SMU Data Science Review

Volume 6
Number 2 *Summer 2022*

Article 21

2022

Fraud Pattern Detection for NFT Markets

Andrew Leppla
Southern Methodist University, alleppla@gmail.com

Jorge Olmos
Southern Methodist University, jorge@jorgeolmos.com

Jaideep Lamba
jaideeplamba@gmail.com

Fraud Pattern Detection for NFT Markets

Andrew Leppla 1, Jorge Olmos 1, Jaideep Lamba 2
1 Master of Science in Data Science, Southern Methodist University,
Dallas, TX 75275 USA
2 Voice Systems Engineering, 1000 Northbrook Dr,
Feasterville-Trevose, PA 19053 USA
{aleppla, jeolmos}@smu.edu
jaideeplamba@gmail.com

Abstract. Non-Fungible Tokens (NFTs) enable ownership and transfer of digital assets using blockchain technology. As a relatively new financial asset class, NFTs lack robust oversight and regulations. These conditions create an environment that is susceptible to fraudulent activity and market manipulation schemes. This study examines the buyer-seller network transactional data from some of the most popular NFT marketplaces (e.g., AtomicHub, OpenSea) to identify and predict fraudulent activity. To accomplish this goal multiple features such as price, volume, and network metrics were extracted from NFT transactional data. These were fed into a Multiple-Scale Convolutional Neural Network that predicts suspected fraudulent activity based on pattern recognition. This approach provides a more generic form of time series classification at different frequencies and timescales to recognize fraudulent NFT patterns. Results showed that over 80% of confirmed fraudulent cases were identified by modeling (recall). For every predicted fraud case, the model was correct 50% of the time (precision). Investors, regulators, and other entities can use these techniques to reduce risk exposure to NFT fraudulent activity.

1 Introduction

Non-Fungible Tokens (NFTs) have emerged to establish ownership and enable transfer of digital assets using the blockchain. Along with cryptocurrencies, the emergence of these new digital assets has spurred a blockchain revolution which has generated much excitement. However, NFTs may be even more chaotic than cryptocurrencies due to much lower barriers of entry. This new asset class has been confusing and particularly hazardous to young or inexperienced investors.

Deficiencies in oversight from the NFT marketplaces and lack of government regulations have exposed investors to large-scale fraud. There are currently no widely

adopted solutions to curb this illicit activity. With the NFT market size approaching $25 billion as of 2021, this is a growing concern. Investors have lost an estimated tens of millions of dollars on various NFT scams [23].

There are generally three categories of scams: *Rug Pull* - this happens before a project gets started; *Wash Trading* - the action of buyers and sellers colluding to artificially inflate the trading volume or value of an asset; *Pump & Dump* - a scheme that promotes an asset leading to heavy retail buying which inflates the price, then fraudsters sell all their low-cost shares causing the stock to crash for everyone else.

Using NFT wallet transactions data, two of the major forms of scams in the NFT space can be classified for prediction, Wash Trading and Pump & Dump. There are underlying patterns in the transactions data that signal potentially fraudulent activity. Patterns over time in NFT price data and buyer-seller networks could indicate Pump & Dumps, Wash Trading, or other fraudulent activity. Until now, a general lack of data labels has prompted researchers to use unsupervised techniques to detect anomalous activity [10, 15, 21]. The main drawback to such approaches is the mere presence of an anomaly does not necessarily constitute fraudulent activity. Using these anomaly-based approaches for classification of fraud will likely inflate the type I error rate.

NFTs are grouped into collections that are generated and sold by specific creators. These collections are often sold on NFT markets such as OpenSea or AtomicHub. Major marketplaces attempt to discern credible collections from problematic ones by tagging them as whitelisted or blacklisted, respectively. Users can report any problematic collections for blacklisting, and collection owners can apply for whitelisting. These inputs are vetted and approved by each marketplace. A whitelisted collection is not immune to fraudulent activity, but it presents lower risk to investors. These lists provide an important source of labels to model potentially fraudulent activity in NFT transactional data.

Scammers are also known to exploit whitelisted collections by creating new collections with similar names that may fool an unwitting buyer to purchase a toxic asset. As an example, one of the top-traded whitelisted collections on AtomicHub is *alien.worlds* with over one million transactions. There are at least 55 blacklisted collections on AtomicHub with similar names: *a11ianworlds, a1ieneworlds, a1ienlwor1ds,* etc. Many of these were reported for blacklisting after only one trade on the AtomicHub market.

A well-established method for pattern recognition and classification is Convolutional Neural Networks (CNN). CNNs are commonly used for image recognition to detect patterns for classification. This approach can be abstracted and leveraged for Time Series Classification (TSC). NFT transactional data can be transformed and fed into such a model to aid in the classification of *blacklisted* or *whitelisted* collections. Transactional data over different time windows (1 hour, 1 day, 1 week, etc.) can be used with multiple CNNs to extract more features and improve

predictions. Prior research focused primarily on distance-based methods that measure similarities in price, volume, or network metrics for anomaly detection [10, 15, 24]. Most of these methods require some form of feature engineering. This research seeks to expand on those metrics and extract deeper features using CNN.

Many thousands of collections on the NFT markets are neither whitelisted nor blacklisted with new unlisted collections being added daily. Buyers are often left on their own to try to verify the credibility of these collections. Features can be extracted from the transaction data of collections that are already blacklisted or whitelisted to predict if these unlabeled collections would be blacklisted or not.

This research aims to use Multiple CNN modeling for fraudulent pattern detection in NFT transaction data to classify collections as *blacklisted* or *whitelisted*. Investors and other entities need better tools to identify legitimate collections or alert them to the possibility of potential scam or fraud. This could lay the groundwork for automated detection of suspicious activity and fraudulent schemes.

2 Literature Review

2.1 NFT and Blockchain Technology

Rehman et al. (2021) define NFTs as "digital assets that are representative of physical or digital creative work or intellectual property including music, digital art, games, gifs, video clips and more" [p.1, 16]. They are mostly based on the Ethereum blockchain but differ from the *fungible* cryptocurrency coins of the same brand name. Non-Fungible (NF) means that a Token (T) is unique, representing one specific object that is not interchangeable or exchangeable with any other NFTs.

NFTs are bought and sold via online markets. Some of the more popular NFT markets are: OpenSea, Axie Infinity, CryptoPunks, AtomicMarket, and PancakeSwap [16].

One enabling technology represented in NFT transactional data is smart contracts. Wang et al. (2021) note that, "Smart contracts enable unfamiliar parties and decentralized participants to conduct fair exchanges without a trusted third party" [p. 4 22]. All NFT transactions are facilitated via smart contracts.

Blockchain addresses are another key concept in NFT transactions. They are a user's unique identifier to buy and sell NFTs, conceptually similar to a bank account [22]. The bank analogy can also be extended to blockchain wallets which hold crypto coins and NFT assets.

Blockchain hash values are used to encode and link transactions such that they are uniquely and publicly traceable back to the creator [14, 22].

2.2 Prevalence of Fraud in NFTs and Cryptocurrencies

Crypto assets can be interpreted by the Securities and Exchange Commission (SEC) as investment contracts, but there are no rules directly addressing digital assets or NFTs. This is not an issue unique to the United States. In the UK, most NFTs and crypto currencies fall under the classification *utility token* or *unregulated token* vs. *regulated token*, thus limiting the power of regulatory bodies to take action against those exploiting this new asset class [9]. The fast-paced nature of the crypto industry makes matters even more challenging, leaving regulation and enforcement many steps behind.

The Wall Street Journal reported that over a six-month span they identified 175 pump-and-dump schemes from 121 digital coins, amounting to a total of $825 million in trading activity. Most of these schemes leveraged social media platforms. For example, one group through a popular messaging app promoted 26 schemes amounting to more than $222 million [18]. One can spot similar schemes in NFT marketplaces where much of the success of an NFT, or collections of NFTs, can be attributed to social media promotion. The most dangerous are where ownership of these assets are in the hands of a few owners, an example is the NFT collection *Loot* where the average price soared more than 31,000% [12]. More recently, the sale of fractionalized ownership of an NFT asset, akin to buying a portion of a piece of art, has boosted the value of assets and exacerbated the problem [12]. Just last year in 2021, a team from the University of Copenhagen examining the NFT wash trades concluded that around 4% of addresses, which account for 2% of the transaction sales, were suspected of wash trading abuse [21].

2.3 Market Manipulation Schemes

The most common fraudulent activities for crypto assets are *Pump & Dump*, *Wash-Trading*, and *Money Laundering* [9]. Many of these are well-known in traditional financial markets [21]. For a *Pump & Dump* there are three phases: the accumulation, the pump, and the dump [10]. Fraudsters accumulate an asset and then promote it for others to buy which artificially inflates its price (the pump), then they sell off the asset (the dump) which causes a massive drop off in price, leaving many victims with major losses. These fraudulent sub-patterns of the time series data are sometimes labeled as *Shapelets*. These *Shapelets* provide information for anomaly detection, time series classification, and fraud identification.

In contrast, wash trading involves one or more individuals colluding with each other to artificially raise the price through bid, auction, or direct sales to between accomplices. Their actions create misleading or false market data about the demand for

an asset [21]. Money laundering, like wash trading, uses a collection of addresses that trade between themselves but in this case is used to obfuscate the origin of illegal funds.

Existing research has often focused on one specific type of fraudulent activity. Kamps, J., & Kleinberg, B. (2018) used 20 days of data and focused on pump and dump schemes [10]. Pelechrinis et al. (2022) examined the network relationship of NFT trades to identify wash trading [15].

2.5 Transaction Network Analysis

Graph analysis of network data between buyers and sellers can provide insights into transactional data. Fraudulent patterns in the data can be explored using several graph network techniques, including Inferred Directed Weighted network analysis. Per Bratanic et al. (2022), "The direction of the relationship will indicate the flow of the money, while the weight will represent the [trading] volume" [3]. Circular or highly convoluted relationships with large weights in the network indicate potential wash trading behavior.

Transactional data can also be used to construct weighted graphs and generate metrics such as edge density, closure, node centrality and node degree distribution to detect patterns in the network over time [15]. Pelechrinis et al. (2022) compare the number of nodes (users) and edges (transactions) in the overall network vs. suspicious sub-networks to provide additional evidence for potentially fraudulent NFT trading patterns [15]. The suspicious sub-networks had a significantly different structure vs. the overall network and random sub-networks.

Hassanzadeh et al. (2012) lay some of the foundation for using these graphical network methods to detect fraudulent patterns, evaluating the relationships between metrics like the number of nodes and edges vs. betweenness centrality. *Betweenness Centrality* of a node is defined as, "the number of shortest paths between all pairs of nodes within that graph that go through that node" [p. 4, 8].

2.6 MCNN Modeling

Cui et al. (2016) propose a distinct method for time series classification using a Multiple Convolutional Neural Network (MCNN) [5]. This approach extracts more generalized features from multiple branches of convolutions at different time scales and frequencies. Previous methods often used *Euclidean Distance* or *Dynamic Time Warping* to deal with phase shifts in signals or use generated *Shapelet* features for classification. The challenge with these methods is that a separate ad-hoc feature extraction is needed to detect patterns for each and every time scale in the data [5]. In contrast, MCNN ingests more raw information that it uses to automatically extract the

pattern features at different scales and frequencies. Lastly, MCNN provides an advantage by identifying complex features and is more computationally efficient [5].

This research aims to find a more generalized NFT fraud detection model that combines many of the features identified in previous research. These features will be further enhanced by an MCNN model that can automatically extract features over varying time scales and windows. NFT collection *blacklist* and *whitelist* labels from the AtomicHub marketplace will enable supervised learning which should improve fraud detection vs. previous unsupervised learning research.

The hypothesis is that these labels and enhanced features will create a more comprehensive model that can accurately detect a broader range of fraudulent activity. This new model can serve as the groundwork for a future tool to be utilized by investors, marketplaces, and regulators to make more informed decisions in NFT markets.

3 Methods

3.1 Data

The dataset was from "Mapping the NFT revolution: market trends, trade networks, and visual features" by Nadini et al. (2021). Regarding the dataset, "Data are downloaded from different APIs, cleaned, and merged … Dataset is about the buyer-seller network only" [p. 1, 5]. An explanation of the columns in the dataset are in the Appendix. The following metrics were calculated from the data and used to derive features for modeling:

- *Volume*: Total count of transactions for a given interval for a given collection
- *Vertex_Count:* Total number of unique addresses transacting in a collection
- *Edge_Count:* Total number of connections between all the vertices
- *Density:* The ratio between the number of edges and the maximum number of edges it can contain. Measures how well a graph is connected.
- *Radius:* The minimum distance among all the maximum distances between all other vertices
- *Diameter:* The absolute maximum distance between any two vertices
- P*eriphery_Count:* The total number of connected vertices or sub-graphs that equal the graph diameter

3.2 Data Preparation

The transaction dataset was grouped by each collection. Given that the original dataset was not labeled, and a labeled dataset is required to implement a Time Series

Classification model, the original dataset was filtered using a list of blacklisted and whitelisted collections from the AtomicHub Marketplace. Of the total 6,283 collections in the dataset, there were 720 whitelisted and 275 blacklisted collections found to have labels. This yielded a total labeled dataset of 995 collections or roughly 15% of the original dataset.

Care was then taken to remove any smart contract addresses from the remaining buyers or sellers; this was a minor data quality issue identified during Exploratory Data Analysis (EDA). Transactions grouped by collection were then sorted in chronological order. At each time step network metrics (Vertex Count, Edge Count, Density, Centrality, Diameter, and Radius) were calculated, thus freezing in time how each collection looked at a given point in time and preventing data leakage.

This dataset, enriched with network data, was then resampled. In some cases, it was down-sampled to decrease the frequency of samples, and in other cases it was up-sampled to increase the frequency. Models used frequencies of 1 min., 10 min., 30 min. and 1 hour, each to be fed into the model. This was done to provide additional structure to the dataset and facilitate feature engineering at different timescales during modeling. To down-sample pricing information for both USD and Cryptocurrencies, the mean price was used for the newly aggregated values. For the generated network metrics, the latest value was taken as the aggregated value at each frequency. Volume was calculated by retrieving the count of transactions at each interval for each frequency. Lastly any missing data was up-sampled by padding or forward filling the last aggregated value. The resampled dataset was then passed through different low pass filters to smooth out each time series and help reveal low-frequency features. Stratified random sampling was used for splitting training and test sequences; this was done to address the imbalance of blacklisted and whitelisted collections.

3.3 MCNN Time Series Classification Modeling

For the task of time series classification, a Multi-Scale Convolutional Neural Network (MCNN) was used. This was a strategy proposed by Cui et al. (2016) that utilizes the concatenation of multiple 1D Convolutional Neural Networks to enhance the process of feature extraction at different time scales and frequencies [5]. Most methods for TSC require the use of *Dynamic Time Warping* which can be sensitive to differing time scales. Additionally, the generation of *Shapelets* or features for time series data is mostly an ad-hoc feature extraction process which is both prohibitively expensive and time consuming. MCNN modeling was selected as it addressed the previously stated drawbacks. The chosen modeling strategy aids in the identification of multiple forms of fraudulent activity such as *Pump & Dump* and *Wash Trading*. Beyond classification, the model also handles pattern detection on different timescales.

The dataset shows patterns of fraudulent activity could appear anywhere from a single day to multiple days to even weeks or months.

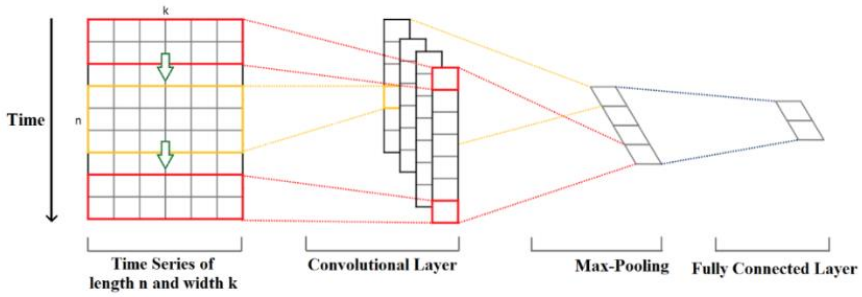

Fig. 1. Time series adaptation to a Convolutional Neural Network [7].

To begin modeling, each of the resampled datasets is run through a 1D CNN as shown in Fig. 1. The shape of the input is an n×k matrix where the size n is the total number of intervals for each time window and k is the number of features (e.g., price, volume, density). This input is convolved on a total of F number of filters (f1, f2, f3, f4) each with a kernel size of K (k1, k2, k3, k4) for each resampled frequency. This 1D CNN uses a RELU activation function. To extract the most important features a 1D Global Max Pooling was used with a pool length of (p1, p2, p3, p4) for each frequency type, respectively. The output was then passed to a 50-neuron dense layer that also uses a RELU activation function with each model having a 0.3 dropout rate to prevent overfitting.

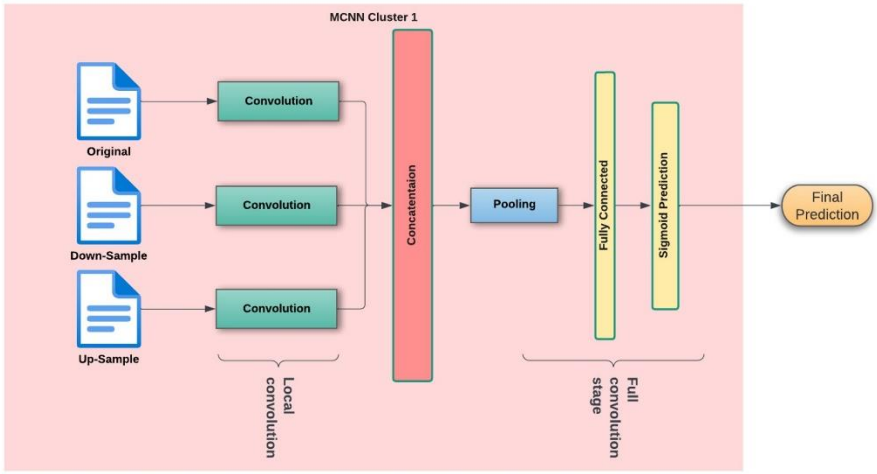

Fig. 2. Multiple Convolutional Neural Network

The output for each of the dense layers are in-essence the features extracted for each of the resampled and smoothed datasets. They are then concatenated into a final classification Neural Network (NN). This final NN is composed on n number of hidden layers with each layer having (s1, s2, s3) number of neurons. The final layer has only 1 neuron with a sigmoid activation function that provides the final prediction for *blacklisted* or *whitelisted*.

3.4 Clustering

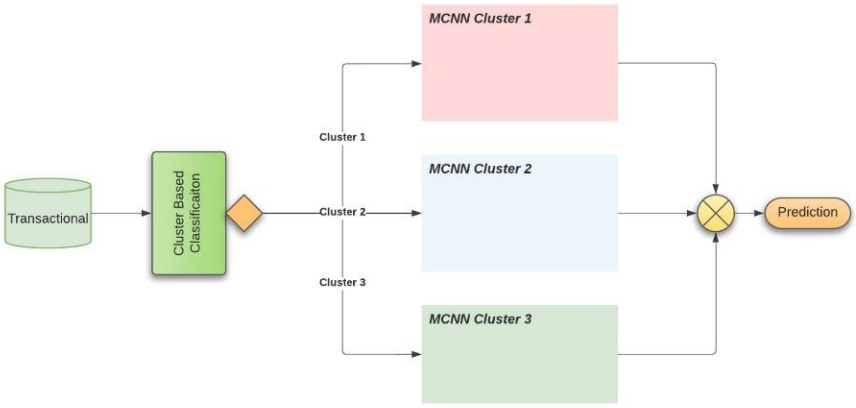

Fig. 3. Clustered Multiple Convolutional Neural Network

Time series clustering was used before the training data were fed into the MCNN to improve model performance. The first 5 days of data for each collection were used with a 30-minute sampling rate. This gave 240 rows of time series data per collection. First, Principal Component Analysis (PCA) was used for dimensionality reduction of the scaled features. Scaling was done so the PCA was based on the correlation between variables rather than the covariance.

K-means was used on the principal component time series data to cluster together collections with similar network patterns over time. Per Amidon (2020), "The most common approach to time series clustering is to flatten the time series into a table, with a column for each time index (or aggregation of the series) and directly apply standard clustering algorithms like k-means" [p. 2, 1].

Care was taken to prevent any data leakage into the test set. PCA was fit only to the training set and was then used to transform the test set. Similarly, k-means was run only on the training set, and the resulting cluster distance formula was then applied to the test set.

3.5 Metrics

Given the imbalance in the target variable classes (approximately 25:75), accuracy was not an appropriate primary metric for this classification problem. Instead, precision and recall were used to evaluate the models. An F1 score that equally weights both precision and recall was used as the primary metric to evaluate final model performance.

Precision and Recall can be tuned for any given model by adjusting the prediction probability threshold. Plotting Precision vs. Recall over the range of probability thresholds from zero to one gives a Precision-Recall Curve. The Area Under the Precision-Recall Curve (PR AUC) was used to evaluate any improvements in classification modeling. After maximizing the PR AUC, the probability threshold was tuned for the final model to maximize the F1 score.

After k-means clustering, each cluster was modeled separately using MCNN. Each cluster model applied a different probability threshold for classification to maximize the overall F1 score.

4 Results

Baseline metrics were generated using a naïve model that predicted all collections as blacklisted (1). These baseline results are summarized in Table 1 below. Models and features that increase PR AUC (Precision-Recall Area Under the Curve) are improving upon this naive classification of blacklisted vs. whitelisted.

Table 1 – Model metrics for baseline naïve classification vs. initial MCNN model.

Metric	Naïve Baseline - All Blacklisted	Initial MCNN Model - Test
PR AUC	0.276	0.520
F1 Score	43.3%	54.6%
Precision	27.6%	47.8%
Recall	100%	63.8%
Accuracy	27.6%	70.3%

Initial MCNN modeling showed promising results using limited features from just one day of trading data for each collection. Per the test results in the rightmost column of Table 1, the model had an F1 score of 54.6% and was able to correctly

identify 63.8% of all blacklisted collections, but it incorrectly predicted collections as blacklisted 47.8% of the time.

The test set PR AUC improved by approximately 88% vs. the naïve baseline model on a relative basis. Based on these preliminary results alone, this research was able to use MCNN modeling for fraudulent pattern detection in NFT transaction data to classify collections as *blacklisted* or *whitelisted*.

PCA and k-means time series clustering were used to further improve model performance. First, PCA was used for dimensionality reduction of the feature space. The first principal component (PC1) had high loading of network metrics (vertex count, max. diameter, etc.) and low loading of price metrics (in USD and crypto coin). In contrast to PC1, the second principal component (PC2) was primarily loaded with price metrics, both USD and Crypto, and third principal component (PC3) was loaded primarily with Time. These results are summarized in Table 2.

Table 2 – Principal Component Feature Loading for Training Set

Feature	**PC1**	**PC2**	**PC3**
Timestep	0.0463	0.0390	**0.7564**
Price_US	-0.0043	**0.9288**	-0.0420
Price_Crypto	0.0016	**0.9289**	-0.0396
volume	0.3589	0.0005	0.1483
Density	-0.5820	0.0045	0.0403
Vertex_count	**0.8445**	0.0323	0.3436
Edge_count	**0.7879**	0.0360	0.3936
Max_diameter	**0.8881**	-0.0239	-0.2616
Max_radius	**0.8699**	-0.0236	-0.2914
Max_periphery	0.5013	-0.0214	-0.3581

PC1 was the principal component most correlated with blacklisted collections in the training set and was used for time series clustering. PC2 and PC3 were also evaluated for clustering but yielded poor results compared to PC1. Collections with similar network patterns over 5 days were clustered together using PC1 and summarized in Table 3. k=5 was optimal, generating three larger-sized clusters. Clusters 4 and 5 had the lowest percentage of blacklisted collections in the training set but were too small for modeling with n=1 and n=9, respectively. These small clusters were combined with Cluster 3 which had a similarly low percentage of blacklisted collections.

The percent of blacklisted collections varied significantly by cluster per Table 3. This class separation between clusters was used to build a basic predictive model to quantify the information gained by clustering. The model simply predicted that a collection was blacklisted if it was in Cluster 1 or whitelisted if it was not in Cluster 1. The results are summarized in Table 4. Compared to the initial results in Table 1, the PR AUC and F1 score were better than the naive model but worse than the initial MCNN model.

Table 3 - k-means clustering on PC1 time series data (k=5)

	Training Set		Test Set	
Cluster ID	**#Collections**	**%Blacklisted**	**#Collections**	**%Blacklisted**
1	387	37.2%	126	40.5%
2	281	20.3%	88	15.9%
3 (4,5)	71 (1,9)	7.0%	32 (1,5)	12.5%

Similar data patterns were found through clustering that helped with class separation and prediction. This is illustrated in Figs. 4a to 3c with the network metric *max_diameter* which is the feature with the highest loading in PC1 per Table 2. The average curves of *max_diameter* vs. time have similar shapes within each cluster. Cluster 1 also shows significant class separation for blacklisted (0, blue) vs. whitelisted (1, red) for both the Training and Test sets. Class separation is also observed to some extent in Cluster 3 and the first 25 timesteps (the first 20 hours) for Cluster 2.

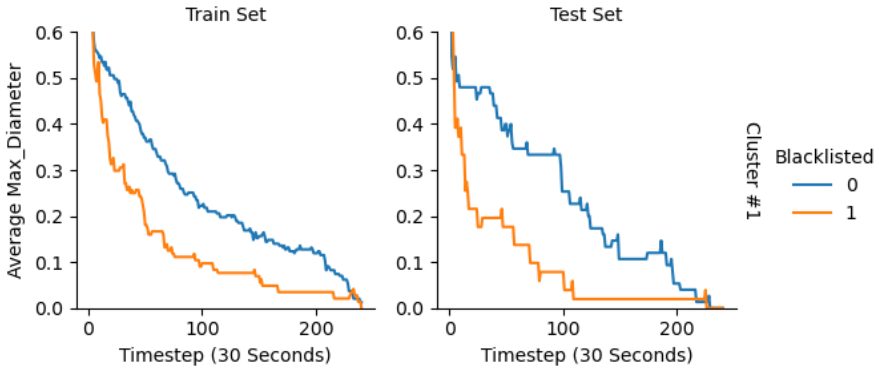

Fig. 4a - Average curves of *max_diameter* vs. time step for k-means Cluster 1 by blacklisted vs. whitelisted and Training vs. Test sets.

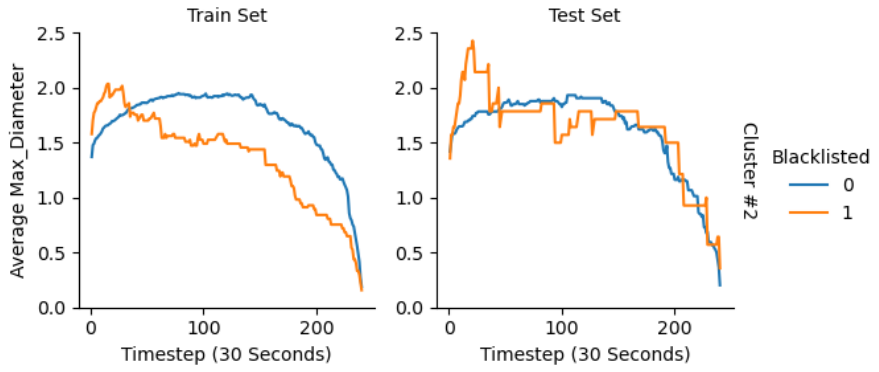

Fig. 4b - Average curves of *max_diameter* vs. time step for k-means Cluster 2 by blacklisted vs. whitelisted and Training vs. Test sets.

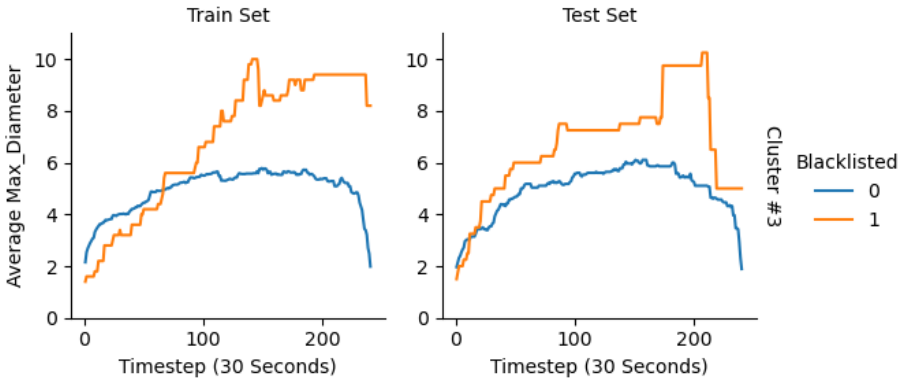

Fig. 4c - Average curves of *max_diameter* vs. time step for kmeans Cluster 3 by blacklisted vs. whitelisted and Training vs. Test sets.

The more consistent patterns and class separation within the k-means clusters likely made it easier to detect signals with MCNN. Each cluster of collections was trained separately with MCNN and then combined into the *Clustered MCNN* final model summarized in Table 4. Compared to the PR AUC values in Table 1, the final model was a 101% relative improvement over the naïve baseline model, and a 6.7% improvement over the *Initial MCNN* model (1 day of data and no clustering). After tuning the thresholds to maximize the F1 score, the final model had a 17% improvement in recall vs. the *Initial MCNN* model. The combination of clustering and MCNN was effective in improving the model.

Table 4 - Model metrics for simple k-means cluster model and final Clustered MCNN model.

Metric	**k-means Model - Test**	**Clustered MCNN - Test**
PR AUC	0.385	0.555
F-1 Score	52.3%	61.2%
Precision	40.5%	49.1%
Recall	73.9%	81.2%
Accuracy	62.2%	71.1%

The final model correctly identified over 80% of the blacklisted collections in the test set. For every predicted blacklisted collection, the model was correct 50% of the time.

This research hypothesized that a comprehensive model could be built to detect and predict a broad range of fraudulent NFT activity using MCNN. Based on the results achieved, the hypothesis has been confirmed.

5 Discussion

5.1 Fraud Risk Reduction

The transactional data from the first day(s) of an NFT collection are useful in predicting if a given collection may become blacklisted or whitelisted. This is a promising result for potential NFT buyers and investors that want to reduce their risk of losing money on fraudulent NFTs. Many thousands of collections on the NFT marketplaces are neither whitelisted nor blacklisted with new unlisted collections being added daily. With a predictive tool similar the one proposed, investors may have more confidence in their own due diligence research and don't have to rely solely on marketplaces to vet collections for them.

5.2 Research Challenges

One challenge in this research was pre-processing the data for more than a few thousand transactions per collection. Many collections had hundreds of thousands of transactions or more. It was computationally expensive to create and analyze the network graphs sequentially for all transactions to generate a time series representation without data leakage. This research addressed this challenge by focusing on the first one to five days of trading for each collection and multi-threading the network calculations. Future research may look for ways to make this network data pre-processing more efficient and/or utilize high performance computing.

Another challenge was the relatively small amount of labeled data available to train a CNN. Only 995 AtomicHub collections in the dataset were found to have labels, of which 739 were used for model training. The class imbalance made this more challenging with only 206 blacklisted collections in the training set. More labeled collections from AtomicHub or other NFT marketplaces would likely improve the model. Timestamps of when collections were reported as blacklisted or approved as whitelisted would be even better for time series modeling and classification. However, that data is not publicly available.

Other supplemental data for the model would be user or wallet level data. Any users or wallets associated with fraudulent activity could be used to help predict suspicious new collections. During EDA prior to modeling, a group of approximately

twenty suspected wash traders was identified that was acting across multiple collections that were not blacklisted.

5.3 Model Limitations and Further Improvements

A limitation of the model is it doesn't predict if a well-established whitelisted collection is suddenly hijacked by fraudsters. The model would have to be modified to ingest fresh data each day to detect any newly developing fraud patterns. The previously mentioned timestamps of when collections are reported as blacklisted would also help inform this model upgrade. Users and traders familiar with the whitelisted collection may notice and report suspicious activity of other users.

An unexpected result of this research was that increasing the number of trading days from 1 to 5 days without clustering did not significantly improve the model. There are diminishing returns as the number of trading days is increased, and the computational load for the network metrics increases significantly as previously mentioned. A potential improvement for the MCNN model with 5 days of trading data would be to change from global max pooling to max pooling (Fig. 1). This would increase the resolution of the data and may help detect more subtle patterns in the data.

For future research, Natural Language Processing (NLP) could be used in several ways to supplement the transactional model developed in this research. First, NLP could be used to detect pump & dump marketing or spam on social media before it manifests in the transactions. It could also help catch the special cases of fraud that exploit similar names to whitelisted collections like the *alien.worlds* example. This could help improve the model performance for any new collections with little or no transactional data.

5.4 Ethics

One ethical issue with fraudulent pattern prediction is that it is unavoidable to incorrectly predict some legitimate collections as blacklisted. For example, the model may find that a limited number of transactions are highly predictive of fraudulent collections. This may make it more difficult for new authentic artists to get traction for their collection. The other issue is incorrectly predicting some fraudulent collections as whitelisted. Any investor tools that use this modeling approach has a duty to make it abundantly clear to users that they are still responsible for doing their due diligence and they should not invest more than they can afford to lose.

Another ethical issue is that fraudsters could use this model to adapt their methods and evade detection. This is a common two-sided problem in fraud detection and prevention. Results of fraud detection studies should be published to inform the

professional community and public at large, but in doing so the information is made available to fraudsters as well.

Transactional data used in the proposed model is publicly available data. However, a bad actor could trace a wallet address to an individual and use the data for nefarious activities. This could include more effective targeted advertising for their pump & dump schemes or other unwanted marketing or publicity. It could even lead to defamation or blackmail. The wallet addresses serve as a first layer of anonymity. If the model was put into production, it should obfuscate the addresses during data ingestion to make the aggregated data more anonymous. This can be accomplished by simply assigned a new random key in place of the wallet address. The anonymized data would then be transformed into the format consumable by the model.

The MCNN model relies on the existing classification of blacklisted and whitelisted collections. The labeled dataset obtained from AtomicHub used for this research primarily relies on user reporting. However, there are few details available on the review process to determine if the claim is valid. The blacklisting and whitelisting process should be transparent and must include checks and balances to prevent exploitation. Users could make false reports to get their competitors blacklisted and hinder growth. Having clear rules in place should help reduce instances of false reporting. Establishing standards and transparency with the process improves the accuracy of reporting and builds confidence in the models trained with these datasets. That should translate to better modeling and fraud detection. Ultimately, a set of reporting standards should be adopted across all NFT markets, but this will likely require regulatory action.

6 Conclusion

In conclusion, a Multi-Scale Convolutional Neural Network (MCNN) model was used to predict if NFT collections were blacklisted or whitelisted based on transactional data. Investors, regulators, and other entities can use these techniques to reduce risk exposure to NFT fraudulent activity. The final model correctly identified over 80% of the blacklisted collections in the test set (recall). For every predicted blacklisted collection, the model was correct 50% of the time (precision). These results could be further improved by max pooling (vs. global max pooling), finding more labeled collections, or getting timestamps of when collections are reported as blacklisted or approved as whitelisted.

The MCNN models extracted features on varying time scales and windows for more automated robust detection of fraudulent activity. This modeling strategy reduces the overhead of feature engineering, and the model inputs used are broad enough that they can be applied to any collection using standard metrics (i.e. Price, Volume, and Graph Network).

Acknowledgments. Jacquelyn Cheun, PhD. – Capstone Professor

References

1. Amidon, A. (2020). How to Apply K-means Clustering to Time Series Data. Retrieved from https://towardsdatascience.com/how-to-apply-k-means-clustering-to-time-series-data-28d04a8f7da3
2. Ante, L. (2021). The non-fungible token (NFT) market and its relationship with Bitcoin and Ethereum. Available at SSRN 3861106.
3. Bratanic, T. (2022). Exploring the NFT transaction with Neo4j: Exploratory graph analysis of 6 million NFT transactions. Retrieved from https://towardsdatascience.com/exploring-the-nft-transaction-with-neo4j-cba80ead7e0b.
4. Chainalysis Team (2022). Crime and NFTs: Chainalysis Detects Significant Wash Trading and Some NFT Money Laundering In this Emerging Asset Class. Retrieved from https://blog.chainalysis.com/reports/2022-crypto-crime-report-preview-nft-wash-trading-money-laundering/.
5. Cui, Z., Chen, W., & Chen, Y. (2016). Multi-scale convolutional neural networks for time series classification. arXiv preprint arXiv:1603.06995.
6. Franceschet, M. (2021). Hits hits art. Blockchain: Res. Appl. 2, 100038 (2021).
7. Granat, M. (2019). How to Use Convolutional Neural Networks for Time Series Classification. Retrieved from https://towardsdatascience.com/how-to-use-convolutional-neural-networks-for-time-series-classification-56b1b0a07a57
8. Hassanzadeh, R., Nayak, R., & Stebila, D. (2012). Analyzing the effectiveness of graph metrics for anomaly detection in online social networks. In International conference on web information systems engineering (pp. 624-630). Springer, Berlin, Heidelberg.
9. Jordanoska, A. (2021). The exciting world of NFTs: a consideration of regulatory and financial crime risks. Butterworths Journal of International Banking and Financial Law, 10, 716.
10. Kamps, J., & Kleinberg, B. (2018). To the moon: defining and detecting cryptocurrency pump-and-dumps. Crime Science, 7(1), 1-18.
11. Kapoor, A. et al. (2022). Tweetboost: Influence of social media on nft valuation. arXiv preprint arXiv:2201.08373.
12. Lee, J. (2021). NFTs Boom Anew as DOG Coin Becomes $550 Million Asset Overnight. Retrieved from https://www.bloomberg.com/news/articles/2021-09-03/nfts-boom-anew-as-dog-coin-becomes-550-million-asset-overnight.
13. Nadini, M., Alessandretti, L., Di Giacinto, F., Martino, M., Aiello, L. M., & Baronchelli, A. (2021). Mapping the NFT revolution: market trends, trade networks and visual features. arXiv preprint arXiv:2106.00647.
14. Park, A., Kietzmann, J., Pitt, L., & Dabirian, A. (2022). The evolution of nonfungible tokens: Complexity and novelty of NFT use-cases. IT Professional, 24(1), 9-14.
15. Pelechrinis, K., Liu, X., Krishnamurthy, P., & Babay, A. (2022). Spotting Anomalous Trades in NFT Markets: The Case of NBA Topshot. *arXiv preprint arXiv*:2202.04013.

16. Rehman, W., e Zainab, H., Imran, J., & Bawany, N. Z. (2021, December). NFTs: Applications and Challenges. In 2021 22nd International Arab Conference on Information Technology (ACIT) (pp. 1-7). IEEE.
17. Senilov, I. (2021). Approaching Anomaly Detection in Transactional Data. Retrieved from https://towardsdatascience.com/approaching-anomaly-detection-in-transactional-data-744d132d524e.
18. Shifflett, S. (2018). Some traders are talking up cryptocurrencies, then dumping them, costing others millions. Retrieved from https://www.wsj.com/graphics/cryptocurrency-schemes-generate-big-coin/.
19. Tao, B., Dai, H. N., Wu, J., Ho, I. W. H., Zheng, Z., & Cheang, C. F. (2021). Complex Network Analysis of the Bitcoin Transaction Network. IEEE Transactions on Circuits and Systems II: Express Briefs.
20. Thompson, P. (2018). Pump and dump in crypto: cases, measures, warnings. Retrieved from https://cointelegraph.com/news/pump-and-dump-in-crypto-cases-measures-warnings.
21. Von Wachter, V., Jensen, J. R., Regner, F., & Ross, O. (2022). NFT Wash Trading: Quantifying suspicious behaviour in NFT markets. arXiv preprint arXiv:2202.03866
22. Wang, Q., Li, R., Wang, Q., & Chen, S. (2021). Non-fungible token (NFT): Overview, evaluation, opportunities and challenges. arXiv preprint arXiv:2105.07447.
23. Woods, R . (2022). A Survey of Fraud Statistics on Top NFT Marketplaces. Retrieved from https://www.cryptopolitan.com/a-survey-of-fraud-statistics-on-top-nft-marketplaces/
24. Yi Cao et al. (2014). Detecting Wash Trade in Financial Market Using Digraphs and Dynamic Programming. In: IEEE Conference on Computational Intelligence for Financial Engineering and Economics. doi: 10.1109/TNNLS.2015.2480959.

Appendix

Explanation of the columns in the dataset [5]:

- *Unique_id_collection*: Unique ID for a given NFT
- *Price_Crypto, Crypto, Price_USD*: Conversion in USD is done with a daily resolution
- *Seller_address, Seller_username, Buyer_address, Buyer_username*: Addresses for sellers and buyers and (when available) their username used on the NFT marketplace
- *Image_url_1, Image_url_2, Image_url_3, Image_url_4*: Url to the digital object associate with the NFT. Given that urls may change over time, first try to download Image_url_1, then Image_url_2, and so on..
- *Datetime_updated, Datetime_updated_seconds*: It identifies the time of the transaction with either a day or second resolution
- *Smart_contract*: Smart contract of the given NFT
- *ID_token*: ID of the NFT asset within a given smart contract

- *Transaction_hash*: hash of the transaction involving a NFT sale
- *Collection*: It corresponds to the collection in which the NFT belongs to
- *Collection_cleaned*: It removes common mispellings in the field Collection. It also uses words in Cleaning_collections.csv to smooth the names. For instance, Aavegotchi renames all collections starting with that string in Aavegotchi. Some unnamed collections are here called Miscellanea
- *Market*: It is where data are downloaded from (so the API).
- *Name*: Title of the NFT listing Description: Description of the NFT listings
- *Permanent_link*: A link that allows to verify the NFT authenticity (valid only for the OpenSea Market)
- *Category*: Category in which the NFT belongs to. Examples are: Art, Games, and Collectible