# Государственное бюджетное общеобразовательное учреждение города Москвы "Школа № 1554"

**Анализ данных национального виктимизационного опроса**

Над проектом работали:

Литовченко Максим Евгеньевич, 11 «Б» класс, ГБОУ «Школа 1554»

Каюмов Вадим Рустамович, 11 «Б» класс, ГБОУ «Школа 1554»

Научные руководители:

Сиухина Арина Алексеевна, ??? курс, мехмат факультет МГУ им. М.В. Ломоносова

Смирнова Ирина Сергеевна, ГБОУ «Школа 1554»

Москва, 2020 г.

# Аннотация

# Данный проект посвящен изучению преступности на территории Российской Федерации. Исходные данные для анализа собраны Институтом проблем правоприменения (ИПП) при ЕУ СПб и представляют результаты телефонного опроса респондентов не моложе 18 лет на основании простой случайной выборки телефонных номеров. Респонденты опрашивались не зависимо от гражданства, но, по понятным причинам, большинство (98.1%) были гражданами РФ. Вопросы делились на анкетные (пол, возраст, социо-демографический профиль), основной («были ли жертвой преступления») и уточняющие вопросы о преступлениях (где, когда и т.п.). Анкета составлена специальным образом профессионалами, учитывая многие особенности, например, специфику русского языка.

##### В результате проделанной работы осуществлена предобработка данных и последующее разбиение на две выборки: лица, являющиеся жертвами преступлений и лица, не являющиеся жертвами преступлений соответственно, далее проведены анализ получившихся выборок, визуализация найденных закономерностей и, исходя из сделанных выводов, построена классификационная модель, предсказывающая вероятность человека стать жертвой преступления исходя из его социально-демографических характеристик.

##### Обоснование выбора темы, актуальность работы, анализ существующих решений

##### Россия – одна из крупнейших юрисдикций мира, имеющая огромные объемы анализируемых данных о работе органов внутренних дел. Однако, ничтожно малым остается объем исследований о работе правоохранителей и о преступности в стране, которые бы проводились независимыми экспертами. До сих пор такие исследования в России проводятся либо на сверхмалых нерепрезентативных выборках, либо собирают очень мало информации (обычно только сведения о самом факте виктимизации), что вызвало в нас интерес к проведению данного исследования.

##### Результаты данной работы могут быть полезны подведомственным органам Министерства внутренних дел Российской Федерации, правозащитным организациям, поскольку подобные исследования дают возможность объективно оценить уровень преступности в стране, так как официальная статистика не только подвержена искажениям, но и не отражает тех случаев преступлений, по поводу которых пострадавшие не обращались в полицию, дают возможность охарактеризовать качество работы полиции.

**Цель и задачи**

# Цель работы – выяснить, как влияют социально-демографические характеристики человека на его возможность оказаться жертвой преступления (виктимность).

# Задачи данного исследования:

* Ознакомиться с форматом данных, осуществить предобработку.
* Сделать анализ данных опроса жителей РФ о преступлениях, жертвами которых они становились.
* Сделать визуализации найденных закономерностей.
* Исходя из выявленных зависимостей, построить классификационную модель, определяющую виктимность человека.

**Методика выполнения работы**

Для того, чтобы определить признаки, влияющие на виктимность человека, был проведен анализ социально-демографических характеристик людей из двух выборок: лиц, являющихся жертвами преступлений, и лиц, не являющихся жертвами преступлений, в результате были по выявленным признакам были составлены идеальные портреты людей, которые с наибольшей вероятностью являются жертвой преступления и не являются соответственно.

В ходе исследования использовался язык программирования Python и такие библиотеки для анализа данных, как Pandas, NumPy. Для импортирования данных использован модуль библиотеки модуль utils высокоуровневого API Keras библиотеки TensorFlow. Для визуализации полученных корреляций и закономерностей (гистограмм, ящиков размаха, и др.) использовались библиотеки Matplotlib, Seaborn. Для построения классификационной модели использовались библиотеки Scikit-learn, классификационная модель CatBoostClassifier из библиотеки CatBoost (Открытая программная библиотека, разработанная компанией Яндекс). Также в ходе исследования были использованы методы математической статистики, в частности медианные и средние значения, межквартильный размах (для исключения выбросов данных), расчет процентных долей выборок.

**Нужна ли гипотеза? Нужно ли пошагово расписывать что мы делали? Прикреплять графики в случае подробного описания? Сделать содержание!**

**Результаты и (?)обсуждение**

**Выводы, сделанные в результате исследования**