Хакатон по обнаружению фишинга

Содержание

Введение

Дисциплина: Применение ML в кибербезопасности
Практическая работа 4: Хакатон на платформе Cyberhackai
Тема: Методы обеспечения безопасности ML
Преподаватель: Юрий Иванов

Оценивание практической работы:

- все участники, получившие оценку работы модели (при точности не менее 85%), получают 10 баллов
- три лучших результата (призовые места) получают максимальный балл за экзамен (4 балла)

Критерии оценивания:

1. Задание считается выполненным, если зафиксирован полученный результат оценки работы модели и точность не менее 85 %
2. Задание считается невыполненным, если результат работы модели не отмечен в личном кабинете на платформе проведения хакатона

Даты проведения:

Начало: **08.10.2025 17:00** по Москве
Окончание и публикация таблицы лидеров: **20.10.2025 23:59** по Москве
Объявление победителей на вебинаре: **22.10.2025**

Глоссарий

Хакатон (англ. hackathon — от hack и marathon) — это мероприятие соревновательного типа, с ограниченным количеством времени и ограниченным техническим заданием

Фи́шинг/фишинговая атака (англ. phishing от fishing «рыбная ловля, выуживание») - один из видов интернет-мошенничества, целью которого является получение доступа к конфиденциальным данным пользователей - логинам, паролям, данным лицевых счетов и банковских карт

Фишинговый сайт — это сайт, который полностью или частично скопирован с оригинального, но таковым не является.
Целью таких сайтов является хищение логина и пароля, который вы используете на оригинальном сайте

Фишинговые ссылки — это ссылки на мошеннические интернет-ресурсы, чаще всего на копии сайтов известных организаций, банков, интернет-магазинов, социальной сети и т. д.

Обычно такие ссылки вам присылают по почте или в личном сообщении, например, в социальных сетях.
Они могут быть:

- прямыми — это ссылки, переходя по которым вы попадаете на сайт, почти ничем не отличающийся от настоящего
- с редиректом (переадресацией), переходя по которым вы перенаправляетесь на другие сайты и, в конечном итоге, попадаете на ресурс мошенников

Фишинг представляет собой простейший способ кибератаки, который, тем не менее, является одним из самых опасных и эффективных

Пример фишинговой атаки:
Вы получили электронное письмо от интернет-магазина, в котором часто делаете покупки: «Подтвердите свой аккаунт, чтобы продолжать пользоваться бонусами».
Вы переходите по ссылке из письма, заново вводите свои личные данные и данные банковской карты.
Затем Вас просят сделать «пробный платеж» на 1 рубль.
В ходе оплаты надо ввести трехзначный код с обратной стороны карты. Как только Вы вводите этот код, Вам приходит сообщение от банка о списании со счета, но не 1 рубля, а 10 000 рублей.

Проблема

На Вашу почту пришло «уникальное предложение забрать приз», «Ваш e-mail был выбран в качестве победителя», «адвокат дальнего родственника из-за рубежа предлагает связаться». Или Вы получаете странную ссылку от друзей в мессенджере… Всем знакомы эти ситуации

Цель задания

Научиться разрабатывать модели для обнаружения и анализа фишинговых сайтов

Задание

Разработать предиктивную модель, способную с высокой долей вероятности классифицировать фишинговые и безопасные сайты по их url адресам.
В том числе с помощью алгоритмов машинного обучения обнаружить различные техники, применяемые злоумышленниками:

- тайпсквоттинг (намеренные опечатки)
- мимикрия под безопасный сайт
- использование ключевых слов в URL и др.

Задание выполняется на платформе Cyberhackai
Участники хакатона получают доступ к фишинговым и безопасным URL адресам, собранным компанией АВ Софт.
Данные представлены в формате csv.

Этапы выполнения задания

Этап 1
Ознакомьтесь с Инструкцией и правилами пользования платформой Cyberhackai

Этап 2
Скачайте наборы данных для хакатона по ссылке

Этап 3
Анализ набора данных
Ознакомьтесь с описанием набора данных

1. Обучающий набор >64000 строк (Таблица 1) — файл train.csv
   В качестве набора данных вам будут предоставлены:
- набор URL адресов
- метки, соответствующие безопасному или вредоносному сайту
2. Тестирующий набор 16000 строк (Таблица 2) — файл test.csv

Содержит только набор URL адресов

3. Образец представления данных в правильном формате 16000 строк — файл sample_submission.csv
   Файл содержит следующие колонки:

- Id - ID записи
- url - URL- адрес сайта (обратите внимание, что url может быть отображен на различных языках, может быть без схемы http/https или представлять собой ip адрес)
- Predicted - Метка для каждого URL (1 и 0)
  (1 — URL- адрес является фишинговым, 0 — URL- адрес является безопасным)

Этап 4
Обучение модели
Поставленная ML задача решается через достижение максимальной метрики на валидационном множестве
Для обучения используйте train.csv
Разбейте файл train.csv на две части: на обучающую и валидационную части
Проведите обучение модели с использованием train.csv

Этап 5
Подготовка ответов
1. Получите первую модель для обнаружения фишинга
2. Выполните предсказание на файле test.csv
3. Сформируйте свой файл ответов в формате sample_submission.csv

Этап 6
Загрузка ответов и оценка работы модели на публичном наборе
1. Перейдите на платформу Cyberhackai
2. Выберите категорию участника “Старшая группа”

Обратите внимание: на главной странице хакатона вы можете:

- *загрузить полученные результаты*
- *посмотреть таблицу загруженных результатов*
- *познакомиться с таблицей лидеров соревнований*

3. Используя персональный токен, перейдите в личный кабинет

4. Загрузите свой файл решения на платформу Cyberhackai для оценки своего результата на публичной части тестового набора

***Этап 7**
Улучшение результата работы модели

При желании улучшить результаты повторяйте действия этапов 4 - 5. Учитывайте, что доступны 2 загрузки данных в сутки

Этап 8
Выбор данных для оценки работы модели
Выберите в личном кабинете на платформе Cyberhackai данные для оценки финального результата до окончания соревнования (можно выбрать только 2 варианта).
В случае, если вы явно не укажете (не выберете), какие решения использовать для приватной оценки, система автоматически выберет 2 лучших решения на основании публичной оценки

Этап 9
Дождитесь окончания соревнования для получения итогового результата

Оценивание модели

Показатель оценки — categorical accuracy
Для оценки используется:
- публичный рейтинг (30% от test)
- приватный рейтинг (70% от test)

Приватный рейтинг будет доступен после закрытия соревнования. Участники должны из всех своих решений отметить 2 решения, которые будут использоваться для расчета приватного рейтинга, в противном случае система автоматически выберет решение с максимальным публичным рейтингом

Полезные советы

- В качестве базового решения задачи участники могут взять следующую тетрадку
- Попробуйте разные алгоритмы: SVM, RandomForest, Catboost или нейронные сети
- Для табличных данных, как правило, лучше подходят ансамблевые и бустинговые методы (RandomForest/Catboost), которые дают возможность получить дополнительно до 10-15% к точности

- Используйте кроссвалидацию или алгоритмы поиска для подбора параметров или для выбора алгоритма. Правильный подбор параметров даст вам еще до 5%

- Для повышения качества модели участникам разрешено использовать сторонние данные, если это не нарушает права третьих лиц