УДК 004.056

МЕТОД ФОРМИРОВАНИЯ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ С ИСПОЛЬЗОВАНИЕМ СОВМЕСТНЫХ КОНФИДЕНЦИАЛЬНЫХ ВЫЧИСЛЕНИЙ

Щетинин Д.С. (Университет ИТМО)

Научный руководитель – кандидат технических наук, Менщиков А.А.

(Университет ИТМО**)**

Введение. В последние годы, наблюдается видимый прогресс в области машинного обучения и нейронных сетей. Эти технологии активно интегрируются в жизни простых людей, в связи с использованием крупными компаниями, а также государство ищет варианты, чтобы стимулировать эту тенденцию. В следствии этого, важной проблемой является безопасность машинного обучения, вычислений, а особенно данных, поскольку именно они наиболее часто могут быть использованы для реализации атак. На текущий момент, эта проблема часто решается путем анонимизации или добавления к данным дополнительного шума [1, 2]. Однако эти способы нельзя назвать эффективными, поскольку удаление большого количества признаков или добавление большого количества шума негативно сказываются на качестве моделей, а в обратных случаях набор может сохранить информацию, достаточную для деанонимизации [2]. Совместные конфиденциальные вычисления позволяют обучать модели машинного обучения на данных, которые не должны быть получены другими участниками, что позволяет обучать модели на большем количестве данных и получать лучшие результаты. Однако, помимо этого преимущества данный подход имеет существенные ограничения, которые обуславливают сферу его эффективного использования [3].

Основная часть. Целью данной работы является сравнение различных методов формирования моделей машинного обучения и разработка собственного подхода, с использованием совместных конфиденциальных вычислений. Суть подхода заключается в том, что при наличии трех или более равных сторон, можно выполнять обучения моделей машинного обучения без необходимости прямого обмена данными. Это возможно, поскольку совместные конфиденциальные вычисления имеют поддержку аддитивной и мультипликативной операций, а не линейный зависимости могут быть аппроксимированы. Учитывая эти факторы, алгоритм обучения при указанном подходе выглядит сложнее и выполняется дольше.

Выводы. Работа включает сравнение различных подходов к созданию моделей машинного обучения, а также описание собственного подхода, который при некоторых условиях способствует увеличению метрик оценки результирующих моделей.

Список использованных источников:

1. Ahuja, Mohit & Belaid, Mohamed-Bachir & Bernabé, Pierre & Collet, Mathieu & Gotlieb, Arnaud & Lal, Chhagan & Marijan, Dusica & Sen, Sagar & Sharif, Aizaz & Spieker, // Opening the Software Engineering Toolbox for the Assessment of Trustworthy AI // CEUR Workshop Proceedings. – 2020. – Vol. 2659. – С. 67 - 70.
2. Prasser, Fabian & Eicher, Johanna & Spengler, Helmut & Bild, Raffael & Kuhn, Klaus. // Flexible data anonymization using ARX-Current status and challenges ahead // Software: Practice and Experience. – 2020. – Vol. 50.
3. Park, Saerom & Kim, Seongmin & Lim, Yeon-sup. // Fairness Audit of Machine Learning Models with Confidential Computing // WWW '22: Proceedings of the ACM Web Conference. – 2022. – С. 3488–3499.

Щетинин Д.С. (автор) _____

Менщиков А.А. (научный руководитель) _____