

CARD COST PREDICTION IN LEGENDS OF
RUNETERRA USING RANDOM FOREST: A MACHINE
LEARNING APPROACH FOR ANALYZING
ELECTRONIC GAME DATA



PREVISÃO DO CUSTO DE CARTAS EM LEGENDS OF RUNETERRA UTILIZANDO RANDOM FOREST: UMA ABORDAGEM DE APRENDIZADO DE MÁQUINA PARA ANÁLISE DE DADOS DE JOGOS ELETRÔNICOS

GUIMARÃES, Pedro Henrique Vieira; SANTOS, Flávia Aparecida Oliveira;
CARVALHO, Jaqueline Corrêa Silva; CARVALHO, Marcos Alberto;
RAMOS, Celso de Ávila; BASTOS, Camila; SOUZA, Patrícia Carolina;
SILVA, Vinícius Duarte Esteves

Pedro Henrique Vieira Guimarães,
UNIFENAS, Brasil

Flávia Aparecida Oliveira Santos, UNIFENAS,
Brasil

Jaqueline Corrêa Silva Carvalho, UNIFENAS,
Brasil

Marcos Alberto Carvalho, UNIFENAS,
Brasil

Celso de Ávila Ramos, UNIFENAS, Brasil

Camila Bastos, UNIFENAS, Brasil

Patrícia Carolina Souza, UNIFENAS, Brasil

Vinícius Duarte Esteves, UNIFENAS, Brasil

Revista Científica da UNIFENAS
Universidade Professor Edson Antônio Velano, Brasil
ISSN: 2596-3481
Publicação: Trimestral
vol. 6, nº. 5, 2024
revista@unifenas.br

Recebido: 08/07/2024
Aceito: 28/08/2024
Publicado: 09/09/2024

URL: <https://revistas.unifenas.br/index.php/revistaunifenas/issue/view/52>

DOI: 10.29327/2385054.6.5-13

ABSTRACT: Artificial intelligence (AI) and machine learning (ML) are revolutionizing data analysis and decision-making in various fields, including electronic games. This study applied the Random Forest technique, a machine learning algorithm, to analyze and model card data from the game Legends of Runeterra (LoR). The main objective was to build a predictive model to estimate a card's cost based on its attributes and explore the distribution of card types in the game. The methodology involved exploratory analysis of a public dataset containing information about LoR cards, using the Python libraries Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn. The Random Forest model was trained and tested on subsets of the data, and its performance was evaluated using the mean squared error (MSE) and the coefficient of determination (R^2). The results revealed insights into the distribution of cost and card types in the game, with most cards having a low cost and "Unit" type cards being the most frequent. The predictive model achieved an R^2 of 0.516, indicating that it explains 51.6% of the variance in card cost. This study demonstrates the potential of machine learning to analyze electronic game data and assist players and developers in making strategic decisions. The conclusions pave the way for future research exploring other machine learning algorithms and card attributes, aiming to improve the accuracy of the predictive model and deepen the understanding of game dynamics.

KEYWORDS: Machine learning, Random Forest, Predictive modeling, Supervised learning, Electronic games.

RESUMO: A inteligência artificial (IA) e o aprendizado de máquina (machine learning) estão revolucionando a análise de dados e a tomada, incluindo jogos eletrônicos. Este estudo aplicou a técnica de Random Forest, um algoritmo de aprendizado de máquina, para analisar e modelar dados de cartas do jogo Legends of Runeterra (LoR). O objetivo foi

construir um modelo preditivo para estimar o custo de uma carta com base em seus atributos e explorar a distribuição dos tipos de carta no jogo. A metodologia envolveu a análise exploratória de um conjunto de dados público contendo informações sobre as cartas do LoR, utilizando as bibliotecas Python Pandas, NumPy, Matplotlib, Seaborn e Scikit-learn. O modelo de Random Forest foi treinado e testado em subconjuntos dos dados. Seu desempenho foi avaliado utilizando o erro quadrático médio (MSE) e o coeficiente de determinação (R^2). Os resultados revelaram insights sobre a distribuição do custo e dos tipos de carta no jogo, com a maioria das cartas apresentando custo baixo e as cartas do tipo "Unidade" sendo as mais frequentes. O modelo preditivo alcançou um R^2 de 0.516, indicando que explica 51.6% da variância no custo das cartas. Este estudo demonstra o potencial do aprendizado de máquina para analisar dados de jogos eletrônicos e auxiliar jogadores e desenvolvedores na tomada de decisões estratégicas. As conclusões abrem caminho para futuras pesquisas que explorem outros algoritmos de aprendizado de máquina e atributos das cartas, visando aprimorar a precisão do modelo preditivo e aprofundar a compreensão das dinâmicas do jogo.

PALAVRAS-CHAVE: Aprendizado de máquina, Random Forest, Modelagem preditiva, Treinamento Supervisionado, Jogos eletrônicos

1 INTRODUÇÃO

A inteligência artificial (IA) tem se destacado como uma área de pesquisa em constante expansão, com aplicações em diversos campos. Em particular, o aprendizado de máquina (machine learning), um subcampo da IA, tem revolucionado a forma como os sistemas lidam com dados e tomam decisões. O aprendizado de máquina permite que computadores aprendam padrões e relações a partir de dados, sem serem explicitamente programados para isso [1].

Uma das técnicas de aprendizado de máquina amplamente utilizada é a Random Forest (Floresta Aleatória), um algoritmo de aprendizado supervisionado que se baseia na combinação de múltiplas árvores de decisão. Cada árvore é treinada em um subconjunto aleatório dos dados, e a previsão final é obtida pela agregação das previsões de todas as árvores. Essa abordagem contribui para a robustez e generalização do modelo, tornando-o menos propenso a overfitting [2,3].

Aprendizado supervisionado é uma abordagem onde o modelo é treinado em um conjunto de dados rotulados, ou seja, dados de entrada emparelhados com as respostas corretas. O objetivo é que o

modelo aprenda a mapear entradas para saídas, permitindo previsões precisas em novos dados.

O aprendizado de máquina está se tornando cada vez mais popular na indústria de jogos eletrônicos devido à grande quantidade de dados disponíveis e ao crescente poder computacional do hardware moderno. Em particular, o jogo Legends of Runeterra (LoR), um popular jogo de cartas digital, oferece um rico conjunto de dados que pode ser explorado para entender melhor as dinâmicas do jogo e desenvolver ferramentas que auxiliem jogadores e desenvolvedores [4].

Este trabalho tem como objetivo principal investigar a aplicação de aprendizado de máquina, utilizando a técnica de Random Forest, para analisar e modelar dados de cartas do jogo LoR. Especificamente, busca-se construir um modelo preditivo capaz de estimar o custo de uma carta com base em seus atributos, como vida e ataque. Adicionalmente, este estudo visa explorar a distribuição dos tipos de carta no jogo. As hipóteses que este trabalho procura responder são:

1. É possível construir um modelo de aprendizado de máquina capaz de prever o custo de uma carta em LoR com base em seus atributos?

2. Quais são os tipos de cartas mais frequentes no jogo?

A relevância deste trabalho reside no potencial de contribuir para o entendimento das dinâmicas do jogo LoR, tanto do ponto de vista dos jogadores quanto dos desenvolvedores. Um modelo preditivo preciso para o custo das cartas pode auxiliar jogadores na construção de decks mais eficientes, enquanto a análise exploratória dos dados pode revelar insights sobre o meta e as estratégias mais eficazes. Além disso, este estudo demonstra a aplicabilidade de técnicas de aprendizado de máquina em jogos eletrônicos, abrindo caminho para futuras pesquisas e desenvolvimentos na área.

2 METODOLOGIA

Este estudo utilizou uma abordagem quantitativa, baseada em análise exploratória de dados e modelagem preditiva, para investigar o conjunto de dados "Legends of Runeterra - All Cards", disponível publicamente no Kaggle (<https://www.kaggle.com/datasets/brandonqilin/legends-of-runeterra>) [5]. Este conjunto contém 1445 cartas do jogo, cada uma com 11 atributos, incluindo região, custo, raridade, tipo, palavras-chave, entre outros. As bibliotecas Python utilizadas para a análise e modelagem foram Pandas, NumPy, Matplotlib, Seaborn e Scikit-learn. O ambiente utilizado para a escrita e execução do código Python foi o Google Colaboratory [6].

Inicialmente, o conjunto de dados foi submetido a um pré-processamento, que incluiu a remoção de duplicatas, a codificação de variáveis categóricas e a imputação de valores ausentes. A codificação de variáveis categóricas foi realizada utilizando One-Hot Encoding para variáveis nominais (como region_1, region_2, keyword_1, keyword_2 e type) e Label Encoding para variáveis ordinais (spell_speed, rarity e card_code). Os valores ausentes nas colunas relevantes foram substituídos pelo valor máximo da coluna + 1, garantindo que não houvesse impacto na distribuição dos dados e que todas as linhas pudessem ser utilizadas na

modelagem.

Após o pré-processamento, realizou-se uma análise exploratória dos dados para identificar padrões e relações entre os atributos das cartas. Foram calculadas estatísticas descritivas, como média e desvio padrão, para as variáveis numéricas, e tabelas de frequência para as variáveis categóricas. Além disso, foram gerados histogramas para visualizar a distribuição de atributos como custo e gráficos de barras para a contagem de tipos de carta.

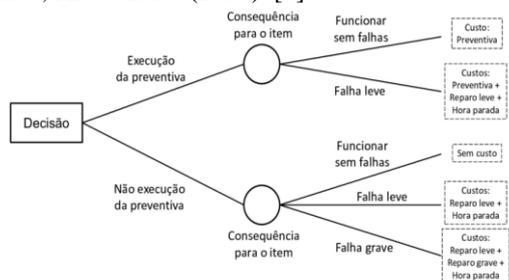
Para a modelagem preditiva, optou-se pela técnica de Random Forest, um algoritmo de aprendizado de máquina supervisionado que se destaca pela sua robustez e capacidade de lidar com dados complexos como demonstrado por [7]. O conjunto de dados foi dividido em dois subconjuntos: 80% para treinamento do modelo e 20% para teste. O modelo foi treinado utilizando o subconjunto de treinamento e, em seguida, avaliado no subconjunto de teste. Para otimizar os hiperparâmetros do modelo, foi utilizada a técnica de Grid Search, que busca a combinação de hiperparâmetros que maximiza o desempenho do modelo.

O desempenho do modelo foi avaliado utilizando duas métricas principais: o erro quadrático médio (MSE) e o coeficiente de determinação (R^2). O MSE mede a média dos erros quadráticos entre os valores reais e os valores previstos pelo modelo, enquanto o R^2 indica a proporção da variância nos dados que é explicada pelo modelo.

Este estudo utilizou um conjunto de dados disponível publicamente, não envolvendo a coleta de dados de participantes humanos. Portanto, não foi necessária a aprovação de um Comitê de Ética em Pesquisa.

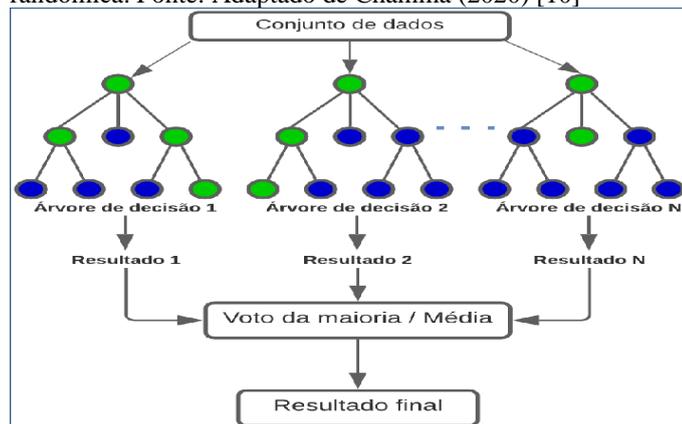
A Random Forest não é uma rede neural, mas sim um conjunto de árvores de decisão. Cada árvore de decisão é um modelo simples que divide o espaço de entrada em regiões e atribui um valor de saída a cada região como demonstrado na figura 1.

Figura 1. Esquemática de uma árvore de decisão. Fonte: Picanço, Ailson Renan & Pelegrina, Guilherme & Torezzan, Cristiano & Silva, Alessandro. (2015). [9]



“Random Forest são um algoritmo de aprendizado que combina várias árvores de decisão aleatórias e agrega suas previsões por meio da média, tornando-se uma ferramenta popular de análise de dados para previsões precisas e robustas.” [8]. A figura 2 mostra o conjunto de dados representado em um esquema de Floresta Randômica.

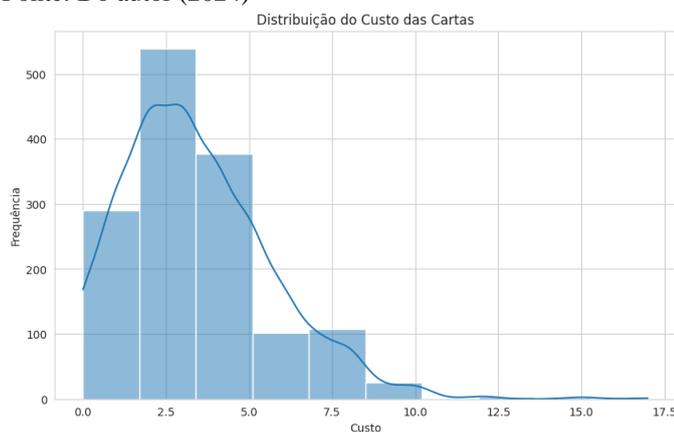
Figura 2. Representação esquemática de uma floresta randômica. Fonte: Adaptado de Chamma (2020) [10]



3 RESULTADOS E DISCUSSÃO

A análise exploratória dos dados revelou insights valiosos sobre as cartas em LoR. A distribuição do custo das cartas, ilustrada na Figura 3, demonstra uma clara tendência: a maioria das cartas tem um custo baixo, entre 0 e 4, com poucas cartas de alto custo.

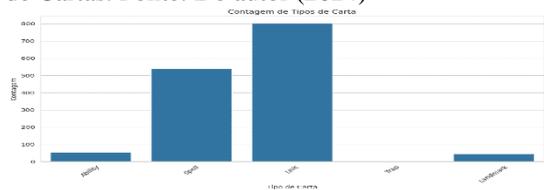
Figura 3. Histograma da Distribuição do Custo das Cartas. Fonte: Do autor (2024)



Essa distribuição assimétrica à direita é comum em jogos de cartas colecionáveis, onde cartas mais poderosas e versáteis geralmente possuem um custo mais alto para equilibrar o jogo.

Em relação aos tipos de carta, o gráfico de barras apresentado na Figura 4 confirma que as cartas do tipo "Unit" (Unidade) são as mais frequentes, seguidas por "Spell" (Feitiço) e "Ability" (Habilidade).

Figura 4. Gráfico em barras da contagem de Tipo de Cartas. Fonte: Do autor (2024)



Essa predominância de unidades é esperada, pois elas são o pilar de qualquer deck, fornecendo poder de ataque e vida para o jogador. As cartas de feitiço e habilidade, por sua vez, oferecem suporte e efeitos adicionais, complementando as unidades.

A modelagem preditiva, utilizando o algoritmo de Random Forest, demonstrou ser eficaz na previsão do custo das cartas. O através da técnica modelo otimizado, obtido de Grid Search, alcançou um coeficiente de determinação (R^2) de 0.516 e um erro quadrático médio (MSE) de 2.99. O R^2 indica que o modelo é capaz de explicar 51.6% da variância no custo das cartas, o que é um resultado promissor considerando a complexidade dos dados e a variedade de atributos que influenciam o custo. Apesar dos resultados promissores, este estudo possui algumas limitações. Em primeiro lugar, a análise exploratória focou apenas em alguns atributos das cartas, como custo e tipo. Uma análise mais aprofundada, considerando outros atributos como raridade, palavras-chave e região, poderia revelar insights adicionais sobre as relações entre as características das cartas.

Em segundo lugar, o modelo preditivo, apesar de apresentar um bom desempenho, ainda possui um erro quadrático médio considerável. Isso indica que há espaço para melhorias no modelo, seja através da utilização de outros algoritmos de aprendizado de máquina, da inclusão de novas variáveis ou da otimização dos hiperparâmetros.

Em trabalhos futuros, pretende-se expandir a análise exploratória para incluir outros atributos das cartas, como raridade, palavras-chave e região, e investigar a correlação entre esses atributos e a taxa de vitória dos decks. Além disso, pretende-se explorar outros algoritmos de aprendizado de máquina, como redes neurais e gradient boosting, para verificar se é possível obter um desempenho ainda melhor na previsão do custo das cartas.

CONCLUSÃO

Este estudo demonstrou o potencial do aprendizado de máquina, especificamente a técnica de Random Forest, para analisar e modelar dados de cartas do jogo Legends of Runeterra. O modelo preditivo desenvolvido, apesar de suas limitações, obteve um desempenho promissor na estimativa do custo das cartas com base em seus atributos. Adicionalmente, a análise exploratória revelou insights relevantes

sobre a distribuição dos tipos de carta no jogo.

Acredita-se que este trabalho contribuiu significativamente para a compreensão das dinâmicas do jogo, oferecendo ferramentas que podem auxiliar tanto jogador na construção de decks mais eficazes quanto desenvolvedores na análise da meta e no balanceamento do jogo. No entanto, reconhece-se que este é apenas um passo inicial em uma área de pesquisa ampla e promissora.

Para trabalhos futuros, vislumbra-se a expansão da análise exploratória para abranger outros atributos das cartas, como raridade, palavras-chave e região, e a investigação da correlação entre esses atributos e a taxa de vitória dos decks. Além disso, pretende-se explorar outros algoritmos de aprendizado de máquina, como redes neurais e gradient boosting, para aprimorar ainda mais a precisão do modelo preditivo.

Em suma, este estudo não apenas responde às perguntas de pesquisa inicialmente propostas, mas também abre um leque de novas questões e possibilidades para futuras investigações. A aplicação de aprendizado de máquina em jogos eletrônicos, como demonstrado neste trabalho, tem o potencial de revolucionar a forma como entendemos e interagimos com esses jogos, e espera-se continuar explorando esse potencial.

REFERÊNCIAS

- [1] Brownlee J. Machine Learning Mastery With Python. 2ª ed. 2020.
- [2] Breiman L. Random Forests. Machine Learning. 45(1), 5-32. 2001.
- [3] Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media. 2019.
- [4] Yannakakis GN, Togelius J. Artificial intelligence and games. Springer International Publishing. 2018.
- [5] Hung B. Riot Games. Legends of Runeterra Dataset. Kaggle.. 2020. [acesso em 25 junho 2024]. Disponível em: <https://www.kaggle.com/datasets/brandonqilin/legends-of-runeterra>.
- [6] McKinney W. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media. 2012
- [7] Probst P, Wright MN, Boulesteix AL. Hyperparameters and tuning strategies for random forest. 2019
- [8] Scornet E, Biau G, Vert J. Consistency of Random Forests. Annals of Statistics, 43, 1716-1741. 2014.
- [9] Picanço AR, Pelegrina G, Torezzan C, Silva A. Um modelo de programação linear inteira para a tomada de

decisão
de manutenção preventiva. 2015.

[10] Chamma W, Batistella D, Crisigiovanni E, Victorino H, Lima V. Aprendizado de máquina aplicado em imagens de satélite para classificação de telhados. 2021.