

УДК 004.056

**РАЗРАБОТКА МЕТОДА ОБЕЗЛИЧИВАНИЯ ПЕРСОНАЛЬНЫХ ДАННЫХ В
ОБУЧАЮЩЕЙ ВЫБОРКЕ ПРИ РАСЧЁТЕ КРЕДИТНОГО РЕЙТИНГА**

Сафиуллин Р.И. (Университет ИТМО)

Научный руководитель – к.т.н., доцент Воробьева А.А.

(Университет ИТМО)

Аннотация. В рамках проведённого исследования рассмотрены применяемые на сегодняшний день методы обезличивания персональных данных, а также приведены их достоинства и недостатки. Разработан алгоритм, который позволяет провести обезличивание персональных данных, не требуя подстройки под конкретную систему машинного обучения, а также приведён пример использования данной модели на тестовом наборе данных (Home Credit Default Risk dataset).

Введение. На сегодняшний день существует множество методов обезличивания персональных данных, в том числе используемых при расчёте кредитного рейтинга. Однако, в контексте машинного обучения их применение зачастую является неэффективным из-за недостаточной итоговой точности обученной модели или из-за необходимости вносить изменения в применяемый алгоритм машинного обучения (который может представлять собой коммерческую тайну). В данной статье представлен алгоритм обезличивания персональных данных, позволяющий обеспечить высокую степень точности обученной модели вне зависимости от применяемого алгоритма машинного обучения.

Основная часть. В предлагаемом методе обезличивания предлагается использовать свойство к-анонимности для обезличивания обучающей выборки и последующего переобучения модели на обезличенных данных. Данная модель не будет содержать персональные данные в соответствии с GDPR-26 и ФЗ-152 – ч.9 ст.3, и может быть передана для обработки и анализа третьим лицам. При корректной настройке и количестве квази-идентификаторов не более 15, данный метод обезличивания демонстрирует лишь несущественное снижение точности переобученной модели.

Выводы. В ходе работы над данной статьёй были рассмотрены существующие методы обезличивания персональных данных и реализован метод обезличивания персональных данных, обеспечивающий значительно лучшие показатели точности при применении в сфере машинного обучения. Его применение может быть обосновано в банковской сфере, где при работе с большими объёмами персональных данных часто требуется передавать часть функционала подрядчикам.

Сафиуллин Р.И. (автор)

Подпись

Воробьева А.А. (научный руководитель)

Подпись