**BC/NW 2018 № 1 (32):11.3**

**РЕШЕНИЕ ЗАДАЧИ КАТЕГОРИЗАЦИИ В СИСТЕМАХ ЗАЩИТЫ ОТ УТЕЧЕК ДАННЫХ**

Мартынова Ю. Ф., Филатов А. В.

В настоящее время технологии, связанные с машинным обучением, внедряются в системы различного назначения. Данные технологии могут быть использованы в системах защиты от утечек данных (DLP-системы).

Системы защиты от утечек данных — программно-аппаратный комплекс, предназначенный для отслеживания и блокирования попыток передачи конфиденциальных данных за пределы информационной системы. DLP-системы строятся на анализе потоков данных, пересекающих пределы информационной системы, защиту которой они обеспечивают.

В рамках анализа потоков данных могут быть успешно применены алгоритмы машинного обучения для классификации угроз. Такой поиск является задачей классификации, которая, в свою очередь, является задачей машинного обучения [2].

При выборе конкретного алгоритма классификации необходимо учесть следующие особенности систем защиты от утечек данных:

1. Данные системы работают с данными различного содержания

2. Архитектурно DLP-система должна быть гибкой, так как может быть внедрена в различные информационные системы

3. Система отслеживает и анализирует данные в автоматическом режиме, собранные ею данные анализируется человеком

Исходя из вышеперечисленных особенностей, для решения поставленной задачи должен быть выбран алгоритм, обучающийся онлайн [2].

Для реализации был выбран алгоритм линейной взвешенной классификации (Multi-Class Confidence Weighted Algorithm) [2]. Применение алгоритма позволяет улучшить ряд характеристик системы: точность, безопасность. Данный алгоритм был реализован в виде аддона для платформы NodeJS для дальнейшего встраивания в WEB-комплекс защиты.

**Литература**

1. Вьюгин В. В. Математические основы теории машинного обучения и прогнозиро-

вания. М.: 2013. — 387 с.

2. Dredze, M., Crammer, K., and Pereira, F. Confidence-Weighted Linear Classification

ICML2008 (Helsinki, Finland), 2008 URL: https://www.cs.jhu.edu/~mdredze/publications/icml\_variance.pdf