

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

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:**

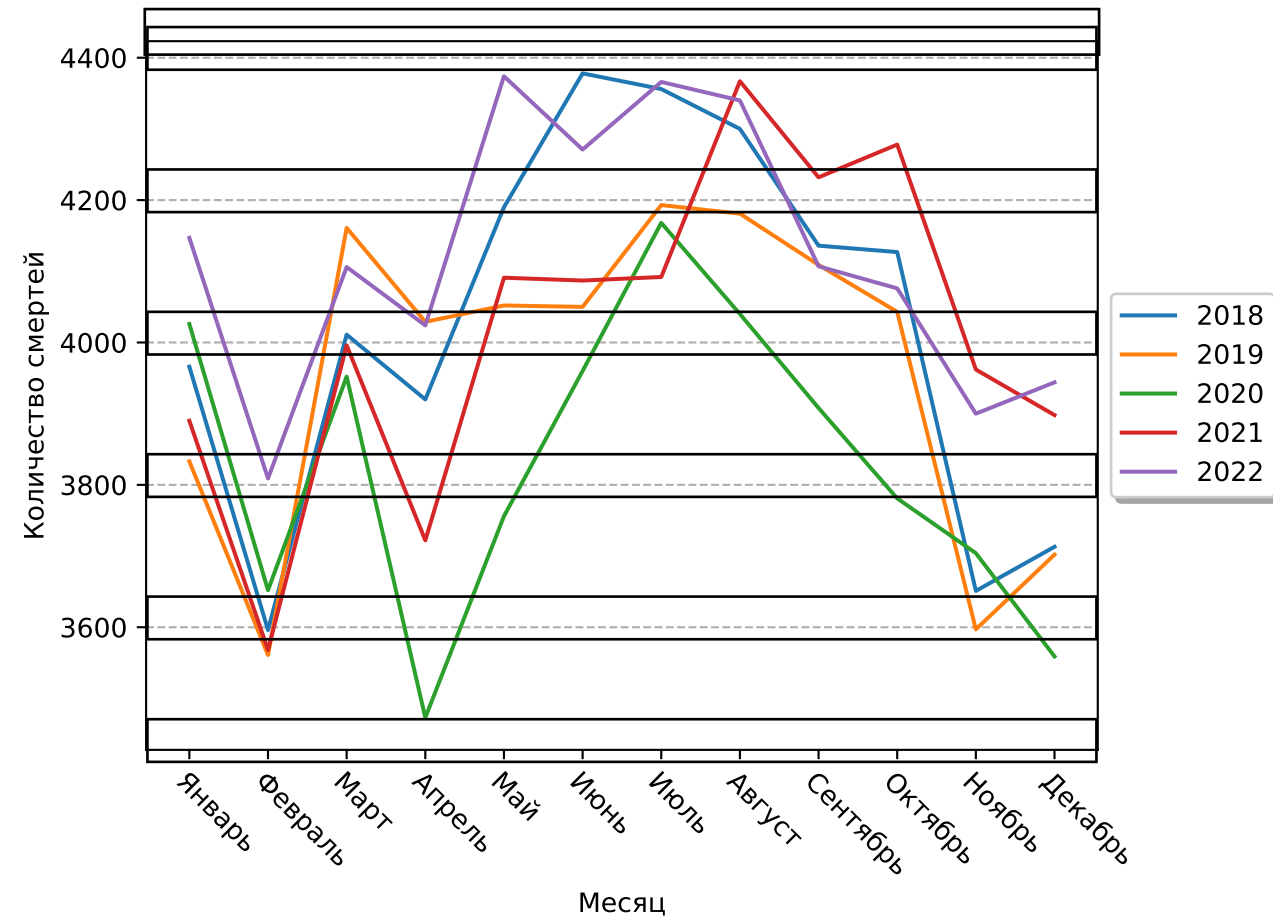
- Провести анализ действий и характеристик, позволяющих распознать паттерны суицидального поведения;
- Классифицировать признаки паттернов суицидального поведения человека;
- Определить метод сбора данных суицидального поведения;
- Разработать метод распознавания паттернов суицидального поведения;
- 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.

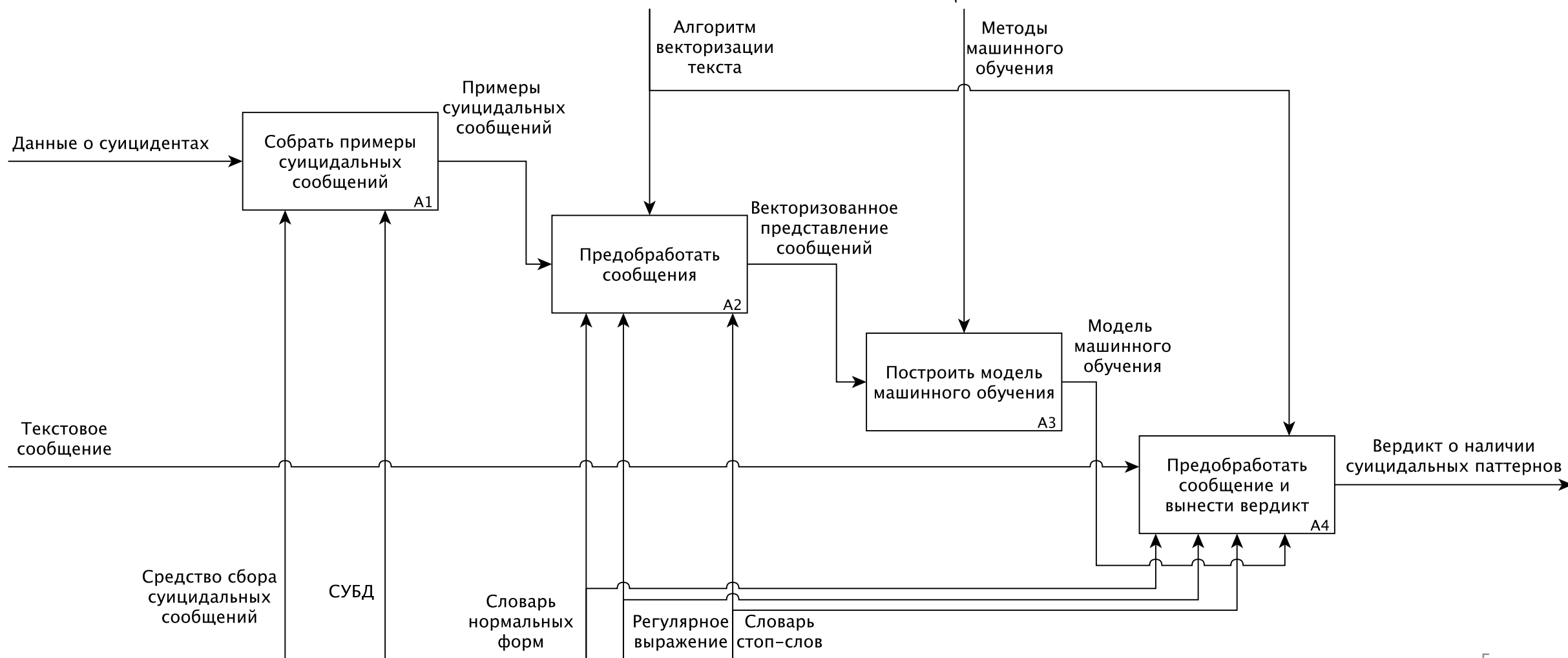
Самый высокий уровень самоубийств наблюдается у людей старше 85 лет.



# Форматы описания признаков

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



# Анализ собранных данных

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

суицидальные сообщения



несуицидальные сообщения



# Анализ собранных данных, несуицидальные сообщения

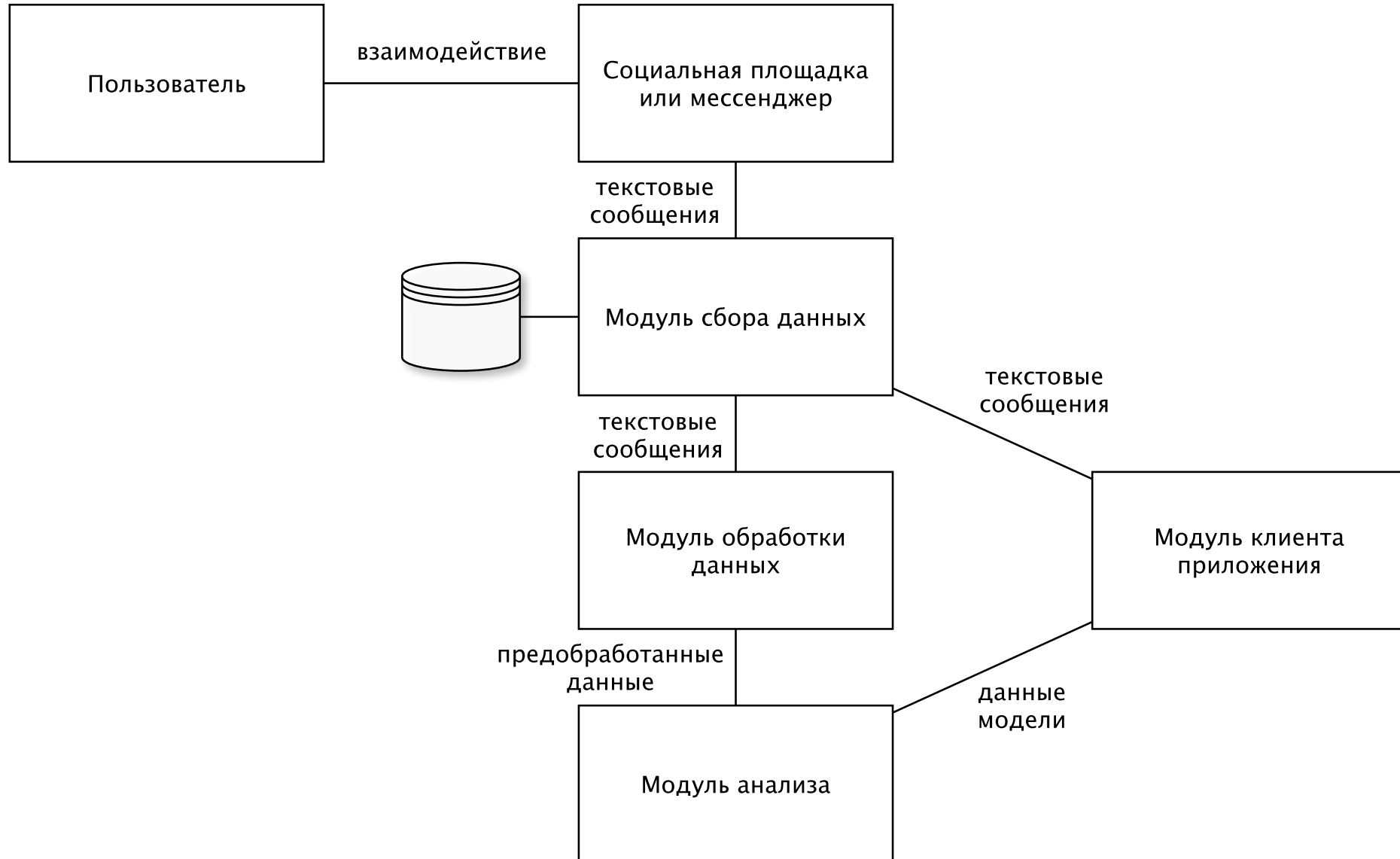


# Анализ собранных данных, суицидальные сообщения





# 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

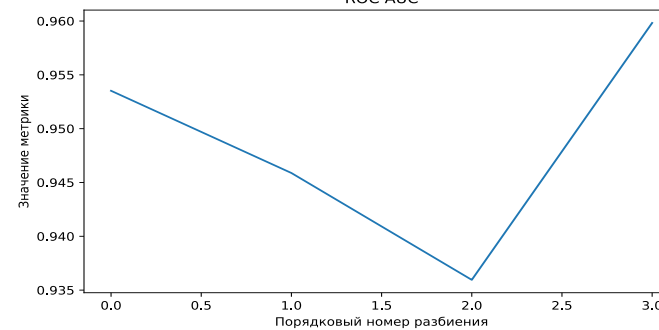
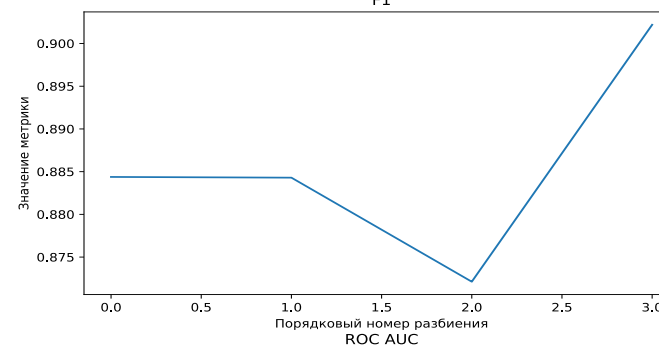
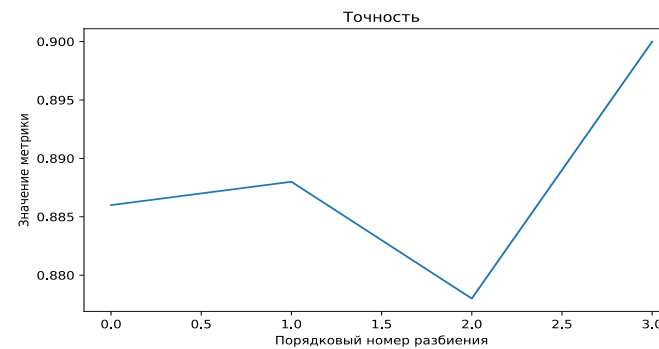
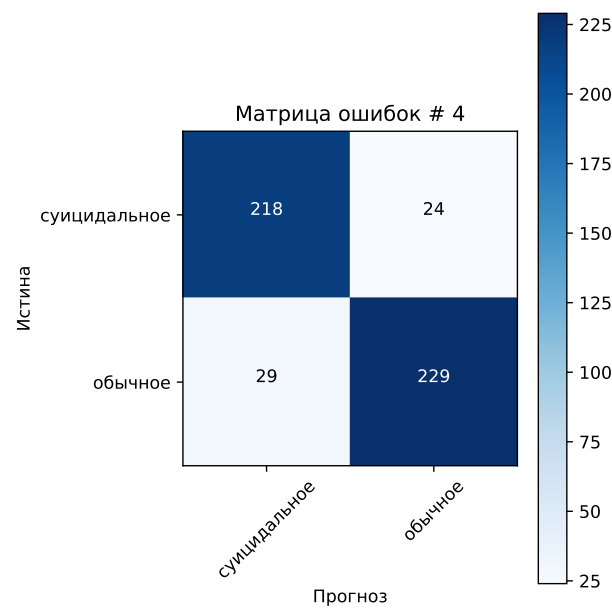
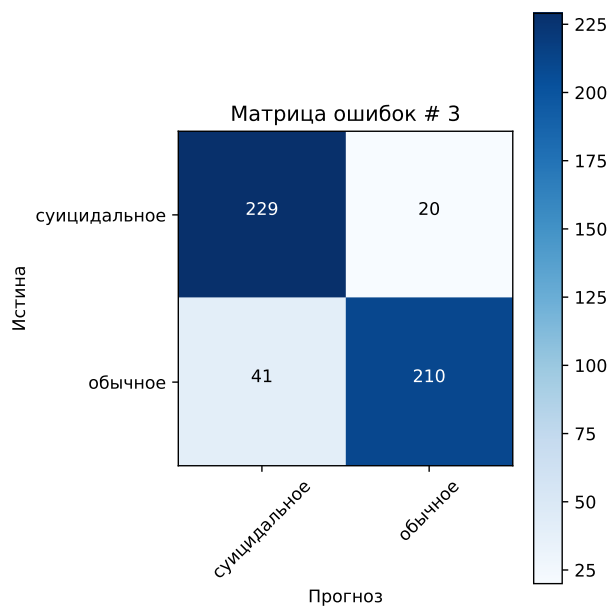
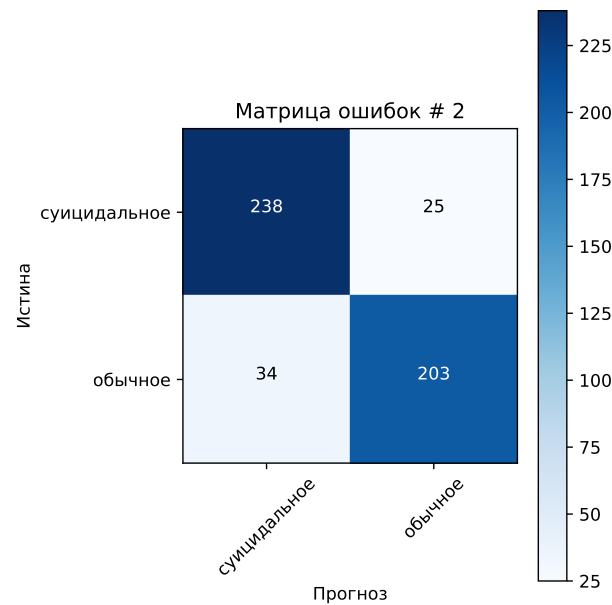
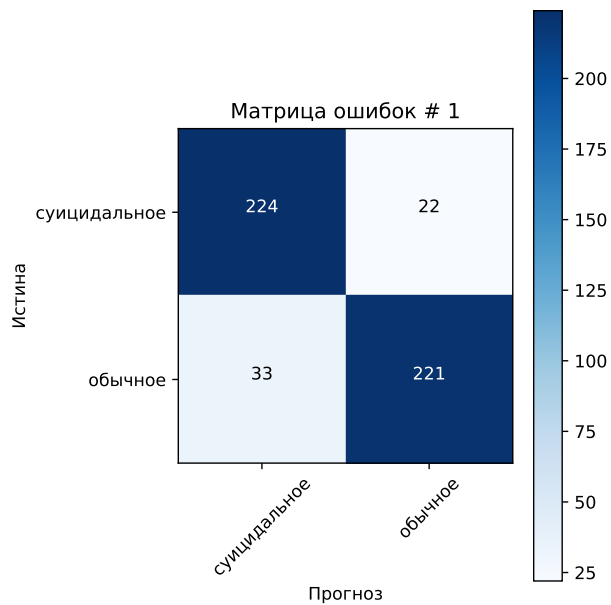
Исследуемые метрики:

- Точность
- F1-мера
- ROC-AUC

# Результаты исследования

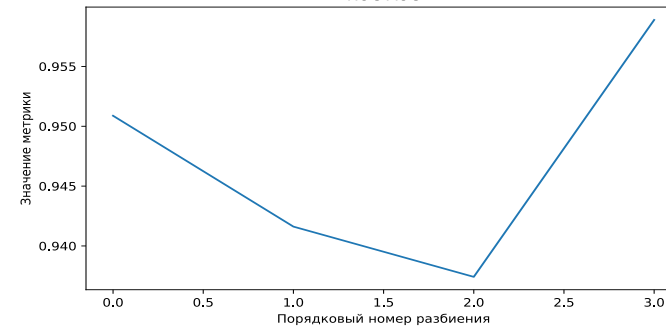
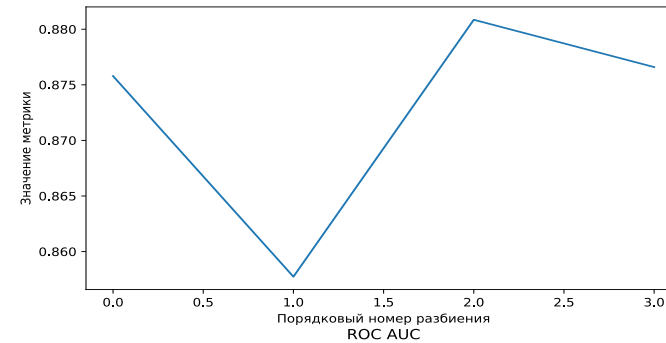
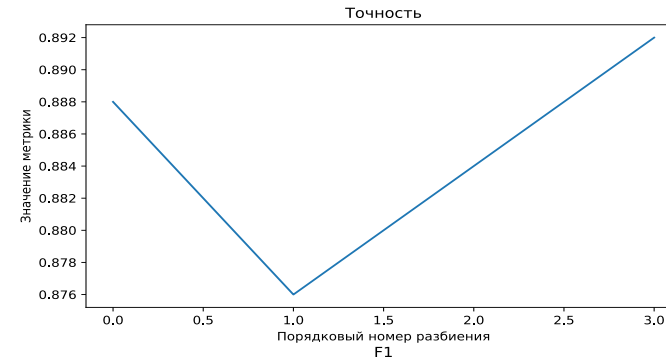
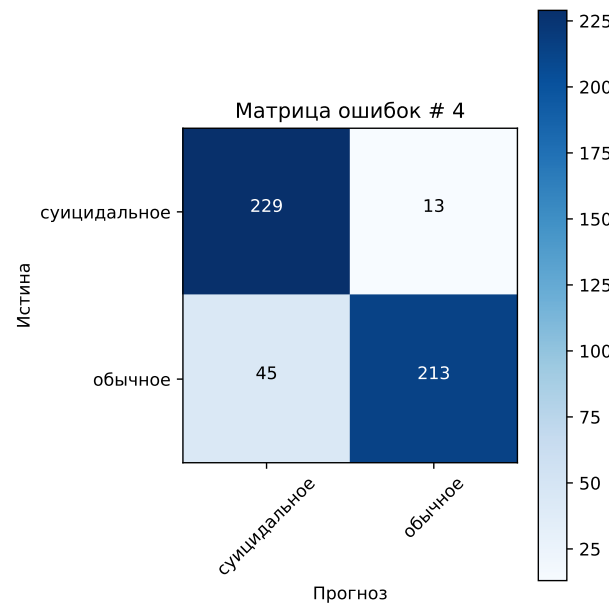
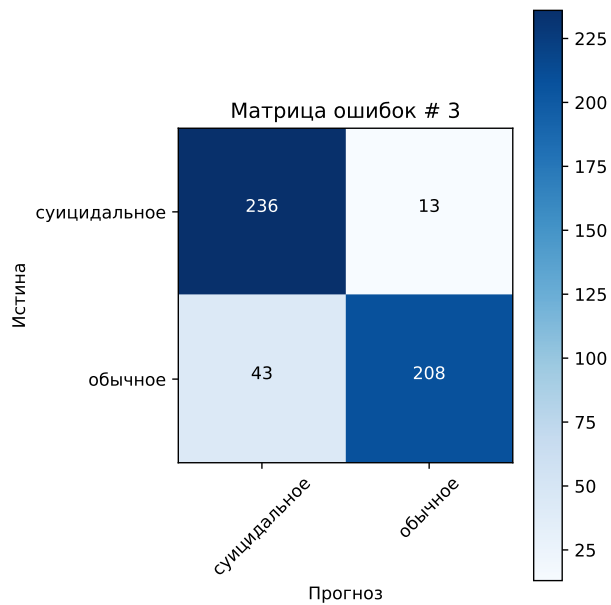
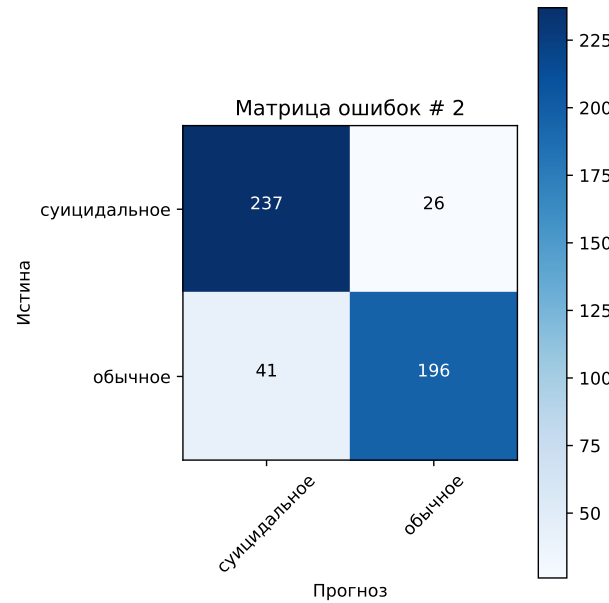
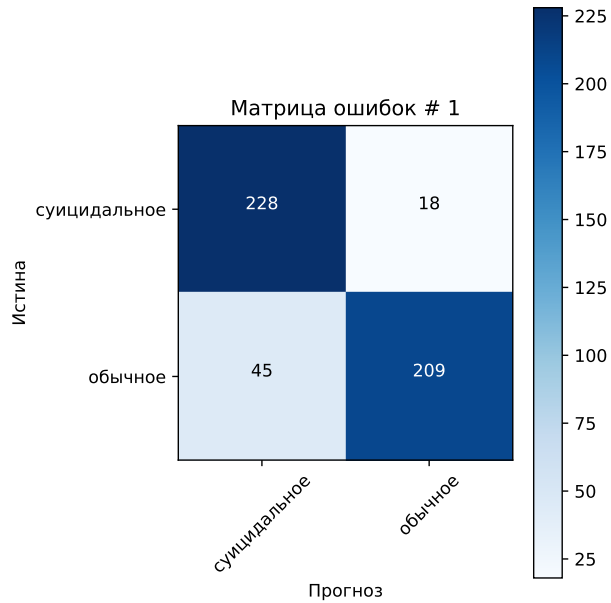
Алгоритм	Векторизация	Точность	F1-мера	ROC-AUC
Градиентный бустинг	“Мешок слов”	<i>0.853</i>	<i>0.844</i>	<i>0.916</i>
	BERT	<i>0.870</i>	<i>0.854</i>	<i>0.929</i>
Случайный лес	“Мешок слов”	<i>0.885</i>	<i>0.873</i>	<i>0.947</i>
	BERT	<b>0.888</b>	<b>0.886</b>	<b>0.949</b>
Метод опорных векторов	“Мешок 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	<i>0.874</i>	<i>0.869</i>	<i>0.942</i>
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



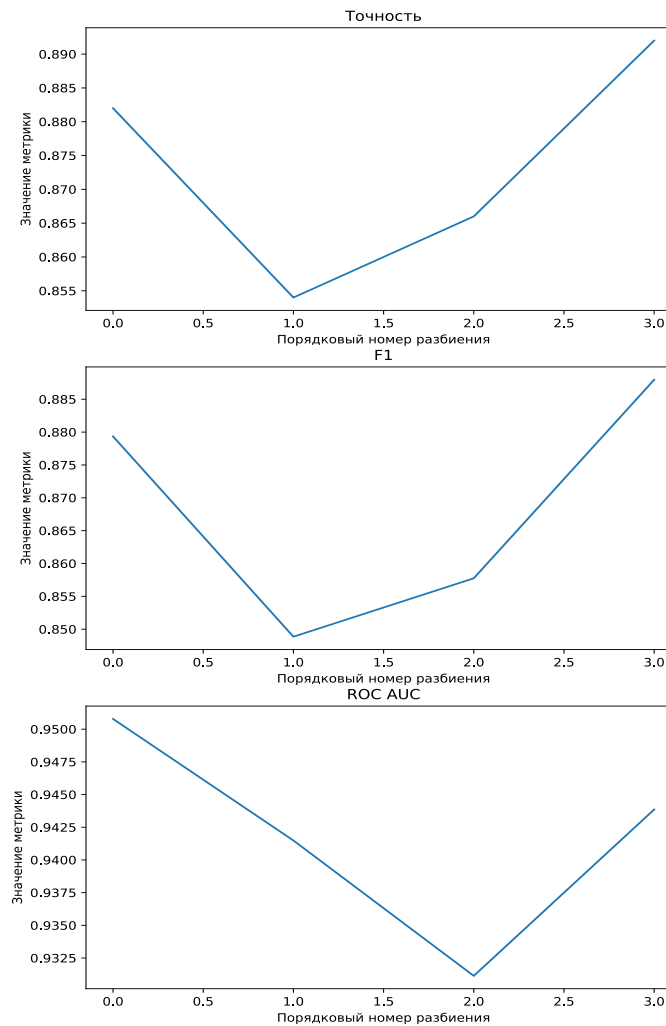
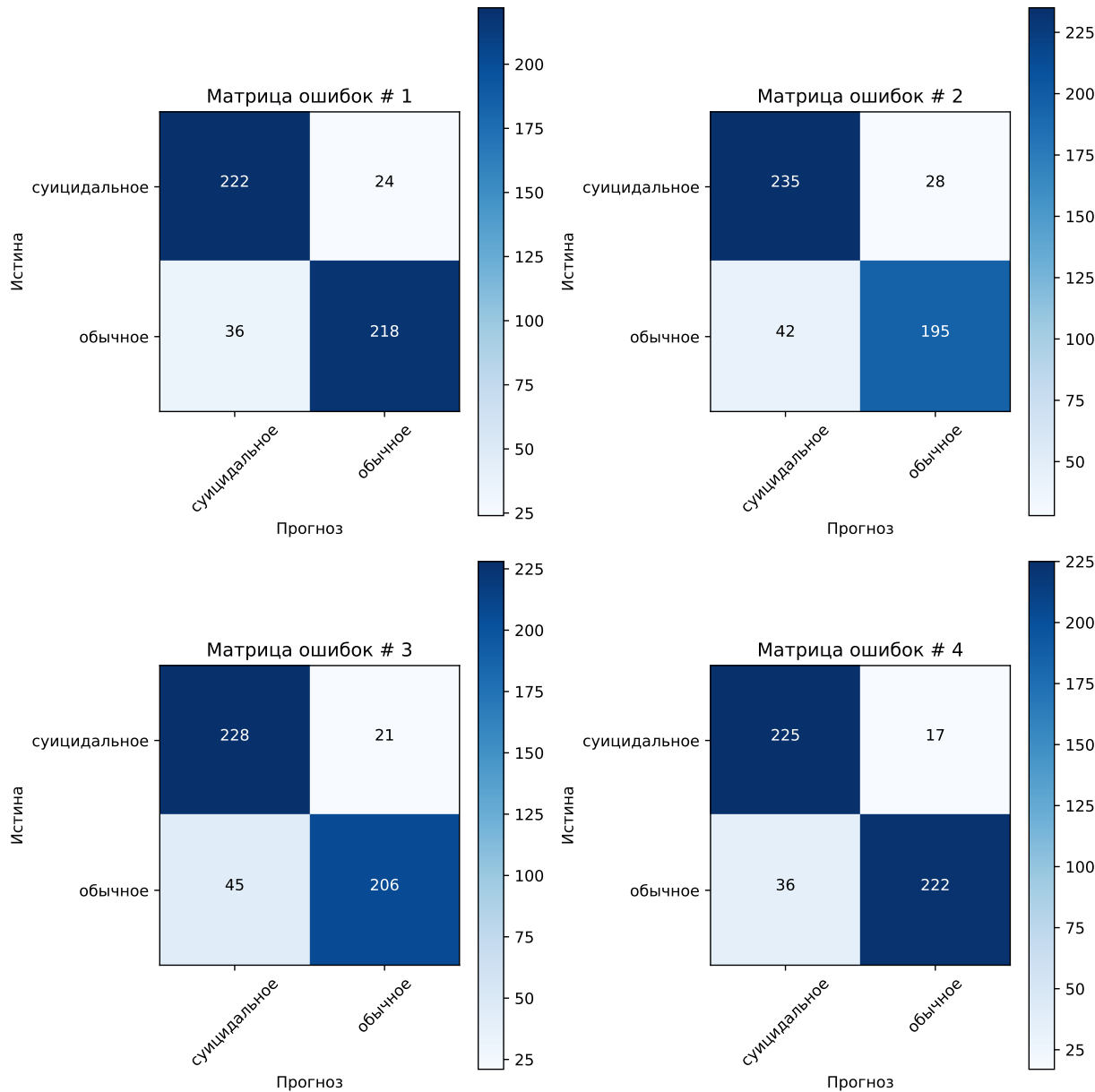
Точность: 0.888  
F1-мера: 0.886  
ROC-AUC: 0.949

# Случайный лес, «мешок слов»



Точность: 0.885  
F1-мера: 0.873  
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 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.