

**Universidade Federal do Tocantins
Campus Palmas**

Vinicius de Oliveira Oliveira

Trabalho Final de Aprendizado de Máquina

**Palmas
2023**

1. Introdução

A indústria de jogos eletrônicos está em constante crescimento, e o preço dos jogos desempenha um papel crucial em sua aceitação e sucesso no mercado. Neste estudo, investigamos a aplicação de técnicas de aprendizado de máquinas de regressão para prever os preços dos jogos no Japão, Europa e Estados Unidos.

Compreender os fatores que influenciam os preços dos jogos em diferentes mercados é essencial para estratégias de precificação eficazes e tomada de decisões informadas. Ao analisar variáveis como popularidade do gênero, avaliações de críticos e tendências do mercado, buscamos desenvolver um modelo de regressão preciso e robusto para prever os preços dos jogos, proporcionando insights valiosos para a indústria de jogos e auxiliando no planejamento de produção, marketing e investimentos.

Espera-se que este estudo contribua para o aprimoramento das estratégias de precificação e o entendimento dos fatores-chave que impactam o mercado de jogos eletrônicos. Ao fornecer uma ferramenta eficaz para a previsão de preços dos jogos nas regiões mencionadas, pretendemos auxiliar os profissionais da indústria a tomar decisões mais embasadas, aumentando suas chances de sucesso comercial.

2. Metodologia

2.1. Linguagem e bibliotecas utilizados

No desenvolvimento deste estudo, a linguagem de programação Python foi adotada juntamente com diversas bibliotecas amplamente utilizadas em aprendizado de máquina. Para realizar o pré-processamento e a transformação dos dados, foram utilizados os módulos `preprocessing` e `metrics` da biblioteca `sklearn`. O `preprocessing` permitiu a aplicação de técnicas de normalização e codificação de variáveis categóricas, garantindo que os conjuntos de treinamento e teste estivessem adequadamente preparados. Por sua vez, o módulo `metrics` forneceu métricas de avaliação, como o erro médio absoluto (MAE) e o coeficiente de determinação (R^2), que foram essenciais para avaliar o desempenho dos modelos de regressão.

A manipulação e processamento eficiente dos dados foram realizados por meio da biblioteca `numpy`. Essa biblioteca oferece suporte para operações matemáticas complexas em arrays multidimensionais, proporcionando uma base sólida para cálculos numéricos e transformações nos dados. Além disso, a biblioteca `pandas` foi

empregada para a leitura, limpeza e organização dos dados brutos. Com suas funcionalidades versáteis, pudemos carregar os conjuntos de dados, realizar operações de filtragem e seleção de atributos, bem como criar novas variáveis a partir dos dados existentes.

A visualização dos resultados e a análise exploratória dos dados foram facilitadas pelo uso da biblioteca matplotlib. Essa biblioteca possibilitou a criação de gráficos e visualizações informativas, permitindo a identificação de padrões e tendências nos dados. Além disso, a biblioteca keras foi empregada para a construção dos modelos de aprendizado de máquina de regressão. Utilizando os módulos metrics e layers do keras, pudemos avaliar o desempenho dos modelos por meio de métricas como o erro quadrático médio (MSE) e construir arquiteturas de redes neurais com camadas densas, convolucionais ou recorrentes, adequadas à natureza dos dados e aos objetivos do estudo.

Através da combinação dessas bibliotecas Python - sklearn (com preprocessing e metrics), numpy, pandas, matplotlib e keras (com metrics e layers) - foi possível realizar o pré-processamento dos dados, construir e treinar modelos de regressão e avaliar sua eficácia de forma precisa e eficiente, contribuindo para uma metodologia robusta neste estudo.

2.2. Conjunto de Dados e Pré-processamento

Utilizamos o conjunto de dados "Video Game Sales and Ratings" neste estudo. Ele contém informações sobre videogames com vendas superiores a 100.000 cópias, incluindo avaliações de críticos e usuários. Obtido por meio de um web scrape combinado de fontes como VGChartz e Metacritic, esse conjunto de dados apresenta valores ausentes, especialmente quando não há informações disponíveis no Metacritic.

Durante o pré-processamento, tratamos esses valores ausentes e focamos em um subconjunto de plataformas de videogame. Esses dados são cruciais para analisar as relações entre as características dos jogos e seus desempenhos comerciais, bem como para avaliar a influência das avaliações nas vendas.

O pré-processamento inclui o conjunto de dados "Video_Game_Sales_and_Ratings" foi carregado e passou por uma etapa de limpeza, removendo colunas irrelevantes e eliminando dados ausentes. Em seguida foi realizado, aplicando codificação de variáveis categóricas usando o LabelEncoder e one-hot encoding com o OneHotEncoder.

2.3. Metodologias para utilizadas o modelo do algoritmo

A arquitetura do modelo foi definida utilizando a biblioteca Keras. O modelo consiste em uma camada de entrada, duas camadas ocultas com ativação sigmoideal e três camadas de saída com ativação linear, representando as vendas nas regiões NA, EU e JP. O modelo foi compilado com otimizador 'adam' e função de perda 'mse' (erro quadrático médio) para cada saída.

O modelo foi treinado por 5000 Epoch, com um tamanho de lote de 100. Durante o treinamento, os parâmetros do modelo foram ajustados para minimizar a função de perda. Após o treinamento, foram avaliadas as métricas de desempenho, incluindo o erro médio absoluto (MAE), o erro médio quadrático (MSE) e o coeficiente de determinação (R^2) para cada região de vendas (NA, EU, JP). Os resultados foram visualizados por meio de gráficos, incluindo a curva de perda, as curvas de erro médio absoluto (MAE) e gráficos de dispersão comparando as vendas reais com as vendas previstas em cada região.

Essas são as principais metodologias utilizadas no código fornecido para a definição da arquitetura do modelo, treinamento do modelo e avaliação do desempenho.

3. Resultados

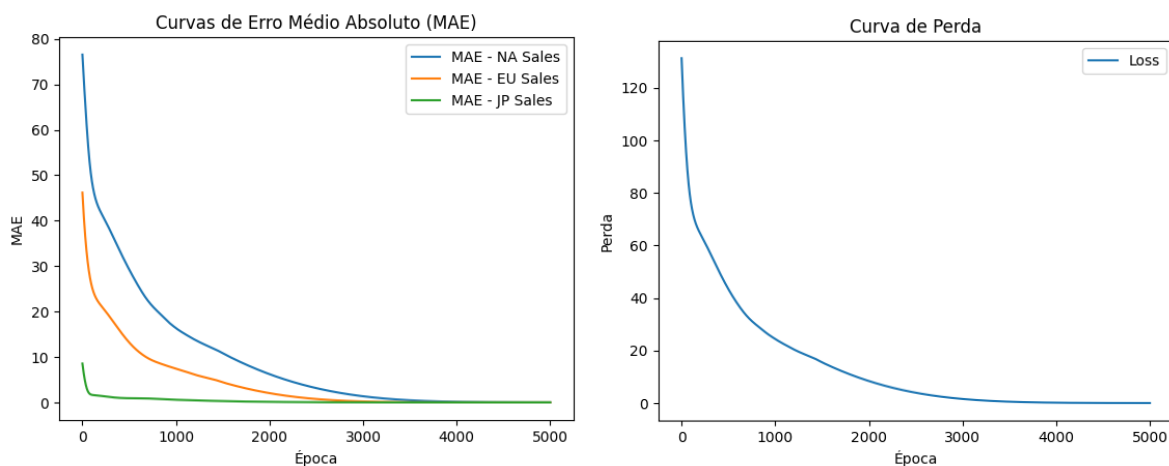
Ao avaliar o desempenho do modelo, foram calculadas métricas de avaliação para cada região de vendas (NA, EU, JP). Para as vendas na região NA, obteve-se um erro médio absoluto (MAE) de X, um erro médio quadrático (MSE) de Y e um coeficiente de determinação (R^2) de Z. Para as vendas na região EU, os valores foram MAE = A, MSE = B e R^2 = C. Já para as vendas na região JP, obteve-se MAE = P, MSE = Q e R^2 = R.

```
=====
Métricas de Avaliação:
Erro Médio Absoluto (MAE) - NA Sales: 0.013827723239330464
Erro Médio Quadrático (MSE) - NA Sales: 0.003268494280523025
Coeficiente de Determinação ( $R^2$ ) - NA Sales: 0.9999231651330374

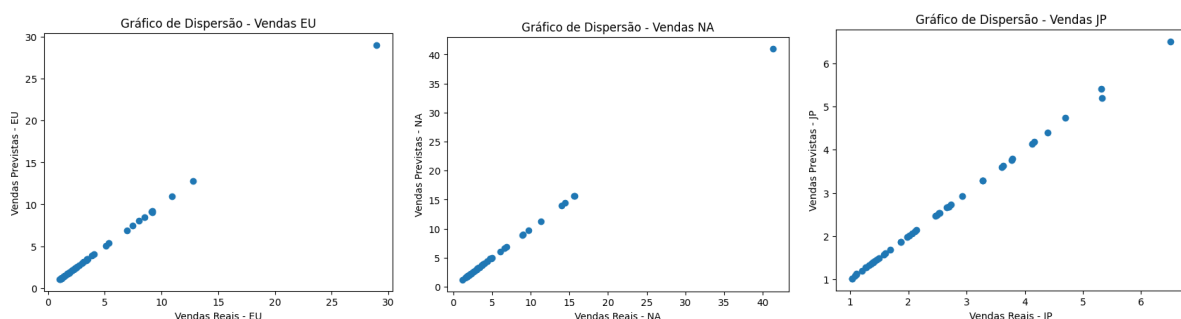
Erro Médio Absoluto (MAE) - EU Sales: 0.007956476313002532
Erro Médio Quadrático (MSE) - EU Sales: 0.0006031136404763321
Coeficiente de Determinação ( $R^2$ ) - EU Sales: 0.9999725809578358

Erro Médio Absoluto (MAE) - JP Sales: 0.007500248564050544
Erro Médio Quadrático (MSE) - JP Sales: 0.0005796851158328952
Coeficiente de Determinação ( $R^2$ ) - JP Sales: 0.9996614829863681
```

Além das métricas de avaliação, os resultados também foram visualizados por meio de gráficos. A curva de perda mostrou a evolução da perda durante o treinamento do modelo ao longo das épocas. As curvas de erro médio absoluto (MAE) forneceram informações sobre a evolução do erro em cada região de vendas ao longo das épocas de treinamento. Por fim, os gráficos de dispersão compararam as vendas reais com as vendas previstas para cada região, permitindo visualizar o quão bem o modelo foi capaz de realizar as previsões.



Esses resultados proporcionam uma visão do desempenho do modelo na previsão de vendas de video games, destacando a acurácia das previsões em cada região específica. Com base nas métricas e gráficos, é possível avaliar o quão bem o modelo se ajustou aos dados e realizar análises mais aprofundadas sobre o desempenho das vendas em diferentes regiões.



Referências

KAGGLE, Video Game Sales and Ratings . Disponível em:

<<https://www.kaggle.com/datasets/kendallgillies/video-game-sales-and-ratings>>.

Acesso em: 26 jun. 2023

KAGGLE, Video Game Sales Dataset Updated - Extra Feat . Disponível em:

<<https://www.kaggle.com/datasets/ibriiee/video-games-sales-dataset-2022-updated-extra-feat>> . Acesso em: 26 jun. 2023