**Алгоритм обучения**

1. **Загрузка данных:**

Используется база данных, содержащая информацию о URL и их метках (Label, безопасный – 0, фишинговый – 1)

1. **Подготовка данных:**

Отбрасывается колонка Domain, так как она не используется для обучения. Выделяются целевые метки (Label) и признаки (остальные колонки).

1. **Разделение данных**:

Данные делятся на обучающую и тестовую выборки в соотношении 70:30. Используется функция train\_test\_split с фиксированным random\_state для воспроизводимости.

1. **Обучение модели**:

Применяется классификатор XGBClassifier из библиотеки XGBoost.

Гиперпараметры:

random\_state=42: для воспроизводимости.

n\_estimators=100: количество деревьев в ансамбле.

Модель обучается на обучающей выборке (X\_train, y\_train).

1. **Оценка модели**:

Прогнозируются метки для тестовой выборки (X\_test). Вычисляется точность (accuracy\_score) и выводится подробный отчет классификации (classification\_report), включающий: Precision, Recall, F1-score и Support для каждого класса.

**Алгоритм определения фишинговых сайтов**

1. **Извлечение признаков из URL**:

URL разбивается на части (с использованием split('/')).

Вычисляются следующие признаки:

**Have\_IP**: Проверка наличия чисел в домене.

**Have\_At**: Наличие символа @ в URL.

**URL\_Length**: Длина URL.

**URL\_Depth**: Количество сегментов в URL.

**Redirection**: Наличие нескольких редиректов (//).

**https\_Domain**: Проверка, начинается ли URL с https.

**TinyURL**: Длина домена меньше 10 символов.

**Prefix/Suffix**: Наличие дефиса (-) в домене.

**DNS\_Record**: Существование DNS-записи домена.

**Web\_Traffic**: Фиксированное значение 1

**Domain\_Age**: Возраст домена в годах (вычисляется с помощью библиотеки whois).

**Domain\_End**: Проверка окончания домена на .com, .org, .net.

**iFrame, Mouse\_Over, Right\_Click, Web\_Forwards**: Устанавливаются в 0.

1. **Классификация URL**:

Признаки, извлеченные на предыдущем этапе, преобразуются в DataFrame.

Выполняется прогноз модели model.predict, возвращающий 0 (безопасный) или 1 (фишинговый).

1. **Результат классификации**:

Если прогноз равен 1, URL классифицируется как **фишинговый**.

Если прогноз равен 0, URL классифицируется как **безопасный**.

Для визуализации предоставляются подробные признаки в формате ключ: значение.