

Avaliação A3

Nomes:

Gustavo Cristhian Almeida Silva - 120112419

David Jacobis Furtado - 12011714

Leonardo Henrique Alves Silva - 12013982

Matheus Henrique Ribeiro - 120117172

Título: Sistema de Recomendação de Filmes

Introdução:

Objetivo: Este documento tem como finalidade apresentar a descrição da avaliação A3, na qual foi sugerido temas para a criação de uma rede neural. Dentre os temas, o grupo em questão optou por fazer um sistema de recomendação de filmes, semelhante à Netflix. As recomendações de filmes baseadas em análise de tendências usando dados obtidos em uma base de filmes, gêneros de filmes, usuários e suas avaliações por filme.

Contexto: Foi solicitado ao grupo a criação de uma rede neural que fizesse recomendações de filmes baseada em perfis de usuários e no histórico de avaliações positivas de cada usuário.

Arquitetura da Rede Neural:

Tipos de camadas: Descrição das camadas da rede (entrada, ocultas, saída).

Hiperparâmetros: Lista dos parâmetros definidos (número de neurônios, taxa de aprendizado, função de ativação, etc.).

Conjunto de Dados:

Origem: Dois datasets. Um contém uma ampla lista de filmes numerados por um ID e seus gêneros. O outro contém os usuários que para eles foram um conjunto de dados fictícios como nomes, sobrenomes, estado de origem, salário e etc, para que seja feita uma melhor análise de um perfil de usuário, separado por seu ID, dados os IDs dos filmes do outro dataset, e a avaliação do usuário para determinado filme.

Pré-processamento: Foram feitas análises de tendências antes do treinamento da rede neural, que serão detalhadas abaixo.

Antes das análises iniciarem foi criada uma seed que salvasse dados de perfil gerados aleatoriamente para que fosse possível associar aos dados dos datasets para fazer as análises de forma que fosse possível agrupar usuários de preferências semelhantes. Todas as etapas foram testadas previamente antes da etapa de treinamento e criação da rede neural.

Treinamento:

Algoritmo de Otimização: Toda a programação foi feita usando a plataforma Google Colab e um Notebook Python para trabalhar os dados e o treinamento da máquina. Bibliotecas usadas: pandas, missingno, matplotlib.pyplot, seaborn, datetime, numpy, faker, sklearn.metrics.pairwise e surprise

Métricas de Avaliação: Foi avaliado o comportamento da rede durante o treinamento a medida que as recomendações ficaram mais assertivas com base nos perfis traçados na análise dos datasets fornecidos. À medida que um perfil é traçado de acordo com idade, renda, estado de origem e a avaliação dos usuários a máquina deveria recomendar filmes que esse grupo de usuário poderia ter interesse levando em consideração tendências analisadas com os dados citados anteriormente.

Resultados:

Desempenho: É possível considerar um desempenho satisfatório da rede neural, onde as recomendações estão condizentes com cada perfil capturado durante seu treinamento.

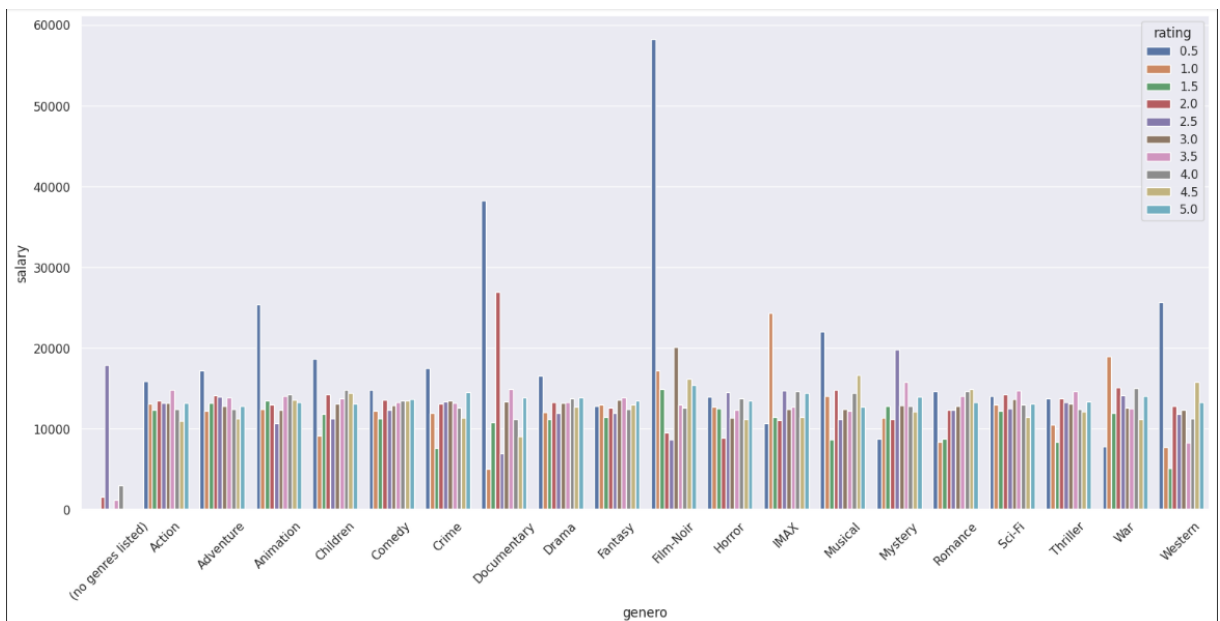
Visualizações:

Exemplo de criação de perfis de usuário:

	movieId	title	genres	userId	rating	timestamp	age	salary	full_name	states
0	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	2	5.0	859046895	31	7431.47	Allison Ward	Rio de Janeiro
1	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	5	4.0	1303501039	14	1200.00	Christopher Holden	Goiás
2	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	8	5.0	858610933	56	2346.89	Keith Young	Minas Gerais
3	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	11	4.0	850815810	12	1200.00	Joshua Smith	Pernambuco
4	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	14	4.0	851766286	58	3556.93	Taylor Lee	Roraima

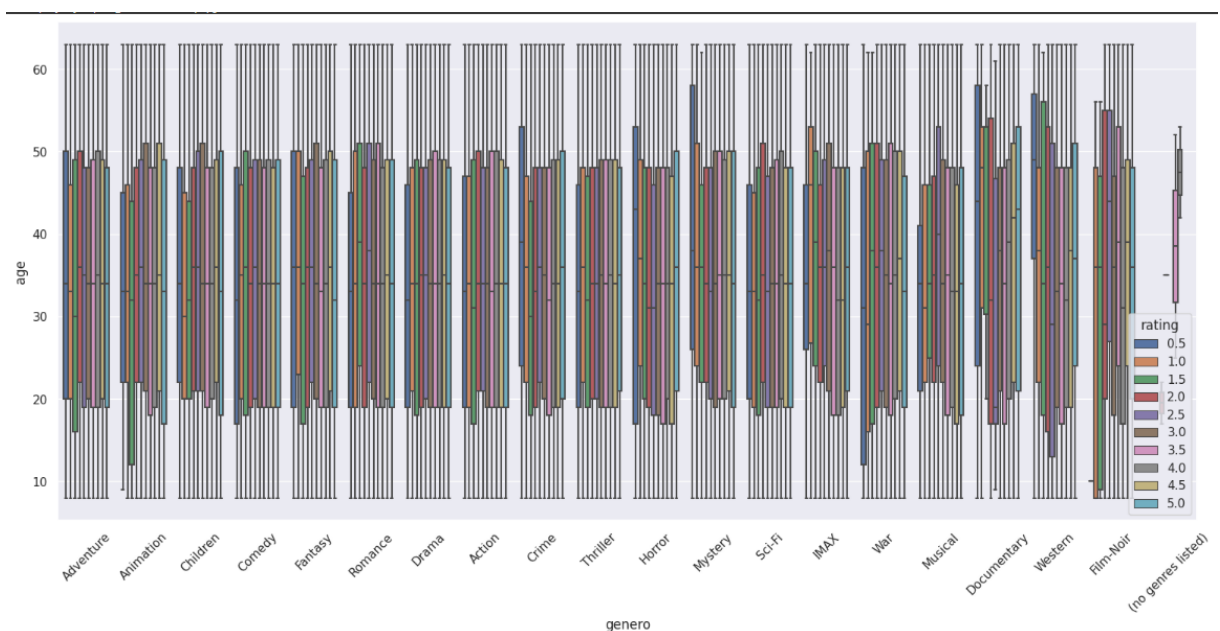
Por meio de uma seed para armazenar e de bibliotecas específicas foram gerados dados aleatórios e construídos perfis que contivessem idade, faixa salarial, estado onde mora e o nome do usuário. Ao separar as avaliações por perfis foi associado um ID de usuário a cada perfil, e sua avaliação para determinado filme. Com essa associação é possível criar tendências para as preferências de cada um.

Gráfico 1:



Nesse gráfico a ideia é ter uma amostra das avaliações dos filmes de acordo com a faixa salarial dos perfis. Nesse exemplo é possível ver que pessoas de maior renda tendem a gostar menos de filmes investigativos ou de documentários. É possível reparar também que notas medianas aparecem mais conforme a faixa salarial aumenta.

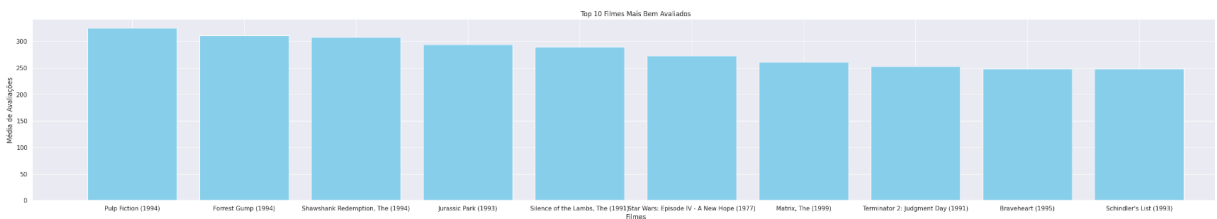
Gráfico 2:



Semelhante ao gráfico anterior, a diferença é que este usa a idade como referência em vez da faixa salarial. Aqui é possível visualizar que pessoas acima dos 50 anos tendem a não assistir ou não avaliar filmes, bem como pessoas abaixo dos 10 anos, com algumas

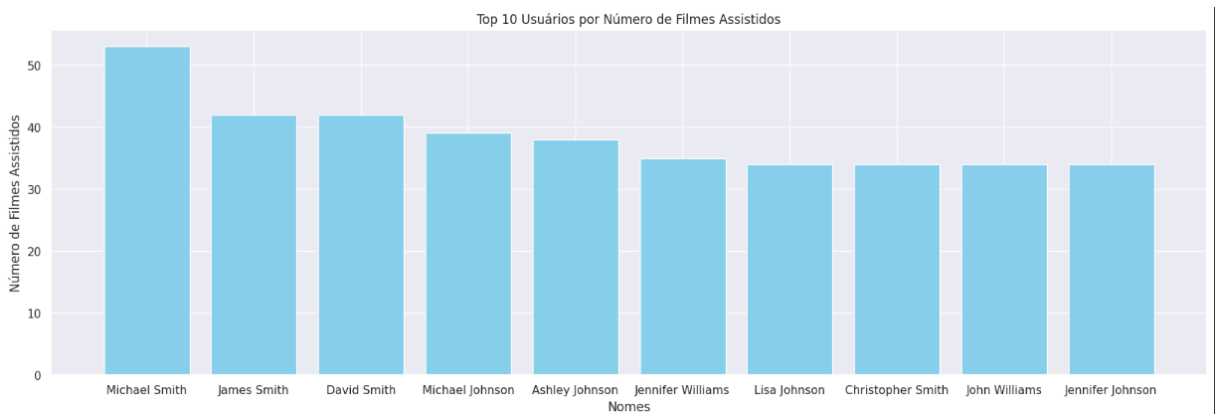
exceções. É possível também notar que a idade não é o melhor parâmetro possível se for avaliada individualmente, dada a pouca variância das avaliações entre os gêneros. Isso sugere que independente da idade, há pessoas que gostam e desgostam de todos os gêneros.

Gráfico 3:



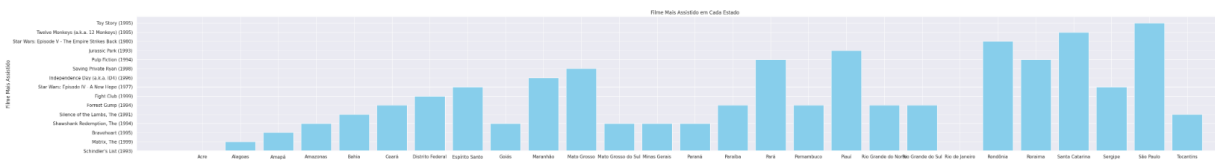
Esse gráfico coleta os 10 filmes mais bem avaliados.

Gráfico 4:



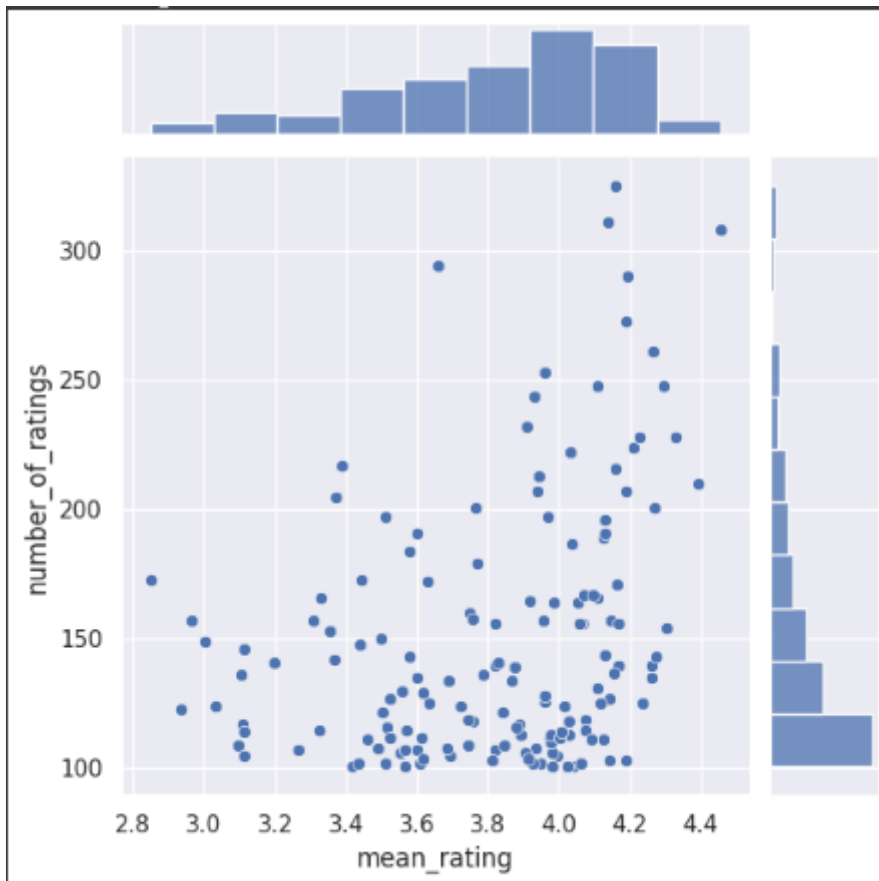
Nesse gráfico a ideia foi coletar quais são os usuários considerados mais engajados, ou seja, os 10 usuários que mais avaliam filmes. Com essa coleta é possível concluir que esses usuários também são os que mais assistem filmes.

Gráfico 5:



Nesse gráfico é possível visualizar quais foram os filmes com maior quantidade de avaliações separadas por estado brasileiro. A ideia é notar se há uma tendência em um determinado estado ter interesse maior por um gênero de filme.

Gráfico 6:



Com esse gráfico é possível ver que há uma concentração maior de avaliações consideradas médias. Dada a dispersão é possível concluir que apesar de muitos usuários terem avaliado os filmes com boas notas, a maior quantidade de avaliações tem notas próximas ao 4,0.

Exemplo de predição de filmes:

	title	rating
0	Austin Powers: The Spy Who Shagged Me (1999)	1.893162
1	Die Hard: With a Vengeance (1995)	1.500000
2	Monty Python's Life of Brian (1979)	1.189320
3	Austin Powers: International Man of Mystery (1...	1.084158
4	Die Hard (1988)	1.081818

Com um algoritmo que escolhia um usuário e um filme previamente citado, foi realizada uma pesquisa no dataset e retornou o resultado acima. O resultado é uma predição de filmes com gêneros semelhantes que também foram avaliados por esse usuário.

Exemplo de taxa de similaridade:

	title	rating	similarity_score
20	Star Wars: Episode IV - A New Hope (1977)	0.811355	1.000000
23	Star Wars: Episode V - The Empire Strikes Back...	0.771930	0.725627
10	Star Wars: Episode VI - Return of the Jedi (1983)	0.968468	0.657868
3	Austin Powers: International Man of Mystery (1...	1.084158	0.541216
58	Pirates of the Caribbean: The Curse of the Bla...	-1.329787	0.428312

Ao escolher o filme Star Wars Episode IV - A New Hope (1977) como base, o algoritmo buscou no dataset filmes de gêneros semelhantes e listou alguns deles com base decrescente no grau de similaridade.

Treinar e avaliar modelo para o usuário:

```
[ ] train_and_evaluate_model_for_user(dbDataset, 7)

RMSE: 1.0987
RMSE on test set for user 7: 1.0987
Predicted rating for movie 5 by user 7: 3.61
```

Essa função tem como objetivo a criação de um sistema de recomendação, onde o modelo é treinado e avaliado para cada usuário com base nos filmes que cada um deles avaliou. Ou seja, essa função ela recebe dois parâmetros, o data frame e o usuário, faz uma filtragem dos dados para um usuário específico, prepara os dados para a modelagem, divide os dados em conjunto para o treinamento e teste, treina o modelo, avalia o modelo e faz a previsão para um filme específico.

Recomendação baseada em histórico:

```
History-based recommendation: [('Incredibles, The (2004)', 3.935052),
```

Por fim temos a função que faz a recomendação baseada no histórico, essa função implementa o sistema de recomendação que utiliza a similaridade entre os itens para fazer previsões sobre a classificação dos filmes não assistidos por um determinado usuário, oferecendo recomendações com base nessas previsões.

Conclusão:

Durante este projeto de IA, exploramos e aplicamos uma variedade de técnicas, algoritmos e métodos para resolver problemas complexos. Mergulhamos em conceitos fundamentais da IA, como aprendizado de máquina, redes neurais, algoritmos de otimização e processamento de linguagem natural. Utilizamos ferramentas e bibliotecas populares para análise e manipulação de dados, modelagem de algoritmos e avaliação de desempenho.

Discussão dos Resultados: Interpretação dos resultados obtidos.

Possíveis Melhorias: Sugestões para melhorar a performance ou a aplicação da rede neural.