

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ Н.Э. БАУМАНА
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)

Факультет «Информатика и системы управления»

Кафедра «Программное обеспечение ЭВМ и информационные технологии»

Выпускная квалификационная работа магистра

Метод распознавания паттернов суицидального поведения человека по текстовым сообщениям

Студент: Якуба Дмитрий Васильевич

Группа: ИУ7-43М

Руководитель: Строганов Юрий Владимирович

Цель и задачи работы

Цель — разработать и реализовать метод распознавания паттернов суицидального поведения человека по текстовым сообщениям.

Задачи:

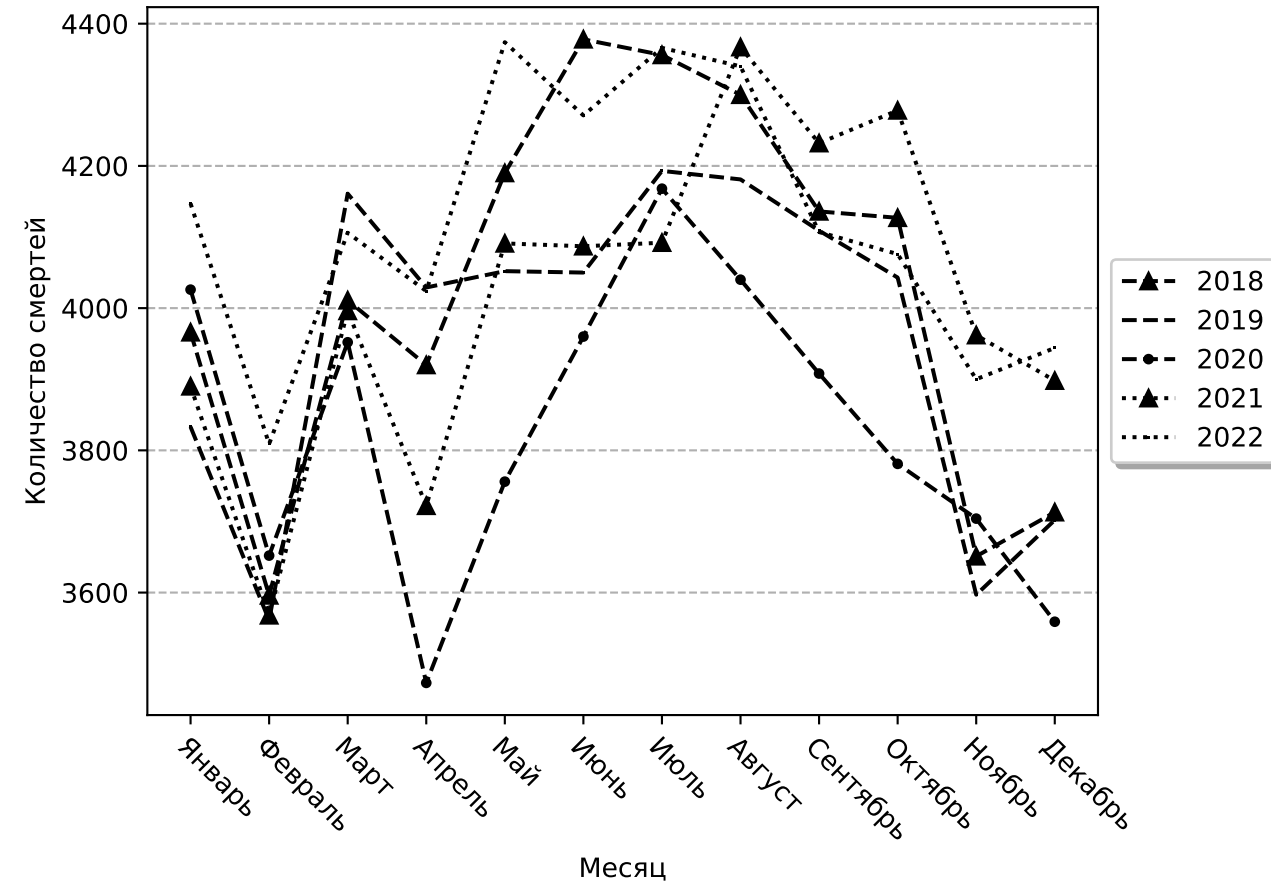
- проанализировать действия и характеристики, позволяющие распознать паттерны суицидального поведения, и классифицировать их;
- разработать метод распознавания паттернов суицидального поведения;
- реализовать разработанный метод;
- сравнить задействованные в методе алгоритмы машинного обучения и дать рекомендации о применимости реализованного метода.

Суицидальная статистика

Каждый год в мире совершается 703 тысячи самоубийств.

В 2021 году уровень самоубийств среди мужчин в 4 раза выше, чем среди женщин.

Самый высокий уровень самоубийств наблюдается у людей старше 85 лет.



Разработки, задействованные в сфере суицидологии

Наименование	Достоинства	Недостатки
Facebook* AI	Автоматизированное средство	Закрытый исходный код
	Обработка большого объема данных	Объект коммерческой тайны
		Недоступно в РФ
Dostoesky	Открытый исходный код	Не предназначен для решения поставленной задачи
ChatGPT	Высокая точность	Дорогостоящая модель
	Широкая сфера применения	Недоступно в РФ



Dostoevsky



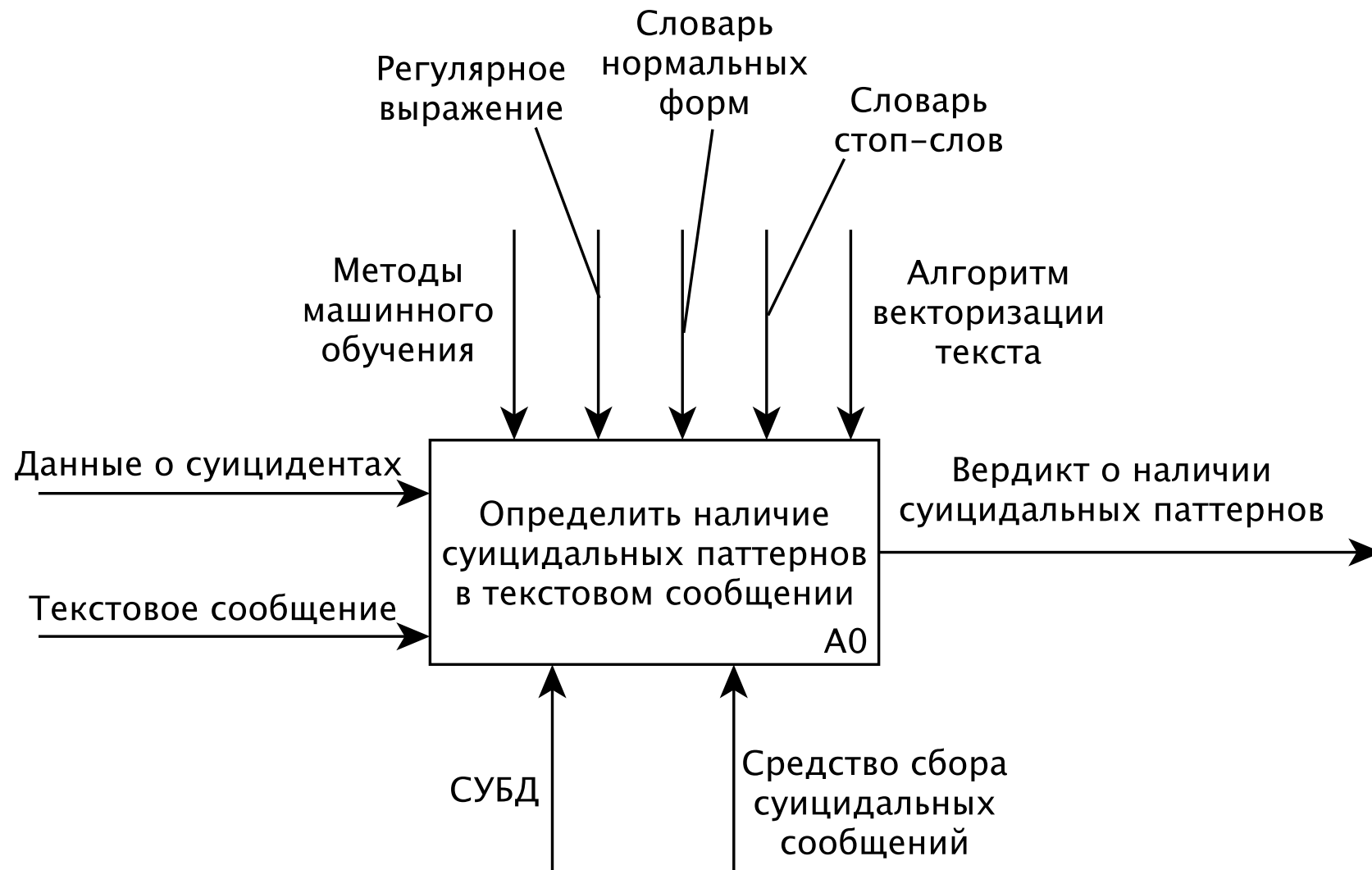
ChatGPT

* сервис, принадлежащий Meta, признанной в РФ экстремистской организацией, деятельность которой запрещена на территории РФ

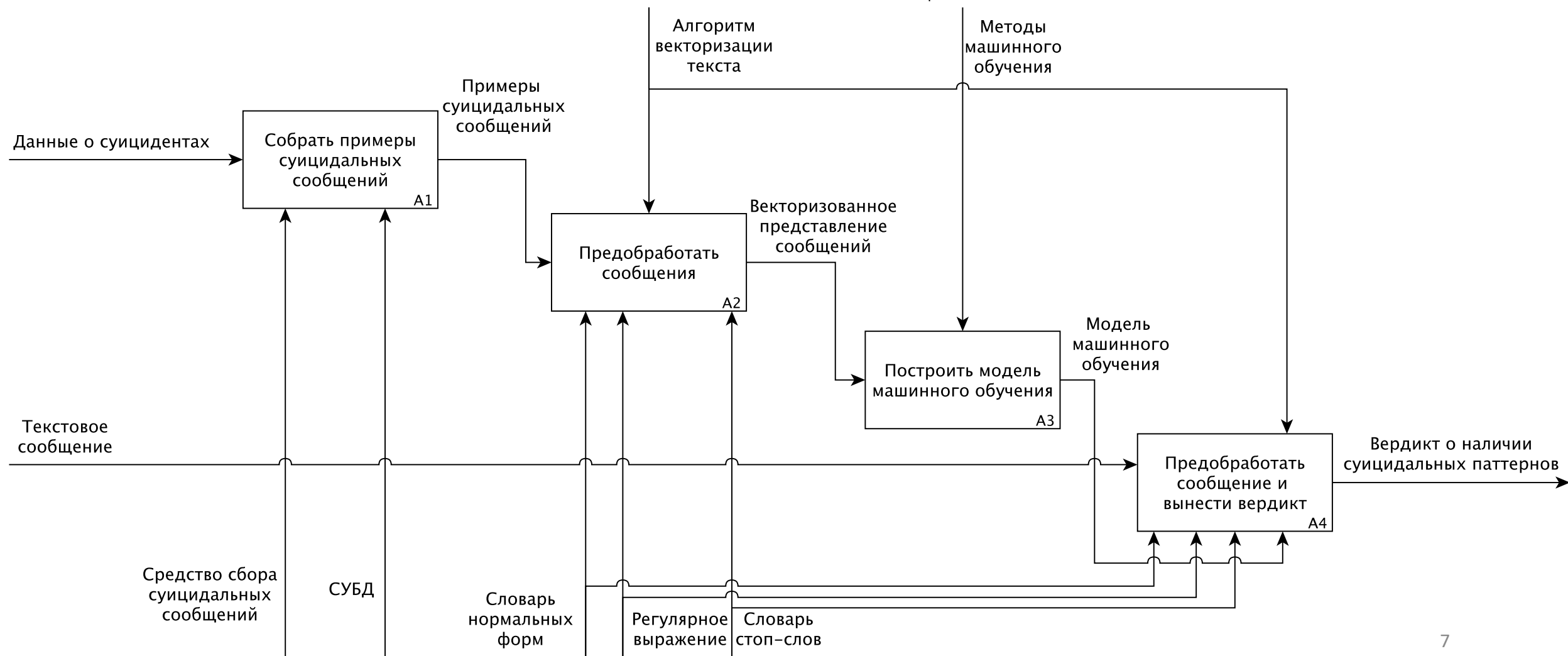
Форматы описания признаков

Признаки	Данные
Аудиальные	аудиофайл
	текстовая расшифровка речи
	эмоциональная карта, аудиофайл / текстовая расшифровка
Текстовые	текстовое сообщение
	текстовое сообщение, эмоциональная карта
Пространственно-временные	дата написания сообщения
	место дислокации автора, дата написания сообщения
Визуальные	видеоряд действий пользователя
	видеоряд действий пользователя, мониторинг контекста происходящего
Физиологические	данные мониторинга уровня стресса
	данные мониторинга уровня кортизола в крови
	данные мониторинга состояния здоровья человека
Биологические	пол пользователя
	возраст пользователя

Метод распознавания паттернов суицидального поведения человека по текстовым сообщениям



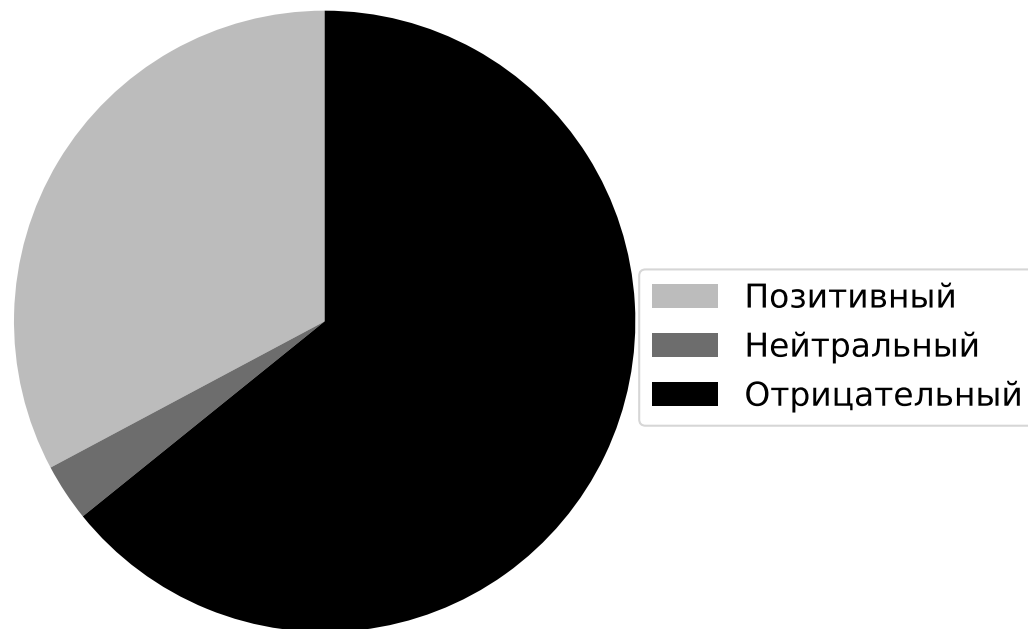
Метод распознавания паттернов суицидального поведения человека по текстовым сообщениям



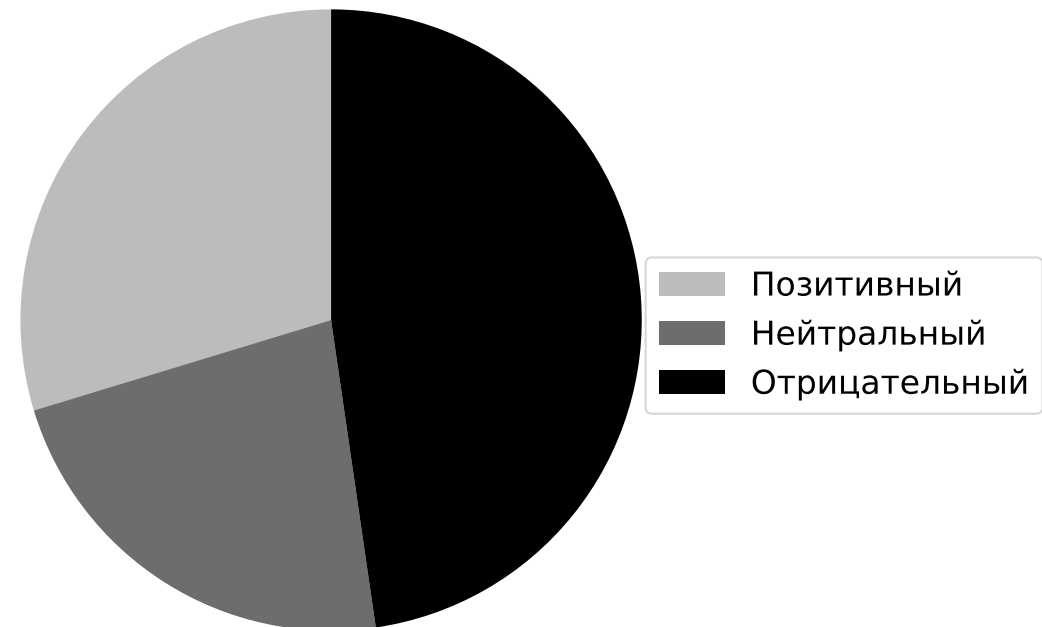
Анализ собранных данных

- Собрано 1 000 суицидальных сообщений;
- Добавлено 1 000 несуицидальных сообщений из датасета обнаружения пресуицидальных сигналов.

суицидальные сообщения



несуицидальные сообщения



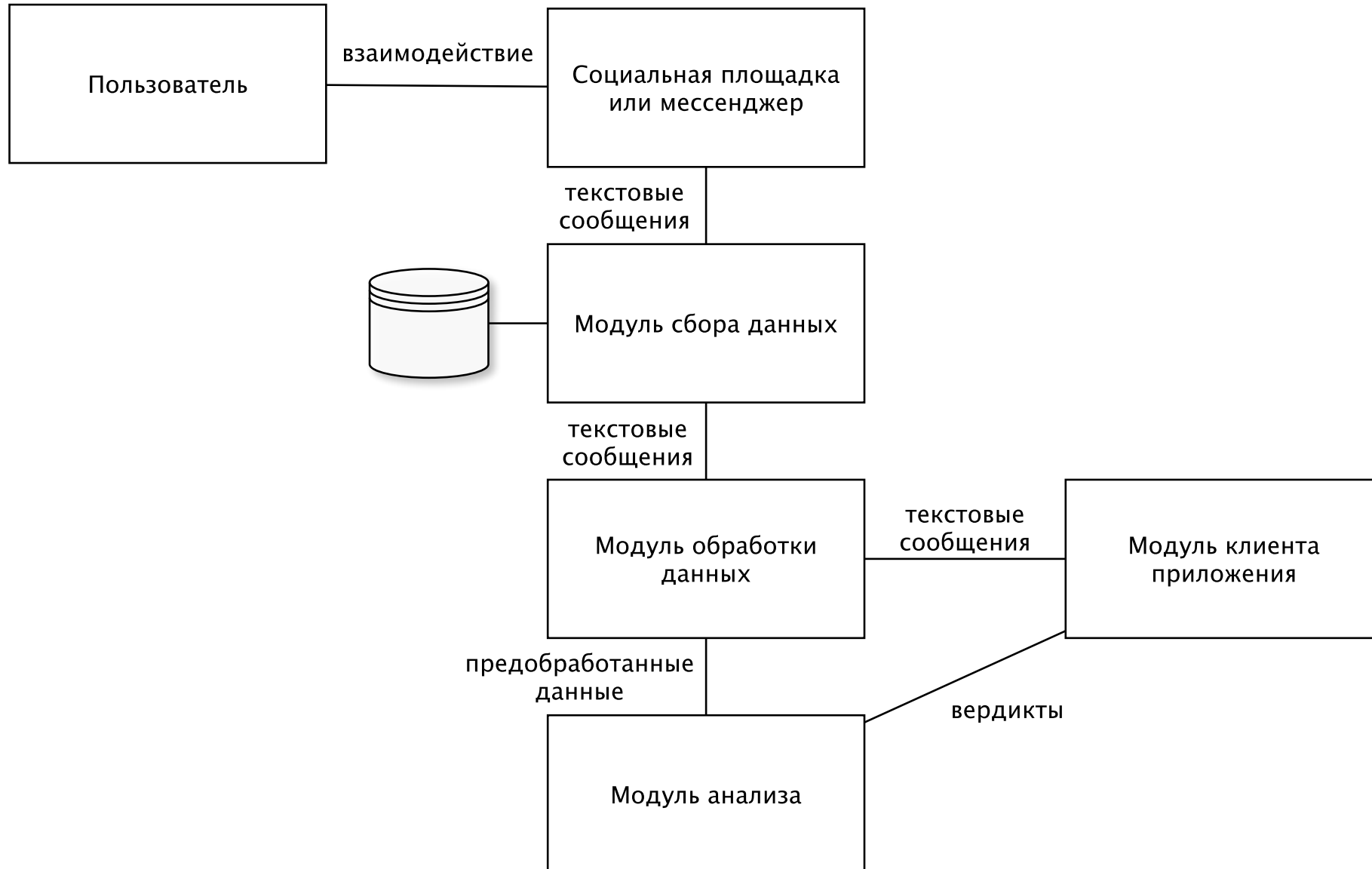
Анализ собранных данных, суицидальные сообщения



Анализ собранных данных, несуицидальные сообщения



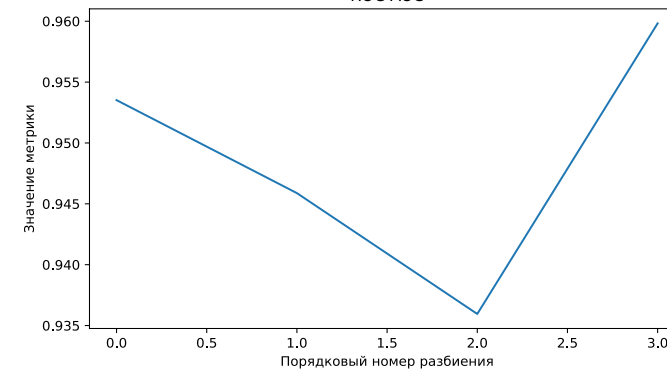
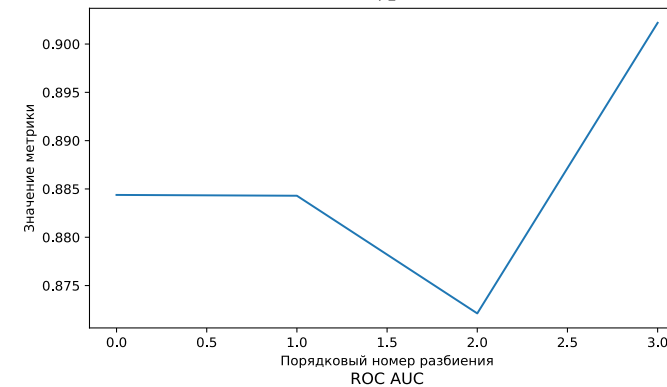
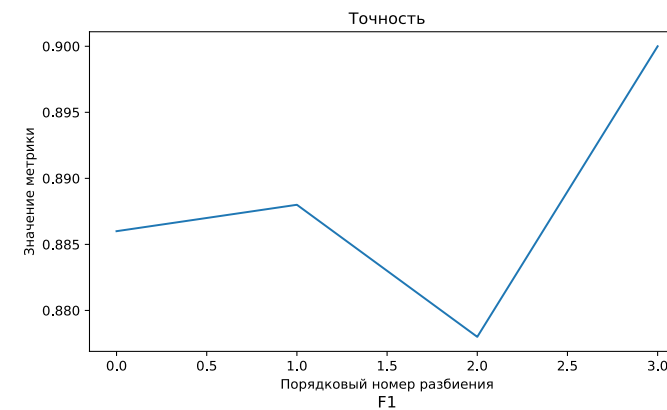
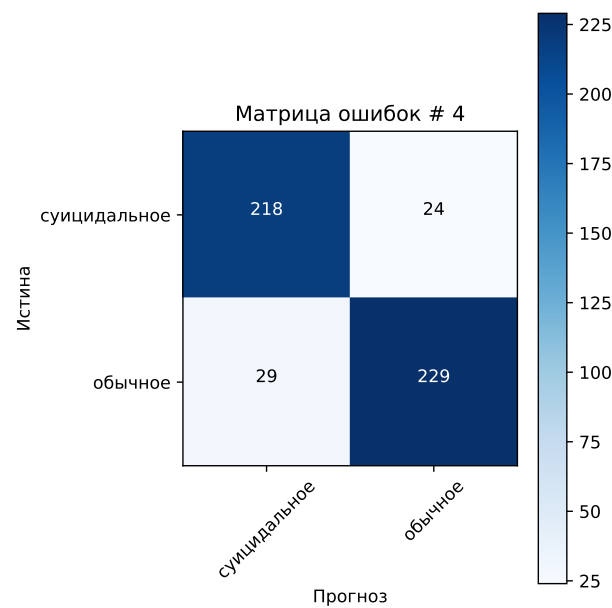
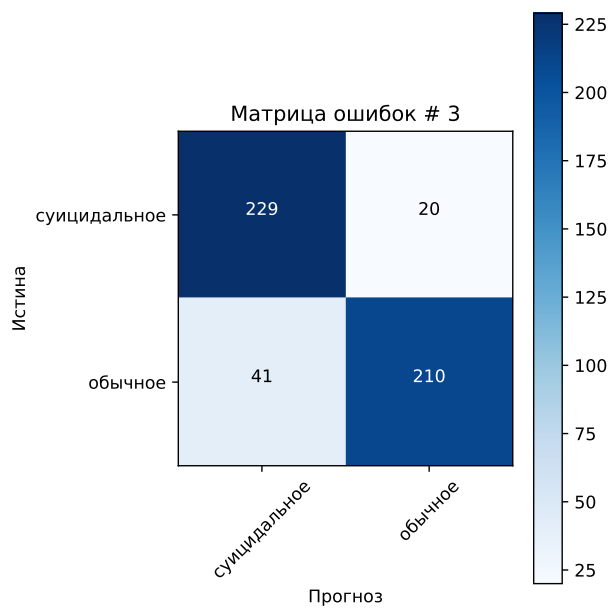
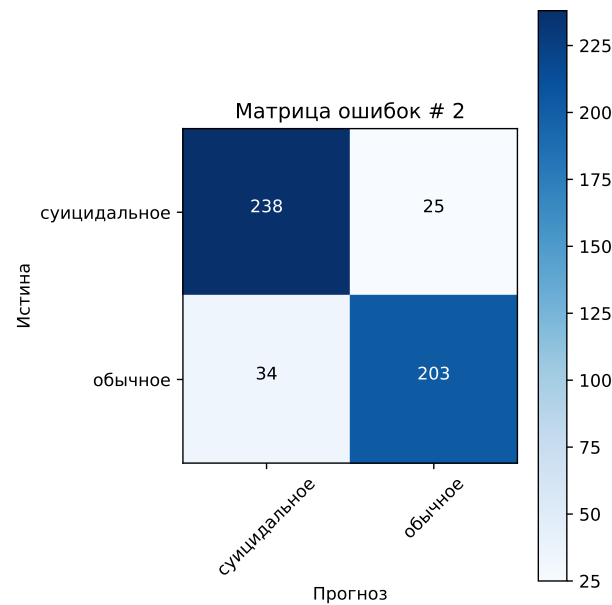
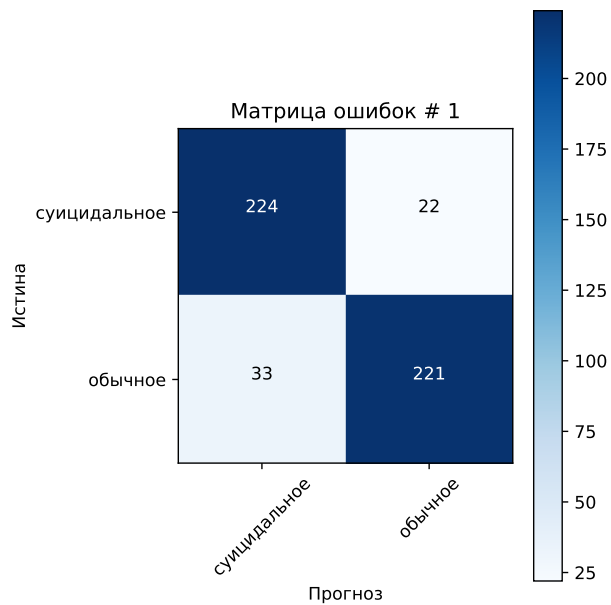
Схема программного обеспечения



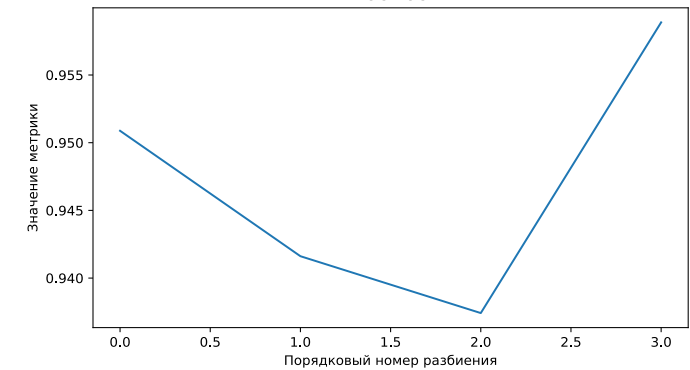
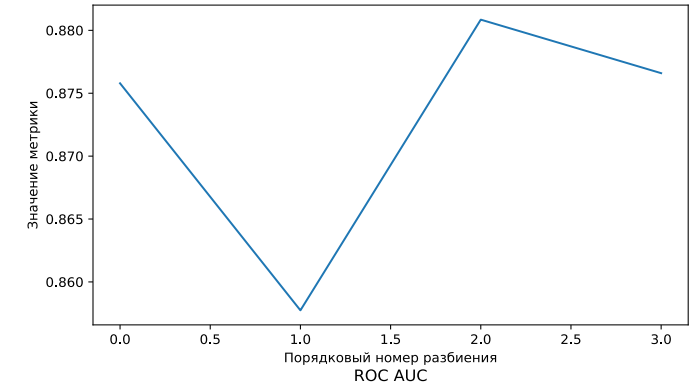
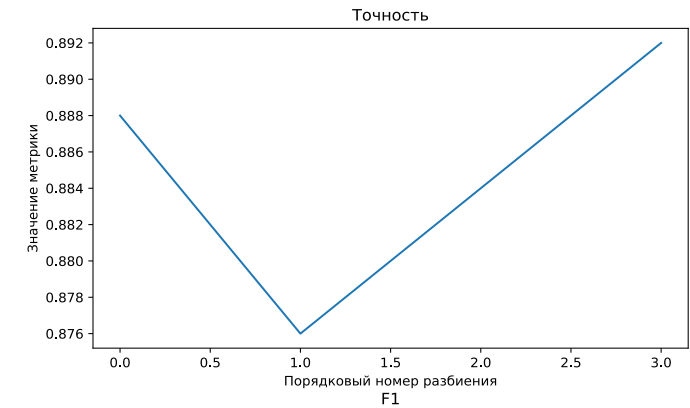
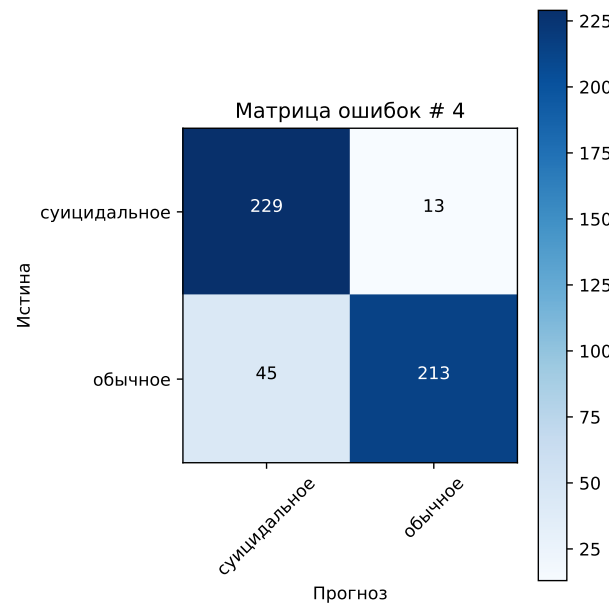
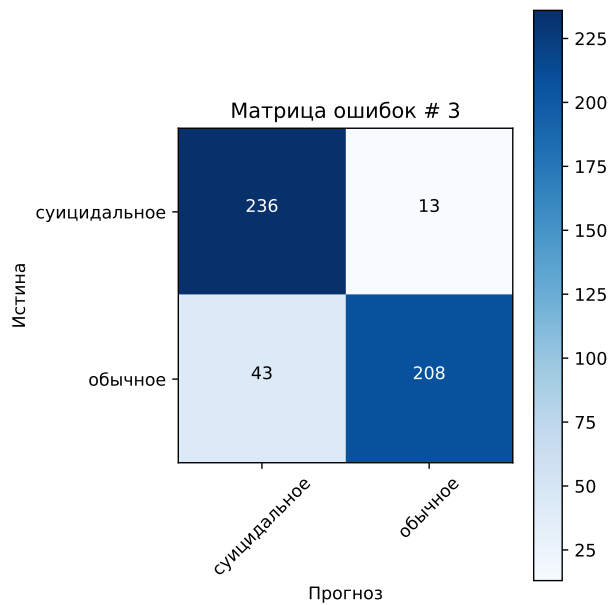
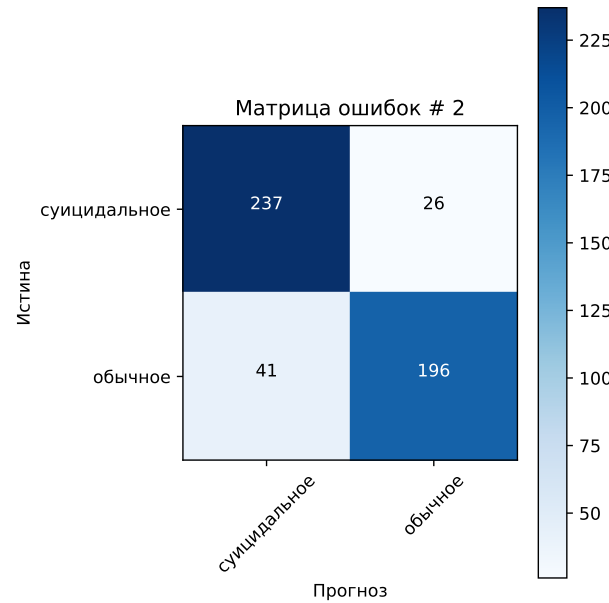
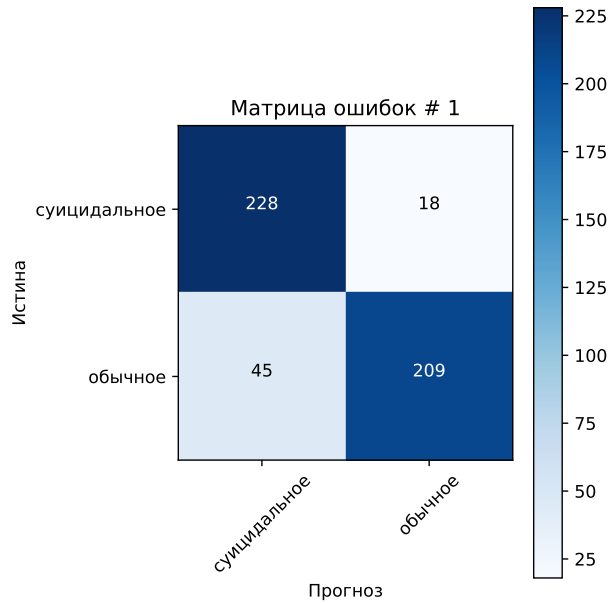
Результаты исследования

Алгоритм	Векторизация	Точность	F1-мера	ROC-AUC	
Градиентный бустинг	“Мешок слов”	<i>0.853</i>	<i>0.844</i>	<i>0.916</i>	
	BERT	<i>0.870</i>	<i>0.854</i>	<i>0.929</i>	
Случайный лес	“Мешок слов”	0.885	0.873	0.947	(2)
	BERT	0.888	0.886	0.949	(1)
Метод опорных векторов	“Мешок слов”	<i>0.847</i>	<i>0.845</i>	<i>0.915</i>	
	BERT	<i>0.862</i>	<i>0.861</i>	<i>0.924</i>	
К-ближайших соседей	“Мешок слов”	<i>0.752</i>	<i>0.704</i>	<i>0.854</i>	
	BERT	<i>0.758</i>	<i>0.737</i>	<i>0.842</i>	
Логистическая регрессия	“Мешок слов”	<i>0.867</i>	<i>0.859</i>	<i>0.937</i>	
	BERT	0.874	0.869	0.942	(3)
Перцептрон	“Мешок слов”	<i>0.838</i>	<i>0.841</i>	<i>0.912</i>	
	BERT	<i>0.853</i>	<i>0.857</i>	<i>0.931</i>	

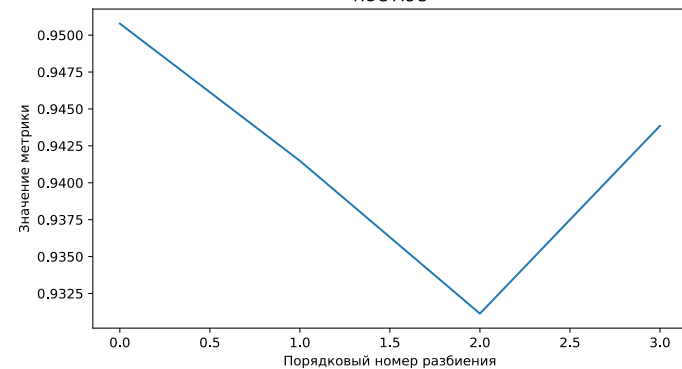
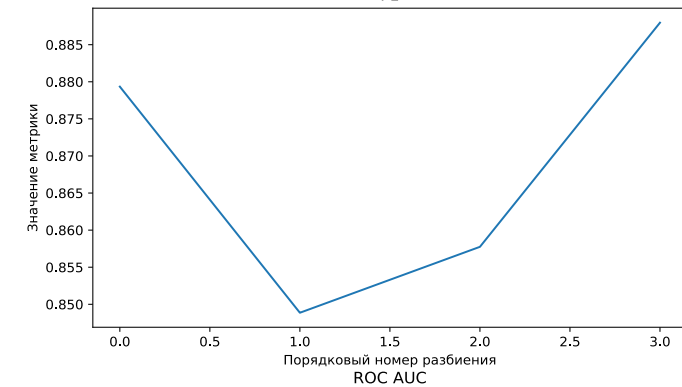
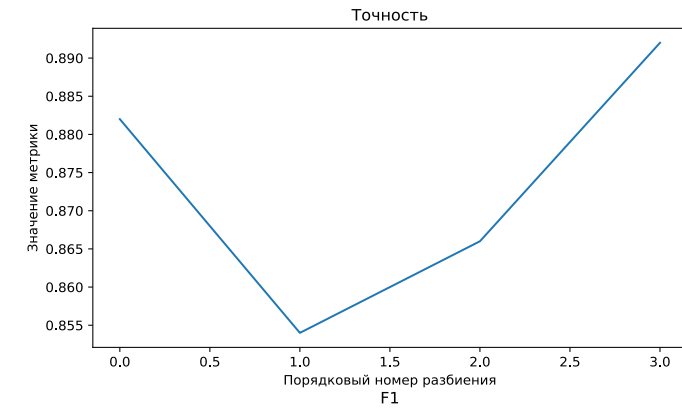
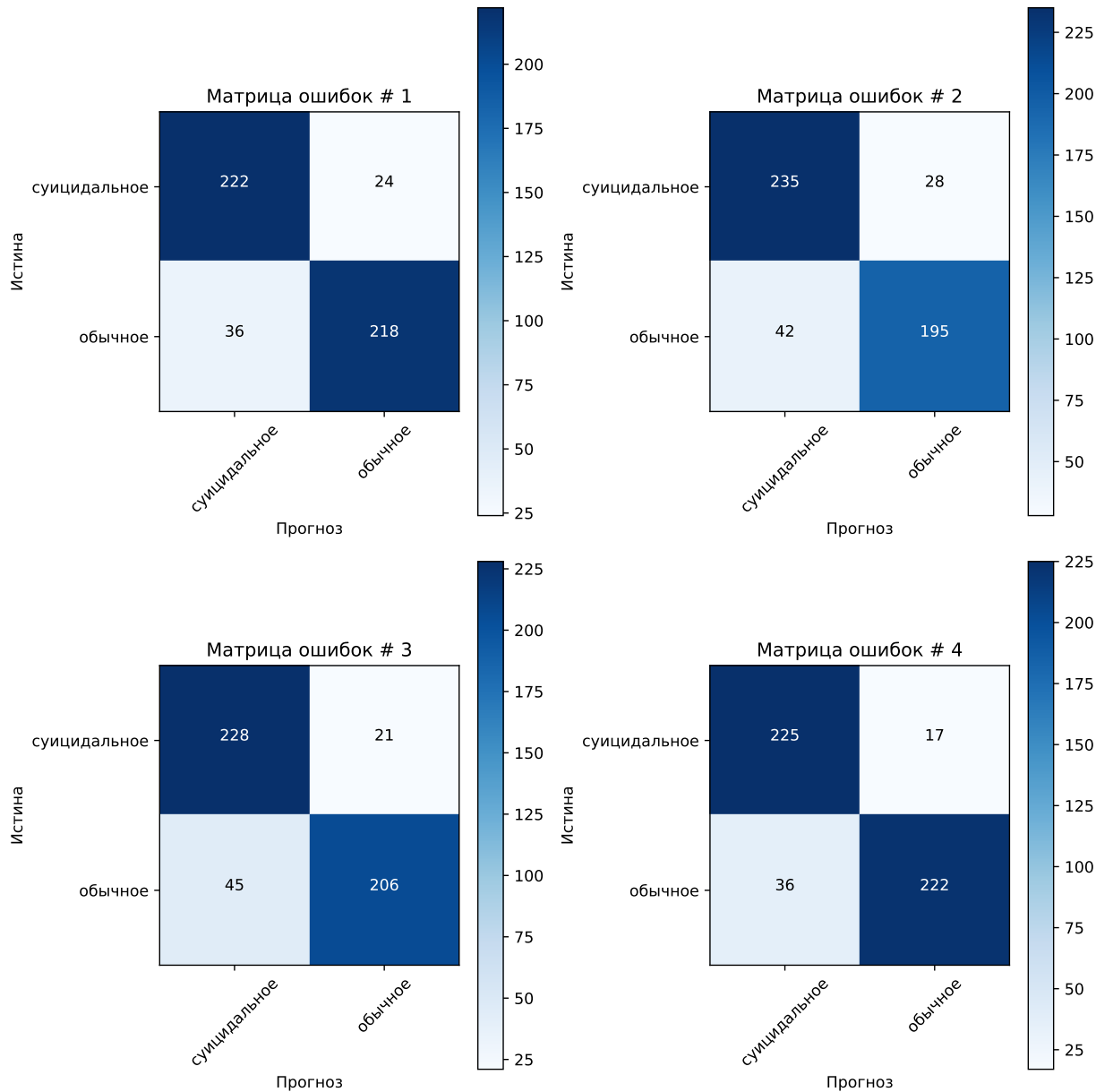
Случайный лес, BERT



Случайный лес, «мешок слов»



Логистическая регрессия, BERT



Результаты исследования

Случайный лес, «Мешок слов» – F1-мера на $\approx 1.4\%$ ниже, чем (1).

Логистическая регрессия, BERT – точность и F1-мера на $\approx 2\%$ ниже, чем (1).

Алгоритм	Векторизация	Точность	F1-мера	ROC-AUC	
Случайный лес	“Мешок слов”	<i>0.885</i>	<i>0.873</i>	<i>0.947</i>	(2)
	BERT	<i>0.888</i>	<i>0.886</i>	<i>0.949</i>	(1)
Логистическая регрессия	BERT	<i>0.874</i>	<i>0.869</i>	<i>0.942</i>	(3)

Заключение

Был разработан и реализован метод распознавания паттернов суицидального поведения человека по текстовым сообщениям.

Были решены следующие задачи:

- проанализированы действия и характеристики, позволяющие распознать паттерны суицидального поведения, и классифицировать их;
- разработан метод распознавания паттернов суицидального поведения;
- реализован разработанный метод;
- сравнены задействованные в методе алгоритмы машинного обучения и даны рекомендации о применимости реализованного метода.

Дальнейшее развитие

- Исследование эффективности использования ансамблевого подхода в решении задачи;
- Исследование применимости алгоритмов нечеткой кластеризации для распознавания суицидальных паттернов;
- Расширение датасета, использование дополнительных признаков;
- Реализовать средство автоматизированного анализа сообщений пользователей в социальных сетях;
- Внедрение в рабочий процесс.