УДК 004.056.5

ОБЕСПЕЧЕНИЕ КОНФИДЕНЦИАЛЬНОСТИ ДАННЫХ В СИСТЕМАХ МАШИННОГО ОБУЧЕНИЯ

Беляев В.В. (Университет ИТМО)
Научный руководитель – Югансон А.Н., к.т.н., доцент (Университет ИТМО)

Аннотация. В работе представлен анализ методов обеспечения конфиденциальности информации, применимых к системам машинного обучения, и представлена архитектура защищенной системы машинного обучения с использованием подхода автоматического машинного обучения и гомоморфного шифрования.

Введение. В настоящее время машинное обучение активно применяется в различных областях, в том числе таких, как, например, банковская сфера или медицина, где осуществляется работа с конфиденциальной информацией. Однако при обучении моделей машинного обучения, специалисты работают непосредственно с подобной информацией, что повышает риски нарушения ее конфиденциальности. В связи с подобными особенностями систем, реализующих технологии машинного обучения, ФСТЭК РФ в 2020 году выделил в отдельные пункты базы данных угроз информационной безопасности угрозы, связанные с технологиями машинного обучения, в частности, угрозы конфиденциальности информации – обучающей информации для машинного обучения и информации о модели машинного обучения. В обоих случаях, основным нарушителем является внутренний нарушитель, то есть сотрудник, взаимодействующий с данной системой. Как показывает статистика компании InfoWatch, в России в 2022 году около 80 процентов всех нарушений информационной безопасности внутреннего характера были предумышленными. К тому же, удаленный режим работы сотрудников, получивший распространение в последнее время, осложняет работу по обеспечению информационной безопасности, поскольку в такой ситуации не представляется возможным полностью контролировать действия сотрудника. Все это свидетельствует о недостаточности настоящих средств обеспечения конфиденциальности информации и необходимости в разработке новых. Для этого будет проведен анализ существующих методов обеспечения конфиденциальности информации, применимых к системам машинного обучения.

Основная часть. В ходе работы проведен анализ существующих методов обеспечения конфиденциальности информации, применимых к системам машинного обучения [1]. Предложена архитектура защищенной системы машинного обучения на основе подхода автоматического машинного обучения и с применением схем гомоморфного шифрования [2]. Предлагаемая архитектура позволяет обеспечить конфиденциальность содержащейся в системе информации как от внутреннего, так и от внешнего нарушителя.

Выводы. Результаты проведенного анализа методов обеспечения конфиденциальности информации были применены при разработке архитектуры защищенной системы машинного обучения. Предложенная архитектура системы будет использована в дальнейшем при разработке ее программной реализации.

Список использованных источников:

1. Xu R., Baracaldo N., Joshi J. Privacy-preserving machine learning: Methods, challenges and directions //arXiv preprint arXiv:2108.04417. – 2021.
2. Clet P. E., Stan O., Zuber M. BFV, CKKS, TFHE: Which one is the best for a secure neural network evaluation in the cloud? // Applied Cryptography and Network Security Workshops: ACNS 2021 Satellite Workshops, AIBlock, AIHWS, AIoTS, CIMSS, Cloud S&P, SCI, SecMT, and SiMLA, Kamakura, Japan, June 21–24, 2021, Proceedings. – Springer International Publishing, 2021. – С. 279-300.