МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ РАДИОФИЗИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ
Кафедра информатики и компьютерных систем

Аннотация к дипломной работе

«Детектирование сетевого трафика на наличие угроз с помощью машинного обучения»

Короткий Никита Денисович

Научный руководитель — ст. преподаватель Бондаренко Ю. А.

Минск, 2025

РЕФЕРАТ

Дипломная работа 44 стр., 9 рис., 1 табл., 13 ист., 1 прил.

СЕТЕВОЙ ТРАФИК, МАШИННОЕ ОБУЧЕНИЕ, КИБЕРБЕЗОПАСНОСТЬ, ОБНАРУЖЕНИЕ ВТОРЖЕНИЙ, КЛАССИФИКАЦИЯ АТАК, PYTHON, SCIKIT-LEARN, XGBOOST, CIC-IDS-COLLECTION, DECISION TREE, SVM, RANDOM FOREST, ADABOOST, MLP

Цель работы — создание платформы для предоставления образовательных услуг, которая обеспечит удобные и эффективные инструменты для всех участников образовательного процесса. Объектом исследования является процесс детектирования угроз в сетевом трафике с использованием методов машинного обучения. Цель работы — исследование и сравнение эффективности различных алгоритмов машинного обучения для задачи детектирования угроз в сетевом трафике.

Проведено исследование и сравнение шести алгоритмов машинного обучения (Decision Tree, Support Vector Machine, Random Forest, AdaBoost, XGBoost, MLP) для классификации сетевого трафика на нормальный и вредоносный. Разработан программный прототип на языке Python с использованием библиотек Pandas, NumPy, Scikit-learn, XGBoost и Matplotlib для проведения экспериментов на общедоступном наборе данных CIC-IDS-Collection. Выполнена предобработка данных, включающая очистку, нормализацию и разделение на обучающую и тестовую выборки. Обучены и протестированы модели, их эффективность оценена по метрикам Accuracy, Precision, Recall и F1-score, а также по времени обучения и скорости предсказания.

Полученные результаты показали высокую эффективность ансамблевых методов, в частности XGBoost и Random Forest, в задаче детектирования сетевых угроз, продемонстрировав лучшие показатели по точности, полноте и F1-мере. Проведенный сравнительный анализ позволяет сделать обоснованный выбор алгоритма в зависимости от приоритетов (скорость, точность, интерпретируемость) при построении систем обнаружения вторжений. Разработанный подход и выводы могут быть использованы для повышения эффективности систем кибербезопасности.

РЭФЕРАТ

Дыпломная работа 44 стар., 9 мал., 1 табл., 13 крын., 1 дад.

СЕТКАВЫ ТРАФІК, МАШЫННАЕ НАВУЧАННЕ, КІБЕРБЯСПЕКА, ВЫЯЎЛЕННЕ ЎВАРВАННЯЎ, КЛАСІФІКАЦЫЯ АТАК, PYTHON, SCIKIT-LEARN, XGBOOST, CIC-IDS-COLLECTION, DECISION TREE, SVM, RANDOM FOREST, ADABOOST, MLP

Аб'ектам даследавання з'яўляецца працэс выяўлення пагроз у сетка-вым трафіку з выкарыстаннем метадаў машыннага навучання. Мэта работы — даследаванне і параўнанне эфектыўнасці розных алгарытмаў машыннага навучання для задачы выяўлення пагроз у сеткавым трафіку.

Праведзена даследаванне і параўнанне шасці алгарытмаў машыннага навучання (Decision Tree, Support Vector Machine, Random Forest, AdaBoost, XGBoost, MLP) для класіфікацыі сеткавага трафіку на нармальны і шкоднасны. Распрацаваны праграмны прататып на мове Python з выкарыстаннем бібліятэк Pandas, NumPy, Scikit-learn, XGBoost і Matplotlib для правядзення эксперыментаў на агульнадаступным наборы даных CIC-IDS-Collection. Выканана папярэдняя апрацоўка даных, якая ўключае ачыстку, нармалізацыю і падзел на навучальную і тэставую выбаркі. Навучаны і пратэставаны мадэлі, іх эфектыўнасць ацэнена па метрыках Accuracy, Precision, Recall і F1-score, а таксама па часе навучання і хуткасці прагназавання.

Атрыманыя вынікі паказалі высокую эфектыўнасць ансамблевых метадаў, у прыватнасці XGBoost і Random Forest, у задачы выяўлення сеткавых пагроз, прадэманстраваўшы лепшыя паказчыкі па дакладнасці, паўнаце і F1-меры. Праведзены параўнальны аналіз дазваляе зрабіць абгрунтаваны выбар алгарытму ў залежнасці ад прыярытэтаў (хуткасць, дакладнасць, інтэрпрэтаванасць) пры пабудове сістэм выяўлення ўварванняў. Распрацаваны падыход і высновы могуць быць выкарыстаны для павышэння эфектыўнасці сістэм кібербяспекі.

ABSTRACT

Thesis 44 pages, 9 figures, 1 table, 13 references, 1 appendix.

NETWORK TRAFFIC, MACHINE LEARNING, CYBERSECURITY, INTRUSION DETECTION, ATTACK CLASSIFICATION, PYTHON, SCIKIT-LEARN, XGBOOST, CIC-IDS-COLLECTION, DECISION TREE, SVM, RANDOM FOREST, ADABOOST, MLP

The object of the research is the process of detecting threats in network traffic using machine learning methods. The aim of the work is to research and compare the effectiveness of various machine learning algorithms for the task of detecting threats in network traffic.

Research and comparison of six machine learning algorithms (Decision Tree, Support Vector Machine, Random Forest, AdaBoost, XGBoost, MLP) were conducted for classifying network traffic into normal and malicious. A software prototype was developed in Python using Pandas, NumPy, Scikit-learn, XGBoost, and Matplotlib libraries to conduct experiments on the publicly available CIC-IDS-Collection dataset. Data preprocessing was performed, including cleaning, normalization, and splitting into training and testing samples. Models were trained and tested, and their effectiveness was evaluated using Accuracy, Precision, Recall, and F1-score metrics, as well as by training time and prediction speed.

The obtained results showed high efficiency of ensemble methods, particularly XGBoost and Random Forest, in the task of network threat detection, demonstrating the best indicators for accuracy, recall, and F1-score. The conducted comparative analysis allows for an informed choice of algorithm depending on priorities (speed, accuracy, interpretability) when building intrusion detection systems. The developed approach and conclusions can be used to improve the effectiveness of cybersecurity systems.