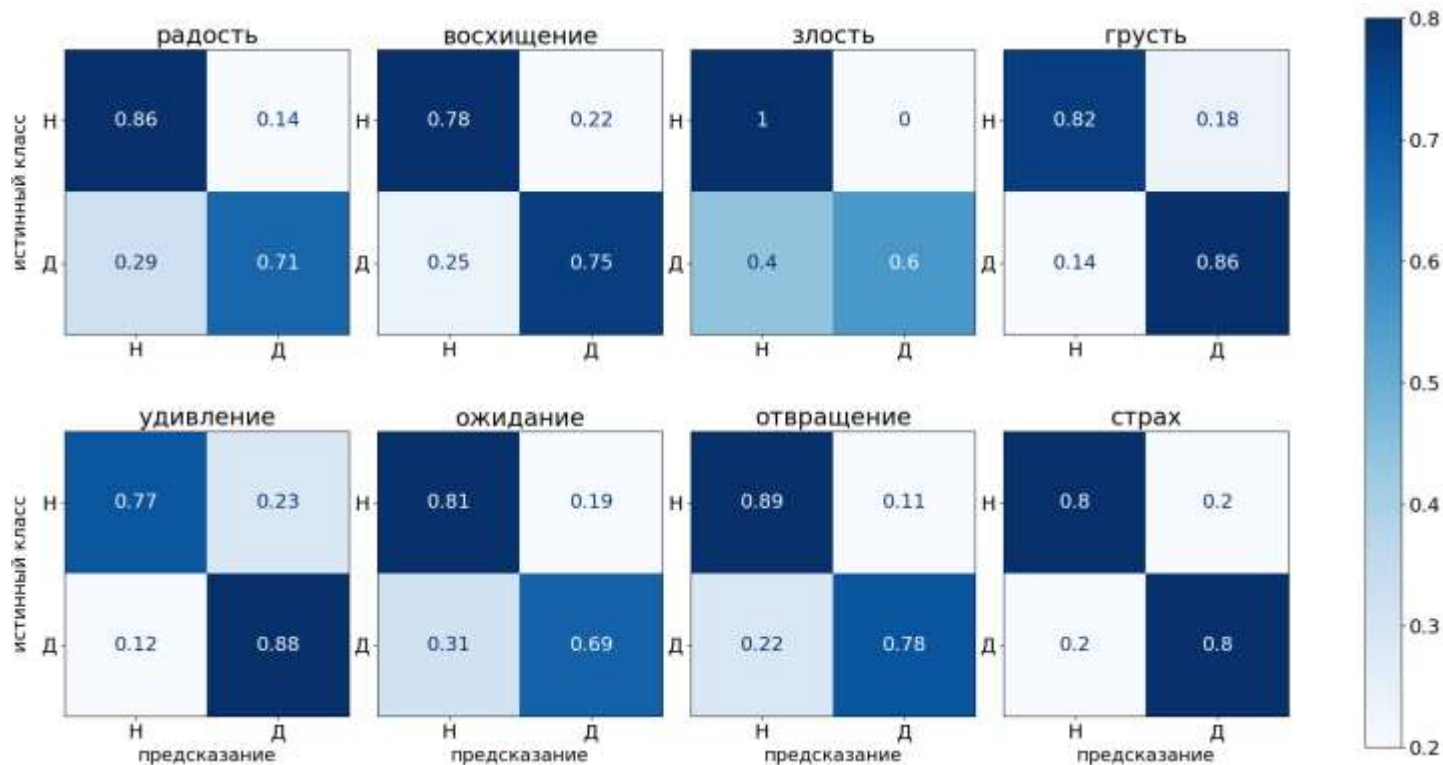
Сравнение различных методов машинного обучения

	Метод				
Метрика	Naive bayes	Naive dict	DP ruBERT	DP ruBERT cedr	ruBERT-tiny cedr
Оценка F1	0.11	0.16	0.41	0.64	0.54
Точность	0.15	0.18	0.56	0.65	0.57

Определение эмоций



Определение эмоций

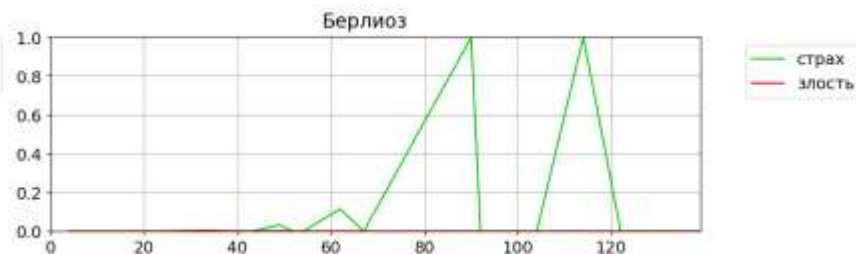
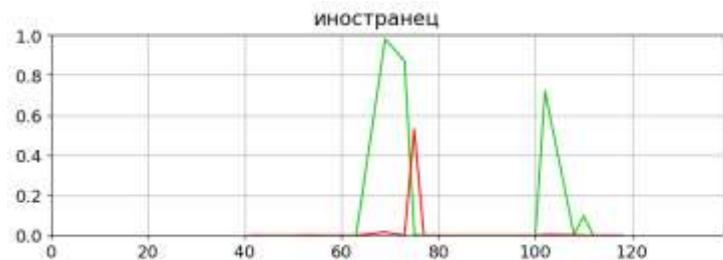
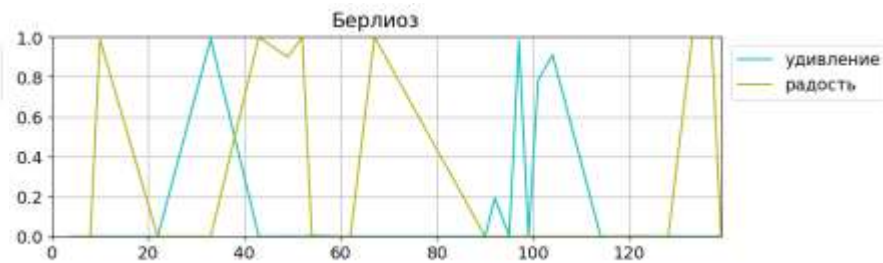
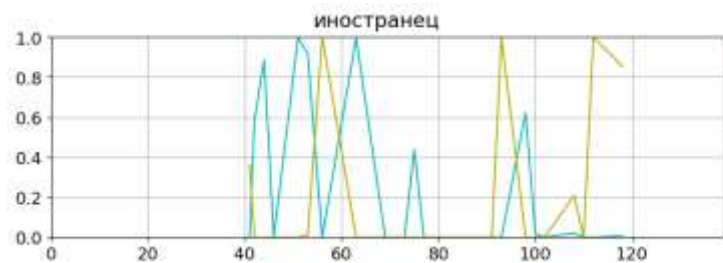


Определение эмоций



Метрики на тестовой выборке								
	Радость	Восхищение	Злость	Грусть	Удивление	Ожидание	Отвращение	Страх
Оценка F1	0.75	0.76	0.67	0.77	0.80	0.74	0.78	0.78
Точность	0.77	0.77	0.75	0.79	0.83	0.77	0.81	0.79
Примеров при обучении	67	32	57	38	62	61	39	59

Готовая программа



Готовая программа



	A	B	C	D	E	F	G	H	I
1		4 – Дайте нарз	нет эмоций	берлиоз	1.4247506e-0	1.217995e-0	2.7771039e-0	7.2011954e-0	2.0687244e-0
2		5 – Нарзану нз	злость	женщина	0.000863816	0.001064185	0.9313898	0.002276112	0.051858265
3		6 – Пиво есть	ожидание	бездомный	0.003404973	0.000344620	0.000104388	0.000619018	0.01814858
4		7 – Пиво прив	нет эмоций	женщина	7.2922717e-0	7.5770953e-0	1.985681e-0	4.0531825e-0	1.1907151e-0
5		8 – А что есть	ожидание	берлиоз	3.991707e-0	2.3821242e-0	2.5377224e-0	6.680884e-0	0.001295207
6		9 – Абрикосов	нет эмоций	женщина	8.191425e-0	8.5886575e-0	2.209254e-0	4.7138994e-0	1.4305116e-0
7		10 – Ну, давай	радость	берлиоз	0.9864103	0.000458337	0.000468560	0.002219610	0.000336849
8		17 – Фу ты чер	страх	редактор	0.000263829	9.116892e-0	0.000552669	0.000217525	0.000137261
9		22 – Нет ни од	нет эмоций	берлиоз	8.0773736e-0	8.605686e-0	2.2293964e-0	4.7180006e-0	1.4386903e-0
10		33 – Ты, Иван,	удивление	берлиоз	0.000127185	0.000220348	0.004969532	0.000271927	0.99084455
11		36 – Извините	удивление		1.8640836e-0	3.9700877e-0	2.9995217e-0	3.2546275e-0	0.96725804
12		41 – Разрешите	нет эмоций	иностранец	0.36044335	0.000550544	0.000130730	0.001251873	9.659151e-03
13		42 – Если я не	удивление	иностранец	1.2059995e-0	3.5793673e-0	2.2331622e-0	1.3214756e-0	0.58388597
14		43 – Нет, вы не	радость	берлиоз	0.9990132	3.7180314e-0	6.583879e-0	0.000177199	6.326427e-0
15		44 – Ах, как ин	удивление	иностранец	1.5638087e-0	4.276424e-0	2.986006e-0	2.1800737e-0	0.88493794
16		46 – А вы согла	нет эмоций	неизвестный	3.0878457e-0	3.0587767e-0	6.2781182e-0	2.1993756e-0	8.581882e-0
17		47 – На все сто	радость		0.99977905	1.5160015e-0	2.6136177e-0	8.437486e-0	2.7466021e-0
18		48 – Изумитель	удивление	бездомный	0.000544700	0.000423242	0.000342494	0.000591112	0.9627348
19		49 – Да, мы не	радость	берлиоз	0.89861935	0.003144345	0.001643397	0.006675275	0.001662723
20		51 – Вы – атеи	удивление	иностранец	2.0344338e-0	3.1792217e-0	2.7773678e-0	8.1523154e-0	0.9972589
21		52 – Да, мы – а	радость	берлиоз	0.99789363	9.631288e-0	4.1028994e-0	0.000518967	3.644346e-03
22		53 – Ох, какая	удивление	иностранец	0.012803487	0.003046781	0.002093744	0.003600264	0.91534823
23		54 – В нашей с	ожидание	берлиоз	2.0568365e-0	4.292772e-0	3.561596e-0	7.260794e-0	0.004147940
24		56 – Позвольте	радость	иностранец	0.9997756	1.5325066e-0	2.641778e-0	8.554094e-0	2.7764813e-0
25		57 – За что это	удивление	бездомный	1.0702015e-0	3.1990923e-0	1.688004e-0	1.3734781e-0	0.6095328
26		58 – За очень	радость		0.9997781	1.5205625e-0	2.6215487e-0	8.46024e-05	2.7539961e-0



Решены следующие задачи:

- Извлечение диалога из текста (выделение слов, указывающих на действующих лиц с точностью 0.95, установление связей между персонажами и высказываниями с точностью 0.81)
- Определение эмоций в предложениях с точностью 0.79

Недостатки полученного решения:

- Невысокая точность определения эмоций
- Вся разметка набора данных для конечного обучения выполнялась одним человеком, а значит, может быть необъективной

Перспективы развития:

- Расширение обучающего набора данных
- Переход на более широкую модель эмоций
- Реализация учета контекста



Спасибо за
внимание!

Дополнительные материалы



Ссылка на данные проекта на github



Дополнительные материалы



класс в обучающем наборе	нет эмоций	24	16	8	19	5	8
	радость	14	39	3	10	2	12
	грусть	5	2	20	5	6	8
	удивление	11	3	3	37	4	15
	страх	11	5	11	11	19	13
	злость	11	7	5	5	5	47
	восхищение	9	14	4	7	4	7
	ожидание	35	6	4	26	2	7
	отвращение	5	9	3	4	3	22
		none	joy	sadness	surprise	fear	anger
		класс в cedr					

Дополнительные материалы



Примеры применения алгоритма определения эмоций к отдельным предложениям

