Blockchain, originalmente proposta por Nakamoto [4], consiste em um banco de dados representados por blocos de informações criptografadas, ligadas em cadeia, em que cada bloco compreende múltiplas transações [2]. Tal tecnologia foi proposta para possibilitar o livre comercio digital, diretamente entre pessoas, livrando-se da necessidade de uma terceira entidade mediadora e, consequentemente, escapando de taxas pelos seus serviços.

Com base nisso, e como prova deste trabalho [4], juntamente com a criação desse banco de dados decentralizado, surgiu a primeira criptomoeda, o Bitcoin. O Bitcoin após sua criação levou alguns anos até sua eventual popularização, sendo marcado por uma expressiva valorização no início de 2021, atingindo a marca histórica, no dia 10 de novembro deste mesmo ano, o incrível valor de 69.044,77 dólares por uma única moeda [1].

Embora não ligados a uma instituição financeira, as criptomoedas seguem o padrão de tokens ERC-20, o que as conferem as mesmas características de uma moeda convencional, como por exemplo, uma moeda qualquer possui o mesmo tipo e valor de qualquer outra moeda [6]. Por outro lado, semelhantes em origem, porém diferentes em propriedades, os NFT's (non-fungible tokens) são um tipo de criptomoeda derivadas de contratos inteligentes da rede blockchain Ethereum [3]. Os NFT's são cunhados no padrão de tokens ERC-721 [5], onde cada token é único, podendo ser distinguido e rastreado por seu tokenId. Em pratica, os NFT's podem ser qualquer tipo de bem digital, sendo os mais comuns trabalhos de arte colecionáveis, objetos virtuais, músicas ou itens dentro de jogos.

Como explicado anteriormente, os NFT's, diferentemente das crypto-moedas, são únicos tanto em identificação como em características, o que torna sua precificação uma tarefa árdua e subjetiva. Neste presente trabalho, motivados pela crescente valorização desses tokens e popularização dessa tecnologia, temos como objetivo utilizar algoritmos de aprendizado de máquina (AM), para estudar a provável relação entre os atributos de um NFT e o seu possível preço.

Como objeto de estudo, utilizaremos um caso particular o Gods Unchained, que consiste em um jogo cujo objetivo é derrotar o seu oponente em uma batalha de cartas. Neste caso, o token NFT são as cartas, essas possuem características únicas bem definidas, como mana, vida, ataque e descrição. Tais individualidades visam favorecer a avaliação de preços, considerando a ausência de literatura cientifica nessa área.

Das ferramentas utilizadas neste trabalho estão a linguagem de programação python e algumas de suas bibliotecas utilizadas na ciência de dados, como pandas, numpy, scikit-learn e Natural Language Toolkit (NLTK). Ademais fez-se uso das bibliotecas Selenium, Beautiful Soup, requests e json, que em conjunto com algumas interfaces de programação de aplicações (API) estruturaram a base para obtenção de dados. Diante disso, o trabalho foi dividido em três etapas, obtenção de dados, pré-processamento e, por fim, aplicações de algoritmos AM.

Neste quadro, utilizando-se de uma API [9], se levantou dados acerca das cartas existentes dentro do jogo, como por exemplo nome, raridade, conjunto, tipo, dentre outras individualidades. Seguiu-se obtendo os preços das cartas já listadas no mercado, preços esses em Ethereum (ETH) e GODS. Essa última, é uma crypto-moeda desenvolvida para transações dentro do jogo. Feito isso, foi realizado uma raspagem de dados da web [10] coletando informações relacionadas a taxa de vitória e porcentagem de baralhos que possuem as cartas analisadas. A fim de filtrar os dados obtidos, se usou como critério de exclusão o fato de algumas cartas não serem colecionáveis (cartas geradas a partir de efeitos de outras cartas). Em síntese, combinando todas essas informações construiu-se o banco de dados a ser trabalhado.

Na etapa de pré-processamento, se separou os dados em três tipos: textuais, numéricos e a descrição das cartas. Seguiu-se aplicando modelos específicos de tratamento. Nos dados textuais, se empregou respectivamente a binarização das informações (representar na forma de 0 ou 1) e conversão delas em um banco de dados. Sob os dados numéricos, foi empregada a normalização, tornando a média zero e desvio padrão igual a 1, seu equacionamento pode ser visto a seguir. Onde x é o valor atual, u a média dos valores, s o desvio padrão e Z é o novo valor.

Z = (x-u)/s

Posteriormente, se realizou o tratamento da descrição da carta. Para tal, o texto foi manipulado removendo caracteres especiais, links, tags e stopwords (palavras de alta frequência e pouca significância). Fazendo uso de modelos linguísticos, extraiu-se o radical das palavras com a finalidade de simplificar sem perder o significado. Seguiu-se vetorizando o texto considerando a frequência de cada palavra nas descrições. Em resolução, se uniu os dados textuais, numéricos e a descrição já tratados de modo a tornar possível a aplicação de modelos a seguir.

Na etapa de aplicação dos modelos de AM, se separou os dados pré processados em treino e teste, para respectivo treinamento e avaliação dos modelos. Nesse contexto, nas ferramentas de avaliação fez-se uso do erro médio absoluto, erro médio quadrático e o coeficiente de determinação também conhecido como R^2. Por fim, se aplicou os modelos de AM.

Dentre os modelos de AM utilizados e entre os que tiveram melhores resultados pode-se citar Random Forest Regressor e Bagging Regressor, obtendo: erro médio absoluto entre 0.260093 e 0.27631; erro médio quadrático entre 0.191451 e 0.212342; e coeficiente de determinação entre 0.869626 e 0.882453.

A rede Blockchain consiste em um banco de dados representados por blocos de informações criptografadas, ligadas em cadeia, em que cada bloco compreende múltiplas transações \cite{Nofer}. Tal tecnologia foi proposta por Nakamoto \cite{nakamoto} para possibilitar o livre comercio digital, diretamente entre pessoas, livrando-se da necessidade de uma terceira entidade mediadora e, consequentemente, escapando de taxas pelos seus serviços. Com base nisso, juntamente com a criação desse banco de dados decentralizado, surgiu a primeira criptomoeda o Bitcoin que após sua criação levou alguns anos até sua eventual popularização, sendo marcado por uma expressiva valorização no início de 2021, atingindo a marca histórica, no dia 10 de novembro deste mesmo ano, o incrível valor de 69.044,77 dólares por uma única moeda \cite{Gecko}.

Embora não ligados a uma instituição financeira, as criptomoedas seguem o padrão de tokens ERC-20, o que as conferem as mesmas características de uma moeda convencional, como por exemplo, uma moeda qualquer possui o mesmo tipo e valor de qualquer outra moeda \cite{ercvinte}. Por outro lado, semelhantes em origem, os NFT's (non-fungible tokens) são um tipo de criptomoeda derivados de contratos inteligentes da rede blockchain. Os NFT's são cunhados no padrão de tokens ERC-721 \cite{erc721}, onde cada token é único, podendo ser distinguido e rastreado por seu identificador. Em prática, os NFT's podem ser qualquer tipo de bem digital, sendo os mais comuns trabalhos de arte colecionáveis, objetos virtuais, músicas ou itens dentro de jogos.

Como explicado anteriormente, os NFT's, diferentemente das crypto-moedas, são únicos tanto em identificação como em características, o que torna sua precificação uma tarefa árdua e subjetiva. Neste presente trabalho, motivados pela crescente valorização desses tokens e popularização dessa tecnologia, temos como objetivo utilizar algoritmos de aprendizado de máquina (AM), para estudar a provável relação entre os atributos de um NFT e o seu possível preço. Como objeto de estudo, utilizaremos um caso particular o Gods Unchained, que consiste em um jogo cujo objetivo é derrotar o seu oponente em uma batalha de cartas. Neste caso, o token NFT são as cartas, essas possuem características únicas bem definidas, como mana, vida, ataque e descrição. Tais individualidades visam favorecer a avaliação de preços, considerando a ausência de literatura cientifica nessa área.

Das ferramentas utilizadas neste trabalho estão a linguagem de programação python e algumas de suas bibliotecas utilizadas na ciência de dados, como pandas, numpy, scikit-learn e Natural Language Toolkit (NLTK). Ademais fez-se uso das bibliotecas Selenium, Beautiful Soup, requests e json, que em conjunto com algumas interfaces de programação de aplicações (API) \cite{APIgods} estruturaram a base para obtenção de dados. Diante disso, o trabalho foi dividido em três etapas, obtenção de dados, pré-processamento e, por fim, aplicações de algoritmos AM.

Neste quadro, utilizando-se de uma API, se levantou dados acerca das cartas existentes dentro do jogo, como por exemplo nome, raridade, conjunto, tipo, dentre outras individualidades. Seguiu-se obtendo os preços das cartas já listadas no mercado, preços esses em Ethereum (ETH) e GODS. Essa última, é uma crypto-moeda desenvolvida para transações dentro do jogo. Feito isso, foi realizado uma raspagem de dados da web \cite{GUdecks} coletando informações relacionadas a taxa de vitória e porcentagem de baralhos que possuem as cartas analisadas. A fim de filtrar os dados obtidos, se usou como critério de exclusão o fato de algumas cartas não serem colecionáveis (cartas geradas a partir de efeitos de outras cartas). Em síntese, combinando todas essas informações construiu-se o banco de dados a ser trabalhado.

Na etapa de pré-processamento, se separou os dados em três tipos: textuais, numéricos e a descrição das cartas. Seguiu-se aplicando modelos específicos de tratamento. Nos dados textuais, se empregou respectivamente a binarização das informações (representar na forma de 0 ou 1) e conversão delas em um banco de dados. Sob os dados numéricos, foi empregada a normalização, tornando a média zero e desvio padrão igual a 1, seu equacionamento pode ser visto a seguir. Onde x é o valor atual, u a média dos valores, s o desvio padrão e Z é o novo valor.

\begin{equation}

Z = \frac{x-u}{s}

\end{equation}

Posteriormente, se realizou o tratamento da descrição da carta. Para tal, o texto foi manipulado removendo caracteres especiais, links, tags e stopwords (palavras de alta frequência e pouca significância). Fazendo uso de modelos linguísticos, extraiu-se o radical das palavras com a finalidade de simplificar sem perder o significado. Seguiu-se vetorizando o texto considerando a frequência de cada palavra nas descrições. Em resolução, se uniu os dados textuais, numéricos e a descrição já tratados de modo a tornar possível a aplicação de modelos a seguir.

Na etapa de aplicação dos modelos de AM, se separou os dados pré processados em treino e teste, para respectivo treinamento e avaliação dos modelos. Nesse contexto, nas ferramentas de avaliação fez-se uso do erro médio absoluto, erro médio quadrático e o coeficiente de determinação também conhecido como R2. Por fim, se aplicou os modelos de AM.

Dentre os modelos de AM utilizados os que tiveram melhores resultados pode-se citar o Random Forest Regressor e o Bagging Regressor, obtendo: erro médio absoluto entre 0.260093 e 0.27631; erro médio quadrático entre 0.191451 e 0.212342; e coeficiente de determinação entre 0.869626 e 0.882453.

[1] https://www.coingecko.com/en/coins/bitcoin

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

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

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

[5] William, E., Dieter, S., Jacob, E., Nastassia, S.: Eip-721: Erc-721 non-fungible token standard. Accessible: https://eips.ethereum.org/EIPS/eip-721 (2018)

[6] Fabian, V., Vitalik, B.: Eip-20: Erc-20 token standard. Accessible: https://eips.ethereum.org/EIPS/eip-20 (2015)

[7] Dowling, M. (2022). Is non-fungible token pricing driven by cryptocurrencies?. Finance Research Letters, 44, 102097.

[8] https://cryptoslam.io/

[9] https://api.godsunchained.com

[10] https://gudecks.com/meta/card-rankings

[11] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html

[12] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

[13] https://www.nltk.org/\_modules/nltk/stem/snowball.html

[14] <https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html>

[12] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

[13] https://www.nltk.org/\_modules/nltk/stem/snowball.html

[14] https://scikit-learn.org/stable/modules/generated/sklearn.feature\_extraction.text.TfidfVectorizer.html

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

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

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

[5] William, E., Dieter, S., Jacob, E., Nastassia, S.: Eip-721: Erc-721 non-fungible token standard. Accessible: https://eips.ethereum.org/EIPS/eip-721 (2018)

[6] Fabian, V., Vitalik, B.: Eip-20: Erc-20 token standard. Accessible: https://eips.ethereum.org/EIPS/eip-20 (2015)

[7] Dowling, M. (2022). Is non-fungible token pricing driven by cryptocurrencies?. Finance Research Letters, 44, 102097.

[8] https://cryptoslam.io/

[9] https://api.godsunchained.com

[10] <https://gudecks.com/meta/card-rankings>

Este é o padrão (formato \LaTeX{} apenas) para a submissão de trabalhos da Categoria 1 do CNMAC, destinados à divulgação de pesquisas em andamento, com resultados preliminares, e pesquisas em nível de Iniciação Científica. \emph{Nesta categoria, os trabalhos devem ser submetidos em Português ou Inglês, em forma de resumo de, no máximo, duas páginas, incluindo-se as referências bibliográficas.} Os \emph{trabalhos submetidos} que \emph{não estiverem de acordo com o formato} apresentado por esse padrão \emph{serão rejeitados} pelo Comitê Editorial do evento, sem análise do mérito científico.

Equações inseridas no resumo devem ser enumeradas sequencialmente e à direita no texto, por exemplo

\begin{equation}

\frac{\partial u}{\partial t}-\Delta u = f, \quad \mathrm{em} \; \Omega. \label{Calor}

\end{equation}

Consulte o arquivo \verb!.tex! para mais detalhes sobre o código-fonte gerador da equação \eqref{Calor}.

Tendo em vista tratar-se de um resumo, sugere-se evitar a inserção de seções, tabelas e figuras. Caso necessária, a inserção de tabela deve ser feita com o ambiente \verb!table!, sendo enumerada, disposta horizontalmente centralizada, próxima de sua referência no texto, e a legenda imediatamente acima dela. Por exemplo, consulte a Tabela \ref{tabela01}.

\begin{table}[H]

\caption{ {\small Categorias dos trabalhos.}}

\centering

\begin{tabular}{ccc}

\hline

Categoria do trabalho & Número de páginas & Tipo do trabalho\\ \hline

1 & 2 & $A$, $B$ e $C$ \\

2 & entre 5 e 7 & apenas $C$ \\

\hline

\end{tabular}\label{tabela01}

\end{table}

A inserção de figura deve ser feita com o ambiente \verb!figure!, ela deve estar enumerada, disposta horizontalmente centralizada, próxima de sua referência no texto, e legenda imediatamente abaixo dela. \emph{Quando não própria, deve-se indicar/referências a fonte.} Por exemplo, consulte a Figura \ref{figura01}.

\begin{figure}[H]

\centering

\includegraphics[width=.475\textwidth]{ex\_fig}

\caption{ {\small Exemplo de imagem. Fonte: indicar.}}

\label{figura01}

\end{figure}

As referências bibliográficas devem ser inseridas conforme especificado neste padrão, sendo que serão automaticamente geradas em ordem de citação no texto. Este {\it template} fornece suporte para a inserção de referências bibliográficas com o Bib\LaTeX{}. Os dados de cada referência do trabalho devem ser adicionados no arquivo \verb+refs.bib+ e a indicação da referência no texto deve ser inserida com o comando \verb+\cite+. Seguem alguns exemplos de referências: livro \cite{Boldrini}, artigos publicados em periódicos \cite{Contiero,Cuminato}, capítulo de livro \cite{daSilva}, dissertação de mestrado \cite{Nofer}, tese de doutorado \cite{Mallet}, livro publicado dentro de uma série \cite{Gomes}, trabalho publicado em anais de eventos \cite{Santos}, {\it website} e outros \cite{CNMAC}. Sempre que disponível forneça o DOI, ISBN ou ISSN, conforme o caso.

\begin{tabular}{c|c|c|c}

\hline

oi ($a$) & Erro médio absoluto ($E1$) & Erro médio quadrático ($E2$) & Coeficiente de determinação ($R2$) \\

\hline

$Random Forest Regressor$ & $0.260093$ $E1$ & $0.191451$ $E2$ & $0.869626$ $R2$

\hline

$Bagging Regressor$ & $0.27631$ $E1$ & $0.212342$ $E2$ & $0.882453$ $R2$

\hline

\end{tabular}