Анализ фишинговых угроз моделями машинного и глубокого обучения

Дан датасет фишинговых ссылок
https://www.kaggle.com/datasets/taruntiwarihp/phishing-site-urls

Необходимо реализовать классификацию данных угроз с использованием машинного и глубокого обучения

Фишинговые атаки являются одной из самых распространённых киберугроз, направленных на кражу личных данных пользователей. В этом проекте вам предстоит разработать систему классификации URL-адресов, способную определять, является ли ссылка фишинговой или безопасной, используя методы машинного и глубокого обучения.

Цели проекта

1. Провести анализ и предобработку датасета с URL-адресами.
2. Выявить и преобразовать признаки, необходимые для обучения моделей.
3. Обучить и сравнить различные модели машинного обучения (например, Logistic Regression, Random Forest, XGBoost).
4. Разработать нейросетевую модель для классификации URL-адресов.
5. Оценить производительность моделей и выбрать оптимальный подход.

Описание датасета:

- Содержит два класса: фишинговые (1) и легитимные (0) URL-адреса.
- Включает набор признаков, таких как длина URL, наличие специальных символов, использование HTTP/HTTPS и другие характеристики.