

Previsão do Sucesso de Filmes: Uma Abordagem Baseada em Grafos e Características Cinematográficas

1st Pedro Augusto Gontijo Moura

Engenharia da Computação

CEFET-MG Campus V

Divinópolis, Brasil

pedroaugustomoura70927@gmail.com

Resumo—A área de análise de dados e ciência de redes tem ganhado destaque na compreensão de padrões em diversas indústrias, incluindo a indústria cinematográfica. Um dos grandes desafios desse setor é identificar quais fatores mais influenciam o sucesso de um filme, tradicionalmente medido pela receita de bilheteria. Este estudo propõe uma abordagem baseada em grafos para analisar as relações entre características-chave dos filmes—orçamento, gênero, diretor e tempo de duração—e seu impacto no desempenho comercial. Por meio da construção e análise de redes de grafos, foram identificados agrupamentos de filmes com atributos semelhantes, além da aplicação de técnicas de aprendizado de máquina para avaliar a importância relativa desses fatores. Os resultados indicam que orçamento e reputação do diretor são os principais determinantes do sucesso nas bilheteiras, com o tempo de duração e o gênero ganhando relevância em contextos específicos. A metodologia proposta fornece insights valiosos para produtores, distribuidores e pesquisadores, auxiliando na tomada de decisões estratégicas na indústria cinematográfica.

Index Terms—Grafos, análise de dados, filmes, aprendizado de máquina, sucesso de bilheteria, orçamento, diretor, gênero.

I. INTRODUÇÃO

O mercado cinematográfico contemporâneo é caracterizado por uma competitividade crescente, onde as decisões de produção e distribuição de filmes dependem, em grande parte, da análise de características que influenciam diretamente seu desempenho. Com o avanço da tecnologia, a coleta de dados sobre o cinema tornou-se abundante, permitindo às empresas, diretores e analistas identificar padrões e comportamentos que definem o sucesso de um filme.

Nesse contexto, esse projeto busca aplicar a teoria dos grafos como uma abordagem analítica no campo do cinema, para compreender as associações entre elementos como orçamento, gênero, diretor e tempo de duração, e o impacto desses fatores no desempenho comercial e crítico de um filme. Grafos, por sua capacidade de representar dados complexos e suas conexões, oferecem uma perspectiva visual e computacional poderosa para identificar associações e padrões que poderiam passar despercebidos em análises convencionais. Por exemplo, ao conectar filmes com base em características compartilhadas, como gêneros ou faixas de orçamento, é possível revelar

agrupamentos naturais e tendências dentro da indústria cinematográfica.

Além disso, o estudo se apoia na ideia de que o sucesso de um filme não é determinado isoladamente por um único fator, mas sim por um conjunto de variáveis interligadas que contribuem para o desempenho final. Assim, esta pesquisa se propõe a explorar como diferentes combinações de características, representadas em um grafo multirrelacional, podem estar associadas ao sucesso de um filme.

Com dados reais e técnicas de análise computacional, este trabalho busca responder perguntas importantes, como: Filmes com orçamentos altos são sempre mais bem-sucedidos? Certos gêneros estão associados a maiores bilheteiras ou melhores avaliações da crítica? O tempo de duração influencia a aceitação do público? Através dessas questões, objetiva-se não apenas entender os fatores que definem o sucesso de um filme, mas também fornecer uma ferramenta que permita visualizar e compreender a complexidade do mercado cinematográfico contemporâneo.

Essa questão é crucial para a indústria do cinema, considerando que o lançamento de um filme demanda grandes investimentos e está associado a elevados riscos financeiros.

II. REFERENCIAL TEÓRICO

A. *Sucesso de Filmes: subjetividade e fatores determinantes*

A receita de bilheteria é frequentemente considerada o principal indicador de sucesso comercial de um filme. Ela reflete a capacidade de atrair espectadores e gerar lucro, sendo influenciada por fatores como marketing, distribuição e apelo do elenco. Filmes que alcançam altas arrecadações são vistos como bem-sucedidos do ponto de vista financeiro. Por exemplo, produções como "Venom" e "A Saga Crepúsculo: Amanhecer – Parte 1" obtiveram receitas significativas, apesar de críticas negativas, evidenciando que a aceitação do público pode superar avaliações críticas desfavoráveis [6]. Considerando isso, esse trabalho teve a receita dos filmes como principal fator determinante para o conceito de sucesso.

Dito isso, o sucesso de um filme é influenciado por uma combinação de fatores que abrangem desde aspectos

de produção até a recepção pelo público. Estudos indicam que elementos como orçamento, gênero, escolha do diretor e duração desempenham papéis significativos nesse contexto. Por exemplo, uma pesquisa exploratória sobre o mercado cinematográfico brasileiro identificou variáveis que influenciam o sucesso financeiro de filmes nacionais, destacando a importância de compreender esses elementos para otimizar o potencial da indústria cinematográfica [7].

- **Orçamento:** Investimentos mais elevados podem resultar em maior qualidade de produção, efeitos especiais aprimorados e campanhas de marketing mais robustas, aumentando a atratividade do filme.
- **Gênero:** Certos gêneros possuem apelo mais amplo ou nichos específicos que podem influenciar o desempenho nas bilheterias.
- **Diretor:** A reputação e o estilo do diretor podem atrair públicos específicos e impactar a percepção comercial do filme.
- **Tempo de Duração:** A duração do filme pode afetar a rotatividade de sessões nos cinemas e a preferência do público, influenciando a receita total.

B. Análise de Grafos: Conceitos e Aplicações

Grafos são estruturas matemáticas compostas por nós (vértices) e arestas (conexões) que representam relações entre entidades. Na análise de redes complexas, os grafos são utilizados para modelar e investigar interações em diversos contextos, incluindo redes sociais, biológicas e de informação. A análise de grandes grafos permite estudar estruturas como componentes, caminhos mais curtos, efeito "small-world" e distribuição dos graus dos vértices, fornecendo insights sobre a conectividade e a importância dos elementos dentro da rede.

Na indústria cinematográfica, a aplicação de grafos possibilita a representação de filmes como nós e suas características compartilhadas como arestas, facilitando a análise de similaridades e padrões. Como feito nesse trabalho, filmes podem ser conectados com base em atributos como gênero, diretor, tempo de duração ou orçamento, permitindo a identificação de comunidades de filmes com características comuns.

C. Técnicas e Algoritmos Utilizados na Modelagem

A detecção de comunidades em grafos visa identificar subgrupos de nós mais densamente conectados entre si do que com o restante da rede. O método de Louvain é uma técnica amplamente utilizada para essa finalidade, permitindo a identificação de agrupamentos naturais dentro da rede.

Além disso, métricas de centralidade, como grau, proximidade e intermediação, são empregadas para avaliar a importância ou influência de um nó dentro do grafo, auxiliando na identificação de filmes que desempenham papéis centrais nas redes de similaridade.

D. Modelos de Aprendizado de Máquina na Análise Cinematográfica

Modelos de aprendizado de máquina, como o Random Forest Regressor, são utilizados para analisar a importância de

diferentes características na predição do sucesso de filmes. Esses modelos permitem avaliar o impacto relativo de variáveis como orçamento, gênero e diretor no desempenho financeiro, fornecendo uma compreensão quantitativa dos fatores que mais influenciam o sucesso cinematográfico.

III. METODOLOGIA

Para implementar a análise associando o sucesso dos filmes a características como orçamento, gênero, diretor e tempo de duração, foi desenvolvido um sistema modular em Python. Esse sistema integra técnicas de pré-processamento de dados, construção e análise de grafos, detecção de comunidades e avaliação da importância dos fatores, oferecendo uma abordagem robusta e escalável para investigar os padrões presentes no dataset. A seguir, detalhamos os principais módulos e os modelos de aplicação empregados.

A. Estrutura e Organização do Sistema

A modularização é um conceito que contribui para uma melhor estruturação, manutenção e expansão do sistema. Ela é chave no desenvolvimento de software, pois possibilita a divisão de um problema complexo em partes menores e mais fáceis de gerenciar. Entre os módulos, têm-se: *read.py*, *graph.py* e *main.py*.

1) *Read.py* - Módulo de Leitura e Pré-processamento

Este módulo é responsável por ler o dataset em formato CSV, converter os dados para os tipos apropriados e aplicar filtros com base em critérios como nota e receita. Adicionalmente, os dados numéricos (por exemplo, receita, orçamento e tempo de duração) são normalizados utilizando técnicas de escalonamento (MinMaxScaler), e as variáveis categóricas (como o gênero primário) são codificadas para possibilitar sua incorporação nas análises.

Algorithm 1 Leitura e Processamento de Dados do Arquivo *read.py*

```

1: Início
   Carregar_Dadosnome_arquivo
2:   Abrir arquivo CSV nome_arquivo para leitura
3:   Ler os dados no formato DataFrame
4:   DataFrame com os dados
   Préprocessar_Dadosdados, min_valor, max_valor
5:   Filtrar dados com receita e orçamento dentro do
   intervalo definido
6:   Remover valores ausentes ou inconsistentes
7:   Normalizar colunas numéricas usando escalonamento
   (MinMaxScaler)
8:   dados normalizados e o objeto de escalonamento
9: Fim
   =0

```

2) *Graph.py* - Módulo de Construção do Grafo

Utilizando a biblioteca NetworkX, cada filme é representado como um nó do grafo, sendo etiquetado com atributos relevantes (como o título e a nota). As conexões (arestas) entre os filmes são estabelecidas com base em uma matriz de

similaridade calculada por meio do cosseno de similaridade. Essa matriz é construída a partir de um conjunto de features que engloba atributos numéricos (Rating, Runtime, Revenue e Budget) e as variáveis dummy referentes aos gêneros. Apenas pares de filmes cuja similaridade ultrapasse um limiar pré-definido e que se encontrem entre os mais próximos (limitando o número de conexões por nó) são conectados, o que garante um grafo interpretável e evita a formação de redes excessivamente densas.

2.1) Módulo de Análise do Grafo e Detecção de Comunidades:

Uma vez construído o grafo, são aplicadas técnicas de análise de centralidade (grau, proximidade e intermediação) para identificar os filmes mais influentes na rede. Além disso, o método de Louvain é utilizado para identificar comunidades de filmes que compartilham características semelhantes. Essa etapa permite a segmentação da rede em clusters, facilitando a identificação de grupos temáticos e regionais e fornecendo insights sobre a distribuição do sucesso no mercado cinematográfico.

2.2) Módulo de Avaliação de Importância dos Fatores:

Para compreender o impacto de cada característica no sucesso dos filmes, foi implementada uma análise de importância utilizando um modelo de Random Forest Regressor. Esse módulo treina o modelo com variáveis como runtime, orçamento, diretor (codificado) e os gêneros (variáveis dummy) e extrai a importância relativa de cada atributo na predição da receita. Os resultados dessa análise fornecem uma visão quantitativa dos fatores que mais influenciam o desempenho financeiro dos filmes.

2.3) Módulo de Visualização e Estatísticas:

A visualização do grafo é realizada por meio do layout de primavera (spring layout) do NetworkX, que organiza os nós de forma a evidenciar agrupamentos e conexões relevantes. Para aprimorar a legibilidade dos rótulos, a biblioteca AdjustText é utilizada, e os nós são dimensionados de acordo com métricas como o grau (indicador da centralidade). Adicionalmente, gráficos descritivos e tabelas (utilizando Matplotlib e Seaborn) são gerados para exibir estatísticas relevantes, como número de nós e arestas, principais medidas de centralidade e distribuições das variáveis de interesse.

3) Main.py - Integração

O módulo principal, responsável por unificar todas as etapas do sistema. Ele coordena a execução, chamando as funções dos diferentes módulos quando necessário e assegurando que o processo aconteça de forma lógica e eficiente. Ao centralizar o controle da execução no main.py, o fluxo do programa se torna mais claro, facilitando a compreensão e a gestão da lógica do sistema.

4) filmes.csv - Base de Dados

A base de dados utilizada nesse estudo, é proveniente de um artigo que realizou um estudo parecido, onde foi feito

a combinação de dois outros conjuntos: "The Hollywood Insider", da Kaggle, e um mais genérico da API da IMDB [4][8]. Esse dataset, contendo 507 filmes do período de 2000 a 2017, foi escolhido devido a qualidade e quantidade de informações presentes no mesmo, já que diversas tentativas de encontrar uma base maior foram frustradas no processo de montagem desse estudo. A partir deste ponto, a ênfase mudou para melhorar a qualidade dos recursos, em vez de simplesmente aumentar a quantidade de dados.

5) Make File

O Makefile é um arquivo utilizado para automatizar a execução de tarefas relacionadas ao projeto, como a execução dos scripts principais, a limpeza de arquivos temporários ou a instalação de dependências. Ele define uma série de comandos organizados em "alvos"(targets) que podem ser executados com o comando make. No caso deste estudo, o Makefile foi configurado para simplificar a execução do código principal (main.py), que integra as funcionalidades dos módulos auxiliares (graph.py e read.py), além de realizar outras tarefas necessárias para o processamento e análise dos dados.

B. Fluxo de Execução e Aplicação dos Modelos

A aplicação dos modelos ocorre por meio do seguinte fluxo:

1) Leitura e Pré-processamento: O dataset é carregado e submetido a processos de conversão, filtragem e normalização. Essa etapa assegura que os dados estejam em formato adequado para as análises subsequentes.

2) Construção do Grafo: A partir do conjunto de dados pré-processado, é calculada a matriz de similaridade e construído o grafo, onde cada nó representa um filme e as arestas representam a similaridade entre eles. O uso de um limite de similaridade e a restrição ao número máximo de conexões por nó garantem que somente as relações mais significativas sejam consideradas.

3) Análise de Centralidade e Comunidades: Com o grafo pronto, são aplicadas análises de centralidade para identificar os filmes mais influentes e o método de Louvain para segmentar o grafo em comunidades. Essas análises ajudam a identificar padrões e agrupamentos que podem estar associados ao sucesso dos filmes.

4) Avaliação de Importância dos Fatores: O modelo Random Forest é treinado para avaliar a importância dos atributos (como orçamento, tempo de duração, diretor e gênero) na determinação da receita dos filmes, fornecendo insights sobre quais características são mais decisivas para o sucesso.

5) Visualização e Interpretação dos Resultados: Os resultados são visualizados por meio de gráficos de rede e tabelas descritivas, permitindo uma interpretação intuitiva dos padrões encontrados e facilitando a comunicação dos insights obtidos com a análise.

C. Aplicações Práticas

Os modelos de aplicação desenvolvidos neste trabalho oferecem benefícios práticos para diferentes públicos:

- **Para produtoras e distribuidoras:** A análise de grafos pode auxiliar na identificação de tendências e preferências do público, permitindo a criação de estratégias de marketing mais direcionadas e a tomada de decisão mais embasada na escolha de investimentos para filmes.
- **Para pesquisadores:** A metodologia apresentada fornece um framework robusto para investigar as relações entre múltiplas variáveis que influenciam o sucesso de filmes, contribuindo para a literatura acadêmica sobre análise de dados no setor cinematográfico.
- **Para desenvolvedores de sistemas de recomendação:** Os insights obtidos a partir da análise de grafos e da importância dos fatores podem ser integrados em sistemas de recomendação, aprimorando a personalização de sugestões para os espectadores com base em características semelhantes e padrões de sucesso.

Esta abordagem modular e integrada permite uma análise abrangente do mercado cinematográfico, destacando as inter-relações entre diferentes características dos filmes e contribuindo para a compreensão dos fatores que impulsionam o sucesso na indústria. Se necessário, ajustes adicionais podem ser realizados para refinar o modelo e adaptar a metodologia a novas fontes de dados ou a outros contextos de análise.

IV. RESULTADOS E DISCUSSÃO

A análise dos resultados foi conduzida com o objetivo de identificar padrões e relações entre características dos filmes e o sucesso, avaliado principalmente pela receita de bilheteria. Por meio da construção de grafos, detecção de comunidades e avaliação quantitativa de fatores de impacto, foi possível observar dinâmicas importantes que destacam as variáveis mais influentes e agrupamentos naturais dentro do dataset.

Nesta seção, é apresentado os principais achados da pesquisa, discutindo tanto as propriedades estruturais do grafo construído quanto os insights obtidos a partir da análise de centralidade, detecção de comunidades e avaliação de importância dos fatores. Os resultados são discutidos com base na literatura e nas hipóteses formuladas, oferecendo uma interpretação abrangente dos padrões observados.

Para destacar os filmes mais representativos, a análise foi realizada em etapas, aumentando iterativamente 100 milhões na receita mínima para consideração dos filmes presentes no dataset [8]. Esse processo permitiu refinar os dados e evidenciar os atributos de maior impacto na receita. Primeiramente, foi explorada a estrutura geral do grafo, destacando estatísticas como o número de nós, arestas e a densidade da rede. Em seguida, discutimos as métricas de centralidade e os filmes que desempenham papéis centrais na rede, avaliando suas características e influências. A segmentação em comunidades por meio do método de Louvain é analisada, com ênfase nos grupos temáticos e nos fatores compartilhados pelos filmes dentro de cada cluster.

Por fim, são discutidos os resultados da análise de importância das variáveis realizadas com o modelo de Random Forest Regressor, evidenciando o impacto relativo de características como orçamento, duração, gêneros e diretor na predição do sucesso financeiro dos filmes. Comparações entre métricas de bilheteria e avaliação crítica também são realizadas, destacando como diferentes medidas de sucesso podem gerar conclusões distintas.

A seguir, estão dispostos os resultados detalhados, juntamente com as respectivas discussões e implicações no contexto da indústria cinematográfica.

A. 1ª Iteração

A análise da primeira iteração abrangeu o grafo completo construído a partir de todos os filmes presentes no dataset. Esta etapa inicial foi essencial para obter uma visão geral das conexões entre os filmes, permitindo identificar comunidades, métricas de centralidade e fatores determinantes para o sucesso na receita dos filmes.

Em cada subseção, têm-se uma análise aprofundada referente a cada iteração específica. Serão apresentados os respectivos grafos e métricas correspondentes a eles. Com isso, os atributos mais relevantes serão destacados, a fim de evidenciar tendências no sucesso dos filmes

Grafo de Filmes - 100% da Base - 1ª Iteração

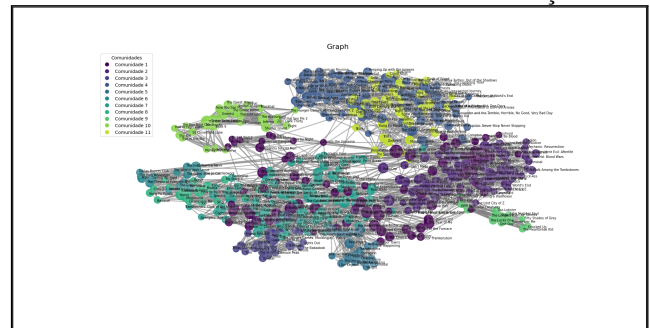


Figura 1. 1ª Iteração com a base total.

1) Análise das Comunidades no Grafo:

A detecção das 11 comunidades revelou agrupamentos de filmes baseados em similaridades de gênero, orçamento e popularidade. Abaixo está um resumo das principais comunidades com exemplos representativos:

- **Comunidade 1 (Ação e Blockbusters):** Filmes de alto orçamento e franquias renomadas, como The Avengers e The Dark Knight. Esses filmes têm alta conectividade com outras comunidades devido a grande orçamento e diretores famosos.
- **Comunidade 2 (Comédias Românticas e Dramas Leves):** Foco em histórias leves, como Crazy, Stupid, Love e The Proposal. Popularidade depende mais de atores e enredo do que do orçamento.
- **Comunidade 3 (Terror e Suspense):** Filmes de baixo orçamento com alto retorno financeiro, como The Conjuring e Paranormal Activity. Altamente concentrados e raramente conectados a outros gêneros.

- **Comunidade 4 (Animações e Filmes Familiares):** Grandes produções voltadas ao público global, como Frozen e Toy Story 3. Alta dependência de marketing e dubladores famosos.

Esses grupos ajudam a compreender padrões de mercado e a relação entre gêneros. Filmes como Hacksaw Ridge e Enchanted funcionam como "pontes" entre comunidades distintas.

Métricas do Grafo - 100% da Base - 1º Iteração

Informações Importantes do Grafo	
Métrica	Valor
Número de Nós	504
Número de Arestas	3414
Grau Médio dos Nós	13.55
Top 3 Filmes por Grau de Centralidade	Robin Hood, Oujia, Origin of Evil, Hacksaw Ridge
Top 3 Filmes por Centralidade de Proximidade	Hairspray, The Bourne Ultimatum, Enchanted
Top 3 Filmes por Centralidade de Intermediação	The Bourne Ultimatum, Hairspray, Enchanted

Figura 2. 1º Iteração com a base total.

O grafo gerado (Figura 1) apresenta um total de 504 nós e 3414 arestas, indicando um alto nível de conectividade entre os filmes. Cada nó representa um filme, enquanto as arestas indicam similaridades ou relações baseadas nos atributos analisados. As cores indicam as 11 comunidades detectadas utilizando o método de Louvain, que foram definidas a partir da modularidade máxima obtida.

2) Análise das Métricas de Centralidade:

As métricas de centralidade foram utilizadas para identificar os filmes mais influentes ou estratégicos dentro do grafo, com base em sua conectividade e posição em relação a outros nós.

- **Centralidade de Grau:** A centralidade de grau mede o número de conexões diretas de um nó. Filmes com alta centralidade de grau, de acordo com a Figura 2, possuem muitas conexões diretas com outros filmes, indicando sua relevância em termos de similaridade com múltiplos títulos. Isso pode ocorrer devido a características amplamente compartilhadas, como gêneros populares ou diretores influentes.
- **Centralidade de Proximidade:** Essa métrica avalia o quão perto um nó está de todos os outros no grafo. Filmes como *Hairspray* e *The Bourne Ultimatum* apresentam alta centralidade de proximidade, o que significa que eles estão em posições estratégicas que permitem acesso rápido a outras partes do grafo. Esses filmes podem ser vistos como hubs conectando subgrupos dentro do grafo.
- **Centralidade de Intermediação:** A centralidade de intermediação indica o número de vezes que um nó atua como ponte em caminhos curtos entre outros nós. Títulos como *The Bourne Ultimatum* e *Enchanted* têm alta

intermediação, evidenciando sua importância em conectar comunidades distintas. Esses filmes são fundamentais para a coesão do grafo, pois promovem a ligação entre diferentes gêneros e públicos.

Essas métricas destacam filmes que não apenas possuem grande apelo ou similaridade com outros títulos, mas também desempenham papéis estruturais importantes no grafo.

Gráfico de Importância - 100% da Base - 1º Iteração

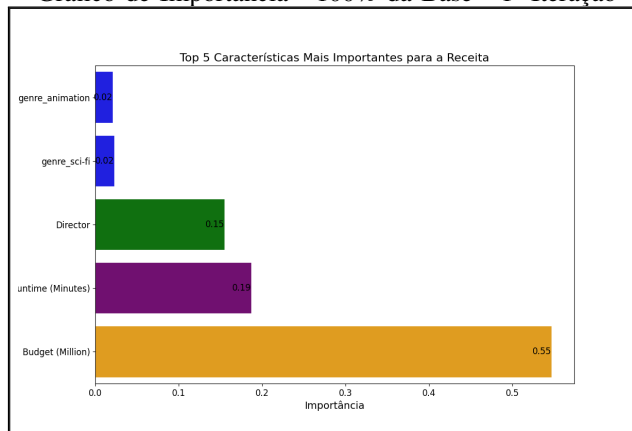


Figura 3. 1º Iteração com a base total.

A análise dos fatores determinantes para a receita (Figura 3), apresentou os seguintes resultados:

- **Orçamento (55%):** O orçamento é a variável de maior impacto, apresentando um peso significativo na previsão da receita. Isso sugere que filmes com altos investimentos tendem a ter maior arrecadação, corroborando estudos que indicam a relação entre produções de grande porte e desempenho comercial expressivo.
- **Tempo de Duração (19%):** O tempo de duração aparece como o segundo fator mais relevante. Esse resultado indica que filmes mais longos podem ter uma relação com maior bilheteria, possivelmente por estarem associados a gêneros que atraem grandes audiências ou por refletirem uma estrutura narrativa mais elaborada.
- **Diretor (15%)** A influência do diretor, demonstra que a reputação e a experiência do cineasta são fatores que impactam o sucesso comercial. Diretores renomados frequentemente possuem um público fiel e são capazes de atrair maior interesse para suas produções.
- **Gêneros (2%)** Por fim, os gêneros cinematográficos tiveram importância relativa menor em comparação aos outros fatores analisados, embora alguns gêneros específicos, como animação e ficção científica, tenham demonstrado alguma correlação com receitas mais elevadas.

Com um domínio muito grande, o orçamento se mostra como fator de maior peso dentre todos os fatores. Tempo de duração e diretor aparecem depois com certa relevância, mas mesmo assim bem atrás do orçamento. Gêneros como ficção científica e animação apareceram, mas com uma importância muito pequena.

B. 2ª Iteração

Nesta segunda iteração, foi filtrada a base de dados na modelagem do grafo para investigar a evolução das conexões entre os filmes e os padrões emergentes ao longo do tempo. A filtragem diminuiu aproximadamente 65% da quantidade de filmes total, passando a ter 174 filmes analisados. O objetivo é avaliar se a estrutura das relações entre as produções cinematográficas se torna mais previsível e se novos fatores emergem como determinantes do sucesso.

A seguir, são detalhados os atributos e informações mais importantes provindas da análise do grafo:

Gráfico de Importância - 100% da Base - 1ª Iteração

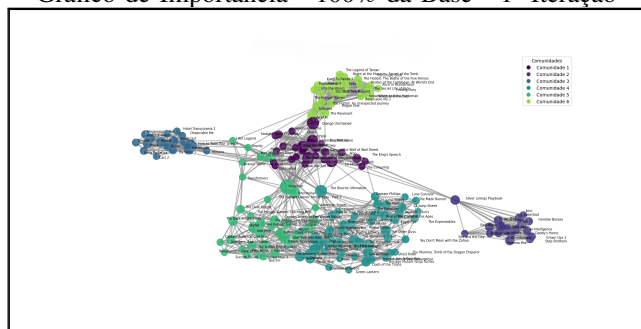


Figura 4. 2ª Iteração com a base total.

A análise estrutural do grafo na segunda iteração revelou um leve aumento das conexões entre os filmes. Em comparação com a primeira iteração, os nós centrais permanecem semelhantes, sendo dominados por filmes de alto orçamento e grandes franquias. No entanto, observamos um crescimento nas conexões entre filmes de gêneros similares, indicando uma tendência de agrupamento mais forte.

O grafo é dividido em seis comunidades, destacadas por cores diferentes. Essas comunidades agrupam filmes que compartilham características semelhantes. São elas:

- **Comunidade 1 (roxo escuro):** Contêm dramas e filmes de apelo crítico forte, como *Silver Linings Playbook* e *The Wolf of Wall Street*.
- **Comunidade 2 (azul escuro):** Possivelmente, filmes de comédia e produções com menor orçamento, como *Juno* e *Superbad*.
- **Comunidade 3 (ciano):** Filmes de animação e familiares, como *Toy Story 3*, *Inside Out* e *How to Train Your Dragon*.
- **Comunidade 4 (verde-azulado):** Parece conter blockbusters e filmes de ação, incluindo *The Dark Knight*, *Jurassic World* e *Transformers*.
- **Comunidade 5 (verde claro):** Provavelmente filmes de aventura e fantasia, como *The Hunger Games*, *The Hobbit* e *Pirates of the Caribbean*.
- **Comunidade 6 (verde mais forte):** Filmes de ficção científica e fantasia, incluindo *Interstellar*, *Inception* e *Gravity*.

Métricas do Grafo - 35% da Base - 2ª Iteração

Informações Importantes do Grafo	
Métrica	Valor
Número de Nós	174
Número de Arestas	1187
Grau Médio dos Nós	13.64
Top 3 Filmes por Grau de Centralidade	Inception, Doctor Strange, Django Unchained
Top 3 Filmes por Centralidade de Proximidade	Inception, The Bourne Ultimatum, Cinderella
Top 3 Filmes por Centralidade de Intermediação	Silver Linings Playbook, Inception, The Bourne Ultimatum

Figura 5. 2ª Iteração com 35% da base total.

O grafo gerado (Figura 4) apresenta um total de 174 nós e 1187 arestas. As métricas de centralidade confirmam essa evolução. O grau médio aumentou levemente, sugerindo que mais filmes estão se conectando diretamente a produções de grande sucesso. A centralidade de intermediação apresentou mudanças, o que mostra que novos filmes, como *Inception*, *Cinderella* e *Silver Linings Playbook*, desempenham um papel intermediário mais relevante na estrutura desse novo grafo.

Gráfico de Importância - 35% da Base - 2ª Iteração

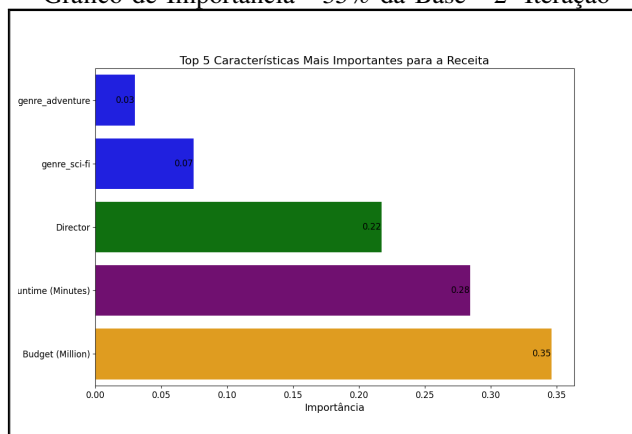


Figura 6. 2ª Iteração com 35% da base total.

A análise do gráfico de importância (Figura 7) mostra alterações relevantes em relação à primeira iteração (Figura 3).

- **Orçamento (35%):** Embora continue sendo o fator de maior peso, o orçamento sofreu uma queda significativa em sua influência sobre a receita. Esse declínio indica que outros fatores começam a desempenhar papéis mais relevantes no desempenho comercial de um filme.
- **Tempo de Duração (28%):** O aumento na importância do tempo de duração sugere que filmes mais longos, muitas vezes associados a narrativas mais elaboradas e experiências cinematográficas mais imersivas, estão sendo valorizados.

- **Diretores (22%):** A relevância dos diretores cresceu, evidenciando que a reputação e a visão criativa de um cineasta podem atrair públicos específicos e agregar valor a uma produção.
- **Gênero de Ficção Científica (7%):** Ficção científica começou a ocupar um espaço mais significativo, consolidando-se como uma categoria de apelo crescente. Esse gênero frequentemente combina elementos de alta produção, inovação tecnológica e narrativas cativantes, características que ressoam com audiências globais.
- **Gênero de Aventura (3%):** Embora tenha o menor impacto entre os fatores listados, a aventura aparece como um gênero sólido, tomando o lugar do gênero de animação no gráfico.

Na segunda iteração, percebe-se uma rede cinematográfica mais conectada e agrupada, com filmes desempenhando papéis estruturais importantes. O orçamento continua sendo o principal determinante de sucesso, mas fatores como o tempo de duração, a reputação do diretor e o gênero ganharam maior relevância. Essas tendências indicam uma transição gradual para uma análise mais diversificada dos elementos que influenciam o sucesso comercial e a influência dos filmes na indústria.

C. 3ª Iteração

Nesta terceira iteração, os dados analisados buscam refinar ainda mais a compreensão das relações entre os filmes, destacando comunidades específicas e características que mais influenciam a receita. Utilizando um grafo filtrado por atributos de receita e avaliação, além das principais métricas de centralidade, obtemos um panorama detalhado da influência de gêneros, diretores e orçamentos no sucesso das produções cinematográficas.

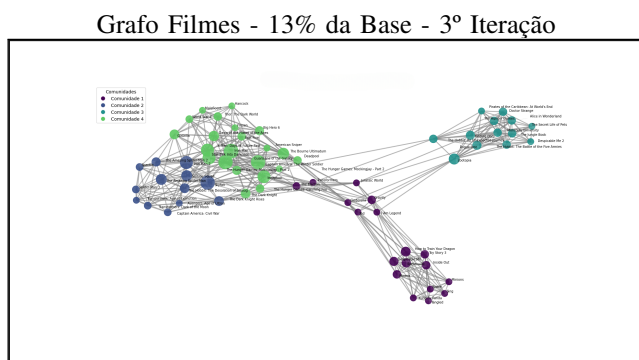


Figura 7. 3ª Iteração com 13% da base

A representação visual do grafo desta iteração é composta por 66 nós e 430 arestas. Os filmes foram divididos em quatro comunidades distintas, representadas por cores:

- **Comunidade 1 (Roxa):** Agrupa filmes com ênfase em animações e conteúdos familiares, como Zootopia e Inside Out.
- **Comunidade 2 (Azul):** Inclui filmes de super-heróis e grandes franquias de ação, como Iron Man e The Amazing Spider-Man 2.

- **Comunidade 3 (Verde):** Focado em produções de ficção científica e aventura, como Guardians of the Galaxy e Inception.
- **Comunidade 4 (Ciano):** Reúne filmes de fantasia e contos de fadas, como Cinderella e Pirates of the Caribbean: At World's End.

Métricas do Grafo - 13% da Base - 3ª Iteração

Informações Importantes do Grafo

Métrica	Valor
Número de Nós	66
Número de Arestas	430
Grau Médio dos Nós	13.03
Top 3 Filmes por Grau de Centralidade	Iron Man, Skyfall, Guardians of the Galaxy
Top 3 Filmes por Centralidade de Proximidade	Guardians of the Galaxy, Inception, Cinderella
Top 3 Filmes por Centralidade de Intermediação	Zootopia, Jurassic World, Cinderella

Figura 8. 3ª Iteração com 13% da base completa

O grafo gerado (Figura 7) apresenta um total de 66 nós e 430 arestas. Desta vez, o grau médio diminuiu levemente em comparação com as duas iterações anteriores, sugerindo que menos filmes estão se conectando diretamente. A centralidade de intermediação apresentou novas mudanças, o que mostra que outros filmes, como *Iron Man*, *Guardians of the Galaxy* e *Jurassic World*, desempenham um papel intermediário mais relevante na estrutura desse novo grafo.

Gráfico de Importância - 13% da Base - 3ª Iteração

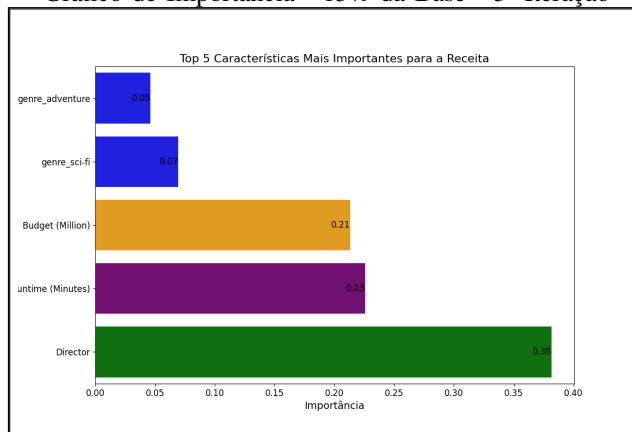


Figura 9. 3ª Iteração com 13% da base

Mais uma vez, uma mudança expressiva no peso de dos atributos em relação a receita:

- **Diretor (38%):** O fator mais importante na determinação da receita, a relevância dos diretores aumentou significativamente nesta iteração. Isso indica que o papel do cineasta não é apenas técnico, mas também estratégico,

visto que nomes consagrados ou inovadores podem atrair audiências leais e investidores, além de agregar valor criativo à produção.

- **Duração (23%):** A influência da duração continua alta, mostrando que filmes com tempos equilibrados tendem a atender melhor às expectativas do público. Narrativas longas, mas bem estruturadas, garantem imersão e impacto emocional, o que pode justificar a correlação positiva entre tempo de duração e sucesso comercial.
- **Orçamento (21%):** Apesar de sua importância histórica, o orçamento caiu para a terceira posição entre os fatores. Essa mudança reflete a tendência de que produções menores, mas estrategicamente posicionadas, podem competir com blockbusters de alto custo, especialmente quando bem dirigidas ou com apelo narrativo diferenciado.
- **Gênero de Ficção Científica (7%):** A ficção científica mantém a mesma presença da 2ª iteração, sendo valorizada por sua capacidade de explorar ideias inovadoras, efeitos especiais de alta qualidade e tramas instigantes que atraem audiências diversificadas.
- **Gênero de Aventura (5%):** Embora tenha o menor impacto entre os fatores listados, a aventura segue como um gênero sólido, tendo um pequeno aumento em relação a iteração anterior. Seu apelo universal e a capacidade de oferecer experiências dinâmicas e emocionantes garantem uma base de público consistente.

Essa análise reforça a ideia de que, na terceira iteração, a influência dos aspectos criativos (como a escolha do diretor) e narrativos (duração e gênero) supera progressivamente os aspectos financeiros (orçamento). Essa transição pode estar alinhada às mudanças no mercado cinematográfico, onde audiências buscam cada vez mais experiências autênticas e inovadoras.

A terceira iteração da análise permitiu um entendimento mais refinado das conexões entre os filmes, suas comunidades e os fatores que influenciam o sucesso financeiro. A representação gráfica, juntamente com as métricas e a identificação de características-chave, reforça a importância de se considerar aspectos estruturais e criativos no planejamento e lançamento de filmes. Essa abordagem oferece insights valiosos tanto para os produtores quanto para os distribuidores do mercado cinematográfico.

D. 4ª Iteração

Nesta iteração, o foco foi analisar a relação entre os filmes e suas características financeiras e qualitativas por meio de uma abordagem gráfica e quantitativa. Diferentemente das iterações anteriores, que enfatizaram aspectos como densidade do grafo e distribuição geral das métricas, nesta etapa o objetivo foi refinar a análise, priorizando conexões específicas entre os filmes, bem como identificar características determinantes para o sucesso financeiro.

Grafo Filmes - 6% da Base - 4ª Iteração

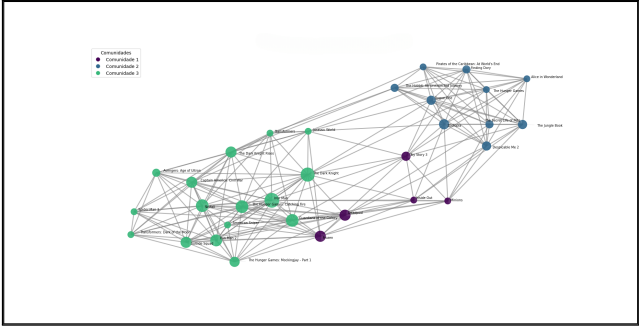


Figura 10. 4ª Iteração com 6% da base total.

A Figura 11 apresenta o grafo segmentado por comunidades, filtrado com base no critério de receita. Comparado aos grafos das iterações anteriores, este traz um refinamento mais evidente ao considerar um número menor de nós (31) e destacar subgrupos bem definidos. Três comunidades foram identificadas:

- **Comunidade 1 (roxa):** Filmes de animação, como *Toy Story 3* e *Frozen*, que mantêm alta receita e popularidade entre o público familiar.
- **Comunidade 2 (verde):** Grandes produções de super-heróis, como *Iron Man* e *The Dark Knight*, que continuam a dominar em termos de influência no grafo e conexão com outros filmes.
- **Comunidade 3 (azul):** Filmes de aventura e fantasia, como *Pirates of the Caribbean* e *The Hunger Games*, reforçando a relevância desses gêneros no mercado cinematográfico.

Métricas do Grafo - 6% da Base - 4ª Iteração

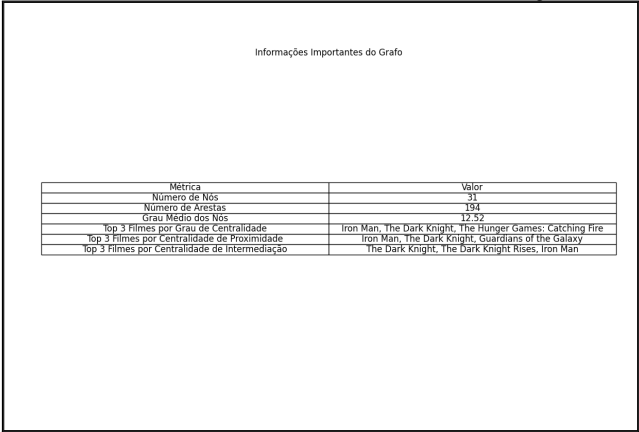


Figura 11. 4ª Iteração com 6% da base total.

O grafo gerado (Figura 10) apresenta um total de 31 nós e 194 arestas. Mais uma vez, o grau médio diminuiu, tendo cada vez menos conexões diretas. A centralidade de intermediação apresentou menos mudanças que anteriormente, mas novos filmes, como *The Dark Knight*, *The Hunger Games* e *The Dark Knight Rises*, desempenham um papel intermediário mais relevante na estrutura desse novo grafo.

Gráfico de Importância - 6% da Base - 4ª Iteração

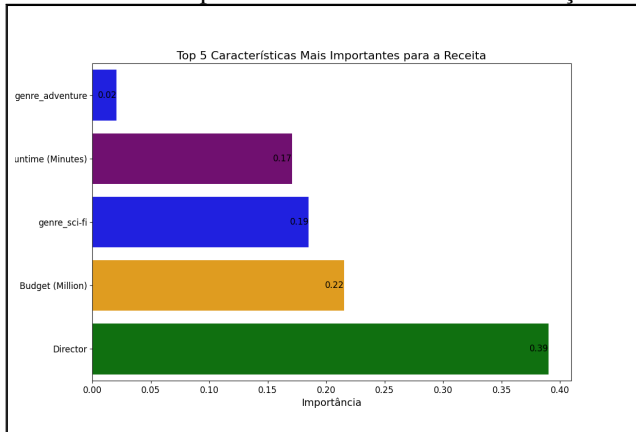


Figura 12. 4ª Iteração com 6% da base total.

- **Diretor (39%):** A característica mais influente, destacando o impacto de diretores renomados na performance dos filmes, similar às observações da 3ª iteração, onde diretores populares conectavam filmes de diferentes gêneros.
- **Orçamento (22%):** Sua relevância reforça a relação direta entre investimento em produção e receita, ecoando padrões vistos na análise da 3ª iteração, onde filmes com altos valores de produção formavam clusters específicos.
- **Gênero ficção científica (19%):** Este gênero aparece como um fator comercialmente atrativo, reforçando a presença de filmes como *Guardians of the Galaxy* e *Iron Man*.
- **Tempo de duração (17%):** Aponta que filmes mais longos, frequentemente associados a narrativas complexas, tendem a ser mais lucrativos.
- **Gênero aventura (2%):** Apesar de voltar a se destacar, o impacto isolado desse gênero é limitado, como já observado na 1ª iteração.

Ao longo das iterações, observamos uma mudança significativa na relevância de determinados fatores. Na primeira etapa, o orçamento era dominante, representando 55% da importância na previsão do sucesso. Contudo, conforme os dados foram refinados, houve um aumento expressivo na influência do diretor (chegando a 39% na última iteração) e do tempo de duração. Esse comportamento pode ser explicado pelo impacto do 'efeito de marca' de diretores renomados, que atraem públicos fiéis e garantem maior engajamento com suas produções [10]. Além disso, filmes com maior tempo de duração frequentemente apresentam narrativas mais elaboradas e experiências cinematográficas imersivas, o que pode contribuir para maior aceitação do público e, consequentemente, melhores bilheteiras [12].

Os resultados obtidos indicam que o orçamento é o principal fator determinante do sucesso financeiro de um filme, o que corrobora estudos anteriores como os de Gupta et al. [3] e Benite de Lucena e Mello [7], que também identificaram uma forte correlação entre investimento e receita. No entanto, ao

longo das iterações, observamos que a influência do orçamento diminuiu, enquanto o impacto de fatores como a reputação do diretor e o tempo de duração aumentou. Isso sugere que, embora um orçamento elevado facilite o sucesso, ele não é o único fator determinante, especialmente em um mercado onde diretores renomados e narrativas bem construídas atraem maior atenção do público.

V. CONCLUSÃO

O estudo apresentado explorou as dinâmicas de sucesso no mercado cinematográfico por meio de uma abordagem baseada em grafos e análise de dados quantitativos. As quatro iterações realizadas forneceram uma visão abrangente sobre os principais fatores que moldam a receita e o desempenho de filmes, considerando tanto suas características qualitativas (gêneros, avaliação e diretores) quanto quantitativas (orçamento, receita e métricas de centralidade).

A. Principais Descobertas

1) Estrutura das Redes Cinematográficas:

Ao longo das iterações, ficou evidente que os filmes não apenas se conectam por similaridade de gênero ou diretor, mas também formam clusters específicos baseados em seu impacto financeiro e popularidade. Foi observado que:

Filmes de super-heróis (*Iron Man*, *The Dark Knight*) são consistentemente centrais na rede, funcionando como hubs que conectam diversos subgrupos. Animações (*Frozen*, *Toy Story 3*) formam uma comunidade estável e separada, refletindo seu apelo específico ao público familiar. Filmes de aventura e fantasia (*The Hunger Games*, *Pirates of the Caribbean*) frequentemente ocupam posições intermediárias, conectando clusters distintos.

2) Impacto das Características Individuais:

A análise das características mais importantes para a receita revelou que fatores como o diretor e o orçamento desempenham papéis cruciais no desempenho financeiro. Em particular:

- **Diretor:** A característica mais determinante, destacando que diretores consagrados agregam valor significativo ao filme, tanto em termos de visibilidade quanto de qualidade percebida.
- **Orçamento:** Confirmando uma relação direta entre altos investimentos em produção e receitas elevadas, especialmente em gêneros como ficção científica e super-heróis.
- **Gêneros específicos:** Ficção científica e aventura, embora populares, mostraram relevância variável, sugerindo que seu impacto depende de combinações com outros fatores (como elenco e marketing).

3) Métricas de Centralidade e Conexões:

O uso de métricas como grau, proximidade e intermediação reforçou a importância de certos filmes na estrutura da rede:

- **Centralidade de Graus:** Indicou os filmes com maior número de conexões diretas, refletindo popularidade ou relevância para o público.

- **Centralidade de Proximidade:** Evidenciou filmes que atuam como facilitadores de acesso a outros clusters, reforçando sua importância estratégica.
- **Centralidade de Intermediação:** Revelou os filmes que servem como "pontes" entre comunidades, desempenhando papel essencial na conectividade geral da rede.

B. Comparação entre Iterações

As quatro iterações mostraram uma evolução clara na abordagem, indo de análises gerais sobre a conectividade dos filmes para insights mais detalhados sobre o impacto de características específicas na receita. Inicialmente, a rede era analisada com um grande número de nós e arestas, o que oferecia uma visão geral da densidade e interconectividade. Nas últimas iterações, refinamos os critérios, identificando comunidades mais claras e analisando aspectos diretamente relacionados ao sucesso financeiro.

Essa progressão metodológica possibilitou:

Uma segmentação mais precisa dos filmes, diferenciando-os com base em gêneros, receita e avaliação. E também, uma análise combinada de métricas qualitativas e quantitativas, conectando fatores subjetivos (diretor e gênero) com resultados objetivos (receita).

C. Implicações do Estudo

Os resultados apresentados têm implicações práticas tanto para a indústria cinematográfica quanto para pesquisadores e analistas de dados:

- **Para produtores e estúdios:** A identificação de características determinantes para o sucesso financeiro, como a escolha de diretores renomados e investimentos adequados, pode guiar decisões estratégicas sobre futuros projetos.
- **Para distribuidores:** A segmentação das comunidades oferece insights sobre como direcionar esforços de marketing para públicos específicos.
- **Para pesquisadores:** A metodologia baseada em grafos demonstra ser uma ferramenta eficaz para investigar dinâmicas de mercado em outras indústrias culturais, como música, literatura ou streaming.

D. Possibilidades Futuras

- Expandir a análise para incluir dados temporais, investigando como tendências de receita e popularidade evoluem ao longo do tempo.
- Incorporar novas variáveis, como premiações e presença em serviços de streaming, para avaliar sua influência na performance dos filmes.
- Explorar a aplicação de algoritmos de aprendizado de máquina para prever o sucesso de filmes com base nas características identificadas.

Em resumo, este estudo reforça que o sucesso de um filme é um fenômeno multifatorial, influenciado por aspectos artísticos, financeiros e de marketing, mas também que se considerarmos os atributos individualmente, temos um destaque

maior para o orçamento e o diretor. A abordagem metodológica, baseada em grafos e análise quantitativa, revelou dinâmicas complexas na indústria cinematográfica e forneceu insights práticos para atores estratégicos nesse mercado. Ao conectar elementos qualitativos e quantitativos, o estudo contribui para um entendimento mais profundo das forças que moldam o sucesso no cinema e abre caminho para novas pesquisas nessa área.

REFERÊNCIAS

- [1] Da SILVA, Michel P. "Material de aula: Analise Assintótica", Apresentação de slides, 2024. Disponível em: <https://ava.cefetmg.br/pluginfile.php/250215/mod-resource/content/8/Aula1.pdf>. Acesso em 15 jan. 2025.
- [2] MOURA, Pedro Augusto G. "Artigo-AEDS" Arquivo Online, 2024. Disponível em: <https://github.com/PedroAugusto08/Artigo-AEDS.git>. Acesso em 15 jan. 2025.
- [3] Gupta, K., Bajpayee, S., Priyadharsini, A. (2019). Movie success prediction. International Journal of Recent Technology and Engineering (IJRTE), 8(3), 5659–5663. Disponível em: https://www.researchgate.net/publication/364105426_Movie_Success_Prediction
- [4] Novaes, C. (2023, October 18). É possível prever o sucesso de um filme por meio do uso de inteligência artificial?
- [5] K Meenakshi et al 2018 J. Phys.: Conf. Ser. 1000 012100. Disponível em: <https://iopscience.iop.org/article/10.1088/1742-6596/1000/1/012100>
- [6] Kelvin. (2025, January 9). Crítica x Público: quem manda no sucesso de um filme? Estado De Minas - Em Foco. Disponível em: <https://www.em.com.br/emfoco/2025/01/09/critica-x-publico-quem-manda-no-sucesso-de-um-filme/>
- [7] Benite de Lucena e Mello, M. F., et al. (2020, Novembro). Fatores de influência sobre as receitas de filmes brasileiros.
- [8] Ciro125. "df_filmes_pre_processada.csv". GitHub. Disponível em: https://github.com/Ciro125/projeto_filme_tera/tree/main/bases_de_dados. Acesso em 15 jan. 2025
- [9] T. Cormen, E. Leiserson, R. L. Rivest, C. Stein. Algoritmos: Teoria e Prática, Editora Campus - Rio de Janeiro - 2002.
- [10] Mariana. (2024, March 19). Diretores de Cinema que Mudaram a Indústria para Sempre. Lucidarium. Disponível em: <https://lucidarium.com.br/diretores-de-cinema-que-mudaram-a-industria-para-sempre/>
- [11] Continente, R. (n.d.). O cinema autoral em Hollywood. Revista Continente. Disponível em: <https://revistacontinente.com.br/secoes/artigo/o-cinema-autoral-em-hollywood>
- [12] Toledo, M. (2024, May 2). Saiba qual a duração ideal de um filme, segundo pesquisa. CNN Brasil. Disponível em: <https://www.cnnbrasil.com.br/entretenimento/saiba-qual-a-duracao-ideal-de-um-filme-segundo-pesquisa/>