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

Выпускная квалификационная работа бакалавра

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:

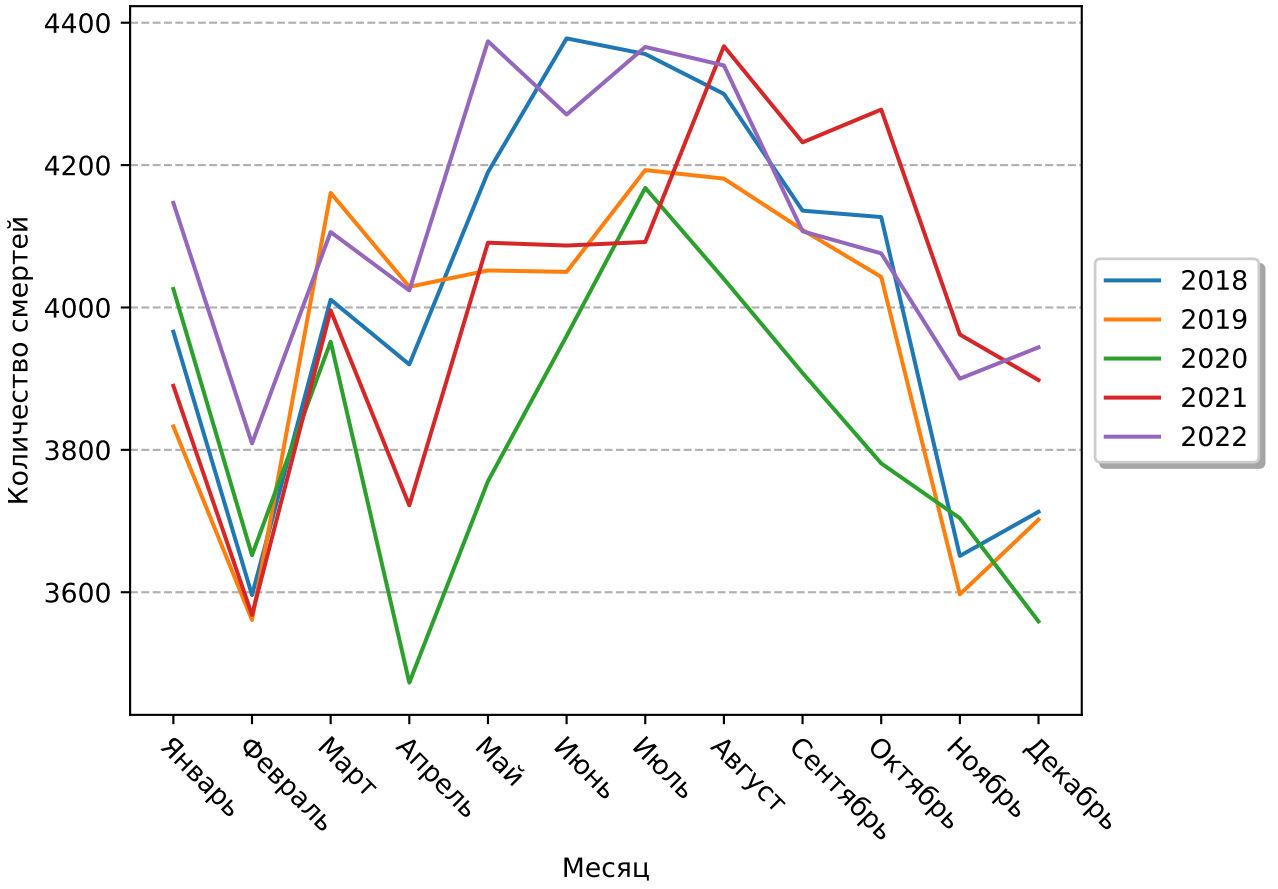
- Провести анализ действий и характеристик, позволяющих распознать паттерны суицидального поведения;
- Классифицировать признаки паттернов суицидального поведения человека;
- Определить метод сбора данных суицидального поведения;
- Разработать метод распознавания паттернов суицидального поведения;
- Реализовать разработанный метод;
- Провести сравнительное исследование задействованных в методе алгоритмов машинного обучения;
- 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.

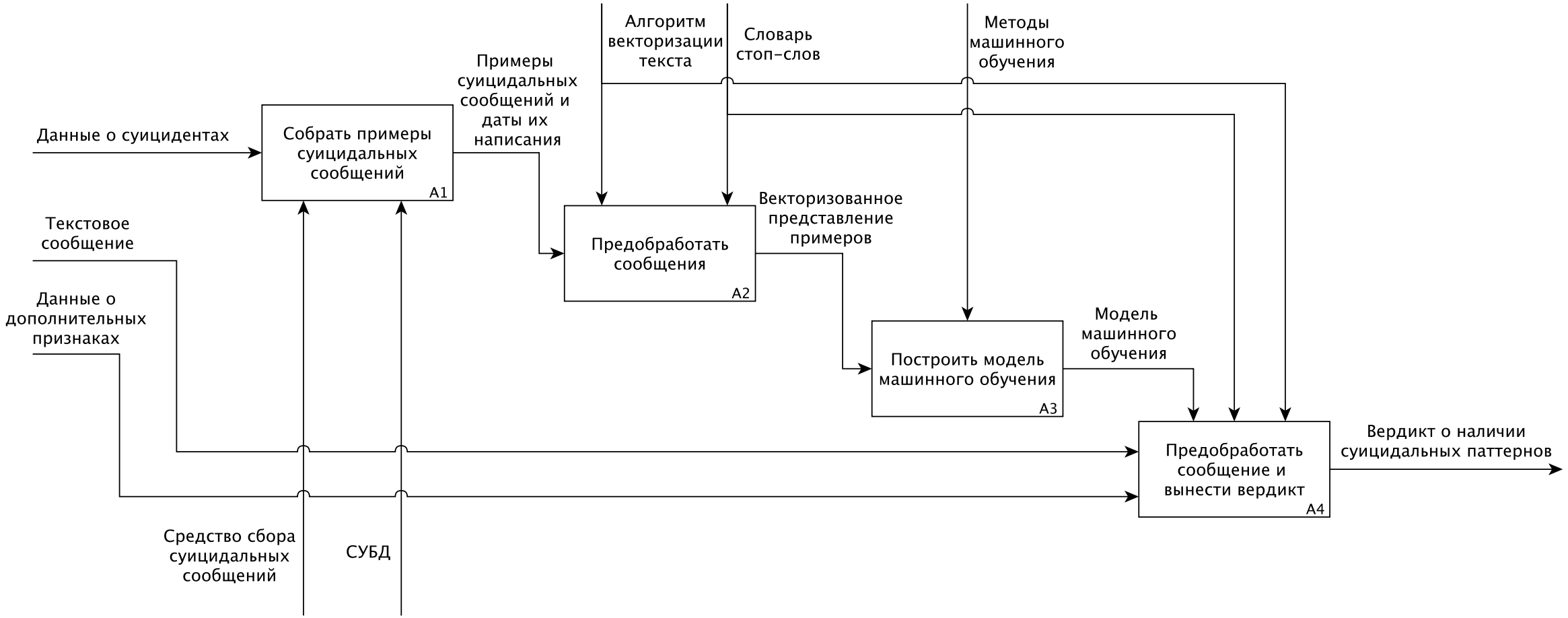
Самый высокий уровень самоубийств наблюдается у людей старше 85 лет.



Форматы описания признаков и методов их обработки

Признаки	Данные	Методы обработки
Аудиальные	аудиофайл	распознавание речи, обработка и анализ текстовых сообщений, анализ характеристик голоса
	текстовая расшифровка речи	обработка и анализ текстовых сообщений
	эмоциональная карта, аудиофайл / текстовая расшифровка	сопоставление эмоциональной карты смысловой нагрузке речи
Текстовые	текстовое сообщение	обработка и анализ текстовых сообщений с использованием методов машинного обучения
	текстовое сообщение, эмоциональная карта	оптимизация модели машинного обучения с использованием эмоциональной карты
Пространственно-временные	дата написания сообщения	соотнесение дат действий пользователя сезонности депрессии
	место дислокации автора, дата написания сообщения	соотнесение контекста происходящего в регионе пользователя его действиям
Визуальные	видеоряд действий пользователя	распознавание эмоций
	видеоряд действий пользователя, мониторинг контекста происходящего	анализ реакций индивидуума на внешние раздражители и жизненные ситуации
Физиологические	данные мониторинга уровня стресса	анализ состояния организма человека и его подверженности стрессам
	данные мониторинга уровня кортизола в крови	
	данные мониторинга состояния здоровья человека	
Биологические	пол пользователя	оптимизация модели машинного обучения с использованием пола пользователя
	возраст пользователя	оптимизация модели машинного обучения с использованием возрастной группы пользователя

Метод распознавания паттернов суицидального поведения человека по текстовым сообщениям

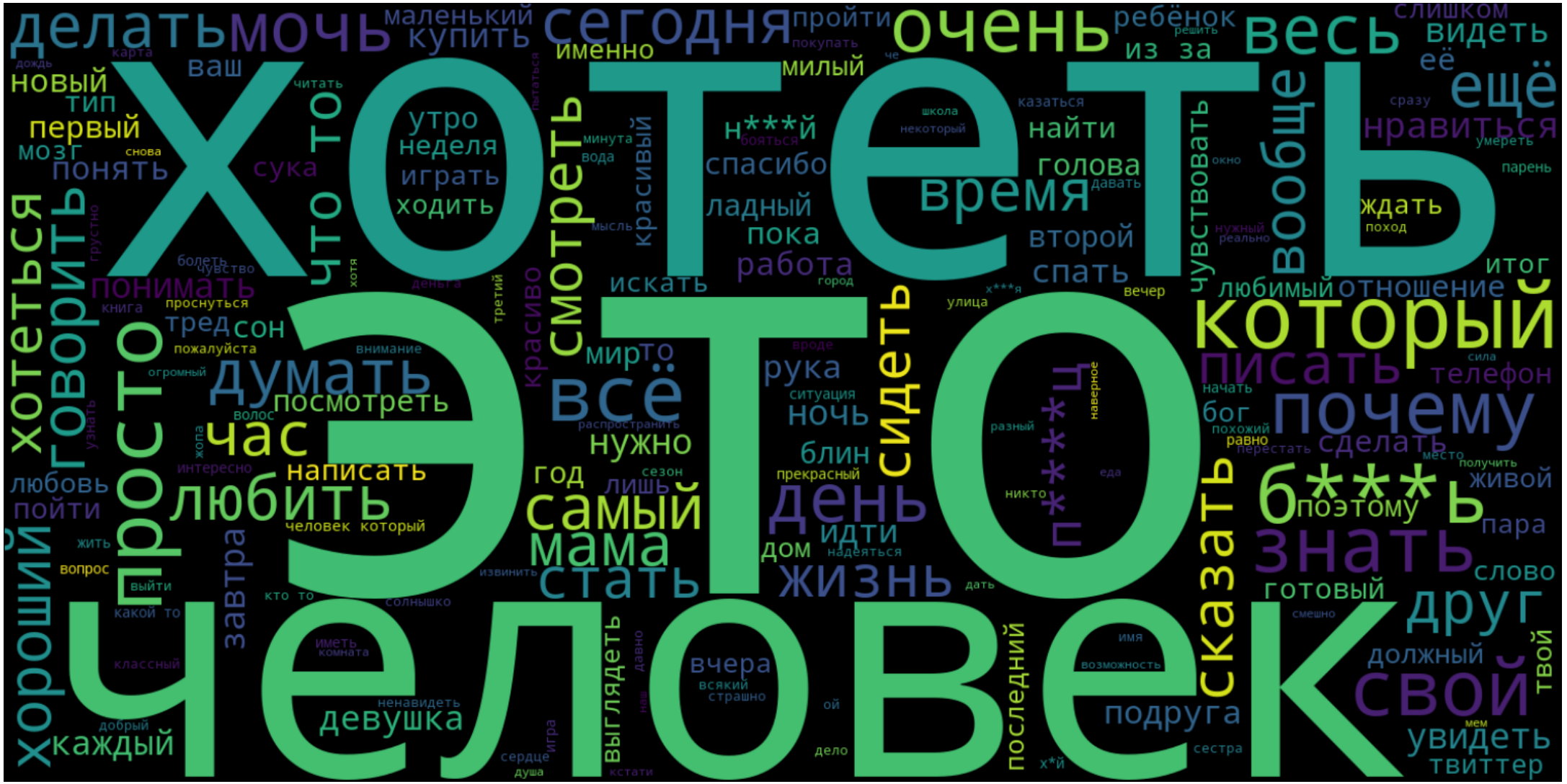


Анализ собранных данных

Всего было собрано 1 000 суицидальных сообщений. К собранным данным было добавлено еще 1 000 несуицидальных сообщений из датасета обнаружения пресуицидальных сигналов.

Тут приведем какую-нибудь частотность и прочее, что еще не добавили в бумажку

Анализ собранных данных, несуицидальные сообщения



Анализ собранных данных, суицидальные сообщения

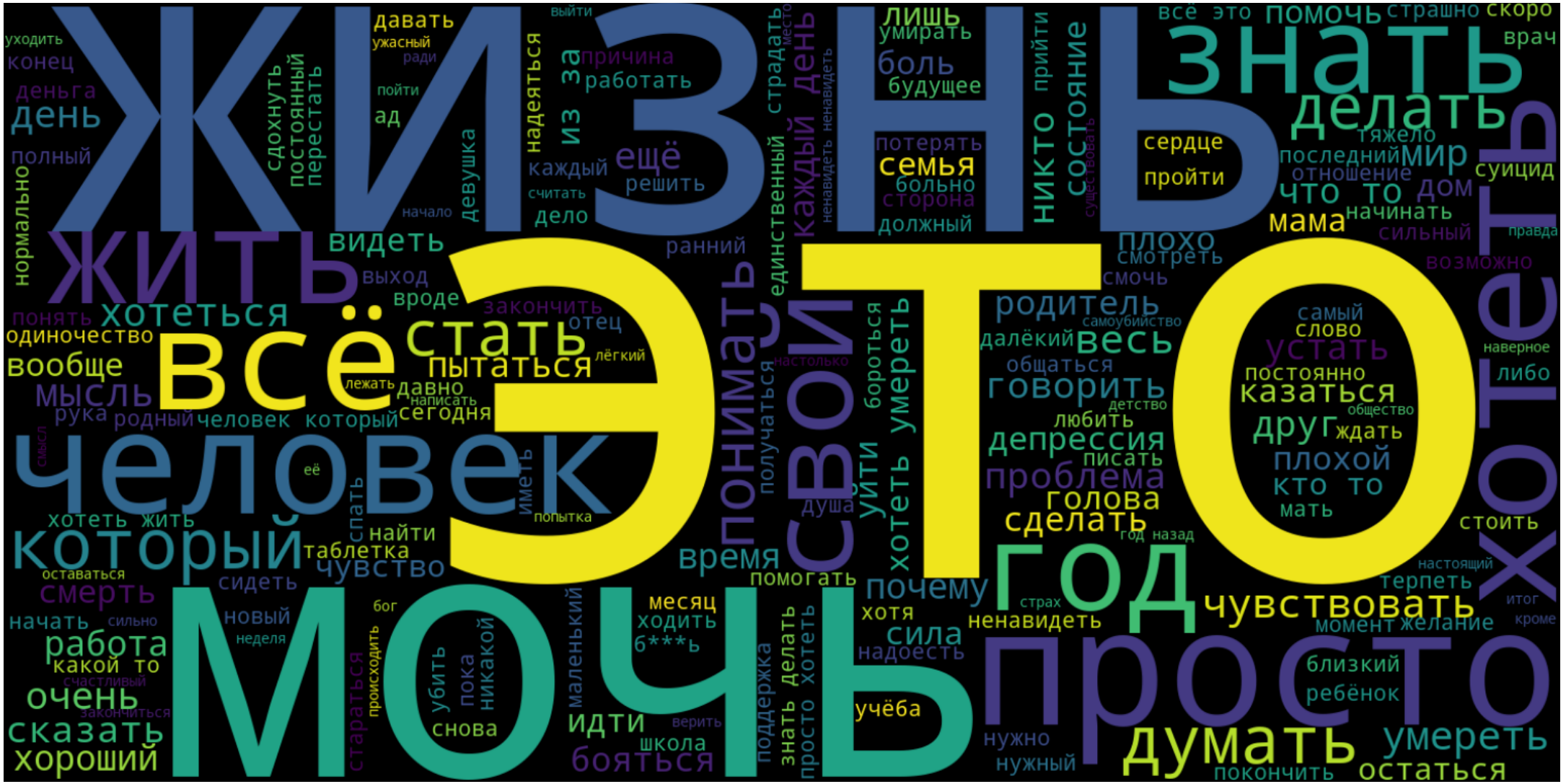
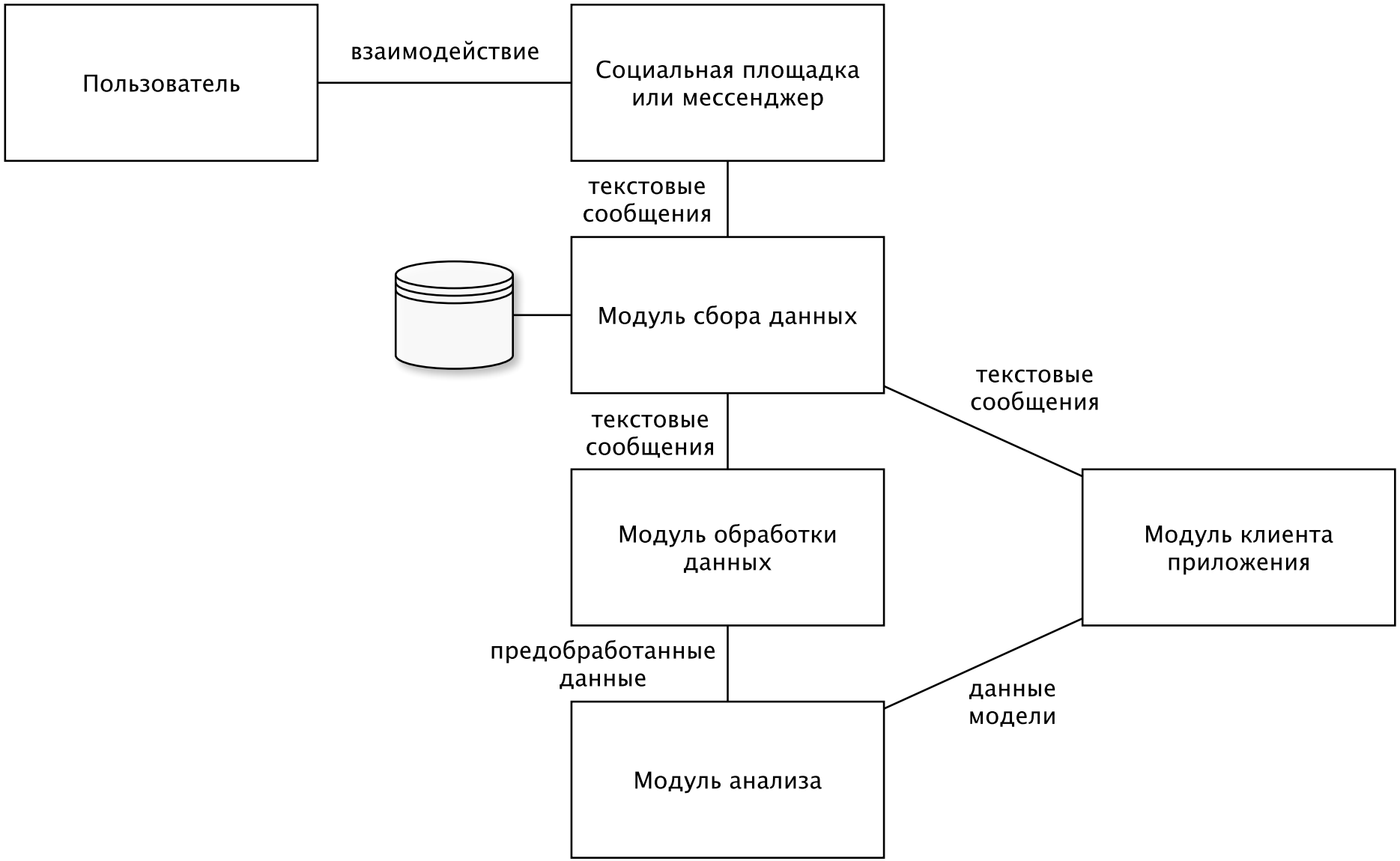


Схема программного обеспечения



Исследование применимости моделей

Рассматриваемые алгоритмы:

- Градиентный бустинг
- Случайный лес
- Метод опорных векторов
- К-ближайших соседей
- Логистическая регрессия
- Перцептрон

Разбиение данных на 4 части, 1 из которых используется в качестве тестовой. Для каждого разбиения строится матрица ошибок, для каждой модели приводится график значений исследуемых метрик.

Сравниваемые методы векторизации:

- ”Мешок слов”
- BERT

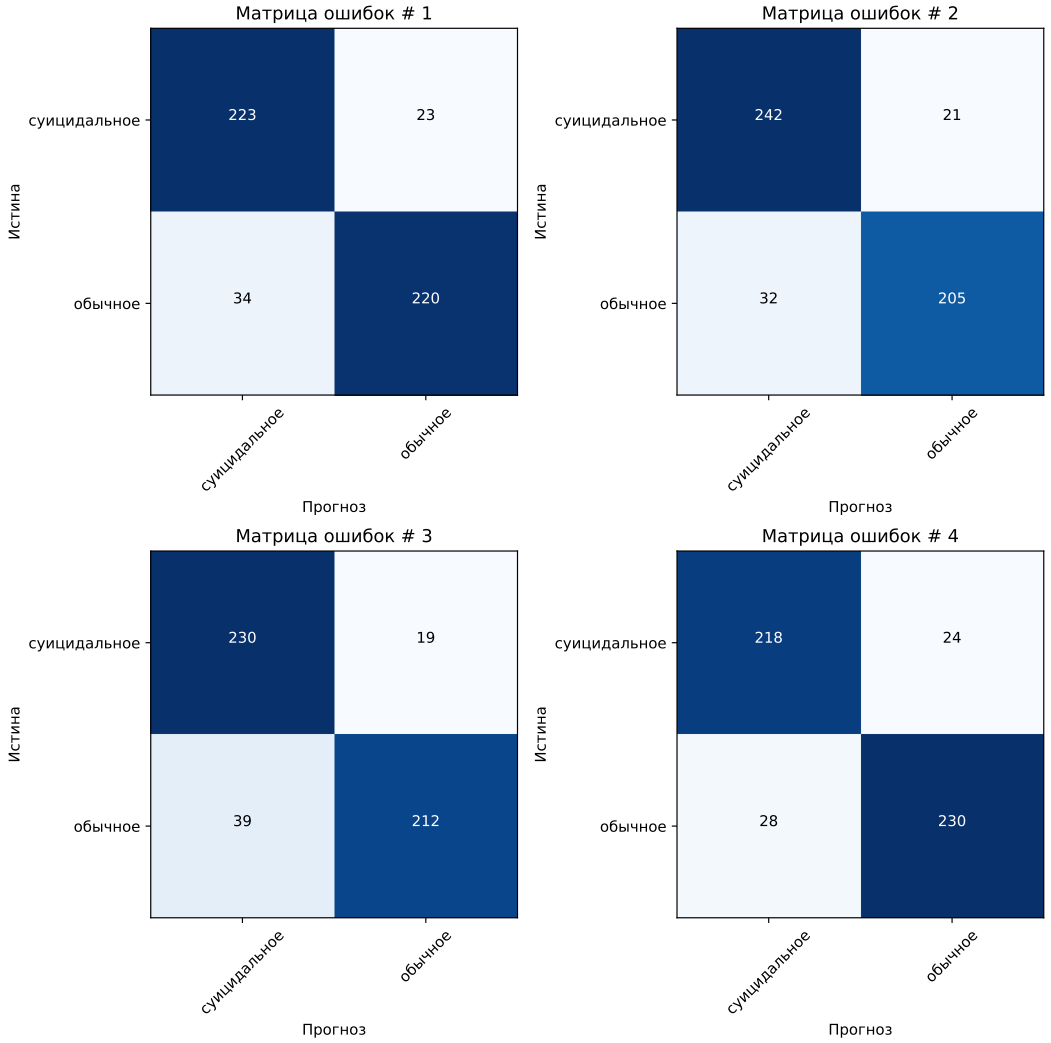
Исследуемые метрики:

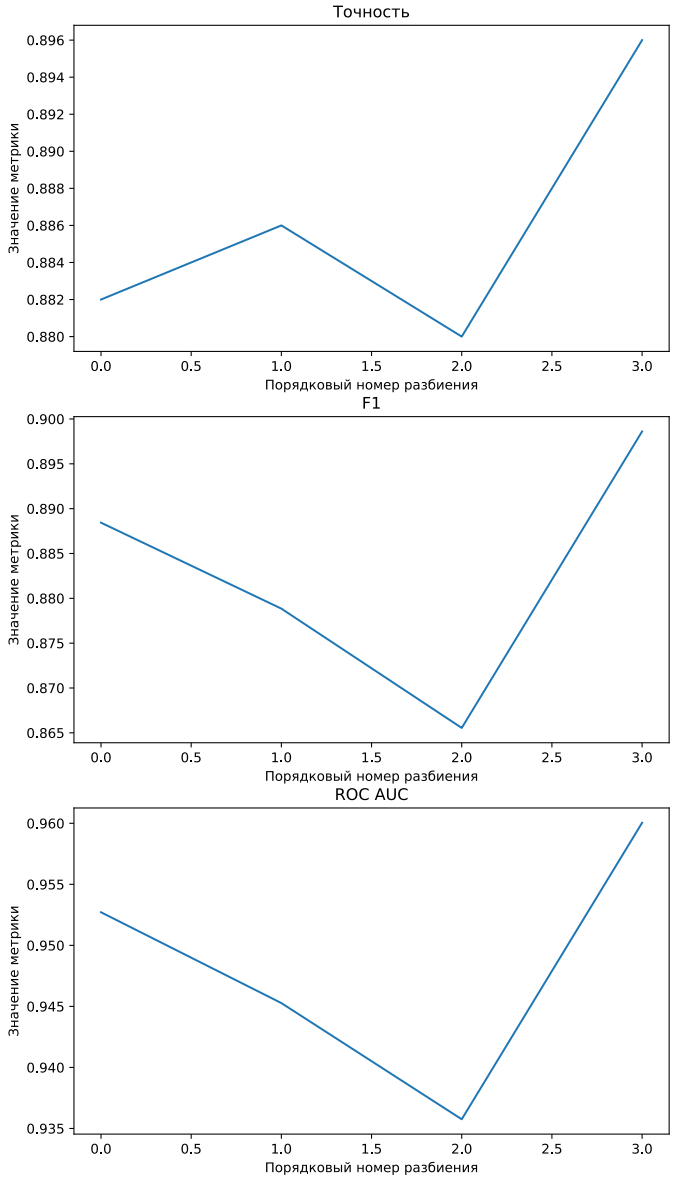
- Точность
- F1-мера
- ROC-AUC

Результаты исследования

Алгоритм	Векторизация	Точность	F1-мера	ROC-AUC
Градиентный бустинг	“Мешок слов”	*0.853*	*0.843*	*0.919*
	BERT	*0.862*	*0.852*	*0.929*
Случайный лес	“Мешок слов”	*0.878*	*0.875*	*0.947*
	BERT	***0.888***	***0.887***	***0.948***
Метод опорных векторов	“Мешок слов”	*0.847*	*0.845*	*0.915*
	BERT	*0.862*	*0.861*	*0.925*
K-ближайших соседей	“Мешок слов”	*0.752*	*0.704*	*0.854*
	BERT	*0.758*	*0.737*	*0.842*
Логистическая регрессия	“Мешок слов”	*0.867*	*0.859*	*0.937*
	BERT	*0.874*	*0.869*	*0.942*
Перцептрон	“Мешок слов”	*0.838*	*0.841*	*0.912*
	BERT	*0.853*	*0.857*	*0.931*

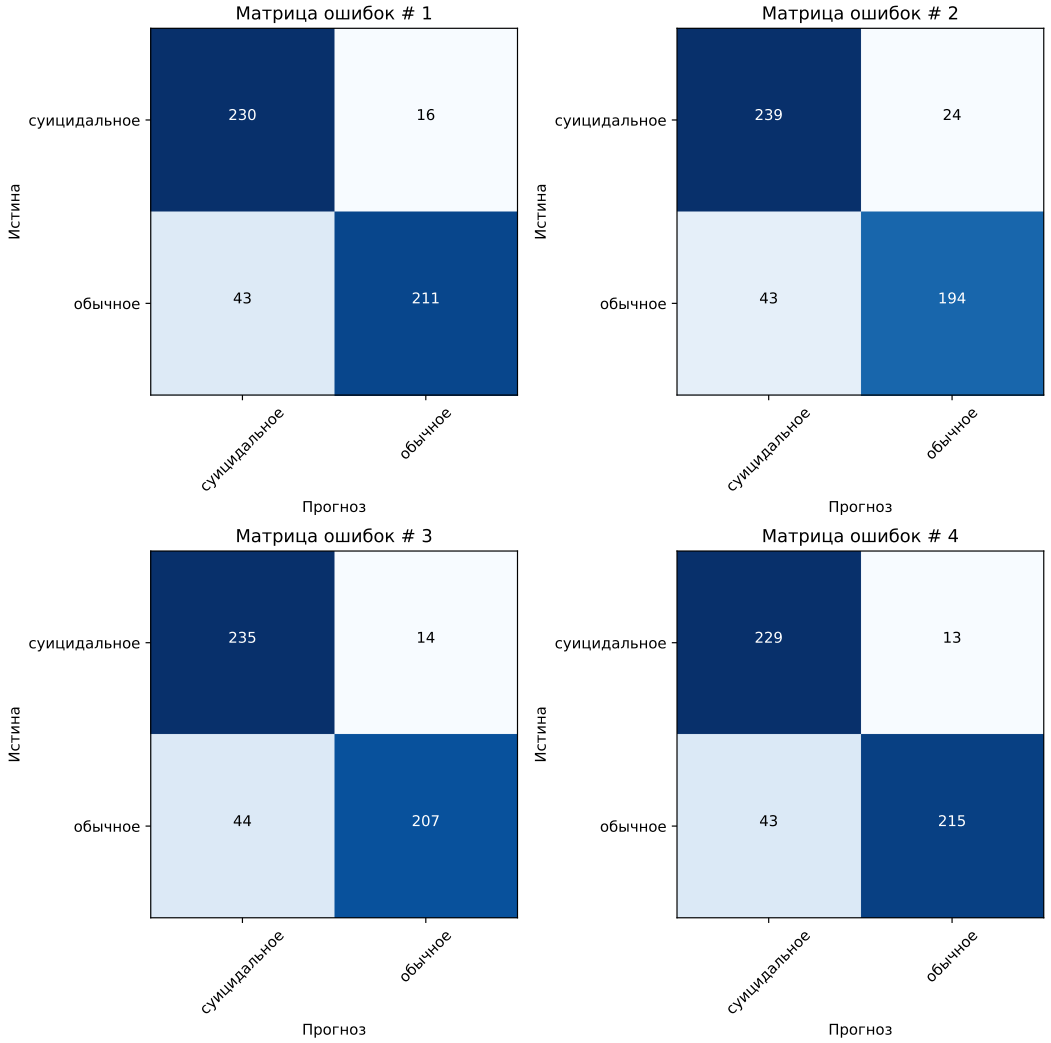
Случайный лес, BERT

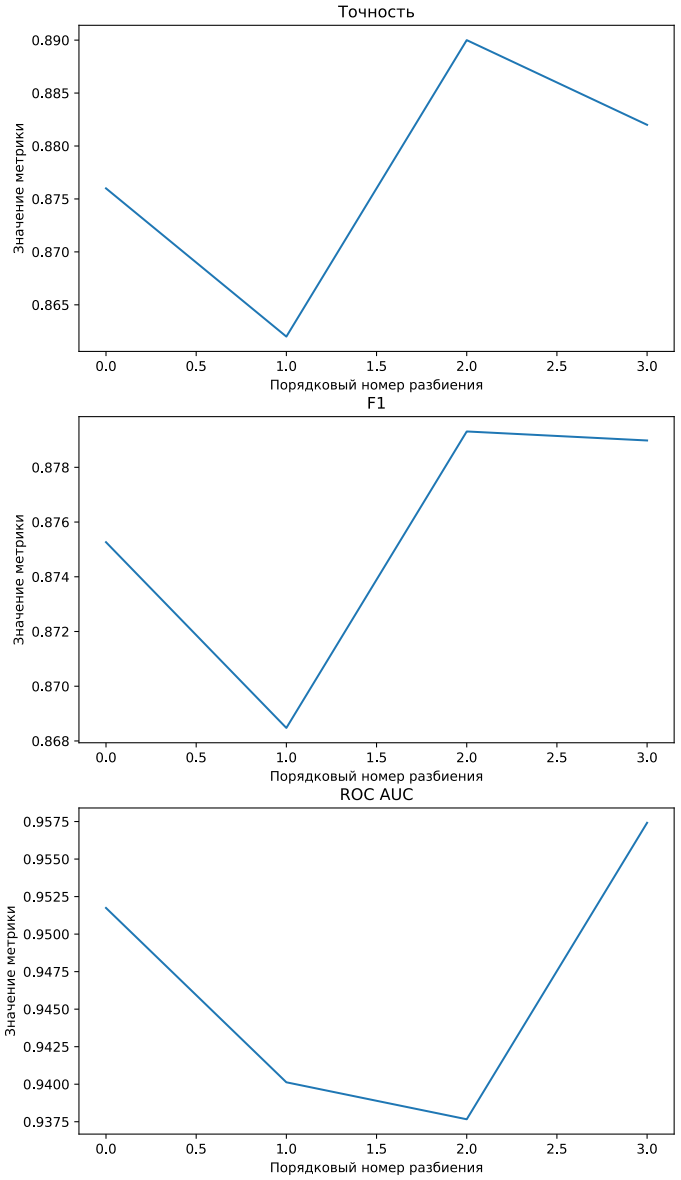




Точность: 0.888
F1-мера: 0.887
ROC-AUC: 0.948

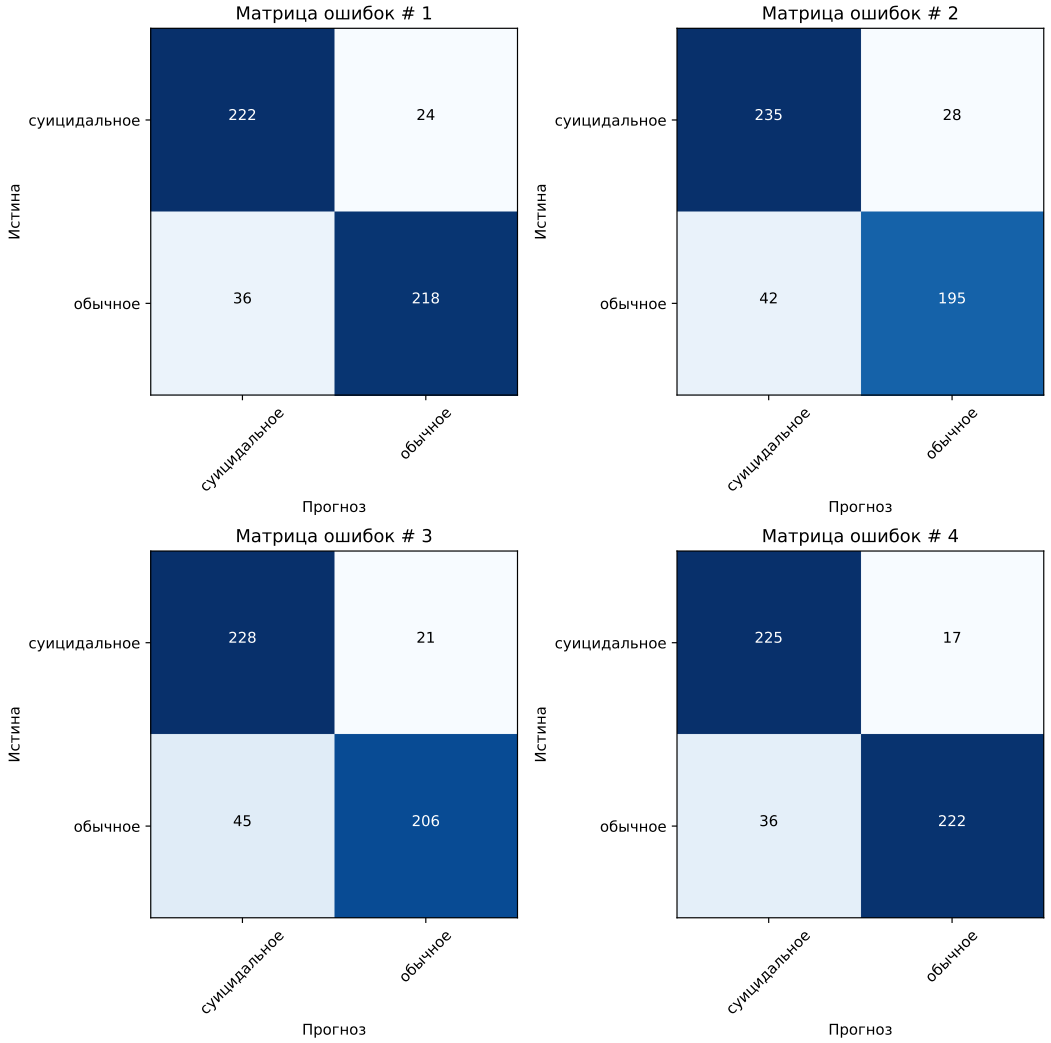
Случайный лес, ”мешок слов”

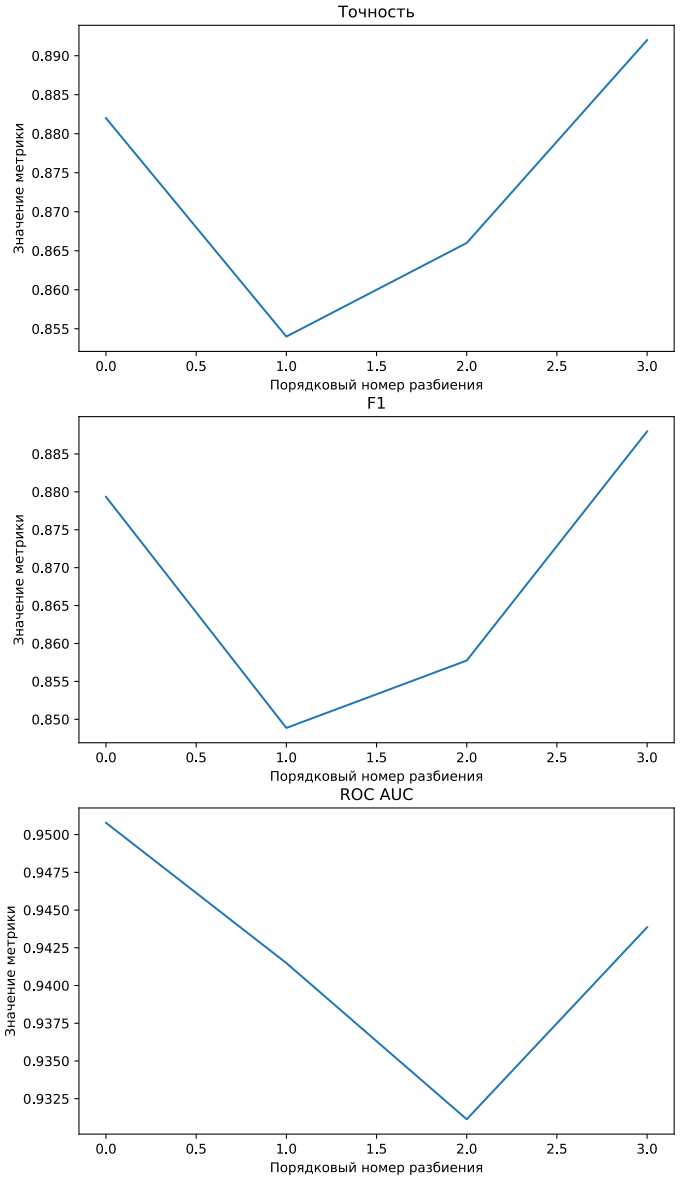




Точность: 0.878
F1-мера: 0.875
ROC-AUC: 0.947

Логистическая регрессия, BERT





Точность: 0.874
F1-мера: 0.869
ROC-AUC: 0.942

Заключение

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

Были решены следующие задачи:

- Проведен анализ действий и характеристик, позволяющих распознать паттерны суицидального поведения;
- Классифицированы признаки паттернов суицидального поведения человека;
- Определен метод сбора данных суицидального поведения;
- Разработан метод распознавания паттернов суицидального поведения;
- Разработанный метод реализован;
- Проведено сравнительное исследование задействованных в методе алгоритмов машинного обучения;
- Даны рекомендации о применимости реализованного метода.

Дальнейшее развитие

- Исследование эффективности использования ансамблевого метода 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.