ML-driven Dynamic Premium Pricing for New Metaverse Products

A B. Tech Project Report Submitted

in Partial Fulfillment of the Requirements

for the Degree of

Bachelor of Technology

by

Priyanka Sachan

(1901CS43)

under the guidance of

Dr. Raju Halder

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY PATNA

PATNA - 801106, BIHAR

CERTIFICATE

This is to certify that the work contained in this thesis entitled "**ML-driven Dynamic Premium Pricing for New Metaverse Products**" is a bonafide work of **Priyanka Sachan** (**Roll No. 1901CS43**), carried out in the Department of Computer Science and Engineering, Indian Institute of Technology Patna under my supervision and that it has not been submitted elsewhere for a degree.

Supervisor: **Dr. Raju Halder**
Associate Professor
Blockchain Technology & Robotics Lab
Department of Computer Science & Engineering
Indian Institute of Technology Patna
Bihar

May, 2023
Patna

Acknowledgement

I would like to express my sincere gratitude to all those who have supported and contributed to the completion of this thesis. First and foremost, I am deeply grateful to my supervisor, Dr. Raju Halder for their guidance, expertise, and continuous support throughout the entire research process. Their valuable insights and constructive feedback have been instrumental in shaping this thesis.

Furthermore, I extend my appreciation to my friends and family for their unwavering encouragement and belief in my abilities. Their support has been a constant source of motivation throughout this journey.

Lastly, I would like to acknowledge the wider academic community and the authors of the references used in this thesis. Their groundbreaking research and insights have significantly contributed to the development of this study.

Although it is not possible to name every individual who has played a part in this endeavour, I am sincerely grateful to each and every one of them. Their collective efforts have enriched the outcome of this thesis.
Thank you.

Priyanka Sachan

Table of Contents

Figures & Tables

Abstract

NFTs, also known as Non-Fungible Tokens, are digital items that possess distinct and unrepeatable visual or audio content. They serve as digital representations of various assets, such as photos, gifs, audio, videos, and other data-based materials. These assets span different categories, including art, in-game goods, and collectibles within the entertainment industry. The allure of NFTs lies in their exclusivity, as each NFT is one-of-a-kind and ownership is authenticated through a digital certificate.

Notably, the global NFT market is worth upwards of 33 billion USD this year.Similarly, $119 million have been stolen through NFT theft (not including the hacks around it) and the bars keep increasing each year.
As the number of NFT hacks continues to grow each month, the importance of insurance becomes prominent in safeguarding the security of NFT creators and owners.

With increasing NFT hacks each month, insurance is prominent to ensure the security of NFT creators and owners . However, the insurance industry has faced a new challenge in providing coverage for risks associated with intangible assets like NFTs, primarily due to the complexities involved in valuing them and the scarcity of relevant data.
In response, we have developed a machine learning model capable of predicting the price of NFTs, even for both existing collections and upcoming ones with limited sales data available.

Introduction

A Non-Fungible Token (NFT) is a type of digital asset that represents ownership or proof of authenticity of a unique item or piece of content. Unlike cryptocurrencies such as Bitcoin or Ethereum, which are fungible and can be exchanged on a one-to-one basis, NFTs are indivisible and cannot be exchanged on a like-for-like basis.
NFTs can represent a wide range of digital assets, including digital art, music, videos, virtual real estate, virtual goods in video games, collectibles, and more.
NFTs come in various forms and dimensions, yet they possess two fundamental interconnected characteristics:

- They embody distinct tangible or intangible assets, whether in the form of physical objects or digital creations. Essentially, an NFT acts as an advanced digital rendition of a "proof of ownership" document, securely encoded within a blockchain through cryptographic techniques.
- Consequently, the ownership record is immutable, impervious to tampering, and the complete history of ownership can always be verified.

Rise of NFTs

NFTs have gained significant attention in recent years, with notable sales of digital artworks and collectibles reaching substantial sums likeBeeple's Everydays: The First 5000 Days sold for $69.3 million in March 2021. Not only did this sale mark the highest-priced NFT transaction to date, but it also triggered a ripple effect across mainstream media, propelling the term "NFT" into households worldwide.
Since then the global non-fungible token market has risen to upwards of ~33 billion USD in 2023.

Risk associated with NFTs

However, the growth in popularity of crypto assets outpaced the infrastructure built to support them leading to multiple malpractices and loss of hundreds of millions in the NFT environment.

Over $100 million in NFTs have been stolen in the period of July 2021 to July 2022 alone without accounting for the hacks related to NFT systems, according to a report "NFTs and Financial Crime" by blockchain analysis firm Elliptic.

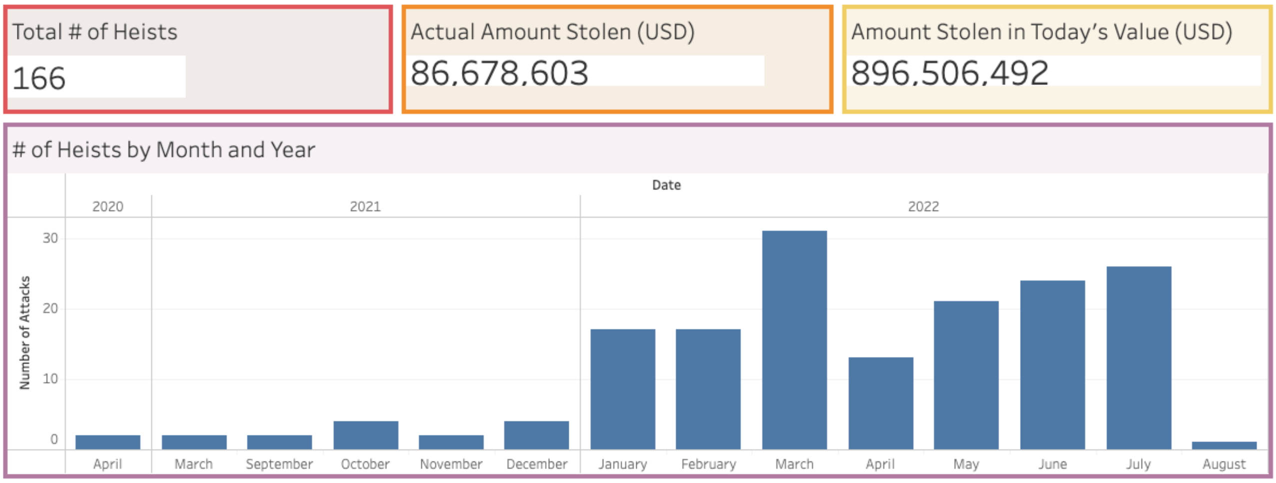


FIGURE 1 - NFT theft data in 2022

A large reason for these hacks despite NFTs having inherited robust security features from the underlying blockchain technology are the multiple security risks in the NFT ecosystem.

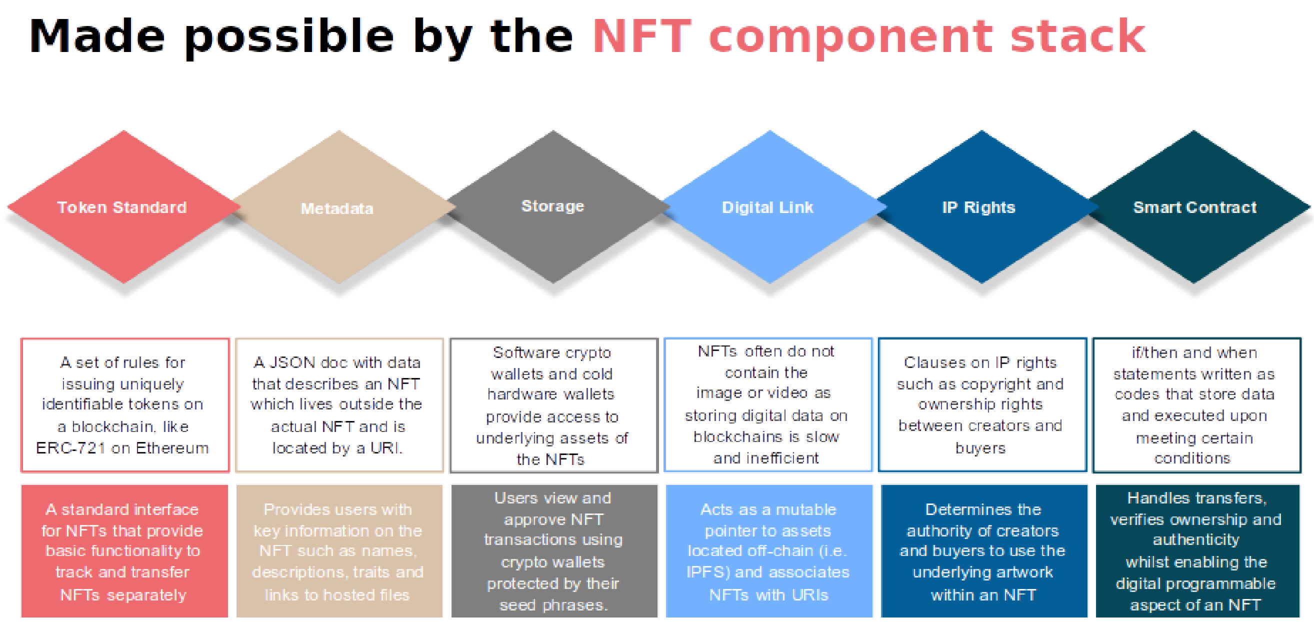


FIGURE 2- NFT Ecosystem

1. Private keys

NFTs represent unique digital assets and, similar to cryptocurrencies, are stored within a blockchain. While anyone can examine the blockchain's record to observe the underlying asset, only the NFT holder possesses the essential "private key" that verifies ownership. Consequently, the NFT holder is recognized as the owner of the underlying asset until the NFT is transferred to another individual's digital wallet. Once an NFT transaction occurs and is assigned to a different private key, the transaction becomes irreversible, adhering to the principle of "Not your private keys, not your NFTs."

This aspect presents one of the primary risks associated with NFTs. Acquiring and possessing an NFT necessitates a digital wallet containing the corresponding "private keys" to conduct transactions on the blockchain. If access to the digital wallet is lost due to forgotten passwords, device damage, or hacking incidents, the NFTs within the wallet can be permanently lost including an incident in 2021 where users of the NFT marketplace "Nifty Gateway" claimed that their entire NFT collections were "stolen."

2. **Asset protection**

   In most instances, the actual asset associated with an NFT, such as an image or audio file, is not stored directly on the blockchain. Instead, the NFT serves as a token providing proof of ownership and contains minimal data, typically only URLs that reference the metadata. Consequently, the images or art showcased in news articles featuring NFTs are not the NFTs themselves.

   However, this setup poses security concerns. If that link is broken or the company storing the asset goes out of business, the owner of the NFT would be left with links to assets or files that no longer exist on the internet. Similar risks emerge if a digital marketplace, storage wallet provider, or server farm involved in an NFT transaction experience financial insolvency or service disruptions that result in damage to the digital files.

3. **Smart Contract Risk**

   Another crucial aspect of NFTs involves the utilisation of smart contracts, which are lines of code stored on the blockchain governing NFT transactions. These smart contracts introduce programmability to NFTs and automate the execution of predefined logic during ownership transfers.

   However, there are risks associated with these smart contracts. They are typically developed by programmers and can be reused across multiple NFT projects. While this practice is legal since the Solidity code used in these contracts is open for public viewing and usage, it introduces potential hazards. For instance, numerous projects have adopted similar smart contracts from established NFT projects like the Bored Ape Yacht Club (BAYC) or widely used smart contract libraries like OpenZeppelin. Nevertheless, this trend of copying contracts brings forth developer risks and the possibility of human errors, which could result in malicious or vulnerable smart contracts that become targets for hackers.

4. **Copyrights**

    Additional risks linked to NFTs encompass concerns regarding the ownership of intellectual property rights associated with the digital asset. What occurs if the creator or seller of the underlying asset fails to adequately secure or validate the required trademarks or copyrights?

Considering these risks, there is a growing need for insurance products that specifically cater to safeguarding digital assets. Having insurance coverage that protects against the theft of NFTs is crucial for the flourishing of this emerging market. Insurance would not only lend legitimacy to this new asset class, but would also compel the entire ecosystem to strengthen its security measures and controls.

Traditional insurance model

An insurance company is interested in the total loss amount L per unit of exposure-to risk *e*, where *L* is the total loss for the *N* claims reported by a policyholder during the exposure period *e*. P&C insurers usually opt for a so-called frequency-severity strategy to price a contract Claim frequency *F* is the number of claims *N* filed per unit of exposure-to-risk *e*. Claim severity *S* refers to the cost per claim and is defined by the average amount per claim filed, that is the total loss amount *L* divided by the number of claims N. The technical price π (or: pure/risk premium) then follows as:

$$\pi = \mathbb{E}\left(\frac{L}{e}\right) \overset{\text{indep.}}{=} \mathbb{E}\left(\frac{N}{e}\right) \times \mathbb{E}\left(\frac{L}{N} \mid N > 0\right) = \mathbb{E}(F) \times \mathbb{E}(S)$$

assuming independence between the frequency and the severity component of the premium.

For an NFT insurance, N (No. of claims) = 0/1 only in the given year
And S (Claim severity) = min(x% of NFT price as declared in policy, Maximum insured amount)
Thus, L (Total loss amount) = 0/S

∴ π (pure/risk premium) = E(claim is made in a given year) ✕ Insured NFT price

E(claim is made in a given year)= P(claim is made in a given year) since maxClaim=1

Thus, traditional insurance models require the probability of a NFT theft/ damage (whatever is covered in the insurance) and the price of the insured NFT.

However, as NFT is non fungible, there is no market price similar to the way a cryptocurrency may evaluate. Thus, underwriters face the problem of lack of data and volatility related to the NFT market that makes it difficult to calculate a premium and coverage price that would keep the balances green.

We intend to solve this model by training a ML model that calculates NFT Valuation at the time of insurance proposal.

Literature Review

- **Mapping the NFT revolution: market trends, trade networks, and visual features**

The primary objective of this study is to demystify the overall structure behind valuation in the Non-Fungible Token (NFT) market while presenting a framework for quantifying its evolution. The paper proposes a linear regression model incorporating these findings as features to predict NFT prices.

Group 1: Network centrality (4)	Group 2: Visual features (5)	Group 3: Sale history (2)
- Degree centrality of seller - PageRank centrality of seller - Degree centrality of buyer - PageRange centrality of buyer	- Five PCA components extracted from the AlexNet vector of the NFT.	- Median price of primary and secondary sales made in the collection of interest. - The prior probability of secondary sale.

FIGURE 3- Features used for NFT Prediction

The study revealed that the median sale price of NFTs within the collection serves as a robust predictor. Notably, the accuracy of the prediction improves when the median of past sale prices is calculated over a recent time window leading up to the primary sale. Additionally, factors such as centrality measures of the buyer and seller in the trader network and visual features associated with the NFT contribute to explaining approximately one-fifth of the variance when utilized together.

The study demonstrates the development of a model with an R2 score of 0.709 for collectibles and 0.408 for metaverse NFTs, setting a baseline for us in predicting the NFT prices in these categories.

- **TweetBoost: Influence of Social Media on NFT Valuation**

The main aim of this research paper is to tackle two important questions:
a) What is the relationship between user activity on Twitter and the prices of NFTs on OpenSea?
b) Can we predict the value of NFTs by leveraging signals obtained from both Twitter and OpenSea data, and identify the features that have the greatest impact on these predictions?

To achieve these objectives, the paper seeks to create one of the initial datasets for NFTs that encompasses data from both OpenSea and Twitter. The authors analyse the growth of NFTs, examine the characteristics of Twitter users promoting NFT assets, and assess the influence of Twitter features on the virality of an NFT. They employ Binary and Multi-classification models to initially predict the profitability of NFTs and subsequently classify profitable NFTs into different price categories. Kapoor et al. concluded that incorporating Twitter data into their feature set resulted in a 6% improvement in model accuracy compared to using solely NFT platform data (such as OpenSea).
This research provides valuable insights into the interdependence of NFT valuation and its popularity on social media, as well as sheds light on additional training strategies and features that we can incorporate into our predictive models.

Factors to consider for NFT Price Prediction

We take inspiration from the older models and the features used for price prediction.

1. **Rarity**

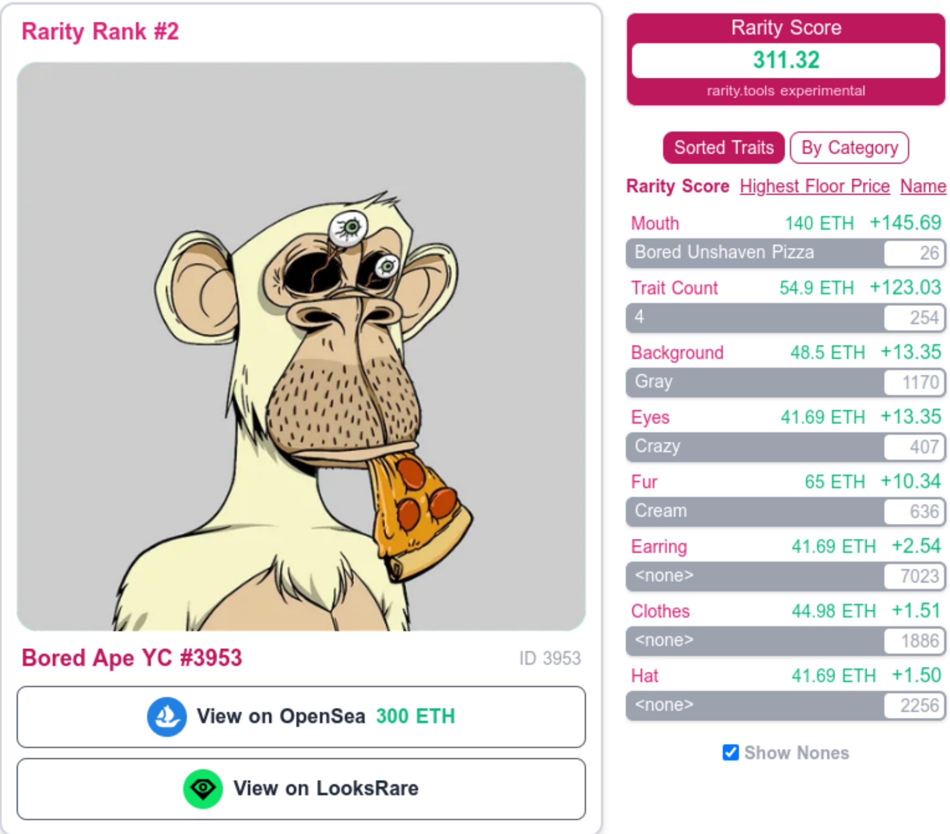


FIGURE 4 - Traits of Bored Ape YC #3953

The scarcity or uniqueness of an NFT often influences its value. A single trait may be the determining factor for an asset's value relative to the rest of the collection - like gold fur in Bored Ape Yacht Club.

NFTs that are one-of-a-kind or part of limited editions tend to attract higher prices.

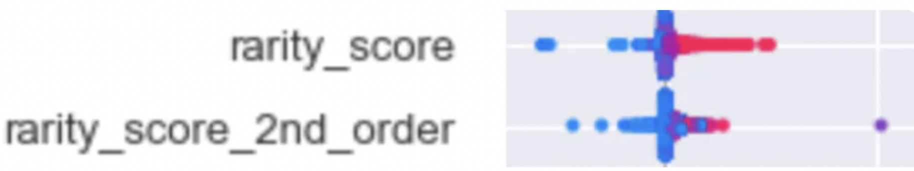


FIGURE 5 - Shapley value to measure importance of rarity score

2. **Historical Sales Data**

   Helps identify value of NFT in the context of other parameters.

FIGURE 6 - Shapley value to measure importance of previous sale price

We use last sale data of the NFT including price and hold time.

3. **Collection sales data**

   Analysing the historical sales data of similar NFTs (NFTs of the same collection here) can provide insights into pricing trends and patterns. Examining past sales and price fluctuations can help identify potential price movements for similar NFTs in the future.

   We use sales statistics of the collection a particular NFT belongs to on the day of its last sale and current day.

4. **Market Conditions**

   The overall market conditions for NFTs, including market sentiment and investor behaviour, can influence pricing. Factors such as market hype, investor confidence, and market trends should be taken into account.

   We use platform and FT market parameters like volume of their transactions at the day of last sale and current day.

5. **Network Centrality**

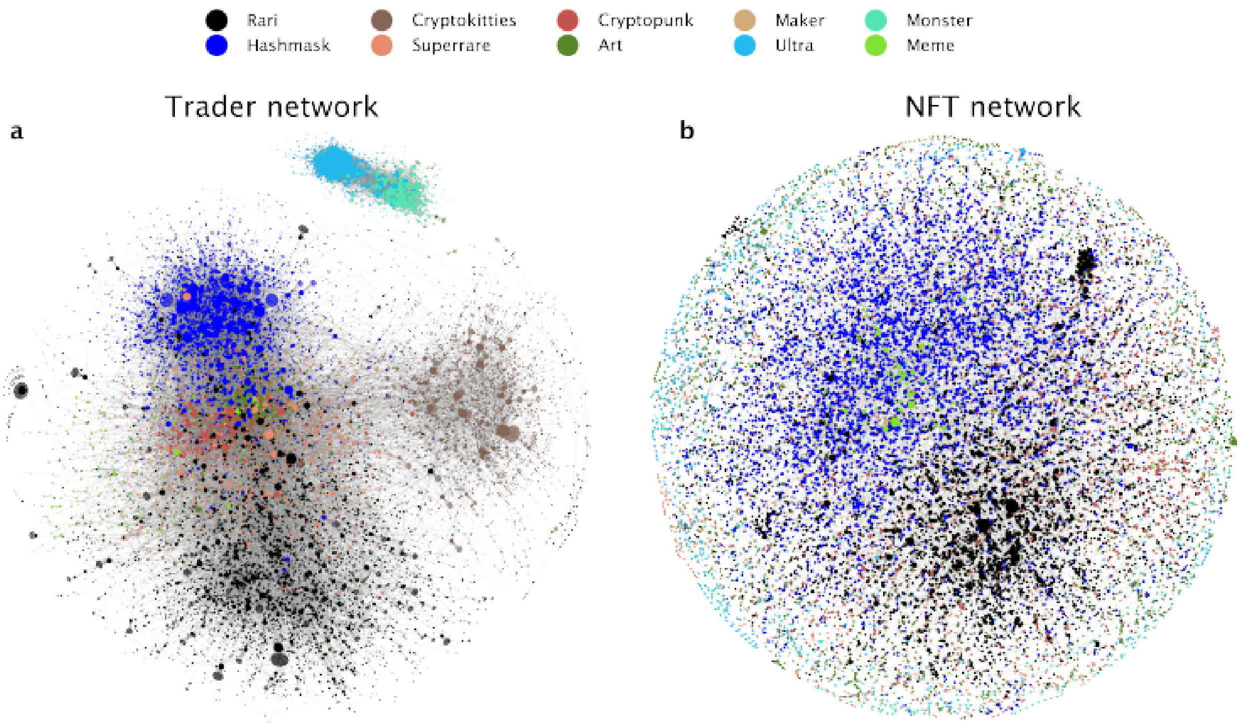


FIGURE 7 - Network centralization

There is an increased likelihood of an NFT being purchased at a more favorable price if it is listed by a seller who actively participates as a collector within the community.

6. **Social Media and Community Engagement**

The level of engagement, discussions, and social media activity surrounding an NFT or its associated community can impact its perceived value. Positive sentiment and active community participation may contribute to higher prices. We intend to use twitter data like follower counts and likes to gauge popularity of the given NFT collection.

There are also other properties like visual aspects of an NFT which we have not considered here.

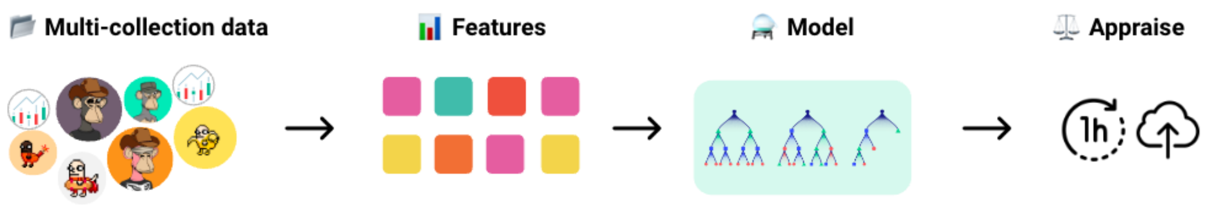
Problem Statement

Our objective is to develop a machine learning model capable of -

- outperforming previous models in accurately predicting the price of NFTs.
- is applicable on new blockchain platforms that it has not been trained on.
- is effective in predicting prices for NFT collections that have recently been minted and are unfamiliar to the model.

Proposed Work

We train NFT sales data from multiple NFT collections on various regression models including and compare their performance on various feature sets.



In layman's terms, our machine learning model understands the relationship between the characteristics of an NFT, past sales of similar items, and the present worth of an NFT. We define the "current value" as an estimation of the selling price that the NFT would likely fetch if it were to be sold on the open market at that exact moment. Now, let's delve into the step-by-step process of how raw NFT data transforms into a valuation.

Data Collection

We collect data from two platforms, namely nonfungible.com and Cryptoslam.io, as part of our data crawling process. We gather sales data of NFTs from a specific category (such as Collectibles, Virtual Land, Art, etc.) by extracting information from the top 1,000 NFT collections listed on nonfungible.com and cryptoslam.io. In addition to the sales data, we incorporate project sales statistics from the day prior to the current sale and the previous sale day. Similarly, we include NFT market statistics and platform statistics from those specific dates. By doing so, we aim to encompass a comprehensive range of past sales information and capture the prevailing market conditions during those periods.

Additionally, we leverage the Alchemy NFT API to calculate rarity scores for large NFT collections.

Data Cleansing and Categorization

Raw data, in its original form, lacks meaningful insights. Therefore, we need to manipulate and refine the data to uncover the specific information that has a significant impact on NFT sale prices.

Data cleaning involves the removal of erroneous, distorted, poorly written, redundant, or inadequate data from a dataset. Data categorization, on the other hand, involves organising data into meaningful categories, facilitating more efficient utilisation and protection of the data.

Our approach is akin to that of experienced traders or collectors who identify patterns based on recent market trends, rarity, and the value of specific traits. By leveraging these patterns, we aim to enhance the predictive power of our model.

NFT Price Prediction using Regression and tree models

We utilize machine learning techniques to train models that can accurately predict transaction prices in the NFT market. These models are built using variables derived from NFT metadata and market indicators, capturing the state of the market at different levels of granularity, such as Ethereum, NFT marketplaces, and specific NFT projects. To ensure the reliability of our predictions, we validate them against data that was not used during the training phase. Additionally, we establish error bounds by comparing our predictions to actual sale prices. The predicted pricing and error bounds offer valuable insights to NFT buyers, sellers, and developers who are involved in the NFT economy. Our valuation approach shares similarities with the high-level principles of art or real estate valuation, although traditional markets rely on simpler linear models with well-established predictor variables based on decades of research and observation. In contrast, the variables influencing NFT prices are still not fully understood.

Evaluation Metrics

1. Mean Absolute Error (MAE)

It is the average of the absolute differences between the actual value and the model's predicted value.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right|$$

2. Mean Squared Error (MSE)

It is the average of the squared differences between the actual and the predicted values.

Lower the value, the better the regression model.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

3. Root Mean Squared Error (RMSE)

It is the average root-squared difference between the real value and the predicted value. By taking a square root of MSE, we get the Root Mean Square Error.

Lower the RMSE value is, the better the model is with its predictions.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

4. R^2 score

R-squared (the coefficient of determination) explains to what extent the variance of one variable explains the variance of the second variable.

$$R^2 = 1 - \frac{SSE}{SST} \quad SSE = \sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \quad SST = \sum_{i=1}^{m}(y_i - \bar{y})^2$$

5. Median absolute percentage error (MAPE)

Captures the 'typical' error. A MAPE of 10% would mean that half of our appraisals have an error of 10% or less.

Performance Analysis

We train our dataset across multiple regression models and compare their accuracy metrics.

Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Random Forest Regressor	709.3202	22958841	4600.8941	0.9315	0.2517	6695.2334
Extra Trees Regressor	704.3915	24079759	4661.5396	0.9284	0.2511	6485.3362
Light Gradient Boosting Machine	907.3626	26230517	4848.7756	0.9243	0.2673	7172.1673
Gradient Boosting Regressor	1117.8579	25287358	4777.9695	0.9233	0.2779	8574.9480
Ridge Regression	1472.9008	26677686	5017.9496	0.9221	0.3414	8366.6302

FIGURE 8 - Collectible price prediction

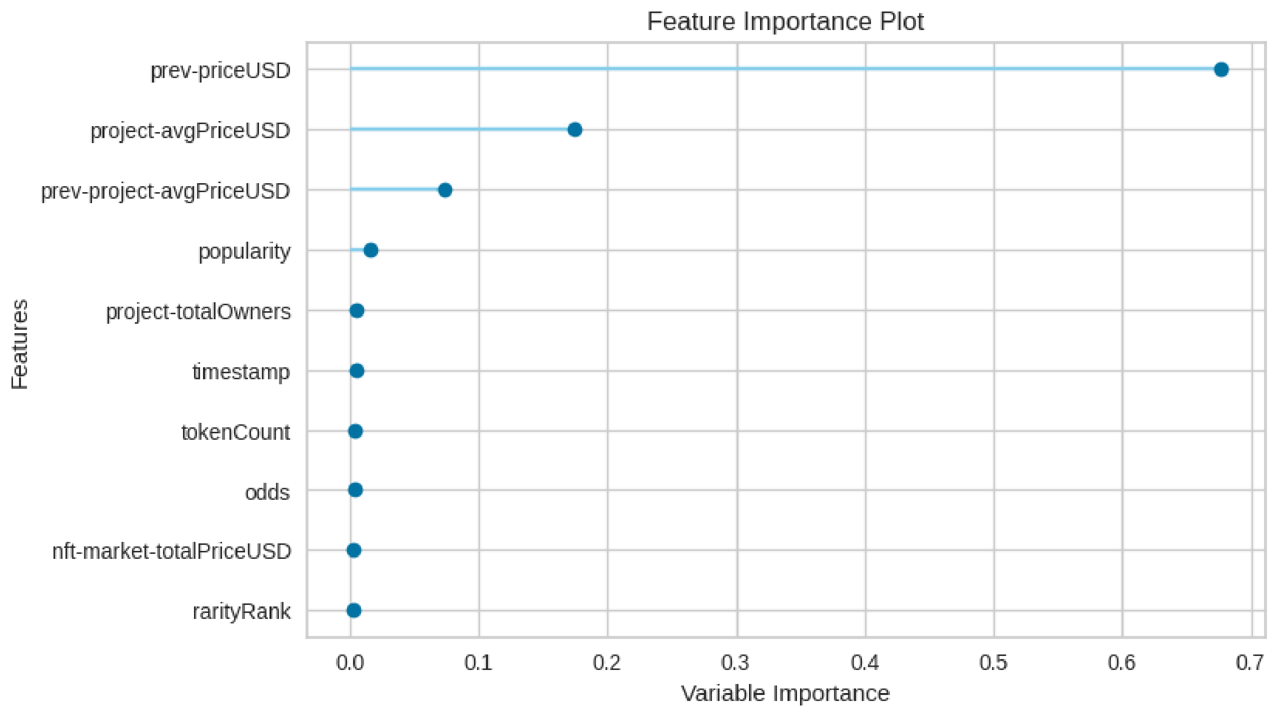


FIGURE 9 - Feature Importance of best model for collectible price prediction

Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Gradient Boosting Regressor	626.7452	11401597	2921.8866	0.6943	0.5187	40.2766
Light Gradient Boosting Machine	611.4642	11883478	3021.7672	0.6710	0.4869	34.4495
Random Forest Regressor	511.3099	12891275	3199.5336	0.6304	0.3443	37.2390
Extra Trees Regressor	471.6145	11932536	3081.8600	0.6213	0.3227	35.0251
Linear Regression	922.6838	13876252	3303.9003	0.6112	0.9230	40.8620

FIGURE 10 - Virtual land price prediction

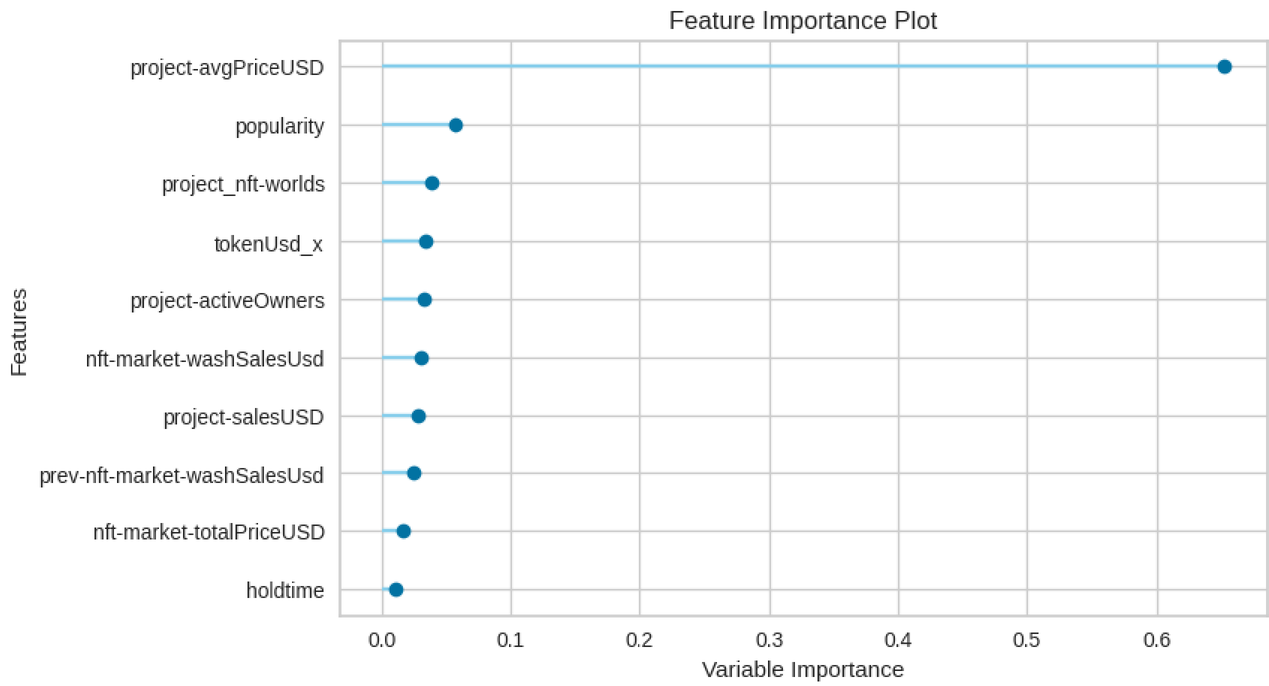


FIGURE 11 - Feature Importance of best model for virtual land price prediction

Without using project tags

Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Random Forest Regressor	711.8802	22978474	4604.2660	0.9314	0.2524	6684.6669
Extra Trees Regressor	702.9161	23951760	4642.7645	0.9297	0.2513	6607.8078
Light Gradient Boosting Machine	912.6332	26504014	4867.2489	0.9236	0.2666	7125.1218
Gradient Boosting Regressor	1124.6946	25375339	4800.9162	0.9232	0.2774	8592.7742
Extreme Gradient Boosting	859.7378	27782101	4997.0570	0.9175	0.2604	7017.0058

FIGURE 12 - Collectible price prediction without project tag

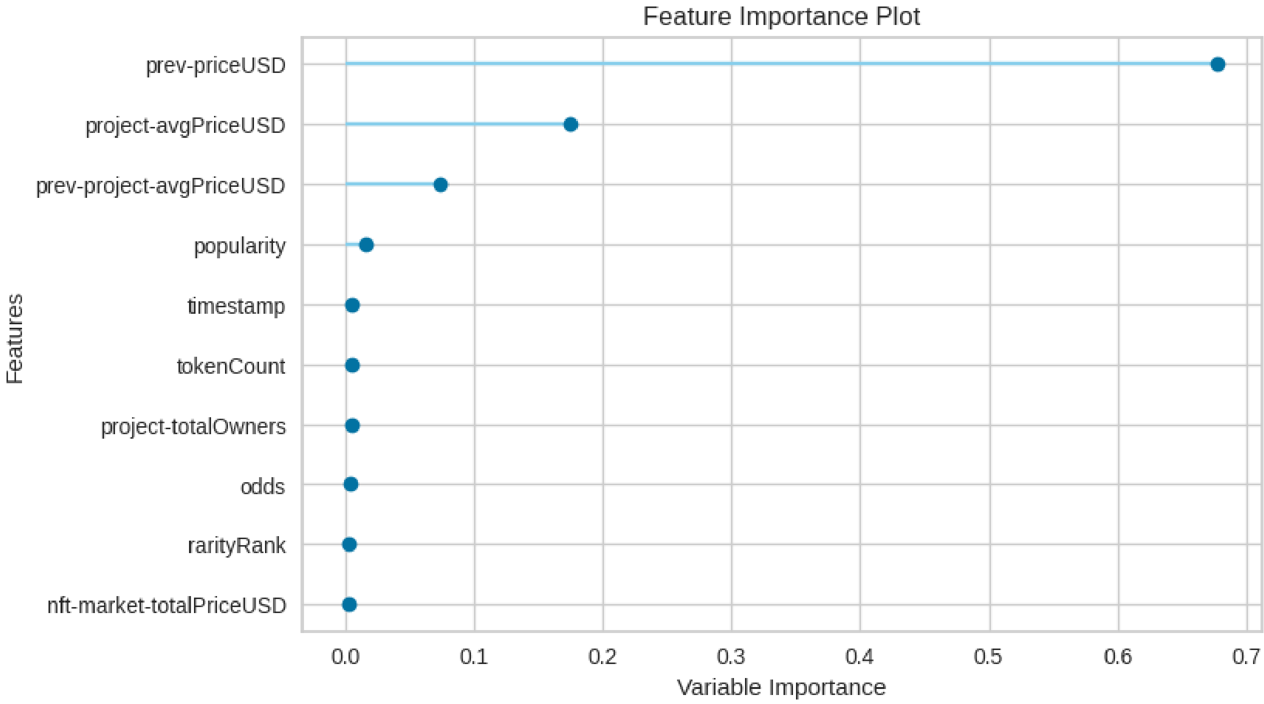


FIGURE 13 - Feature Importance of best model for collectible price prediction without project tag

Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Gradient Boosting Regressor	632.6916	11927538.6773	3012.5837	0.6758	0.5115	58.5009
Light Gradient Boosting Machine	614.9954	12003045.3371	3035.6851	0.6677	0.4907	33.6721
Random Forest Regressor	517.9080	13080713.9005	3224.5371	0.6244	0.3493	37.6660
Extra Trees Regressor	480.2553	12331325.4398	3147.4282	0.6130	0.3278	34.4074
Linear Regression	912.3871	13935363.8005	3313.5526	0.6086	0.8878	41.3928

FIGURE 14 - Virtual land price prediction without project tag

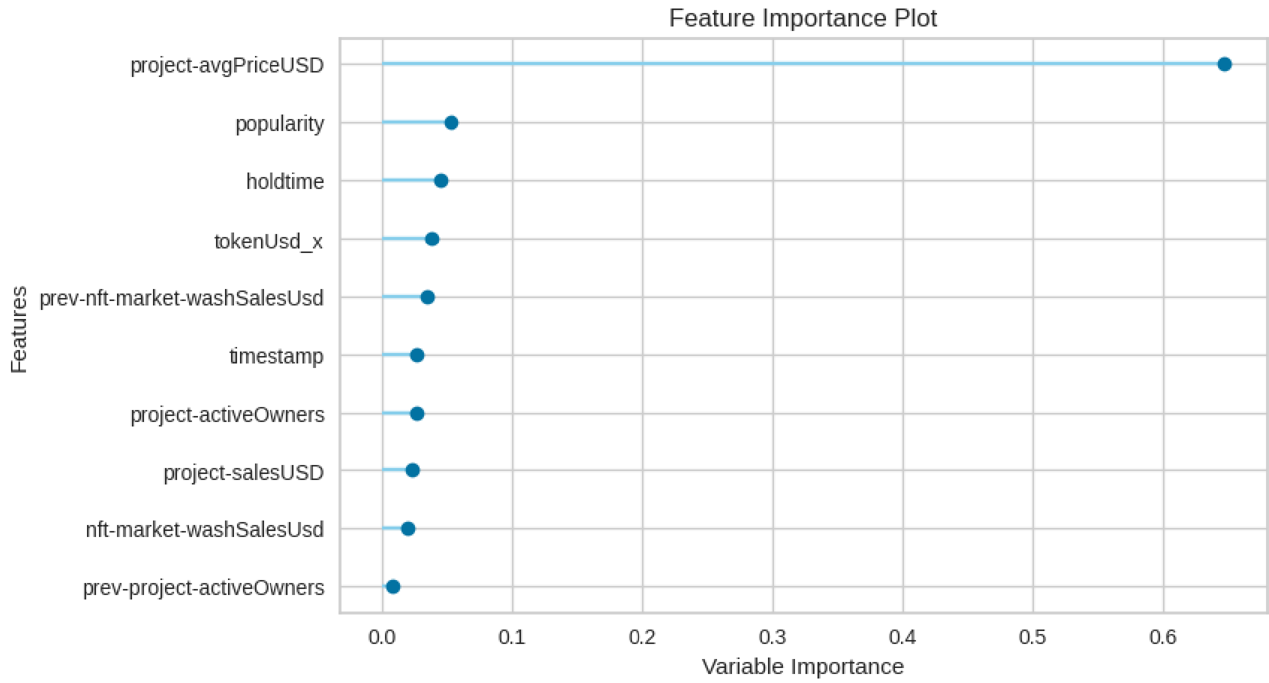


FIGURE 15 - Feature Importance of best model for virtual land price prediction without project tag

The accuracy metrics of the models, whether including or excluding the project tag, remain consistent in both categories. This suggests that our model does not rely heavily on the specific project information, as most of the relevant details can be inferred from the project sales statistics from the previous day and the day of the last sale. Consequently, this indicates that the model can effectively predict NFT prices for relatively new collections without requiring any modifications to its feature sets.

Without using platform tags

Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Random Forest Regressor	708.3057	22878531.3014	4593.3616	0.9317	0.2520	6740.2495
Extra Trees Regressor	700.6764	23225325.7952	4586.3492	0.9313	0.2510	6363.2383
Light Gradient Boosting Machine	907.3626	26230517.8841	4848.7756	0.9243	0.2673	7172.1673
Gradient Boosting Regressor	1116.8176	25457823.4606	4791.2227	0.9228	0.2780	8574.9511
Ridge Regression	1472.9008	26677686.7746	5017.9496	0.9221	0.3414	8366.6302
Linear Regression	1481.3862	26727428.9856	5019.5872	0.9220	0.3424	8331.7898

FIGURE 16 - Collectible price prediction without platform tag

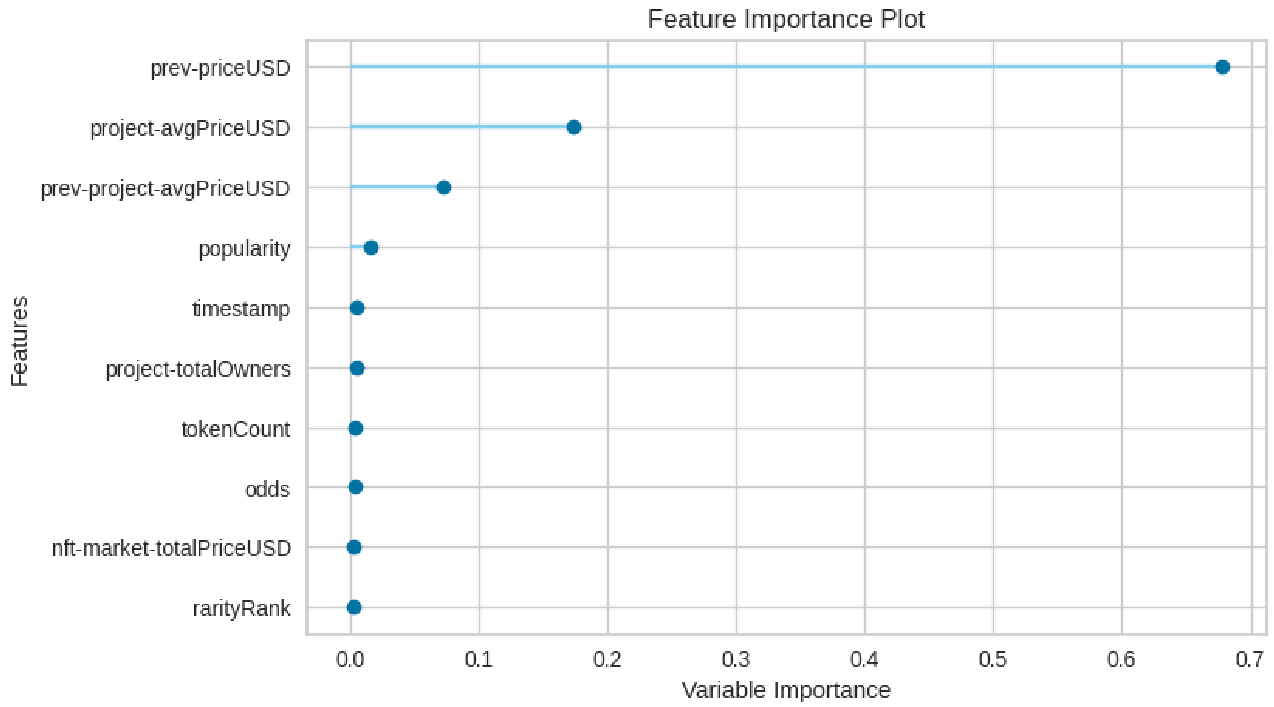


FIGURE 17 - Feature Importance of best model for collectible price prediction without platform tag

Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Gradient Boosting Regressor	629.4873	11498794	2951.4804	0.6886	0.5195	40.2782
Light Gradient Boosting Machine	611.4642	11883478	3021.7672	0.6710	0.4869	34.4495
Random Forest Regressor	512.1551	12888217	3170.7433	0.6355	0.3457	37.9231
Extra Trees Regressor	467.8332	11917598	3076.5595	0.6204	0.3219	33.4409
Linear Regression	922.6942	13876283.	3303.9087	0.6112	0.9230	40.8603

FIGURE 18 - Virtual land price prediction without platform tag

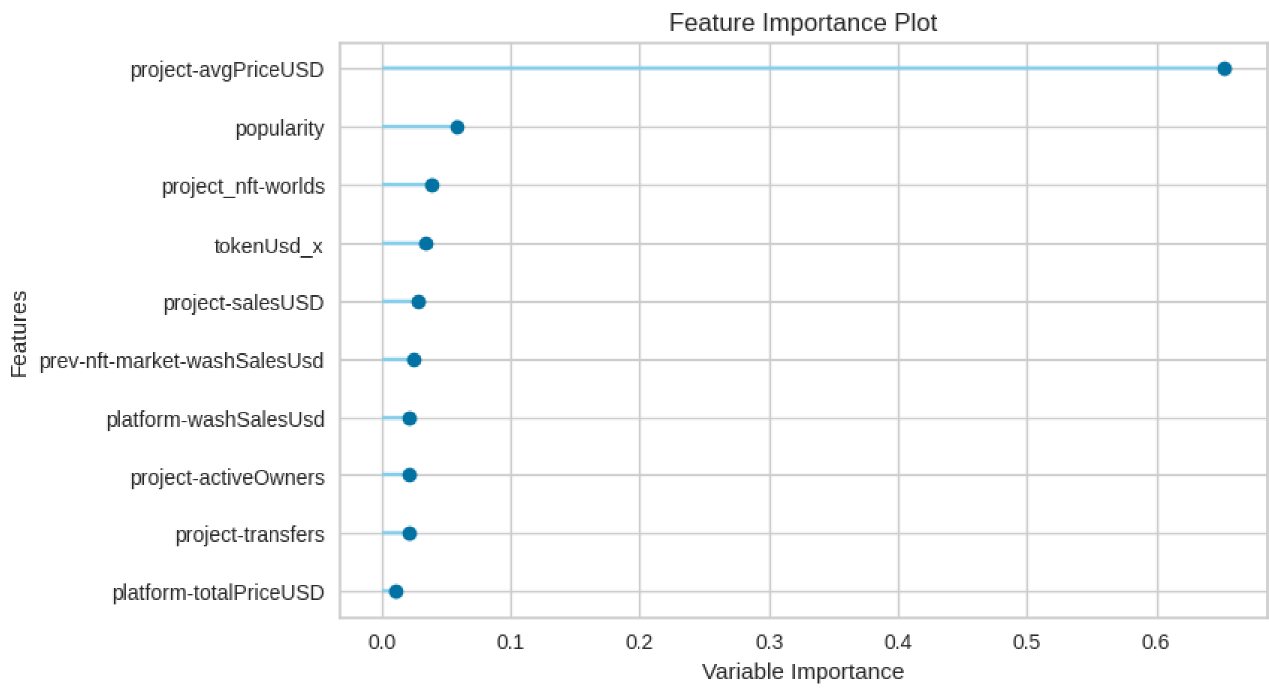


FIGURE 19 - Feature Importance of best model for virtual land price prediction without platform tag

Likewise, the inclusion of platform tags in our model has minimal significance, accounting for only approximately 1% of its overall importance. The majority of the relevant information has already been captured through the platform statistics from the previous day of sale and the day of the last sale. Hence, our model is capable of accurately predicting NFT collections even in newer platforms, particularly considering the growing prominence of layer 2 platforms in the current landscape.

Conclusion

In summary, our model provides accurate predictions of NFT prices for both new collections and established ones with adequate data. This empowers actuarial professionals by equipping them with precise valuation information and enabling them to navigate market competition effectively through informed business intelligence. As a result, our model contributes to the development of a sustainable NFT insurance model that aligns with market dynamics and requirements.

Overall, NFT price prediction brings valuable insights, risk management, transparency, and efficiency to the dynamic NFT market, benefiting investors, collectors, and the ecosystem as a whole.

References

1. M. Nadini and L. Alessandretti and F.D. Giacinto and M. Martino and L.M. Aiello and A. Baronchelli. (2021)
   Mapping the {NFT} revolution: market trends, trade networks, and visual features.
   Springer Science and Business Media {LLC}

2. X. Zheng (2022)
   Multimodal Learning for Improved NFT Price Prediction
   2022 IEEE International Conference on e-Business Engineering (ICEBE), Bournemouth, United Kingdom, 2022 pp. 74-79.
   doi: 10.1109/ICEBE55470.2022.00022

3. Alon, I., Bretas, V. P., & Katrih, V. (2023)
   Predictors of NFT Prices: An Automated Machine Learning Approach.
   Journal of Global Information Management (JGIM), 31(1), 1-18.
   http://doi.org/10.4018/JGIM.317097

4. Balcioglu, Yavuz. (2022)
   Prediction with Machine Learning and comparison of land prices in the metaverse universe

5. Al-majed, Rasha & Abualkishik, Abedallah & Ibrahim, Amer & Mourad, Nahia. (2023)
   Forecasting NFT Prices on Web3 Blockchain Using Machine Learning to Provide SAAS NFT Collectors.
   Fusion: Practice and Applications. 10. 55-68. 10.54216/FPA.100205.