When Big Data Meets NFT:
Challenges, Impacts, and Opportunities

Qinuo Chen, Boston College, USA

Jingyao Guo, New York University, USA*

Bocheng Wei, University of Connecticut, USA

Bangcheng Li, Pioneer Academy, USA

Jack Michael Kelly, Hightstown High School, USA

ABSTRACT

Non-fungible tokens (NFTs) are unique digital assets that are based on smart contracts of blockchain technology and traded via cryptocurrencies. They became known to the public in 2021 and kept growing rapidly. Until the second quarter of 2022, the total volume of NFTs has reached $12.22 billion. However, since the NFT market is still in its early stage, there are limited studies on this topic. In this paper, the first purpose is to analyze the overall market structure and volatility, characteristics of top NFTs on famous marketplaces, and future trends of NFTs. The next focus is to summarize current research trends about the concept of NFT, investigate the challenges faced by researchers, and provide current data collection or feature extraction techniques that are frequently utilized to solve those challenges.

KEYWORDS

Artificial Intelligence, Big Data Analytics, Blockchain, Cryptocurrency, Decision-Making Science, Deep Learning, Machine Learning, Metaverse, NFT

INTRODUCTION

On March 11, 2021, NFT artwork "Everydays: The first 5000 Days" by digital artist Beeple was sold for $69 million in the auction at Christie's. This work is the collage of all the works done and published by Beeple in the past 5000 days, and its final auction price made it become the most popular NFT with the highest value (Kapoor et al., 2022). After that, NFT market started to be known to the general public, and people began to discuss "What is NFT, and why it has such a high value?" NFT, also known as Non-fungible token, is the ID or certification of a digital asset in metaverse. Unlike the currencies people use to trade, NFT is not interchangeable, and every token is unique. Its uniqueness shows its scarcity as well as explaining why it has potential to appreciate. Also, NFT is traded via

*Corresponding Author

cryptocurrencies such as Ethereum. Through blockchain technology, its ownership will be certified to make sure that the process is decentralized and transparent; everyone will know who owns which NFTs at what time and the price of each resell (Wang et al., 2021). Until the second quarter of 2022, the total volume of NFTs has reached $12.22 billion (Akamo, 2022). However, since NFT market is nascent, there are only limited amount of researches. Therefore, having a comprehensive and deep market analysis is critical and helpful to fully unveil the mask of NFT market. Certain challenges include NFT's relative nature of unpredictability and irregular volatilities.

However, with the help of technologies and big data, many current challenges can be effectively solved. Big data represents velocity (the high speed of data collecting and processing), volume (the large amount of data), and variety (different types of data such as image, text, and gif etc) (Laborde, 2020). Two challenges described above: the high volatility and limited research of NFT market, can be solved, or at least mitigated, by big data. Through big data, researchers are able to build a comprehensive database that is powerful enough to be used to as the basic of the following analytical tasks, such as training models and performing sentimental analysis via machine learning and deep learning and making predictions based on the results. The predictive analysis aims to solve the challenge of the relatively unpredictable nature. By finding out patterns of the price flow, trading frequency, as well as social media hypes, researchers can gain a more exhaustive understanding of the market structure as well as the insights behind the irregular changes.

This paper aims to analyze both the current and future NFT market structure and trend as well as the motivations, rationales, methods, and challenges in current research that have been done. The first focus will be the overall analysis of NFT market, including characteristics of top NFTs on OpenSea, noticeable market volatilities, and future market trends. The second part will be the summary of current research trends, investigation of why these trends prevail, and potential challenges involved. Finally, the paper will discuss the data collection and integration methods, bid data analytics, as well as decision-making and predictive models and analysis by machine learning and deep learning to further investigate the NFT market and solve significant challenges.

BACKGROUND

The concept of Metaverse was suggested in the science fiction "Snow Crash" by Neal Stephenson approximately 30 years ago, and it has become more and more widespread in the field of technology as the rise of Blockchain, Artificial Intelligence, and Cloud Computing (Yang et al.,2020). Since blockchain technology provides the Metaverse with the chance to and associates the economic system in the physical world people live in with the virtual world, it triggers the interest of many famous companies. For instance, Facebook, the company that later renamed as Meta, opened up their own virtual reality world of avatars "Horizon Worlds" in the United States and Canada. Also, they use "Zuck Bucks" as their own currency in the virtual world (Nambiampurath, 2022). NFTs, are also one of the rapid-developing in the world of metaverse, and it is the representative of asset ownership. It is based on blockchain technology and traded through cryptocurrency such as Ethereum and Solana. Zhang et al. (2021) suggests that NFTs are minted via smart contracts, which are digital contracts stored on a blockchain, to certify asset ownership and manage the transferability. This process is completely decentralized without the participation of the third-party agency. Its mechanism is similar to that of the vending machine. Customers choose which product they want to buy on the machine and pay the money. The whole transaction process is based only on some computer codes rather than other agencies. As a result, the trading process of NFTs are inseparable from the concept of blockchain technology, cryptocurrency, and Metaverse.

In the last decades, the main investment people know is stocks or bonds. As the field of digital art develops and the rise of one of the earliest NFT collection "CryptoPunks" by Larva Lab, people start to be attracted by a nascent market: NFTs. Since it is not a mature market that investors are familiar with, its market value has been largely debated. White et al. (2022) proposes that media can

exacerbate the expectations of investors through hyping and highlighting the "celebrity effect", and they can also serve as the information gatekeeper by educating investors by providing concrete and informative facts. Moreover, Jain et al. (2022) associates NFTs' market value with the fluctuations of other financial assets and the public's Google search volume. Additionally, Kapoor et al. (2022) investigates the market value of NFTs based on its popularity on Twitter. Furthermore, many researchers are curious about the differences between NFT market under normal condition and NFT market during the global pandemic. They indicate that NFTs tend to be more coherent with other major assets such as currencies and stock market during Covid-19 outbreak (Umar et al., 2022).

Big data and Artificial Intelligence have been effectively used in enhancing the exhaustiveness of the limited researches on NFTs and extenuating the unpredictability nature of NFT market. Piyadigama & Poravi, (2022) proposes an idea of established an AI-based recommendation system to predict the investors' preference and choices based on big data and recommend them with NFTs that they are mostly likely to buy. By doing so, they aim to boost the volume of NFT investment by enlarging the number of potential buyers. Furthermore, as described before, Jain et al. (2022) tends to relate the market value and popularity of NFTs with search data from Google Trends, which is a popular source big data research. More details, interpretations, and analysis of these research trends will be provided in the section "Current Research Trend".

MAIN FOCUS OF THE ARTICLE

Mapping the NFT Market

As NFT becomes prevalent, there are more and more emerging marketplaces, such as OpenSea, Axie Marketplace, and Rarible, to trade NFTs of different kinds. Among these marketplaces, OpenSea is both the largest and the most popular platform with over 1.2 million active users and a total of $20.37 billion in sales (Kapoor et al., 2022). The Top NFTs Ranking on OpenSea records frequently-traded and high-valued NFTs based on their total volumes, floor prices, and the number of owners. The Ranking is calculated on a "last 24 hours", "last 7 days", "last 30 days", as well as "all-time" basis. Since the "last 24 hours" ranking tends to have a lot of irregular and sudden changes, and the results from the "last 30 days" and "all-time" rankings are too macroscopic, the result of this paper is mainly based on observations about the "last 7 days" ranking from June 22, 2022 to June 27, 2022. The summarized characteristics of top NFTs, obvious volume volatilities, as well as the analysis of how some trending NFTs indicate the future market trends of worth-invested NFTs are listed and further interpreted below.

Characteristics of Top NFTs

What must be prioritized first is that having a large and mature social media community on Twitter and Discord is one of the most significant features for becoming "a Top NFT". Social media has always the most dominant information-sharing platforms during the era of Internet, so the high degree of prevalence of NFTs on different social media can enhance both the social influences and advertising scope, which will raise the value of NFTs effectively. Table 1 provides the Twitter and Discord statistics of Top 15 NFT on Opensea. Except "Clone X" that does not have an official Twitter account, all NFTs on the ranking have more than 110K followers. Furthermore, among thirteen NFTs with Discord communities, six of them have more than 100K members, and eleven of them have more than 60K members.

Second, the notion of "celebrity effect" is also an important factor of becoming "top". Famous people have more abilities to influence others, and people, especially the fans, tend to imitate what celebrities have done. Therefore, some NFTs created by famous artists or actors have more chances to become prevalent because the prestige of some creators, especially those with extensive influences or great reputations, can attract more traders. Among the top NFTs in the ranking, many of them are

Table 1. Twitter and Discord Statistics of Top 15 NFT on OpenSea

NFT Name	Twitter followers (by June, 2022)	Discord members (by June, 2022)
CryptoPunks	233K	65779
Bored Ape Yacht Club	967.2K	206314
Primates	298.3K	181065
Otherdeed for Otherside	209.7K	131474
Mutant Ape Yacht Club	967.2K	206314
Doodles	347.9K	70137
Clone X - X Takashi Murakami	N/A	N/A
Art Blocks Curated	150.4K	43717
Azuki	309.7K	131250
SNEAKER HEADS OFFICIAL	303.2K	80396
goblintown.wtf	123.5K	N/A
Moonbirds	200.5K	17688
LonelyPop	109.1K	64029
Dooplicator	347.9K	70137
Yeah Tigers	110.2K	221534

Source: (Twitter, 2022) (Discord, 2022)

created by celebrities such as well-known artists. For instance, "Clone X - X Takashi Murakami" is created by RTFKT studio in collaboration with a world-famous Japanese artist Takashi Murakami. Under far-ranging publicity and Murakami's international influences, the popularity this collection outstrips all the other NFT collections RTFKT has created and become one of the most valuable NFT on OpenSea, which always remains in the top 10 of the ranking. Third, NFTs with additional benefits like club membership or other privileges and bonuses like free NFT airdrops are definitely more popular and worth-invested because it renders the buyers more chances to gain both financial and social profits. NFT Airdrops mean that the buyers of NFTs from certain collections will get an additional NFT from another collection for free. Membership means that the investors of certain NFTs may get the membership access to some private clubs with many other authoritative NFT buyers, which will be a great chance to enlarge social networks. For example, according to nfttalk.media (2022), every holder of the collection "the Doodles" will get one vote per Doodle, and the vote can be used to make important decisions for any ideas or derivative proposals, such as team hiring or charity donation, in the official forum established by "the Doodles". Also, every holder can get a Dooplicator NFT (the airdrop) for free for every Doodles NFT they hold. Fourth, the category "Art", on OpenSea contains the most high-valued and frequently-traded NFTs. Within this category, the profile photos NFT occupies both the largest percentage and popularity. Among the Top 20 NFTs based on the daily ranking on June 22, 2022, seventeen of them belongs to the "Art" category, and fourteen of them are NFT profile photos. Its popularity started with the earliest NFT collection "CryptoPunks" created by Larva Lab and remained its dominance until now. Finally, people tend to be more willing to bet on NFT launched by credible company because it can minimize the risks.

To be more specific, among the top 50 of the ranking, there are six NFTs (Bored Ape Yacht Club, CryptoPunks, Mutant Ape Yacht Club, Otherdeed for Otherside, Meebits, and Bored Ape Kennel Club) from the same company, which is Yuga Lab. What is more surprising is that four of them (Bored Ape Yacht Club, CryptoPunks, Mutant Ape Yacht Club, and Otherdeed for Otherside) always remain in the top 4 on the all-time NFT Top Ranking, as shown in figure 1. Thus, Yuga Lab is no doubt

Figure 1. Top 4 NFTs on OpenSea based on all-time volume (OpenSea, 2022)

the best-known and popular blockchain technology company that develops NFTs. Its success keeps attracting more and more investors and forms a virtuous cycle that it will always be the investors' first choice. Nowadays, it almost forms a monopolistic phenomenon of the most valuable NFT artworks.

Obvious Volume Volatility: A Risk Factor

Due to the limited amount of researches have been done on NFTs, it is hard to predict the relatively irregular volume volatility of the market. However, since all NFTs on OpenSea are traded using cryptocurrencies such as Ether (ETH) and Solana, it is reasonable to assume that the fluctuation of the market of cryptocurrencies will impact that of the NFT market as well as the possibility of volatility spillover. Antecedent research Ante (2021) aimed to explore the crossover effects between NFT pricing and cryptocurrencies pricing because users from the two markets tend to overlap. They analyzed data including the volume of NFT sales in US Dollars, the prices of Ether (ETH) and Bitcoin (BTC) in USD, as well as the number of blockchain wallets holding or interacting with NFTs on a particular day. Also, they utilized a VAR framework to find out the interrelationship between all three variables. In their results, though the prices of NFTs do not affect those of Bitcoin and Ether, the prices of cryptocurrencies impact those of NFTs. It conforms to the assumption made earlier because the drop of the value of cryptocurrencies will negatively influence buyers' purchasing power and enlarge the risks, which lowers the value of NFTs. Besides Bitcoin and Ether, the Solana chain is even more unstable. Based on the analysis of the Top 50 NFTs in the ranking on OpenSea from June 22 to June 27, 2022 shown in table 2, eight NFTs dropped over fifty places. Among these Top 50 NFTs, only twelve NFTs are based on the Solana chain, but eight of them dropped to different degrees. Also, among the eight dramatically-dropped NFTs just mentioned, half of them are Solana-based. Thus, compared to Ether, it is considered a risker choice. Overall, the different degrees of instability, including sudden and irregular surges and plummet, of all cryptocurrencies is one of the most significant risk factors of the NFT Market.

Additionally, the high gas fees is also a critical risk factor. Due to the considerable amount of resources and energies involved in smart contracts and during the minting process, a high gas fee is required to compensate the miners. As the value of NFT rises, the gas fee also increases. Thus, when balancing the risk and benefits of investing in NFTs, whether the buyers are able to earn profits more than the gas fee they paid is also a concern (Wang et al., 2021). Finally, since NFT market is nascent, many related companies do not have enough experiences and solid foundation or technology. Thus,

Table 2. Obvious volume volatility in Top 50 NFTs on June 22, 2022

NFT Name	Chains	06/22/2022 Ranking	06/27/2022 Ranking
Yeah Tigers	Solana	#15	#32
Okay Bears	Solana	#16	#17
Midnight Panthers	Solana	#23	outside of top 100
PENGSOL	Solana	#37	#88
Baked Beavers	Solana	#46	outside of top 100
Crypto Cowboy Country	Solana	#49	outside of top100
Solana Monkey Business	Solana	#50	#65

Source: (OpenSea, 2022)

the risk may include the half-formulate and immature technology, prototype nature of the products, as well as the regulatory inexperience.

Market Trend Analysis of Future NFTs

NFT market became well-known to the general public since the launch of the earliest NFT collection "CryptoPunks" by the end of 2017. In March 2021, its daily exchange volume exceeded 10 million US dollars, making NFTs prevail in the art market as well as opening up the age of digital arts. Until July 2022, 71 percent of NFT market has been dominated by digital artworks (Nadini et al., 2021). However, today, as the market grows more and more rapid, it has no longer been restricted to the world of digital artworks but stepped into other fields: digital fashion is one of the most trending direction with lots of potentials.

In May 2019, Iridescence Dress, as shown in figure 2, the first and unprecedented digital and blockchain based dress created by the digital fashion studio Dapper Lab and The Farricant, was released and sold for 9500 dollars. Johanna Jaskowska, the filter artist on Instagram, was the model for this digital dress. Though its price seemed insignificant compared to the whopping price of today's NFT artworks and avatars, it was still unparalleled in 2019, and it opened up the field of digital clothing in the NFT market. Instead of being based on virtual characters, NFT clothing is based on the investor's real appearances. The digital fashion studio will utilize 3D modeling and film special effects to the real photos provided by the buyers to digitally tailored the clothes. Also, the digital clothing will be transferrable, which means the owners can change it to other apps, experiences, or backgrounds. Most importantly, the digital clothes is one-of-a-kind, and fully owned by the buyers because NFTs cannot be copied (Gharegozlou, 2019).

Another recent instance in the digital fashion industry will be "CryptoKicks" created by Nike and RTFKT with the goal of achieving digital transformation. CryptoKicks's total volume has reached 6.3K on OpenSea by June 2022 (OpenSea, 2022). The first NFT sneakers they released in this collection was "Dunk Genesis CryptoKicks" enabling the buyers to change the look of their CryptoKicks through the EVO X, the RTFKT Skin Vials, as showed in figure 3. Additionally, another noticeable trends in NFT market is the gradual prosperity of metaverse-related NFTs. Since July 2020, the total transactions of the metaverse or game category surged to 44 percent of the whole market (Nadini et al., 2021). Until now, the most well-known collection related to metaverse is "Otherdeed for Otherside" by Yuga Labs. Each Otherdeed represents different environments and sediments within the Otherside, the metaverse of Yuga Labs on the Ethereum blockchain. Furthermore, similar to one characteristic of top NFTs described in the previous section, most trending NFTs have benefits or bonuses besides its own value. In other words, to be trendy in the future markets, NFT cannot be based only on hype, and it cannot just be a bet that investors have to guess its random rises or falls like a stock. It has to

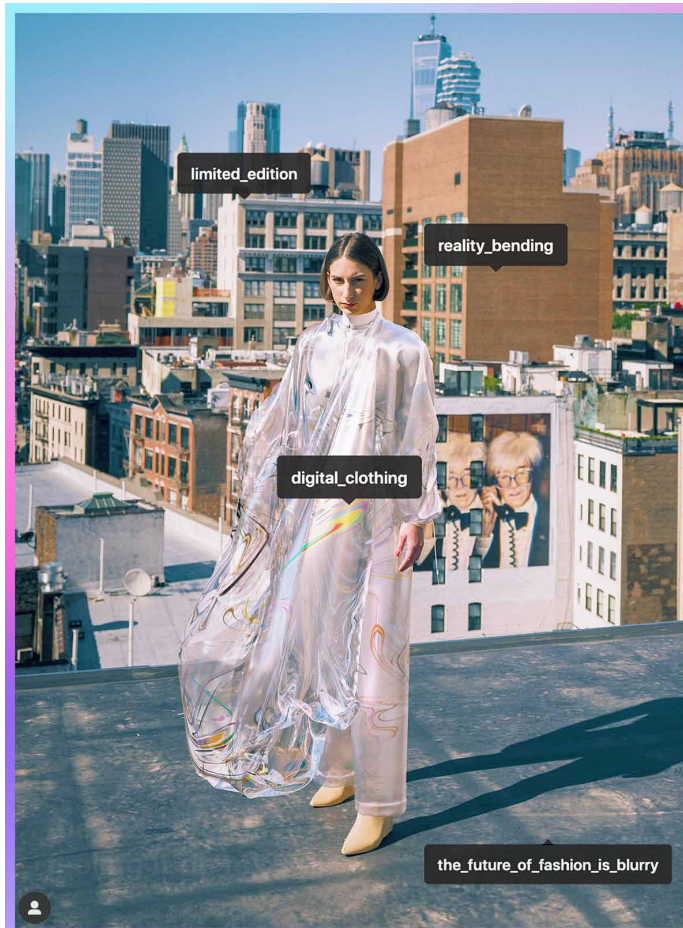**Figure 2. Iridescence Dress, the first NFT outfit, by Dapper Lab and The Farricant (Gharegozlou, 2019)**

Figure 3. "Dunk Genesis Cryptokicks" and EVO X created by Nike and RTFKT (OpenSea, 2022)

provide the buyers with some benefits as guarantees of the investment in order to be competitive. IDZ technologies' new NFT called "Asset 0" is an intriguing example. IDZ has its unique way of calculating the value of the NFT. Rather than based the value of the NFT on hype or fate, the value of Asset 0 is calculated by its own formula. Also, IDZ provides the buyers with a liquidity guarantee that all the Asset 0 holders will be able to sell their asset back to IDZ, which ensures that they can always find the buyer if they want to resell their NFT with no losses (OpenSea, 2022). Therefore, instead of dominating only the art market, the future trends of NFT market become more and more related

to technologies and virtual world, which keeps refreshing people's imagination of the limitations of the physical world and technology.

Current Research Trend

Cryptocurrencies, AI, and Metaverse

One of the current research trends is to explore the way to combine blockchain technology involving cryptocurrencies with artificial intelligence (AI) to further develop the possibilities of the metaverse, which is a three-dimensional virtual reality based world that enables users to interact with each other as digital avatars. With the advent of the era of Web 3.0, which dominates by the machine-based network system of human cooperation and understanding of data, the metaverse also prevails. Nowadays, the rise of the metaverse has drawn attentions of many technology researchers and companies, influenced many industries like fashion and health care, as well as shaped the general public's mentalities and lifestyles such as how they entertain. Similar to the physical world that people live in now, the metaverse has its own mature and self-perpetuating economy of producing and consuming digital assets using digital currency. For example, "The Sandbox" founded by Arthur Madrid was one of the earliest and the most popular utilization of metaverse. It enables users to trade, interact, and compete with each other using their own currency "Sand", which serves as NFT-liked tokens to prove users' ownership of their lands (IVANONTECH, 2021). However, different from the physical world, the metaverse is full of unknowns whose potentials and limitations are hard to precisely predict. Also, due to the large amount of transaction volume, the economy of metaverse needs to be based on decentralization in order to achieve an effective trading process. As a result, blockchain, as a decentralized ledger, will bring metaverse more innovations and possibilities. As demonstrated in the research Yang et al. (2020), the blockchain can be both the authentication of digital assets and the privacy guarantee of all transactions that happened in the metaverse due to its transparent nature. Also, cryptocurrencies such as Bitcoin and Ethereum based on the blockchain can serve as the official currency in the virtual world during trade and exchange. Furthermore, AI can also bring significant contributions. By using supervised learning, unsupervised learning, and reinforcement learning to train models, AI can perform analytical tasks, help people to make decisions, and enhance the process of digital creation, which is the basic production process in metaverse. Therefore, if blockchain and AI can be successfully and effectively combined with the metaverse, both authentication and digital economy establishment can be achieved.

Market Analysis, Price Return, and Risk Management

Since NFT is a nascent market without comprehensive researches, one of the most significant current research focus is to explore the market structure and future trend of NFTs as well as the accompanying risks. Firstly, Nadini et al. (2021) aims to evaluate the extent each different category (Arts, Collectibles, Music, and Virtual Worlds etc) of NFTs contributes to the whole market and the common visual features NFTs from the same collection share. Regarding the market structure, they found that though arts have dominated the market most of the time, other categories, such as Gaming and Virtual Worlds, start to thrive gradually. Regarding the visual features, they concluded that NFTs from the same collection tend to share similar traits. Their purpose of researching both factors is to improve the accuracy of the sale predictivity of NFTs; by doing so, it can reduce the uncertainties of the volatile market and minimize the risks. More importantly, they also examine the interactions and networks between NFT traders. They found that the transaction contributions from the top 10 percent traders of the market is the same as the other 90 percent users. It indicates that the NFT markets are dominated by a small number of expert users, but some potential risks might arise due to the possibility that the prices of NFT index are likely to be bid up by those specialized investors. Secondly, Wang et al. (2021) also demonstrates the properties of the overall NFT market

plus the future opportunities. They summarize the characteristics of NFTs as verifiability, transparent execution, availability, tamper-resistance, usability, atomicity, and tradability, and they state that future opportunities include flourishing gaming and virtual worlds industry and enhancing people's understanding of metaverse. Additionally, another recent trend focus mainly on risk management and price return is to analyze the connectedness and spillover effects between the returns of NFTs and other financial assets such as stocks, gold, or currencies. Thirdly, Abaron & Demir (2021) suggests that NFTs are mostly independent from other financial assets such as stock, gold, and currencies. As a result, only a small number of NFT fluctuations are due to other financial assets. However, Ante (2021) shows that the price and volatilities of Ethereum will result in changes in NFT markets. Therefore, the stability of Ethereum market should be an important factor to consider in determining the NFT investment decisions. However, though the main focus of market analysis and risk management is to gain a more exhaustive understanding of the market in order to reduce the potential risks, risks are still an inevitable and the most significant challenge due to the relative unpredictability and illiquidity nature of NFTs in the early stage.

NFT, Global Pandemic, and Social Changes

It is undeniable that the global pandemic has brought earth-shaking changes to the how the world process and how people live. Also, it has severely affected the stock market, such as the severe plummet in the financial markets during March 2020. Thus, one recent research trend is related to analyze how Covid-19 affect the fluctuation of NFT market, and how it impacts the connectedness between NFT market and other financial markets. Firstly, not only does the research Abaron & Demir (2021) explores the relationship between NFT market and other financial markets under the normal situation, it also investigates that connectedness during the global pandemic as well as comparing the similarities and differences between the pre-pandemic period and on-going pandemic period. They found that similar to the normal time, NFTs during the Covid-19 also behave differently with other financial assets, except the Ethereum. What is significantly distinct is that instead of having a negative return before the global pandemic, both NFTs and Ethereum become assets with the highest positive return among Gold, Oil, Bonds, and USD index etc. Also, though NFTs and Ethereum have an positive correlation before 2020, their connectedness becomes opposite after the pandemic. Furthermore, rather than experiencing market crash like the stock market, NFT market tends to gain benefits from the pandemic. Its role has been shifted from risk spillover transmitter during the pre-pandemic period to risk spillover receivers during the on-going pandemic. Secondly, Umar et al. (2022) examines how social media coverage influence the return and volatility of NFT market during the global pandemic. Such research trend perfectly combines NFT, Covid-19, and social changes by investigating how NFT is impacted by the pandemic and the subsequent social changes it brings. The researchers found that while the price return increases during the Covid-19 crisis, which conforms to the result from Abaron & Demir (2021) described above, the volatility spillover of NFTs, especially those from the category Arts, also increases. As a result, NFT market is greatly affected by the increasing number of media coverages during the global pandemic. Since NFTs are based largely on hype, social media is the most significant platform that may shape the direction of its future development. For instance, Kapoor et al. (2022) investigates the influence Twitter brings to NFTs and its valuation. Moreover, White et al. (2022) explores the characteristics of news related to the pandemic and how the intensity and tone of the media coverage lead to further market activities.

Issues, Controversies, and Solutions

Data Collection, Sources, and Integration

To analyze the NFT market, gathering data and information from different sources is inevitable, as shown in Table 3. First, in order to get data and information about NFTs, OpenSea API is widely used. OpenSea API provides developers with high-speed access to their database as well as fetching specific

data and information about NFTs on their marketplace, such as the assets, collections, historical and current prices, and creators. Second, to explore the reason behind the high volatilities behind NFT market, Binance API, which connects developers to their servers, is needed to get the fluctuations and trading volumes of other financial assets such as currencies and stocks. Overall, Binance can offer access to over 300 digital and fiat currencies. Third, since social media popularity is a significant element to evaluate an NFT, many researches use Twitter API, which enables developers access to tweets, direct messages, spaces, lists, and users etc. By doing so, the researches can fetch data and information about Twitter users' attitudes, comments, or demographic information toward certain NFT or NFT related events.

Besides three data sources mentioned above, there are also other useful platforms such as Google Trends, Yahoo Finance, and dune analytics. First, Google Trends can provide users with the overall search volume on Google. By using PyTrends, users can get the search data as well as measuring interest in a particular NFT collection or NFT related events. Second, besides Binance, Yahoo Finance can also be utilized to fetch financial asset-related data such as the price and fluctuation of Bitcoin or Ethereum. Those factors are important for the analysis of connectedness, spillover, and volatility. Third, Dune Analytics, a data platform that provides investors or researches available on-chain crypto data. By using this tool, NFT market researchers are able to get NFT transaction data on different blockchains such as Ethereum, Polygon, and Solana. Thus, by doing so, researchers can effectively find similarities and differences between NFTs based on different chains.

After collecting data from different sources above, it is indispensable to integrate the multiple database together to set foundation for the further analysis and comparison. For instance, Twitter data can be associated with OpenSea data. The first step is to extract NFT assets from certain collection and fetch the information related to the same group of NFT assets from Twitter as well. Possible features on Twitter that worth extracting are related tweets as well as numbers of likes, replies, and hashtags. Next, it is significant to analyze correlation between NFT popularity on both OpenSea and Twitter. On Twitter, NFT popularity can be determined by aggregating all the features; on Opensea, NFT prevalence can be measured by calculating the difference between the floor price and current price as well as the degree of fluctuations in its historical price. After that, researchers can compare if the degree of popularity on both platforms correlate with each other. Furthermore, the second instance related to database integration is to connect the OpenSea data with Binance or Yahoo Finance data. From OpenSea, NFT prices can be extracted, and from Binance and Yahoo Finance, the price of other financial assets such as stock, Bitcoin, Ethereum, and US Dollar can be fetched. Then, the price flows from these elements can be analyzed to find any spillover, connectedness, or disparity.

Feature Extraction With Big Data Analytics

When analyzing the numerous data sets collected from various sources, it is necessary to apply Feature Extraction technique to find out the most valuable features. Different Feature Extraction methods should be chosen based on the data type that is to be examined, which can be seen in Table 4. For articles and papers which contains large amount of human language text, Natural Language Processing (NLP) is undoubtedly among the most cited approachs. NLP is an artiftial intelligence procedure which asks the computer to learn, investigate and extract the meaning of human language efficiently and smartly. After deeply mining the text data, NLP will be able to derive the most important and frequently-mentioned features out of the corpus. Though training an NLP model might take quite some time due to the complexity of human language, it is still worthy since its performance on text data mining is considerably impressive. When it comes to image data, Image Recognition technique is then to be introduced to let the computer identify the key features in an image. By splitting the given images into pixels and training the system to study the location of desired features on the sample images, all images will get the label of the features that are extracted from them. Beyond what is mentioned so far, there are also many other powerful

Table 3. Papers using Data Collection, Sources, and Integration techniques

Author(s) (Year)	Research Domain	Main Focus	Methodology	Models	Attributes/Features
Kapoor et al. (2022)	Influence of social media on NFT valuation	NFT Asset Valuation	Temporal Analysis, Correlation Analysis of Signals, Image Recognition, Machine Learning, Deep Learning	Logistic Regression, SVM, Random Forests, LightGBM, XGBoost, Convolutional Neural Network (CNN)	**Twitter**: Listed count, has nft in username or not, number of likes, number of replies, number of hastags; **Opensea**: Is presale or not, verified asset or not, bid withdrawn or not, number of bid entered, number of offer entered, number of transfer
Jain et al. (2022)	Search Trends, Public Market Data, Linear Regression and Recurrent Neural Networks	NFT Appraisal Prediction	Machine Learning, Deep Learning, Time Series Analysis	Linear Regression, RNN, LSTM, GRU	Opening date of NFT; average price of the NFT as of opening date; the unique token id of NFTs; closing price of ETH token, Bitcoin, gold, S&P value, Dow Jones value, Nasdaq 100 value, Microsoft, Apple, Netflix, Tesla, Amazon, Meta; Relative google search volume for collection name on a scale from 0-100; -1,0,1 indicating bad news, no news, and good news respectively; A measure of network traffic, which indicates the transaction fee of purchase; Average price of the NFT as of the next day
Aharon et al. (2022)	Return connectedness, COVID-19, Spillover	NFTs and asset class spillovers during COVID-19	Static and Dynamic Analysis, Time Series Analysis	Time-Varying Parameter Vector Autoregressions (TVP-VAR) approach	Risk and return of gold, equities, currencies, bonds, cryptocurrencies

Source: (OpenSea, 2022)

Feature Extraction methods, including Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) etc. Most of these approaches extract the most dominant factors by reducing the dimensionality of the data and then removing the random noisy features. These approaches are commonly used in finding the patterns of the data and they show good performance in the fields of face recognition and image compression.

Decision-Making Models and Predictive Analysis

After the initial investigation of input data and obtaining the key features, the next step is to find the connections among all the features and make prediction on random new data. In Table 5, what is most commonly used nowadays would be Machine Learning algorithms, which contains a large family of models. As the name suggests, Machine Learning models are developed to let machine learn the patterns in the input data, give labels to do the classification based on these patterns, and then make predictive decision. The approaches are divided into three basic categories: Supervised Learning, Unsupervised Learning and Reinforcement Learning, depending on the types of the input data and output results. Ranging from simple algorithms such as Linear Regression, Support Vector Machine (SVM), Random Forests (RF) to complicated Deep Learning models including Convolutional Neural Network (CNN), Long Short-term Memory (LSTM) and Gated Recurrent Unit (GRU), the flexibility of the models rises when the complexity goes up.

Table 4. Papers using Feature Extraction techniques

Author(s) (Year)	Research Domain	Main Fouces	Methodology	Models	Attributes/Features
Nadini et al. (2021)	NFT market trends, trade networks, and visual features	NFT Market Analysis	Data Aggregation, Network Analysis, visual Feature Extraction, Deep Learning, Image Recognition, Predictive Analysis	Modularity, Convolution Neural Network (CNN), Linear regression, Principal Component Analysis (PCA), AdaBoost	transactions of NFT purchases; types of cryptocurrencies; Historical NFT sales; digital artworks images/urls
Kapoor et al. (2022)	Influence of social media on NFT valuation	NFT Asset Valuation	Temporal Analysis, Correlation Analysis of Signals, Image Recognition, Machine Learning, Deep Learning	Logistic Regression, SVM, Random Forests, LightGBM, XGBoost, Convolutional Neural Network (CNN)	**Twitter**: Listed count, has nft in username or not, number of likes, number of replies, number of hastags; **Opensea**: Is presale or not, verified asset or not, bid withdrawn or not, number of bid entered, number of offer entered, number of transfer
Zhang et al. (2021)	Invention of a NFT verification method for distributed AI system	Comparison method invention	Quinnester sampling law, Maximum-minimum interval comparison	Threshold for repetition rate, CNN-based AI model	NFT pictures and annotation box

Source: (OpenSea, 2022)

But meanwhile, the interpretability will drop. Apart from the Machine Learning models, another strong method for Predictive Analysis is Time Series Analysis. It is widely operated when time is a significant variable of the input data. By exploring the data of different time period using Time Series Analysis approach, one can clearly see the trends and patterns over time. With these seasonal trends, it is then possible to predict the result of future events and make reliable decision now. Models such as Vector Autoregressive (VAR), Cross-wavelet Transform and Time-Varying Parameter Vector Autoregressive (TVP-VAR) are the most popular ones in the research of NFT market and they have a highly favorable performance in data visualization and prediction. Besides these two types of technical models, there also exist other financial models, such as Capital Asset Pricing Model (CAPM), Sharpe Ratio, Alpha and Beta and so on. These methods are frequently imported to research the risk and return of NFT market and NFT-related company, which can be used to point out whether the market is profitable or not and further influence the decision of investment.

CONCLUSION

Current research about NFT with Big data analytics is very limited due to it is a new topic. Former articles and references are not sufficient and therefore becomes the limitation of this paper. Among all the available research, popular topics include the interrelationship between NFT, Cryptocurrency, Metaverse, and AI, NFT Market Analysis, Price Return, and Risk Management, and NFT, Global Pandemic, and Social Changes. Within those research, many techniques, such as NLP, time series analysis, machine learning/deep learning for classification, etc, are frequently used. However, current research also have several limitations. First, current databases are limited, and data records are small.

Table 5. Papers using Decision-Making Models and Predictive Analysis techniques

Author(s) (Year)	Research Domain	Main Fouces	Methodology	Models	Attributes/Features
Kapoor et al. (2022)	Influence of social media on NFT valuation	NFT Asset Valuation	Temporal Analysis, Correlation Analysis of Signals, Image Recognition, Machine Learning, Deep Learning	Logistic Regression, SVM, Random Forests, LightGBM, XGBoost, Convolutional Neural Network (CNN)	**Twitter**: Listed count, has nft in username or not, number of likes, number of replies, number of hastags; **Opensea**: Is presale or not, verified asset or not, bid withdrawn or not, number of bid entered, number of offer entered, number of transfer
Jain et al. (2022)	Search Trends, Public Market Data, Linear Regression and Recurrent Neural Networks	NFT Appraisal Prediction	Machine Learning, Deep Learning, Time Series Analysis	Linear Regression, RNN, LSTM, GRU	Opening date of NFT; average price of the NFT as of opening date; the unique token id of NFTs; closing price of ETH token, Bitcoin, gold, S&P value, Dow Jones value, Nasdaq 100 value, Microsoft, Apple, Netflix, Tesla, Amazon, Meta; Relative google search volume for collection name on a scale from 0-100; -1,0,1 indicating bad news, no news, and good news respectively; A measure of network traffic, which indicates the transaction fee of purchase; Average price of the NFT as of the next day
Pinto-Gutiérrez et al. (2022)	Factors that draw investors' attention to and hype the NFT market	Connection between NFT, cryptocurrency and investor's attention	Time Series Analysis, Wavelet Coherence Analysis	Vector Autoregressive (VAR) Model, Cross-wavelet Coherence, Augmented Dickey-Fuller (ADF) Test	NFT, CryptoPunk, Bitcoin, Ethereum and Decentraland attention; Return of Bitcoin, Ether, VIX, Gold, S&P 500, CryptoPunk and Decentraland
Mazur et al. (2021)	Analyze the risk and return of NFT-based startups	Risk and return of NFTs	Sharpe ratios and market-adjusted returns, Standard Event Study	Capital Asset Pricing Model (CAPM), Alphas and Betas	NFT listing characteristics, risk, and raw return; NFT risk-adjusted performance; NFT alphas and betas; NFT price index
Aharon et al. (2022)	Return connectedness, COVID-19, Spillover	NFTs and asset class spillovers during COVID-19	Static and Dynamic Analysis, Time Series Analysis	Time-Varying Parameter Vector Autoregressions (TVP-VAR) approach	Risk and return of gold, equities, currencies, bonds, cryptocurrencies

Source: (OpenSea, 2022)

Second, current data collection tools mainly depend on APIs, while it is risky if those tools are not functional in the future. Third, feature extractions are limited since many features have not been measured or detected due to the absence of using big data analytics. For instance, it is challenging to collect large volumes of data across multiple databases. Moreover, different data types, such as image, voice, video, and gif, are difficult to process during feature extractions. Heretofore, there are multiple

possible solutions for those limitations. First, by using CI/CD to automate the data collection, the data collection process can become more efficiently. Second, computational needs can be satisfied through cloud-based computational environment, such as AWS. Third, if APIs are not working, Web Scrapper can be an alternate way to reduce the risk. Fourth, object detection/image recognition can be utilized to extract features from images; voice recognition can be employed to fetch features from music and other types of voice assets; video analysis/motion recognition can be used to extract information from videos.

ACKNOWLEDGMENT

REFERENCES

Aharon, D. Y., & Demir, E. (2022). NFTs and asset class spillovers: Lessons from the period around the COVID-19 pandemic. *Finance Research Letters*, *47*, 102515. doi:10.1016/j.frl.2021.102515

Akamo, A. (2022). *NFT trading volume slumps by 64% in Q2 2022*. https://nairametrics.com/2022/07/22/nft-trading-volume-slumps-by-64-in-q2-2022/

Ante, L. (2022). The non-fungible token (NFT) market and its relationship with Bitcoin and Ethereum. *FinTech*, *1*(3), 216–224. doi:10.3390/fintech1030017

Gharegozlou, R. (2019). *Dapper Labs x The Fabricant x Johanna Jaskowska SS19 iridescence*. https://medium.com/dapperlabs/dapper-labs-x-the-fabricant-x-johanna-jaskowska-ss19-iridescence-3ebe00e76b06

IVANONTECH. (2021). *What is The Sandbox, the SAND Token and LAND?* https://academy.moralis.io/blog/what-is-the-sandbox-the-sand-token-and-land

Jain, S., Bruckmann, C., & McDougall, C. (2022). *NFT Appraisal Prediction: Utilizing Search Trends, Public Market Data, Linear Regression and Recurrent Neural Networks*. arXiv preprint arXiv:2204.12932.

Kapoor, A., Guhathakurta, D., Mathur, M., Yadav, R., Gupta, M., & Kumaraguru, P. (2022). *Tweetboost: Influence of social media on nft valuation*. arXiv preprint arXiv:2201.08373.

Laborde, R. (2020). *The Three V's of Big Data: Volume, Velocity, and Variety*. https://blogs.oracle.com/health-sciences/post/the-three-vx27s-of-big-data-volume-velocity-and-variety

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. *Scientific Reports*, *11*(1), 1–11. doi:10.1038/s41598-021-00053-8 PMID:34686678

Nambiampurath, R. (2022). *Meta Launching Currency for Use in Horizon Worlds VR App*. https://beincrypto.com/meta-currency-horizon-worlds-vr-app/

NFTTALK Media. (2022). *The Doodles NFT Success Story*. https://medium.com/nfttalk/the-doodles-nft-success-story-fd117556be23#:~:text=A%20Doodle%20NFT%20holder%20can,from%20sales%20of%20the%20NFTs)

OpenSea. (2022). *IDZ*. https://opensea.io/collection/idz

OpenSea. (2022). *RTFKT x Nike Dunk Genesis CRYPTOKICKS*. https://opensea.io/collection/rtfkt-nike-cryptokickst

Piyadigama, D. R., & Poravi, G. (2022). *Exploration of the possibility of infusing Social Media Trends into generating NFT Recommendations*. arXiv preprint arXiv:2205.11229.

Umar, Z., Abrar, A., Zaremba, A., Teplova, T., & Vo, X. V. (2022). The Return and Volatility Connectedness of NFT Segments and Media Coverage: Fresh Evidence Based on News About the COVID-19 Pandemic. *Finance Research Letters*, *49*, 103031. doi:10.1016/j.frl.2022.103031 PMID:35669177

Wang, Q., Li, R., Wang, Q., & Chen, S. (2021). *Non-fungible token (NFT): Overview, evaluation, opportunities and challenges*. arXiv preprint arXiv:2105.07447.

White, J. T., Wilkoff, S., & Yildiz, S. (2022). *The role of the media in speculative markets: Evidence from non-fungible tokens (NFTs)*. Available at SSRN 4074154.

Yang, Q., Zhao, Y., Huang, H., Xiong, Z., Kang, J., & Zheng, Z. (2022). Fusing blockchain and AI with metaverse: A survey. IEEE Open Journal of the Computer Society.

Zhang, T., Song, M., Sui, Y., Chen, H., & Tan, J. (2021, December). A NFT verification method for distributed AI system. *Journal of Physics: Conference Series*, *2132*(1), 012017. doi:10.1088/1742-6596/2132/1/012017

Qinuo Chen is an undergraduate student at Boston College. She is currently working as a BI Specialist at IR Big Data & AI Lab. Her research interest mainly focuses on Business Operation and Management, Communication and Media Studies, Big Data Analytics, Social Media Analytics, and Artificial Intelligence.

Jingyao Guo is a Master's candidate of Mathematics in Finance at NYU Courant. Currently she is working as a Project Manager at IR Big Data & AI Lab. Her research focuses on Machine Learning, Deep Learning, Big Data Analytics, Quantitative Analysis, and Risk Management.

Bocheng Wei is an undergraduate student at University of Connecticut. His research focuses cover a broad range of practical orientations, including Information System, Database Management, Machine Learning and Big Data Analytics. He is currently working as an IT Specialist at IR Big Data & AI Lab with concentrated in UI Design, Database Management, Feature Engineering, and Big Data Processing.

Bangcheng Li is currently a student at Pioneer Academy in New Jersey. His research focuses on Machine Learning, Deep Learning, and Big Data Analytics. He dedicated himself to joining and contributing to the R&D team of IR Big Data & AI Lab in multiple research and business projects, such as Big Data Applications for Covid-19, Smart Community Management, and Business Analysis for NFT using AI, since 2020.

Jack Kelly is currently a student at Hightstown High School in New Jersey. His research focuses on Machine Learning, Deep Learning, and Big Data Analytics. Beginning from 2022, he dedicated himself to joining and contributing to the R&D team of IR Big Data & AI Lab for multiple projects associated with computational science and technological management.