Proposal #17a: Ethereum-based Metaverses Analysis and Models

Stefano Notari 1039001
Giuseppe Massimo 1038803

January 2023

1 Introduction

The project is focused on the study of the Decentraland's metaverse, the goal is to develop a model capable of approximate NFTs' price (for instance LAND).

1.1 What is Decentraland?

Decentraland is a virtual reality platform powered by the Ethereum blockchain. It allows users to create experience and monetize content and applications in a virtual world. Decentraland is a decentralized, community-driven platform where users can buy and build on virtual land, creating unique and immersive experiences. The platform's economy is driven by its own cryptocurrency, the MANA token, which is used to purchase and trade virtual real estate, as well as access premium content and services.

1.2 Dataset used to train the model

This project [1] was done with the collaboration of EPFL-École polytechnique fédérale de Lausanne that provided us the dataset build by WhaleAnalytica [4]. The dataset was constructed by the transactions' history stored on Ethereum blockchain (used by Decentraland). In particular, the dataset for each transaction identifies the date, token id and the sell price.
More than this information, it's also available:

- the trend of keyword "metaverse" on google and twitter and "decentraland" on twitter
- statistics of ethereum's price and MANA (the native token of Decentraland)

1.3 Correlation analysis

The main measures of correlation are Pearson, Spearman, and Kendall. The main characteristics of them are:

- Pearson Correlation measures the linear association between two variables and ranges from -1 to 1. A value of 1 means that there is a perfect positive linear relationship between the two variables, a value of -1 means a perfect negative linear relationship, and a value of 0 means no linear relationship.
- Spearman's Rank Correlation is a non-parametric measure of the association between two variables. It measures the monotonic relationship between two variables, regardless of whether the relationship is linear or not. It also ranges from -1 to 1.
- Kendall's Tau Correlation is also a non-parametric measure of association between two variables. It measures the ordinal association between two variables, the value of Kendall's Tau ranges from -1 to 1.

In order to verify whether there was an effective correlation between some data and the price we chose the Spearman's method because it doesn't assume a specific distribution and can be used for both linear and not linear relationship of the data (so we discarded Pearson). Moreover, because our data doesn't have an ordinal association we have also discarded Kendall.

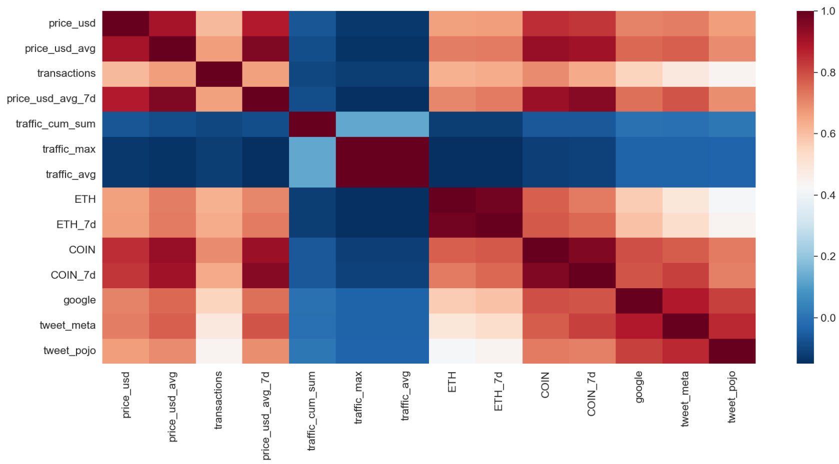

Figure 1: Spearman correlation on the dataset's fields

As we can see in the Figure 1, Spearman's method found a good correlation (red) between some data like transactions, ETH, ETH_7d, COIN, COIN_7d, google, tweet_meta e tweet_pojo and the price (price_usd), while there is no correlation (blue) between traffic's data (traffic_cum_sum, traffic_max e traffic_avg) and the price.
The parameters about google and twitter have the following meaning:

- google: Popularity of the Metaverse trend.
- tweet_pojo and tweet_meta: Popularity of project (Decentraland) and general topic "metaverse".

After having found the correlated data it was created a features matrix (with only the data that have a positive correlation) and the array of ground truth (usd_price).

2 Selection of features and tuning of model

We chose the library XGBoost because it is a gradient boosting algorithm that builds a set of decision trees. So it is easier to interpret and to understand the model's decision. Moreover, it is possible to check the weights the model gives to the input features.

2.1 Tuning of the model

Our model is a XGBRegressor, so the parameters that must be tuned are the following:

- objective : Objective function for XGBRegressor.
- n_estimators: Number of trees.
- max_depth: Maximum depth of a tree. Increasing this value will make the model more complex and more likely to overfit. 0 indicates no limit on depth.
- min_child_weight: Minimum sum of instance weight (hessian) needed in a child. If the tree partition step results in a leaf node with the sum of instance weight less than min_child_weight, then the building process will give up further partitioning. In linear regression task, this simply corresponds to minimum number of instances needed to be in each node. The larger min_child_weight is, the more conservative the algorithm will be.
- gamma: Minimum loss reduction required to make a further partition on a leaf node of the tree. The larger gamma is, the more conservative the algorithm will be.
- subsample: Subsample ratio of the training instances. Setting it to 0.5 means that XGBoost would randomly sample half of the training data prior to growing trees. and this will prevent overfitting.
- learning_rate: Step size shrinkage used in update to prevents overfitting. After each boosting step, we can directly get the weights of new features, and learning_rate shrinks the feature weights to make the boosting process more conservative.

Most of this parameters have a continuous domain so we have decided to select a subset of possible value in order to find the best possible outcome.
We have optimized both the R2 accuracy and neg_mean_squared_error because if we optimize only one of them the results of the metric not chosen worsen.
In the Table 3 are shown the first results obtained.

Feature	Weight
COIN	1342.0
tweet_meta	1069.0
ETH	1008.0
tweet_pojo	889.0
ETH_7d	742.0
COIN_7d	673.0
transactions	598.0
google	293.0

Table 1: Weights for each feature, useful for price approximation

Metric	Value
MSE	10350177.111952
RMSE	3217.169115
MAPE	295.005692
MAE	2078.792938
R2	0.753113

Table 2: Accuracy

Table 3: Summary of the first obtained results

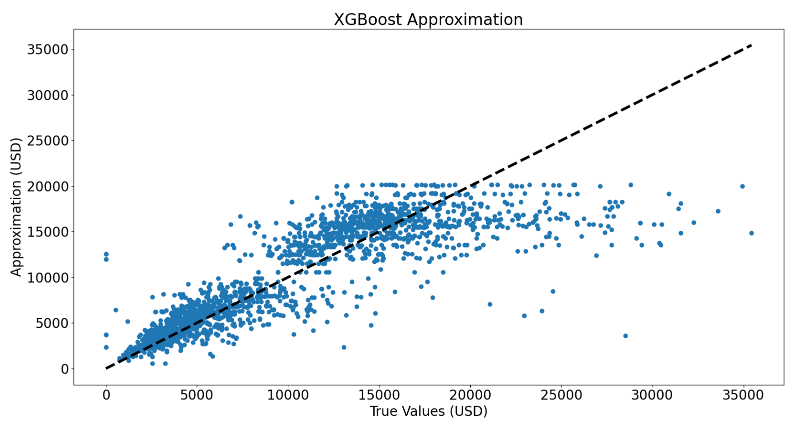

Figure 2: Model's approximations with data that have positive correlation

As can be seen in the Figure 2 there is a saturation around the 20000 dollar, so in order to resolve this issue we tried to feed the model with the position's features and some metadata about the NFTs.

2.2 Metadata and position's features

Since this analysis is focused on the price approximation of virtual estates, it is better to take into consideration also the position in the map of the NFT (as coordinates). This was already proven by a similar analysis on the Sandbox [3] [2], as in the real world the price of estates are affected by the position and nearest to a point of interest, the same can be said for virtual estates. The features that we are going to add are the following:

- x and y: the coordinates for the position of Metaverse's estates,the center (0,0) is the square where the Fashion Week 2022 took place.
- is_road and is_plaza: Type of NFT sold , it can be a simple land if both values are set to 0.
- estate_size: Size of a sold land.
- traffic_avg and traffic_max: respectively, the average of access to the metaverse and the maximum number of simultaneously active users

The results obtained adding the described features are the following:

Feature	Weight
x	5566.0
y	4851.0
COIN	1852.0
ETH	1533.0
tweet_pojo	1483.0
tweet_meta	1412.0
transactions	1364.0
ETH_7d	1231.0
COIN_7d	1206.0
google	518.0
estate_size	393.0
traffic_avg	25.0
traffic_max	24.0

Table 4: Weights for each feature, useful for price approximation

Metric	Value
MSE	8569326.736402
RMSE	2927.341240
MAPE	0.641279
MAE	1681.290057
R2	0.795592

Table 5: Accuracy

Table 6: Results obtained through the position and metadata features

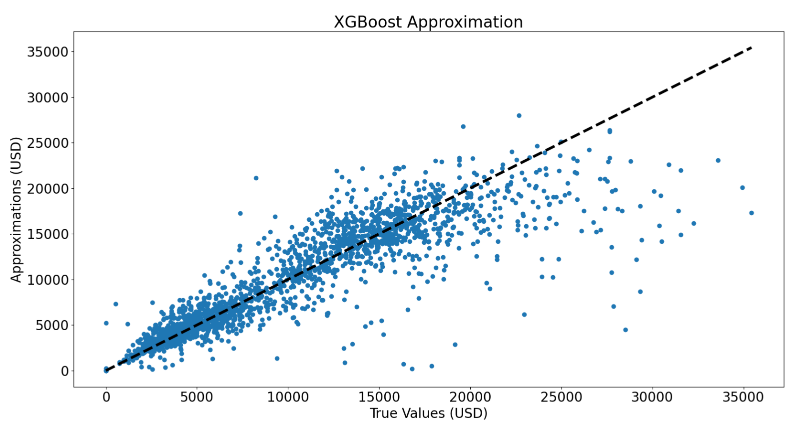

Figure 3: Model's approximations with the position and metadata features

2.3 Analysis on the data with traffic average

Since in the dataset the traffic_avg was monitored only from July 2022 we tried to run the model only on the sub-dataset that had this information. With this constraint we obtained a dataset with 114 elements and then, we got the following results:

Feature	Weight
x	647.0
y	408.0
transactions	365.0
tweet_pojo	267.0
ETH	199.0
tweet_meta	183.0
COIN	169.0
ETH_7d	141.0
google	108.0
traffic_avg	105.0
traffic_max	85.0
COIN_7d	63.0
estate_size	8.0

Table 7: Weights of features on the sub dataset with information of traffic_avg

Metric	Value
MSE	538120.923158
RMSE	733.567259
MAPE	0.142254
MAE	505.881903
R2	0.515291

Table 8: Accuracy

Table 9: Results obtained with tuned parameters

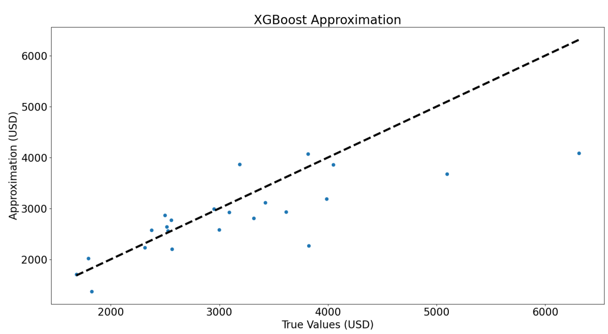

Figure 4: Approximation of the model

As we can see the features of position still the most relevant and the traffic's features the least relevant. There are some switch of some features but nothing of significative.

3 Conclusion and future developments

Analyzing the weights of features obtained in the Table 9, we can see that the features of the LAND's position (x and y) have the greater weights that means they have a greater relevance in the price approximation.
We can conclude that the users pay a lot of attention on LAND's position before the purchase. Moreover, the purchase is also influenced by the daily price of COIN, ETH and the poplarity of the keywords "Decentraland" and "Metaverse" on google and twitter and also by the number of transactions.
We have observed that the traffic's features are not relevant in the price approximation that could mean the Decentraland is little frequented and the main activity is aimed to the speculation.
Our model has obtained a R2 score of 0.80 and a MAPE of 0.64.
Finally, as future developments one could:

- analyze prices in relation to whether or not a point of interest is nearby, such as a casino or an art gallery and the proximity of an event.
- repeat the analysis with a dataset with more entries on traffic parameters and see if the results change.
- use other learning models (LSTM).

References

[1] *Proposal #17a: Ethereum-based Metaverses Analysis and Models*
https://virtuale.unibo.it/mod/page/view.php?id=930332

[2] *LAND Ho! Digital Real Estate in the Metaverse as an Emerging Asset Class Scientific Report*
https://www.blockchainresearchlab.org/2022/08/30/
land-ho-digital-real-estate-in-the-metaverse-as-an-emerging-asset-class-scientific-repo

[3] *Land Valuation in the Metaverse: Location Matters*
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3932189

[4] *WhaleAnalytica*
https://metaverse.whaleanalytica.com/?meta=decentraland

[5] *BRL-Research-Note-2-An-analysis-of-LAND-in-The-Sandbox-Q4-2022.pdf*
https://www.blockchainresearchlab.org/wp-content/uploads/2020/
05/BRL-Research-Note-2-An-analysis-of-LAND-in-The-Sandbox-Q4-2022.
pdf