Detailed project plan

NFT Fraud Detection System

Project Supervisor: Dr. Yuen, John T.H.

Chan Tsz Hei, Victor (3035692060)

Choi Yik Ho, Edward (3035684415)

2 October 2022
Table of Contents

1. Introduction .................................................................................................................................................. 3
   1.1. Background ............................................................................................................................................. 3
   1.2. Existing solutions ................................................................................................................................. 3
       1.2.1. Optic engine .................................................................................................................................. 3
       1.2.2. Alchemy API ................................................................................................................................ 3
   1.3. Industry partner - DTTD ..................................................................................................................... 3
2. Objectives ..................................................................................................................................................... 4
   2.1. Systematic NFT Scam detection ........................................................................................................... 4
   2.2. Discover patterns of scams .................................................................................................................. 4
3. Project Methodology .................................................................................................................................... 4
   3.1. Phase 1: System/infrastructure Design ............................................................................................... 4
       3.1.1. Data Collection ............................................................................................................................... 5
       3.1.2. Dashboard ....................................................................................................................................... 5
   3.2. Phase 2: Model Training and Optimization .......................................................................................... 6
       3.2.1. Machine Learning Approach ....................................................................................................... 6
       3.2.2. Classification States ...................................................................................................................... 6
       3.2.3. Variables to Consider .................................................................................................................... 6
       3.2.4. Validation agent ............................................................................................................................ 7
5. Timeline ....................................................................................................................................................... 7
6. Proposed work distribution table ................................................................................................................ 7
7. Reference ....................................................................................................................................................... 8
8. Appendix ....................................................................................................................................................... 9
1. Introduction

1.1. Background
With many well-known artists and musicians jumping on board, Non-Fungible Tokens (NFTs) are quickly growing in popularity as a means of purchasing and selling digital artwork [1]. However, the widespread theft and fraud that these blockchain-based collectibles are fueling are difficult for artists and collectors to keep up with.

Established and emerging artists alike have expressed dissatisfaction about their publicly accessible works being turned into NFTs and peddled to buyers over the past year [2]. Due to the pseudo-anonymity offered by public blockchains, a new breed of opportunists has evolved who tokenize the efforts of others into NFTs without their consent.

Artwork theft, artist impersonation, and the absence of NFT collection verifications are all issues without clear-cut solutions. NFT collectors and artists are at the mercy of centralized services to provide much-needed protection. To ensure the safety and fairness of NFT transactions, we suggest an NFT classification system to detect theft, impersonations and scams.

1.2. Existing solutions
As of the time of writing, there are some tools for combating against fraudulent NFTs. Yet, due to certain reasons, they may not be the best solutions. Optic engine and Alchemy API are two examples.

1.2.1. Optic engine
The Optic engine is a copymint detection engine processing millions of NFTs minted each day, adding up to a total of 2TB metadata. They are compared to existing NFT collections in respect of visual similarities, including flips and color changes [3]. Optic then reveals a percentage showing how likely the artwork is a counterfeit based on the results. However, copyminting is allowed by some projects to increase their popularity. For example, the Creative Commons Zero (CC0) license, a tool moving artwork to the public domain by allowing the owner to give up their copyrights, was used by Goblintown to encourage the emergence of derivatives [4].

1.2.2. Alchemy API
Alchemy offers a public API that helps to filter spam NFTs. They provide a list containing the addresses of smart contracts on Ethereum having been classified as spam. The API also checks whether a particular NFT originates from any of the suspicious contracts in the list. Currently, the API has identified more than 5000 spam contracts [5]. However, the classifications are still in beta testing and contain many false positives and lack accuracy upon DTTD’s investigation.

1.3. Industry partner - DTTD
The proposed system will be created in collaboration with our industry partner, DTTD (Dotted Company Limited), a local start-up developing a mobile-first NFT social platform, which will offer the technical infrastructure and data for the development [6].

DTTD aims to provide a safe and user-friendly platform for web3 natives and new users alike, so even inexperienced NFT users can feel safe using the app without having to worry about scams. To achieve this, the classification system will incorporate into the DTTD infrastructure to enable functions such as spam filtering in user feeds.
2. Objectives
Given the problems caused by fraudulent NFTs, this project aims to build an NFT scam detection system and find the patterns or relevant scams.

2.1. Systematic NFT Scam detection
As existing solutions are not holistic enough and only focus on narrow aspects of NFTs, there is a need for a comprehensive classifier that considers more factors. The focus of the project is to create a pipeline/system to support the integration of an online learning classifier to correctly detect and flag suspicious NFT collections using a wide range of data sources (to be discussed at the classifier section) [7].

2.2. Discover patterns of scams
After the development of NFT contract classifiers, we also aim to discover the contributing factors and variables that determine the legitimacy of NFT contracts or even owners. By establishing such correlations, the information could be used by NFT users to spot problematic contracts.

3. Project Methodology
This project will be divided into two phases. The first phase will focus on laying the infrastructural foundation (such as data collection, building dashboards) to pave the way for the second phase, which is dedicated to model training and optimization.

3.1. Phase 1: System/infrastructure Design
Fig. 1 Workflow of the system design (subject to change)

The focus of phase 1 will be infrastructure design. Fig. 1 describes how data are fetched, processed, transmitted and used to train models.
3.1.1. Data Collection

<table>
<thead>
<tr>
<th>Steps</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Amazon Kinesis Data Streams (KDS), a serverless real-time data streaming service, fetches blockchain live data including ERC721 transferred, transaction time and marketplace from blockchain tracking platforms such as etherscan [8].</td>
</tr>
<tr>
<td>2</td>
<td>Lambda functions process the record fetched by KDS, converting the record format to a desired one. Functions can keep running regardless of the status of users’ devices due to their serverless nature.</td>
</tr>
<tr>
<td>3</td>
<td>KDS delivers the reformatted data to the designated Amazon S3 bucket, such data are later exported to DynamoDB, a scalable NoSQL database applying encryption at rest to secure sensitive data and reduce the operational burden involved [9].</td>
</tr>
<tr>
<td>4</td>
<td>Lambda function reads and exports data from DynamoDB to csv files in S3.</td>
</tr>
<tr>
<td>5</td>
<td>Python web scraping code built in a local IDE is uploaded to lambda as a lambda function, which is connected to the dynamicDB using boto3, an AWS SDK for python [10].</td>
</tr>
<tr>
<td>6</td>
<td>Lambda function stores the web scraping output to DynamoDB, which are then exported to csv files in S3.</td>
</tr>
</tbody>
</table>

3.1.2. Dashboard

![DashboardFig](image)

*Fig. 2 A prototype of the proposed dashboard using dummy data*

An internal dashboard will be developed, serving as the front-end of the pipeline. It will be used by DTTD to monitor and diagnose the models by manually overriding classifications if needed, and to create label new data.

Owing to the security concerns arising from an online dashboard connected to DTTD’s database, the deployment approach will be decided with more discussion with the company.
Initially, it will serve as an interface for validators to label unclassified NFTs. As the model matures, the interface will also allow override of classifications from the model output. As a result, incremental learning of the classifier can be achieved with an iterative approach of training and evaluation.

3.2. Phase 2: Model Training and Optimization

In the second phase of the project, the training and optimizing of a classifier and validation agent will be the main objective.

3.2.1. Machine Learning Approach

The flagging of problematic contracts requires manual labeling, and most contracts have unknown safety parameters, so a semi-supervised approach is justified. This approach combines a small amount of labeled data with a large amount of unlabeled data during training. Semi-supervised learning falls between unsupervised learning (with no labeled training data) and supervised learning (with only labeled training data) [11].

As there are new NFT contracts every day, an online learning approach is suitable to iteratively update the model with new information. When data becomes available in sequential order and is used to update the best predictor for future data at each step, an online learning approach is more appropriate as opposed to batch learning techniques which generate the best predictor by learning on the entire training data set at once.

3.2.2. Classification States

Currently, DTTD is relying on simple heuristics to classify NFT contracts into classes as

<table>
<thead>
<tr>
<th>Classification</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safe/Normal</td>
<td>Contracts deemed safe</td>
</tr>
<tr>
<td>Suspicious</td>
<td>E.g. Potential theft, phishing</td>
</tr>
<tr>
<td>Scam</td>
<td>E.g. Impersonation, malicious contracts</td>
</tr>
<tr>
<td>Unknown</td>
<td>For other unlabeled contracts</td>
</tr>
</tbody>
</table>

Our classifier aims to improve the classification by creating subclassifications to provide more explainable labels, such as theft under suspicious class, and impersonation under the scam class.

3.2.3. Variables to Consider

As we opt for a comprehensive approach, the following potential variables may be included to train our model:

a. Detect abnormal collection transaction activities (e.g. wash trading)

b. Detect “Copyminting” with text/image-based similarity detection

c. Social legitimacy: Twitter followers, discord members and activeness

d. Detect anomalies in collection homepage: (phishing, suspicious wallet connection prompts)

e. Collection source code: open source, bytecode size, correctly implemented methods, ERC1155/ERC721 token standards

f. Classification from other services
3.2.4. Validation agent
Besides training the NFT classifier, we plan to incorporate a validation agent that attempts to spot potentially misclassified collections from classifier output, in order to enhance the data labeling and model diagnoses process. By learning from patterns of previous reclassified collections, it would flag such collections for validators from DTTD to manually verify in the dashboard.

4. Conclusion
The popularity of NFTs has drastically changed how people transact assets, but it has also given rise to many artwork scams and thefts, which is undermining the benefits of creators and investors, as well as the stability of the NFT ecosystem.

Considering this, this project will serve as a detection system for fraudulent NFTs and catch the patterns of problematic contracts. The potential outcomes are expected to identify fraudulent tokens and prevent unauthorized activities.

5. Timeline

<table>
<thead>
<tr>
<th>Time Period</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>2022 Sept</td>
<td>Detailed project plan</td>
</tr>
<tr>
<td></td>
<td>Project research</td>
</tr>
<tr>
<td></td>
<td>Design website</td>
</tr>
<tr>
<td>2022 Oct</td>
<td>Process data</td>
</tr>
<tr>
<td></td>
<td>Build infrastructure</td>
</tr>
<tr>
<td>2022 Nov</td>
<td>Train and deploy prelim classifier</td>
</tr>
<tr>
<td>2022 Dec</td>
<td>Detailed interim report</td>
</tr>
<tr>
<td>2023 Jan</td>
<td>Finish Preliminary implementation</td>
</tr>
<tr>
<td></td>
<td>First presentation</td>
</tr>
<tr>
<td>2022 Feb</td>
<td>Optimize model</td>
</tr>
<tr>
<td>2022 Mar</td>
<td>Further code refinement</td>
</tr>
<tr>
<td>2022 Apr</td>
<td>Project exhibition &amp; final report</td>
</tr>
</tbody>
</table>

6. Proposed work distribution table

<table>
<thead>
<tr>
<th>Task</th>
<th>Chan Tsz Hei</th>
<th>Choi Yik Ho</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model training</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Model evaluation</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Web scraping</td>
<td>60%</td>
<td>40%</td>
</tr>
<tr>
<td>Website</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Data visualization</td>
<td>60%</td>
<td>40%</td>
</tr>
<tr>
<td>Report writing</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Project exhibition video</td>
<td>40%</td>
<td>60%</td>
</tr>
</tbody>
</table>
7. Reference


8. Appendix

Fig 3. Screenshot of DTTD in App store

Fig 4. Tentative procedure of data import/export and model training