УДК 004.021

СРАВНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ ДЕЙСТВИЙ ЗЛОУМЫШЛЕННИКОВ В БОЛЬШИХ ДАННЫХ СИСТЕМЫ ЦИФРОВЫХ ДВОЙНИКОВ

Татаров Д.А., Перфильев В.Э., Менщиков А.А. (Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

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

(Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Данная работа посвящена сравнению методов машинного обучения. Экспериментальное сравнение произведено на больших данных системы цифровых двойников. Метод с наибольшей точностью может впоследствии быть использован для обнаружения действия злоумышленников в финансовых организациях.

Введение. С каждым годом растет количество успешных кибератак и появляются новые схемы мошенничества. Для обнаружения случаев мошенничества финансовым организациям требуется собирать, обрабатывать и хранить большие объемы данных. Применение методов машинного обучения вместе с большими объемами данных могут позволить эффективно обнаруживать и предотвращать мошенничества в финансовых организациях. Заблаговременное обнаружение действий злоумышленников имеет важное значение для обеспечения безопасности клиентов финансовых организаций. Чем раньше сотрудники финансовых организаций смогут обнаружить подозрительную активность, тем быстрее они смогут на нее среагировать, например ограничив доступ к банковскому счету, что позволит минимизировать убытки как для организации, так и для клиента. Внедряя ряд схем обнаружения мошенничества, возможно обеспечить необходимую защиту и избежать значительных финансовых и репутационных потерь.

Основная часть. В работе был проведен эксперимент по выявлению метода машинного обучения, показывающего наибольшую точность обнаружения мошеннических действий в больших объемах данных. Для получения больших объемов данных была использована система цифрового двойника финансовой организации. Данная система полностью повторяет необходимую функциональность финансовой организации, и способна воспроизводить различные состояния системы, в том числе состояние атаки на целевую систему. С помощью данной системы были получены необходимые для тестирования методов машинного обучения объем входных данных.

Для сравнения были выбраны наиболее популярные методы машинного обучения как с учителями, так и без. Сравнение методов было произведено экспериментальным образом. Сгенерированный набор данных путем One-Hot Encoding преобразован в числовое представление и разделен на тренировочную и тестовую выборки. Наилучший метод определен по совокупности времени обучения, классификации и показателям Accuracy, Precision, Recall & F1. Также для всех методов машинного обучения построены графики AUC.

Выводы. В работе представлено несколько экспериментов на больших данных. Необходимые данные получены с помощью системы цифровых двойников финансовой организации. На основании проведенных экспериментов выявлен наилучший метод машинного обучения для обнаружения мошеннических действий злоумышленников в финансовых организациях.

Татаров Д.А. (автор) Подпись

Перфильев В.Э. (автор) Подпись

Менщиков А.А. (автор) Подпись

Менщиков А.А. (научный руководитель) Подпись