Estimativa de preços de NFT’s utilizando algoritmos de aprendizagem de máquina

Autores

Isaac Miranda Camargos Email: isaacmirandacamargos@gmail.com

Nicole Maia Argondizzi Email: nicolemaiaargondizzi02@gmail.com

Leandro Cruvinel Lemes Email: leandro.lemes@uftm.edu.br

Resumo

O presente estudo compara os algoritmos Random Forest Regressor, Bagging Regressor, Gradient Boosting Regressor e Decision Tree Regressor, quanto à estimativa de preço das cartas NFT's do jogo Gods Unchained, considerando suas características. O estudo foi dividido em três etapas: coleta, pré-processamento e modelagem dos dados. As métricas utilizadas foram o erro médio absoluto (EMA), erro médio quadrático (EMQ) e coeficiente de determinação (R2). Os resultados obtidos da precificação em Ethereum e GODS foram tecnicamente iguais para os correspondentes modelos. Com relação aos modelos, o Random Forest Regressor e Bagging Regressor obtiveram os melhores resultados nas duas precificações, sendo seus respectivos valores para ETH: EMA 0.124±0.008 e 0.130±0.009; EMQ 0,045±0.006 e 0.049±0.007; R2 0.773±0.028 e 0.752±0.034. Ademais, os seguintes valores para GODS: EMA 0,125±0.007 e 0.1320.008; EMQ 0,044±0.005 e 0.049±0.006; R2 0.778±0.025 e 0.753±0.030.

Resumo (abstract)

The present study compares the algorithms Random Forest Regressor, Bagging Regressor, Gradient Boosting Regressor and Decision Tree Regressor, in estimating the price of NFT's cards from the game Gods Unchained, considering their features. The study was divided into three stages: data collection, pre-processing and modeling. The metrics used were the mean absolute error (AME), mean square error (EMQ) and coefficient of determination (R2). The results obtained from pricing on Ethereum and GODS were technically the same for the corresponding models. Regarding the models, Random Forest Regressor and Bagging Regressor obtained the best results in both pricing, with the respective values for ETH: EMA 0.124±0.008 and 0.130±0.009; CMS 0.045±0.006 and 0.049±0.007; R2 0.773±0.028 and 0.752±0.034. Furthermore, the following values for GODS: EMA 0.125±0.007 and 0.1320.008; CMS 0.044±0.005 and 0.049±0.006; R2 0.778±0.025 and 0.753±0.030.

Introdução

Neste estudo, o objeto de estudo trata das cartas NFT’s do jogo Gods Unchained. Em jogos envolvendo o comercio entre jogadores podem existir manipulações de mercado e monopólios, desse modo, casos em que indivíduos com grande capital compram todos os itens de um determinado tipo e os revendem por altos valores não são incomuns. As estimativas de preços geradas por modelos de aprendizagem de máquinas têm como finalidade guiar e auxiliar os jogadores, sugerindo valores justos e atuais de compra e venda com embasamento científico.

Interessado no livre comércio digital de bens e serviços, em 2008, Satoshi Nakamoto em seu artigo “Bitcoin: A Peer-to-Peer Electronic Cash System” propõem um sistema de armazenamento de dados descentralizados, a Blockchain, e, como prova de seu trabalho, cria a primeira criptomoeda, o Bitcoin (Nakamoto, 2008).

Neste contexto, a rede Blockchain consiste em um banco de dados criptografados, composta por blocos de informações ligadas em cadeia, os quais cada bloco compreende múltiplas transações (Nofer, Gomber, Hinz & Schiereck, 2017). Ademais, o Bitcoin consiste em uma aplicação prática da tecnologia da Blockchain e segue o padrão de tokens ERC-20 (Vogelsteller & Buterin, 2015), padrão este que confere a moeda virtual as mesmas características de uma moeda convencional.

Por outro lado, os NFT’s (non-fungible tokens), cunhados no padrão de tokens ERC-721 (Entriken, Shirley, Evans & Sachs, 2018), são outra categoria de criptomoeda, também criada usando a tecnologia Blockchain. Os NFT’s se distinguem das criptomoedas tradicionais, pois cada um de seus tokens é único, podendo assim ser distinguido e rastreado por seu identificador. Além disso, em prática, os NFT’s podem ser qualquer tipo de bem digital, como imagens, músicas ou vídeos, os mais comuns são trabalhos de artes colecionáveis, objetos virtuais ou itens em jogos.

Nesta pesquisa, motivados pela crescente valorização dos tokens NFT’s, temos como objetivo comparar os algoritmos Random Forest Regressor, Bagging Regressor, Gradient Boosting Regressor e Decision Tree Regressor, quanto à estimativa de preço das cartas NFT's do jogo Gods Unchained, considerando suas características. Como fonte de dados, utilizaremos o jogo já citado que consiste em um jogo de cartas, onde cada carta possui atributos e efeitos únicos. Ademais, nosso objeto de estudo são as cartas NFT’s no jogo, descartando as outras cartas.

Métodos

O estudo foi dividido em três etapas: coleta de dados, pré-processamento e modelagem dos dados. A linguagem de programação utilizada foi Python e as bibliotecas empregadas foram: requests, json, selenium e BeutifulSoup para coleta de dados; pandas, numpy, nltk, re, string e sklearn para o pré-processamento; por fim, novamente, sklearn para aplicação e avaliação dos modelos de aprendizagem de máquina (Pedregosa et al., 2011).

Neste contexto, a coleta de dados se iniciou obtendo informações básicas sobre as cartas, como identificador (*id*), nome, efeito, divindade, raridade, tribo, mana, ataque, vida, tipo e coleção, para isto utilizou-se da API disponibilizada pela Immutable X (Conolly, Ferguson & Ferguson, 2018). Em seguida, foram coletadas informações acerca dos preços das cartas disponíveis em mercado, número de vendas semanais, variação de preços semanais, tendência, porcentagem de baralhos que contem a carta, porcentagem de vitória de baralhos que contem a carta e a porcentagem de vitória de baralhos que contem a carta contra baralhos que não contém a carta, para isto foi utilizado da mesma API. Ademais, os valores das cartas foram obtidos em Ethereum e GODS, sendo está última é uma criptomoeda desenvolvida pelos próprios criadores do jogo.

Encerrada a coleta de dados, o pré-processamento se iniciou organizando e concatenando os dados obtidos pelas diferentes fontes, nessa etapa foram removidas as informações duplicadas e descartadas as cartas que não são comercializáveis no jogo. Foram removidas as cartas que possuem preços extremos, outliers, usando o boxplot. Os dados de texto, com exceção do nome e efeito da carta, foram discretizados e, após isto, foram codificados usando o esquema de codificação “one-hot”. Este sistema codifica e cria uma coluna binaria para cada dado discretizado e retorna uma matriz destas colunas para todos os dados. Os valores de ataque, vida, mana, tendência e número de vendas das cartas foram normalizados subtraindo a média e dividindo pelo desvio padrão. Por último, os preços das cartas foram convertidos em dólar com base na cotação do dia 26/05/2022 das moedas Ethereum e GODS e tratou-se a assimetria aplicando a função ln(x+1), onde x é o preço de cada carta.

Além disso, o efeito da carta, por possuir um grande volume de texto, foi tratado da seguinte forma: foram removidos caracteres especiais, quebras de linha e pontuações; foram retiradas as stopwords (palavras de alta frequência e pouca significância); foram radicalizadas as palavras; e, por fim, foi vetorizado o texto já tratado criando-se uma coluna para cada palavra distinta existente e atribuindo valores a cada correspondente descrição das cartas, seguindo a expressão abaixo.

Onde (wi,j) é valor atribuído para a palavra i na descrição j, (tfi,j) é a frequência da palavra i no texto j, (dfi) é o número de documentos que contenham a palavra i e, por último, (N) é o número total de descrições.

Na etapa de modelagem dos dados, aplicamos os modelos Random Forest Regressor, Bagging Regressor, Gradient Boosting Regressor e Decision Tree Regressor para estimar os preços dos NFT’s. As métricas utilizadas foram o erro médio absoluto (EMA), erro médio quadrático (EMQ) e o coeficiente de determinação (R2).

Analise e discussão de resultados

Os resultados obtidos foram divididos, entre a estimativa de preços das cartas em ETH e a estimativa de preços das cartas em GODS. Desse modo, na Figura 1 constatamos os resultados para o ETH.

Figura 1: Métricas avaliadas dos modelos na estimativa de preços das cartas NFT’s em ETH.

Ressaltando que quanto menor o EMA e o EMQ melhor, por outro lado, quanto maior o R2 melhor. Observou-se que para a precificação em Ethereum o melhor modelo, com os menores EMA, EMQ e maior R2 foi o Random Forest Regressor, sendo seus valores respectivamente 0.124±0.008, 0.045±0.006, 0.773±0.028. Ademais, o modelo com o maior EMA foi o Gradient Boosting Regressor, com 0.150±0.009. Por fim, o modelo com maior EMQ e menor R2 foi o Decision Tree Regressor sendo eles, respectivamente, 0.074±0.015 e 0.624±0.084.

Os resultados para o GODS podem ser constatados na Figura 2. Foi constatado que o modelo com melhores resultados foi o Random Forest Regressor, com EMA, EMQ e R2 respectivamente 0.125 ± 0.007, 0.044 ± 0.005, 0.778 ± 0.025. O modelo com maior EMA foi o Gradient Boosting Regressor, com 0.152 ± 0.007. E, novamente, o modelo com maior EMQ e menor R2 foi o Decision Tree Regressor sendo eles, respectivamente, 0.073 ± 0.014 e 0.627 ± 0.079.

Os resultados adquiridos também podem ser consultados nas Tabelas 1 e Tabelas 2, que acrescentam informações acerca dos desvios padrões. Com base nisto, foi analisado que o modelo com menores desvios padrões foi o Random Forest Regressor, nas duas estimativas de preço.

Figura 2: Métricas avaliadas dos modelos na estimativa de preços das cartas NFT’s em GODS.

|  |  |  |  |
| --- | --- | --- | --- |
| **ETH** | | | |
| Modelos | EMA | EMQ | R² |
| Bagging Regressor | 0,130 ± 0,009 | 0,049 ± 0.007 | 0,752 ± 0,034 |
| Random Forest Regressor | 0,124 ± 0,008 | 0,045 ± 0,006 | 0,773 ± 0,028 |
| Gradient Boosting Regressor | 0,150 ± 0,009 | 0,056 ± 0,008 | 0,715 ± 0,026 |
| Decision Tree Regressor | 0,147 ± 0,014 | 0,074 ± 0,015 | 0,624 ± 0,084 |

Tabela 1: Resultados obtidos dos modelos analisados em ETH.

|  |  |  |  |
| --- | --- | --- | --- |
| **GODS** | | | |
| Modelos | EMA | EMQ | R² |
| Bagging Regressor | 0,132 ± 0,008 | 0,0491 ± 0,006 | 0,753 ± 0,030 |
| Random Forest Regressor | 0,125 ± 0,007 | 0,0442 ± 0,005 | 0,778 ± 0,025 |
| Gradient Boosting Regressor | 0,152 ± 0,007 | 0,0573 ± 0,007 | 0,712 ± 0,029 |
| Decision Tree Regressor | 0,148 ± 0,013 | 0,0739 ± 0,014 | 0,627 ± 0,079 |

Tabela 2: Resultados obtidos dos modelos analisados em GODS.

Os resultados obtidos da precificação em Ethereum e GODS foram tecnicamente iguais para os correspondentes modelos. Com relação aos modelos, o Random Forest Regressor e Bagging Regressor obtiveram os melhores resultados nas duas precificações, com os respectivos valores para ETH: EMA 0.124±0.008 e 0.130±0.009; EMQ 0,045±0.006 e 0.049±0.007; R2 0.773±0.028 e 0.752±0.034. Ademais, os seguintes valores para GODS: EMA 0,125±0.007 e 0.1320.008; EMQ 0,044±0.005 e 0.049±0.006; R2 0.778±0.025 e 0.753±0.030.

Conclusão

Analisando comparativamente os modelos inferem que a aprendizagem de máquina pode auxiliar jogadores assertivamente na compra e venda de suas cartas NFT’s.

No que diz as limitações do presente estudo, para realizar a estimativa de preços foi utilizado informações de cartas já estabelecidas no jogo, como número de vendas semanais, variação de preços semanais e tendência, ou seja, o modelo deste estudo se limita as cartas já existentes, sendo necessárias alterações para estimar preços de cartas a serem lançadas.

Para futuras pesquisas sugere-se: a estimativa de preços utilizando outros algoritmos, inclusive redes neurais; incluir a etapa de hipertunagem de parâmetros dos modelos avaliados; estimar valores futuros para os ativos da coleção, considerando a variável tempo e a liquidez do ativo. Ademais, diante dos resultados propõem-se a criação de um programa para realizar a estimativa de preços das cartas existentes de modo simplificado e de fácil utilização, para que os jogadores consigam analisar quantitativamente valores justos e atuais de compra e venda das cartas. Por fim, a aplicação destes métodos para outros jogos ou coleções de NFT’s.

Referências

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. Decentralized Business Review, 21260.

Nofer, M., Gomber, P., Hinz, O., & Schiereck, D. (2017). Blockchain. Business & Information Systems Engineering, 59(3), 183-187.

Vogelsteller, F., & Buterin, V. (2015, November 19). EIP-20: Token Standard [Review of EIP-20: Token Standard]. EIP-20: Token Standard. https://eips.ethereum.org/EIPS/eip-20

‌

‌

Entriken, W., Shirley, D., Evans, J., & Sachs, N. (2018, January 24). EIP-721: Non-Fungible Token Standard [Review of EIP-721: Non-Fungible Token Standard]. https://eips.ethereum.org/EIPS/eip-721

‌Connolly, A., Ferguson, J., & Ferguson, R. (2018). immutable x api [Review of immutable x api]. Immutable X. https://www.immutable.com

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.