Национальный исследовательский ядерный университет

МИФИ

Кафедра 42 «Криптология и кибербезопасность»

# «Обнаружение внутреннего нарушителя путём выявления стрессового состояния пользователя»

**Исполнитель:** Султанов Азамат

*студент гр. Б16-505*

**Научный руководитель:** Когос К.Г.

к.т.н.

# Актуальность

Выявление стрессового состояния на основе биометрических показателей пользователя информационной системы позволяет обнаруживать внутреннего нарушителя*

*:

1. Работа «Insider Threat Detection Based on Users' Mouse Movements and Keystrokes Behavior» (Yessir H., 2017 г.);
2. Работа «On the Possibility of Insider Threat Detection Using Physiological Signal Monitoring» (Abdulaziz A., 2014 г.);
3. Работа «An Application of Data Leakage Prevention System based on Biometrics Signals Recognition Technology» (Hojae L., 2013 г.);
4. Работа «Inside the Mind of the Insider: Towards Insider Threat Detection Using Psychophysiological Signals» (Yessir H., 2016 г.)

# Цель работы

Оценить возможность выявления стрессового состояния пользователя на основе анализа взаимодействия с клавиатурой и мышью.

# Сбор данных

6 сценариев:

- нормальное поведение (найти в браузере ответы на вопросы, поработать с таблицами, ответить на письма – в условиях неограниченного времени)
- поведение нарушителя (взломать архив, подделать документ, скрытно поделиться информацией – в условиях ограниченного времени)

События мыши:

1. Клик левой кнопкой
2. Клик правой кнопкой

События клавиатуры:

1. Специальные символы (esc, alt, caps lock,…)
2. Биграммы
3. Триграммы

# Признаки

| № | Название | Определение |
|---|---|---|
| 1 | Время удержания | время  между нажатием и отпусканием одной и той же клавиши |
| 2 | Время «полёта» | время между нажатием одной клавиши и нажатием другой клавиши |
| 3 | Время задержки | время между нажатием одной клавиши и отпусканием другой клавиши |
| 4 | Интервал | время между отпусканием одной клавиши и нажатием другой клавиши |
| 5 | Отпускание-отпускание | время между отпусканием одной клавиши и отпусканием другой клавиши |
| 6 | Частота нажатий* | количество использования клавиши в минуту |
| 7 | Скорость движения мыши** | расстояние (в пикселах), пройденное курсором за минуту |

*  —  рассчитан только для спец. символов

**  —  не использован в данном исследовании

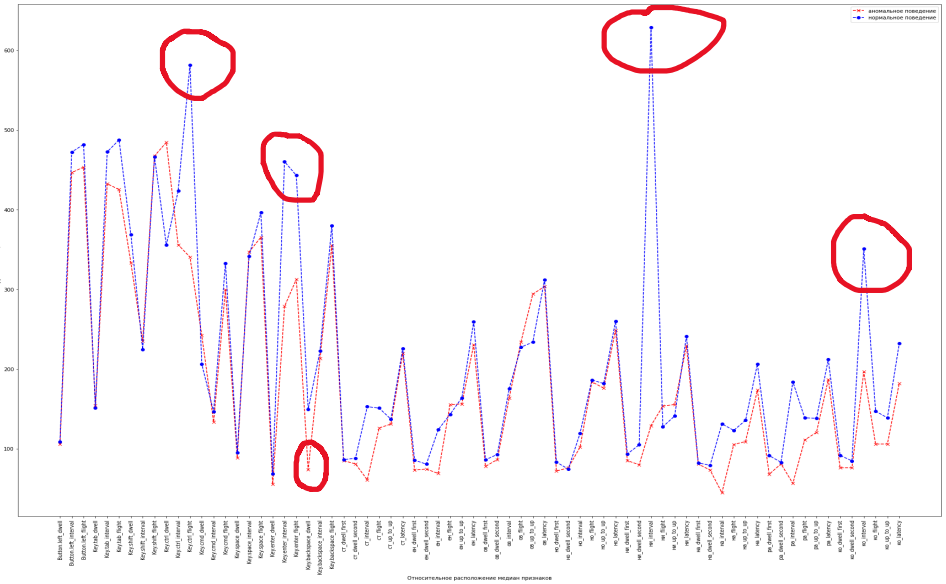
# Предобработка данных – часть 1

Шаги*:

1. Удаление признаков редко встречающихся событий клавиатуры и мыши, например события нажатия клавиши alt;
2. Удаление признаков с маленьким значением стандартного отклонения;
3. Заполнение пустот признаков медианами.

* - Размерность пространства признаков уменьшилась со 191 до 150;

# Предобработка данных – часть 1



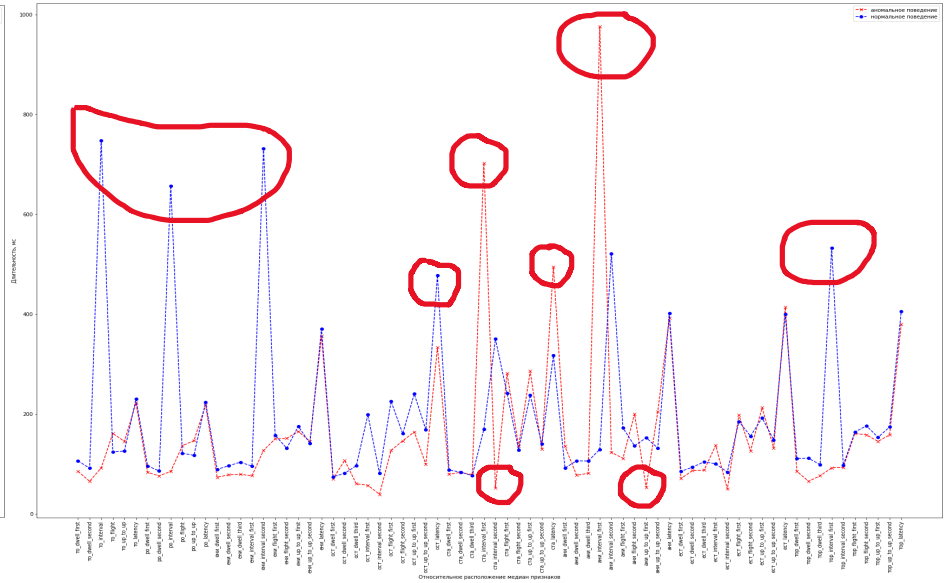
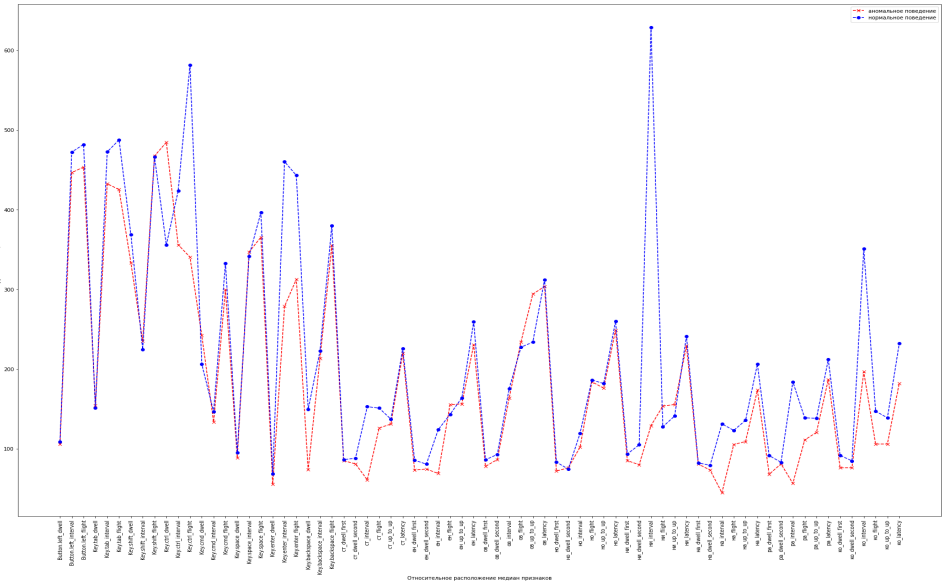



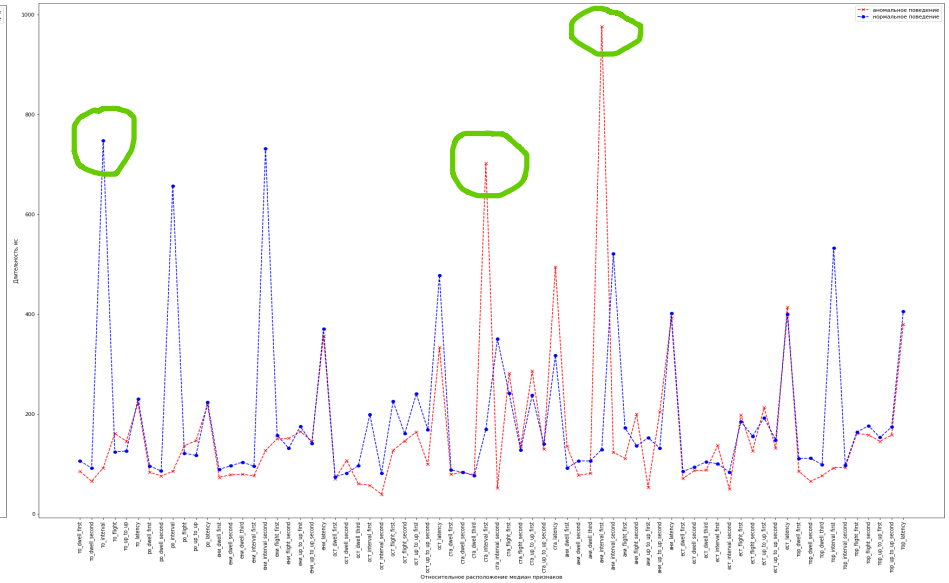

Рис 1,2 — усреднённые значения признаков, рассчитанные для категорий нормального и аномального поведений по отдельности

*: синий цвет — нормальное поведение, красный цвет — аномальное поведение.

** — выделенные красным цветом зоны на рисунках показывают наиболее информативные признаки

# Предобработка данных - часть 2
# Выделение информативных признаков*

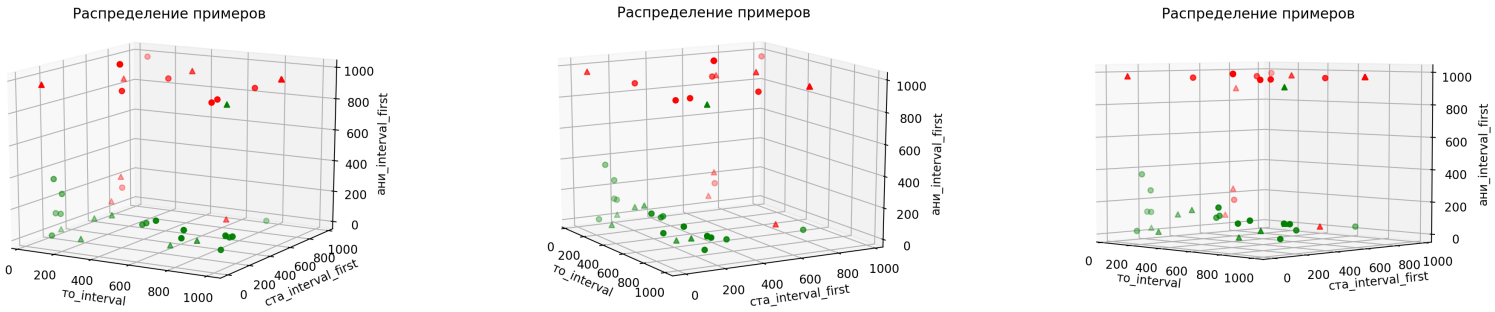



* — использован алгоритм SelectKBest на основе критерия хи-квадрат из библиотеки sklearn

** — зелёным цветом выделены признаки, отобранные в результате выделения наиболее информативных признаков

# Предобработка признаков – часть 2




- Аномальное поведение в обучаемой выборке
- Нормальное поведение в обучаемой выборке
- Аномальное поведение в тестовой выборке
- Нормальное поведение в тестовой выборке

Рис 3 — Распределение примеров датасета в пространстве выделенных признаков

# Результаты классификации

| № | Алгоритм | Точность (нет стресса) (обучение / тестирование) | Точность (есть стресс) (обучение / тестирование) | Полнота (нет стресса) (обучение / тестирование) | Полнота (стресс) (обучение / тестирование) | Точность (Accuracy) (обучение / тестирование) |
|---|---|---|---|---|---|---|
| 1 | LR | 1.00 / 0.70 | 1.00 / 0.83 | 1.00 / 0.88 | 1.00 / 0.62 | 1.00 / **0.75** |
| 2 | k-NN | 0.92 / 0.73 | 1.00 / 1.00 | 1.00 / 1.00 | 0.91 / 0.62 | 0.95 / **0.81** |
| 3 | RF | 1.00 / 0.80 | 1.00 / 1.00 | 1.00 / 1.00 | 1.00 / 0.75 | 0.91 / **0.88** |
| 4 | MLP | 1.00 / 0.70 | 1.00 / 0.83 | 1.00 / 0.88 | 1.00 / 0.62 | 1.00 / **0.75** |
| 5 | GB | 1.00 / 0.70 | 1.00 / 0.83 | 1.00 / 0.88 | 1.00 / 0.62 | 1.00 / **0.75** |

LR Логистическая регрессия

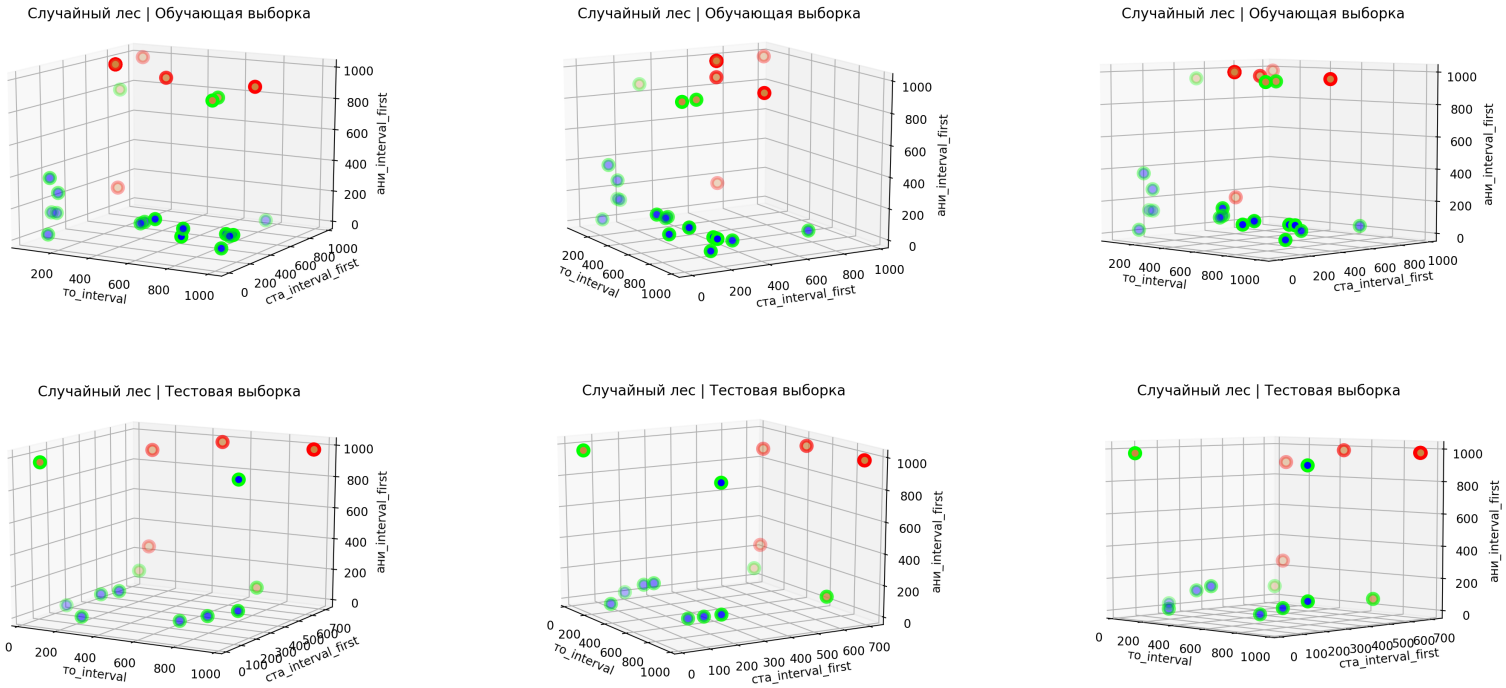
k-NN Метод k-ближайших соседей

GB Градиентный бустинг

RF Метод случайного леса

MLP Многослойный перцептрон

# Результаты классификации




ИП - истинно положительный

ИО - истинно отрицательный

ЛО - ложно отрицательный

ЛП - ложно положительный

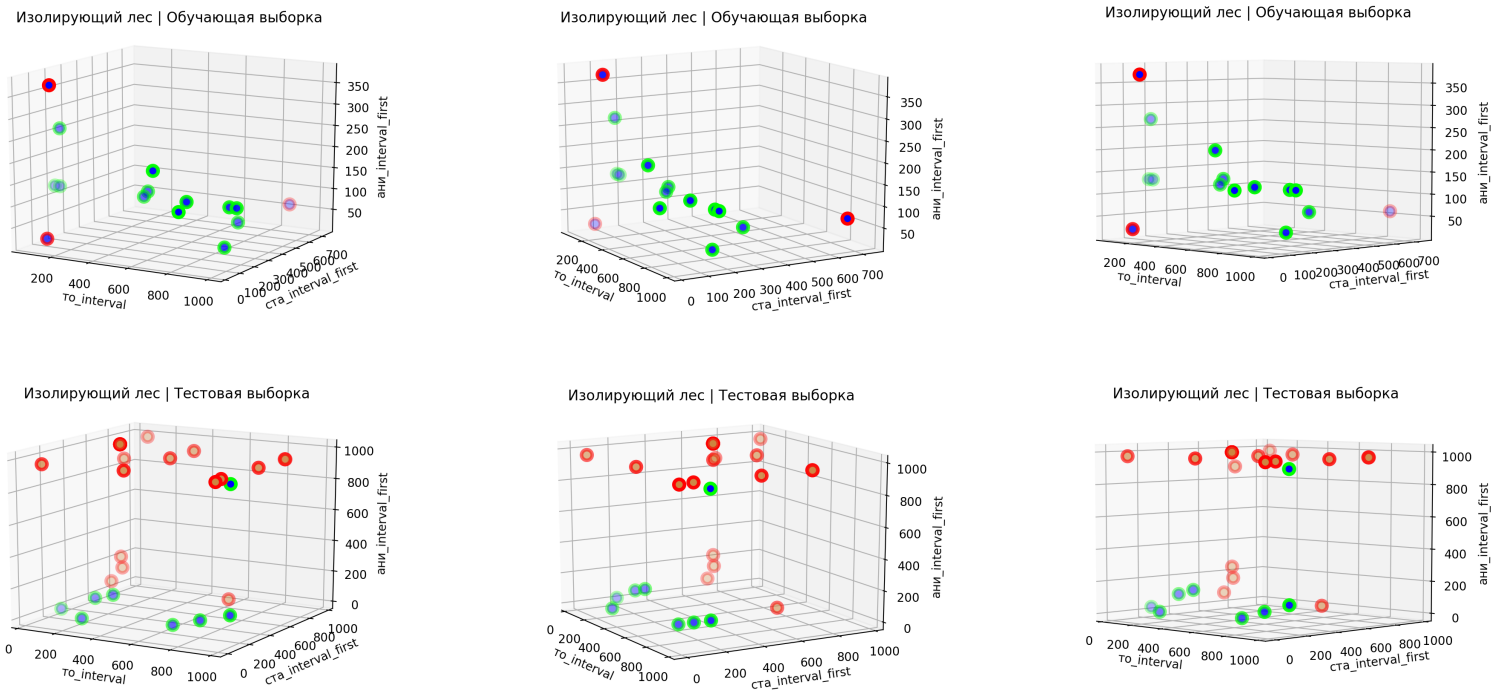
Рис 4 — визуализированные результаты лучшего классификатора

# Результаты обнаружения аномалий

| № | Алгоритм | Точность (нет стресса) | Точность (есть стресс) | Полнота (нет стресса) | Полнота (стресс) | Точность (Accuracy) |
|---|---|---|---|---|---|---|
| 1 | RC | 0.60 | 0.91 | 0.75 | 0.83 | **0.81** |
| 2 | OCSVM | 0.00 | 0.75 | 0.00 | 1.00 | **0.75** |
| 3 | IF | 1.00 | 1.00 | 1.00 | 1.00 | **1.00** |
| 4 | LOF | 0.88 | 0.96 | 0.88 | 0.96 | **0.94** |

| | | | |
|---|---|---|---|
| RC | Метод робастной ковариации | IF | Метод изолирующего леса |
| OCSVM | Одноклассовый метод опорных векторов | LOF | Локальный уровень выброса |

# Результаты обнаружения аномалий




ИП - истинно положительный

ИО - истинно отрицательный

ЛО - ложно отрицательный

ЛП - ложно положительный

Рис 5 — визуализированные результаты лучшей модели обнаружения аномалий

# Результаты исследования

- Проанализированы существующие методы обнаружения внутреннего нарушителя с использованием биометрических показателей на основе алгоритмов машинного обучения с учителем и без учителя
- Реализованы процессы накопления данных, предобработки данных, обучения и оценки моделей классификаторов и моделей обнаружения аномалий
- Наилучшие результаты получены для моделей на основе алгоритмов случайного леса (Точность - 88%) и изолирующего леса (Точность – 100%)