УДК 004.8

Н. Р. Григорук

АНАЛІЗ МОЖЛИВОСТЕЙ ВИКОРИСТАННЯ НАЇВНОГО БАЙЄСОВОГО АЛГОРИТМУ У ЗАДАЧАХ КІБЕРБЕЗПЕКИ

Вінницький національний технічний університет

Анотація

Проведено аналіз застосування наївного байєсового алгоритму (НБА) у задачах кібербезпеки.

Ключові слова: кібербезпека, штучний інтелект, машинне навчання, методи машинного навчання, наївний байєсовий алгоритм.

Abstract

The article analyzes the application of the naive Bayesian algorithm (NBA) in cybersecurity problems.

Keywords: cybersecurity, artificial intelligence, machine learning, machine learning methods, naive Bayesian algorithm.

Вступ

У сучасному світі зростання технологічних досягнень супроводжується збільшенням небезпек у кіберпросторі. Актуальність штучного інтелекту та машинного навчання у контексті кібербезпеки стає все більш очевидною, оскільки ці технології здатні ефективно аналізувати великі обсяги даних і виявляти нові загрози. Наївний байєсовий алгоритм (НБА) є одним із основних методів машинного навчання, який забезпечує швидку класифікацію даних на основі ймовірності. Його простота та ефективність роблять його цінним інструментом у кібербезпеці, зокрема для виявлення шкідливого програмного забезпечення та аналізу мережевого трафіку. Завдяки своїй здатності адаптуватися до нових загроз, НБА допомагає підвищити рівень захисту інформаційних систем. У цій роботі буде детальніше розглянуто його застосування в кібербезпеці.

Результати дослідження

Наївний байєсовий алгоритм — це метод машиного навчання, що ґрунтується на теоремі Байєса, яка оновлює ймовірність гіпотези на основі нових даних. Основна ідея алгоритму — обчислення ймовірності належності об'єкта до певного класу, виходячи з частоти появи ознак. Ключове припущення — незалежність ознак, що спрощує обчислення, хоча на практиці це не завжди відповідає дійсності. [1]

Алгоритм ефективно використовує частотний аналіз для швидкої класифікації даних і є простим у реалізації, що робить його корисним для великих наборів даних. Алгоритм наївного байєса може швидко адаптуватися до нових загроз завдяки оновленню наборів даних. Наприклад, додавання нових даних про шкідливі програми дозволяє алгоритму "вчитися" на нових загрозах без необхідності повного перенавчання. Це важливо для кібербезпеки в умовах швидкої еволюції загроз. [2] Однак його точність може знижуватися через припущення незалежності ознак, що є його основним недоліком.

Наївний байєсовий алгоритм знаходить широке застосування в різних аспектах кібербезпеки завдяки своїй простоті, швидкості та ефективності. Цей алгоритм виконує важливу роль у класифікації загроз і шкідливого програмного забезпечення, адже дозволяє аналізувати різні види загроз, таких як віруси та трояни, на основі їхніх характеристик. Наприклад, використання даних про відомі зразки шкідливого програмного забезпечення дозволяє навчити модель, здатну класифікувати нові зразки за їх ознаками, такими як шляхи доступу до системи, використовувані порти та поведінку в мережі. Це значно пришвидшує виявлення нових загроз та зменшує час реагування на атаки.

Ще однією важливою функцією НБА є аналіз мережевого трафіку для виявлення аномалій, що може свідчити про злом або зловмисну активність. Алгоритм може навчитися розпізнавати звичайний трафік певної організації, а потім використовувати цю інформацію для виявлення незвичних патернів, які можуть свідчити про атаки типу DDoS або спроби несанкціонованого доступу. Завдяки своїй простоті, НБА легко інтегрується в існуючі системи моніторингу трафіку. [3]

Наївний байєсовий алгоритм є одним із ефективних підходів для виявлення шкідливих веб-сайтів.

Він використовується для класифікації сторінок на основі певних характеристик URL, таких як орфографічні помилки, наявність надмірної кількості спеціальних символів або підозрілих HTML-тегів. У проаналізованому дослідженні було розроблено систему, яка використовує наївний байєсовий алгоритм для аналізу як безпечних, так і шкідливих сторінок, що дозволяє виявляти загрози на етапі доступу користувача до сайту. [4] Тестування на вибірці з 7,000 веб-сайтів показало високу точність моделі, яка досягла 99.5% у розпізнаванні шкідливих сторінок, що підтверджує її ефективність для застосування у реальних умовах кібербезпеки. НБА, завдяки своїй швидкості та низьким вимогам до ресурсів, є ідеальним для використання в реальному часі, проте його точність може страждати через залежності між ознаками. Інші алгоритми, такі як K-Nearest Neighbors і Support Vector Machine , хоч і забезпечують вищу точність у складних завданнях, потребують більше часу на навчання та можуть бути менш ефективними при обробці великих обсягів даних. Вибір алгоритму залежить від специфіки задачі, проте НБА залишається потужним інструментом у кібербезпеці, коли важливі швидкість і простота. [3]

В оглянутому дослідженні було проаналізовано приклад класифікації кіберзагроз на основі глобальних мережевих атак, де для аналізу атак, зокрема програм-здирників, у часовому проміжку 2020–2023 років використовувався метод наївного Байєса. [2] У процесі класифікації було проведено дослідження різних типів загроз, таких як фішинг, DDoS-атаки та програми-здирники. Результати дослідження показали, що метод наївного Байєса дозволяє ефективно класифікувати атаки та виявляти закономірності, що сприяє підвищенню рівня мережевої безпеки та запобіганню вразливостям.

Наївний байєсовий алгоритм демонструє високу ефективність у кібербезпеці. Його швидкість і простота роблять його придатним для роботи з великими обсягами даних, однак припущення незалежності ознак може знижувати точність в умовах корельованих даних. Незважаючи на цей недолік, НБА залишається корисним інструментом для реальних кібербезпекових задач.

Висновки

Наївний байєсовий алгоритм є ефективним інструментом у кібербезпеці завдяки своїй простоті, швидкості та здатності адаптуватися до нових загроз. Його висока точність у класифікації шкідливого програмного забезпечення та аналізі мережевого трафіку підтверджує його практичну цінність.

Проте, обмеження алгоритму, пов'язані з припущенням незалежності ознак, можуть знижувати точність в умовах корельованих даних. Подальші дослідження можуть зосередитися на вдосконаленні НБА шляхом інтеграції з іншими методами або застосуванням нових підходів до навчання, що підвищить його ефективність у боротьбі з еволюційними загрозами кіберпростору.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Berrar D. Bayes' Theorem and Naive Bayes Classifier. Encyclopedia of Bioinformatics and Computational Biology. 2019. P. 403–412. URL: https://doi.org/10.1016/b978-0-12-809633-8.20473-1 (date of access: 25.09.2024).
2. Sandi Mutia A., Irawan I., Juliane C. Global Network Cyberattack Classification Using Naive Bayes Method Time Range 2020 – 2023. ASTONJADRO. 2024. Vol. 13, no. 2. P. 587–596. URL: https://doi.org/10.32832/astonjadro.v13i2.15683 (date of access: 25.09.2024).
3. Predicting Cybersecurity Risk Severity Using Bayesian-Based Machine Learning. ISACA. URL: https://www.isaca.org/resources/isaca-journal/issues/2020/volume-6/predicting-cybersecurity-risk-severity-using-bayesian-based-machine-learning#:~:text=Step%201:%20Calculate%20the%20Conditional (date of access: 25.09.2024).
4. Magdacy Jerjes A. Z. A., Dawod A. Y., Abdulqader M. F. Detect Malicious Web Pages Using Naive Bayesian Algorithm to Detect Cyber Threats. Wireless Personal Communications. 2023. URL: https://doi.org/10.1007/s11277-023-10713-9 (date of access: 25.09.2024).

Григорук Надія Романівна – студентка групи 2БС-22б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: grigoruknadia15@gmail.com

Науковий керівник: ***Гарнага Володимир Анатолійович*** – канд. тех. наук, доцент кафедри захисту інформації, Вінницький національний технічний університет, м. Вінниця

Hryhoruk Nadiia - student of group 2BS-22b, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: grigoruknadia15@gmail.com

Supervisor: ***Harnaha Volodymyr*** – Cand. Sc. (Eng), Assistant Professor of information technologies and computer engineering, Vinnytsia National Technical University, Vinnytsia.