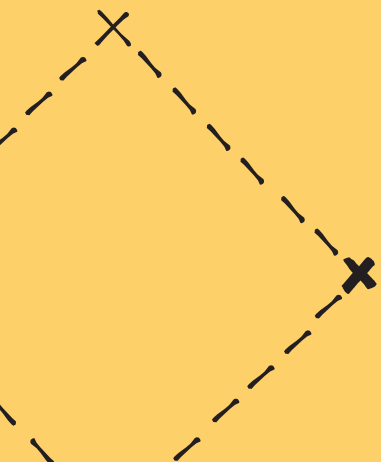


**Estimador de  
bilheteria  
de produções  
cinematográficas  
brasileiras:  
uma abordagem  
do aprendizado  
de máquinas**

Julia Taunay Perez,  
Everton Rodrigues Reis  
e Davi Noboru Nakano

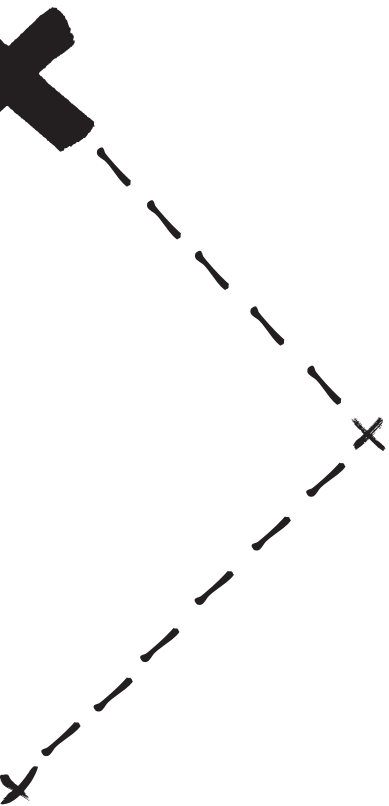


## Estimador de bilheteria de produções cinematográficas brasileiras: uma abordagem do aprendizado de máquinas

**1** Economista formada pela Universidade Estadual Paulista (Unesp), é atualmente doutoranda em engenharia de produção pela Escola Politécnica da Universidade de São Paulo (Poli/USP), com pesquisa centrada no desenvolvimento de instrumentos financeiros voltados para as indústrias criativas. É também sócia-fundadora da Lumen, consultoria de técnicas de inteligência artificial. Foi pesquisadora do projeto Creativeworks, em parceria com a Queen Mary University of London, auxiliando empreendimentos criativos a desenvolverem seus modelos de negócios.

**2** Cientista de dados em um fundo de investimentos e mestrando pela Poli/USP. Bacharel em ciência e tecnologia e em engenharia de gestão pela Universidade Federal do ABC (UFABC), já trabalhou como cientista de dados também na CVC Corp. Suas áreas de pesquisa são sistemas multiagentes e aprendizagem de máquina.

**3** Professor associado do Departamento de Engenharia de Produção da Escola Politécnica da USP. É pesquisador do Observatório da Inovação e Competitividade do Instituto de Estudos Avançados da USP e pesquisador visitante da Queen Mary University of London (Network Centre). Foi pesquisador visitante na University of Alberta, no Canadá, e coeditor do periódico *Production*, da Associação Brasileira de Engenharia de Produção. Trabalha em projetos nacionais e internacionais com instituições de ensino norte-americanas e europeias, tendo diversos artigos publicados em periódicos e congressos nacionais e internacionais. Sua pesquisa focaliza os seguintes temas: economia criativa e indústrias criativas; métodos para inovação no desenvolvimento de produtos e serviços; e ensino de engenharia.



## 1. Introdução

O mercado cinematográfico nacional alcançou, em 2017, uma renda de R\$ 2,7 bilhões, levando mais de 181 milhões de telespectadores às salas de exibição. No entanto, quando observamos a fatia dos produtores nacionais nesse montante, a realidade não se mostra tão pujante. Os 160 títulos nacionais lançados no mesmo ano foram assistidos por cerca de 17 milhões de telespectadores e geraram aproximadamente R\$ 240 milhões, o que representa menos de 9% do total da bilheteria (ANCINE, 2017).

A dificuldade em atrair público mediante a concorrência hollywoodiana não é exclusividade do cenário nacional. Em 2015, na União Europeia, os títulos lançados por estúdios americanos ficaram com 64% da audiência (KANZLER; MILLA, 2016). Nesse sentido, formuladores de políticas públicas em todo o mundo discutem a implementação de subsídios e cotas de telas como forma de mitigar os efeitos da concorrência norte-americana sobre a indústria cinematográfica nacional (MESSERLIN; PARC, 2014). A cadeia de valor da indústria cinematográfica se divide em três atividades principais: produção, distribuição e exibição (LEGOUX et al., 2016; MICHEL; AVELLAR, 2014). A relação econômica entre esses elos é mediada por contratos, cujos termos devem refletir a forma na qual as partes percebem e mitigam riscos.

Como no mercado norte-americano as atividades de produção e distribuição são verticalmente integradas, de modo que os maiores estúdios (Disney, Sony, Fox, Paramount, Universal, Warner e DreamWorks) produzem e distribuem os filmes (LEGOUX et al., 2016), poucos estudos buscaram entender o conflito de interesses existente entre tais agentes. No Brasil, há indícios de que as atividades de produção e distribuição não são verticalizadas. Segundo dados da Agência Nacional do Cinema (Ancine), no ano de 2016, das produtoras que atraíram 80% do público de longas-metragens nacionais (Rede Record, DiamondBack, Paris Produções, Lereby, Glaz Entretenimento, Camisa Listrada, Fraiha Produções e Filmland International), apenas duas atuam em atividades de distribuição<sup>4</sup>.

---

<sup>4</sup> A DiamondBack tem contratos de codistribuição com estúdios americanos e a Paris Produções distribui seus filmes em parceria com a Downtown Filmes.

Seguindo a montante na cadeia, a relação entre distribuidor e exibidor é envolta de conflitos. Na elaboração do contrato, o distribuidor pretende maximizar os rendimentos do filme, ao passo que a exibidora pretende maximizar tal rendimento em uma óptica de portfólio. Assim, enquanto o distribuidor eleva receita ao aumentar o tempo de exibição dos filmes nas salas de cinema, o exibidor, que ganha uma participação da bilheteria, intenciona ter mais lançamentos para atrair público (CHISHOLM et al., 2015).

Diversos estudos apontam que o desempenho do filme na pré-estreia e/ou na primeira semana de lançamento é determinante do seu sucesso (CHEN; XU; ZHANG, 2016; KARNIOUCHINA, 2011; LEGOUX et al., 2016; SAWHNEY; ELIASHBERG, 1996). No Brasil, essa problemática também se mostra presente; na média, os filmes nacionais arrecadam 48% de suas receitas na primeira semana. Ademais, a concentração da distribuição em *majors*, como apontam Michel e Avellar (2014), reduz a competitividade dos atores nacionais. Nesse sentido, o desenvolvimento de um modelo que estime o desempenho futuro de bilheteria de um filme expande a compreensão dos determinantes e da dinâmica da demanda brasileira de filmes nacionais, e permite elaborar estratégias tanto no âmbito dos atores da cadeia (produtoras, distribuidoras e exibidoras) como da política pública, aumentando a competitividade das produções nacionais.

Contudo, como a elevada incerteza da demanda do setor reflete em um comportamento de bilheteria que apresenta variância ilimitada (WALLS, 2005), os modelos lineares de predição, como o dos mínimos quadrados, mostram-se pouco aderentes às características da demanda cinematográfica. Esse cenário fez com que houvesse uma tendência de desenvolvimento de estudos que buscam encontrar métodos mais robustos de análise, com aprendizagem computacional (CHEN; XU; ZHANG, 2016; DELEN; SHARDA, 2010; GHIASSI; LIO; MOON, 2015; GUO; ZHANG; HOU, 2015; HUR; KANG; CHO, 2016; LEE et al., 2018; LIU; ZHAO, 2016; SHARDA; DELEN, 2006; ZHANG; LUO; YANG, 2009).

Assim, o presente estudo buscou aplicar técnicas de aprendizagem computacional em uma base de dados composta de 403 filmes lançados e exibidos entre 2009 e 2016, constituída a partir de diferentes fontes (Ancine, IMDb, AdoroCinema, Ministério da Justiça e Grupo Globo), com vistas ao desenvolvimento de um modelo preditivo de bilheteria. As variáveis explicativas selecionadas foram tempo de projeção, gênero, classificação indicativa, tipo de produtora, tempo de produção, sazonalidade, sequência, salas no lançamento e grau de fama do elenco.

Para as produtoras, espera-se que a compreensão de tais parâmetros possa auxiliar as etapas iniciais de desenvolvimento das produções nacionais (PACKARD et al., 2016).



Para as distribuidoras, tal instrumento permite maior eficiência na escolha do filme a ser distribuído e dá subsídio à definição de melhores termos contratuais com as exibidoras. Essas últimas também se beneficiarão ao conseguirem maior segurança e redução da incerteza do seu portfólio de filmes a serem exibidos.

## 2. Determinantes da demanda

Em estudos empíricos, a demanda do cinema é usualmente representada por duas variáveis: audiência e bilheteria. Enquanto a audiência refere-se ao público que assiste um filme em uma sala de exibição, a bilheteria refere-se ao montante arrecadado com a exibição.

Diversos fatores nos ajudam a compreender o comportamento da demanda, em termos tanto de público quanto de renda: características dos filmes, condições econômicas e demográficas, cenário competitivo, condicionantes financeiros, condições de distribuição e lançamento, rede e reputação do elenco e equipe técnica, críticas e boca a boca, como mostra o **Quadro 1**.

**Quadro 1 – Lista das variáveis explicativas da demanda no cinema**

	AUDIÊNCIA	BILHETERIA
<b>CARACTERÍSTICAS DOS FILMES</b>		
<b>Classificação indicativa</b>	CLEMENT; WU; FISCHER (2014)	SAWHNEY; ELIASHBERG (1996) COLLINS; HAND; SNELL (2002) BASUROY; CHATTERJEE; RAVID (2003) TERRY; BUTLER; ARMOND (2003) CHANG; KI (2005) DELMESTRI; MONTANARI; USAI (2005) WALLS (2005) SHARDA; DELEN (2006) BREWER; KELLEY; JOZEFOWICZ (2009) NELSON; GLOTFELTY (2012) KIM; PARK; PARK (2013) CLEMENT; WU; FISCHER (2014) DERRICK; WILLIAMS; SCOTT (2014) CRAIG; GREENE; VERSACI (2015) PACKARD et al. (2016)

<b>Gênero</b>	CLEMENT; WU; FISCHER (2014)	SAWHNEY; ELIASHBERG (1996) COLLINS; HAND; SNELL (2002) TERRY; BUTLER; ARMOND (2003) CHANG; KI (2005) DELMESTRI; MONTANARI; USAI (2005) WALLS (2005) SHARDA; DELEN (2006) BREWER; KELLEY; JOZEFOWICZ (2009) TREME (2010) NELSON; GLOTFELTY (2012) KIM; PARK; PARK (2013) CLEMENT; WU; FISCHER (2014) DERRICK; WILLIAMS; SCOTT (2014) CRAIG; GREENE; VERSACI (2015) PACKARD et al. (2016)
<b>Origem da produção</b>	CLEMENT; WU; FISCHER (2014)	MCKENZIE; WALLS (2013) CLEMENT; WU; FISCHER (2014)
<b>Sequência</b>	CLEMENT; WU; FISCHER (2014)	SAWHNEY; ELIASHBERG (1996) COLLINS; HAND; SNELL (2002) BASUROY; CHATTERJEE; RAVID (2003) TERRY; BUTLER; ARMOND (2003) CHANG; KI (2005) WALLS (2005) SHARDA; DELEN (2006) BREWER; KELLEY; JOZEFOWICZ (2009) KIM; PARK; PARK (2013) MCKENZIE; WALLS (2013) CLEMENT; WU; FISCHER (2014) CRAIG; GREENE; VERSACI (2015) PACKARD et al. (2016)
<b>Efeitos especiais</b>		SAWHNEY; ELIASHBERG (1996) SHARDA; DELEN (2006)
<b>CONDIÇÕES ECONÔMICAS E DEMOGRÁFICAS</b>		
<b>População</b>		NELSON; GLOTFELTY (2012)
<b>Índice de Preços</b>		BREWER; KELLEY; JOZEFOWICZ (2009)
<b>Renda per capita</b>		BREWER; KELLEY; JOZEFOWICZ (2009) NELSON; GLOTFELTY (2012)



<b>Sazonalidade</b>	CLEMENT; WU; FISCHER (2014)	TERRY; BUTLER; ARMOND (2003) KARNIOUCHINA (2011) CLEMENT; WU; FISCHER (2014) KARNIOUCHINA (2011) DERRICK; WILLIAMS; SCOTT (2014)
---------------------	--------------------------------	--

#### CENÁRIO COMPETITIVO

<b>Competição por tela e/ou por audiência</b>		SHARDA; DELEN (2006) CLEMENT; WU; FISCHER (2014) KARNIOUCHINA (2011) KIM; HONG; KANG (2017)
---	--	--

#### CONDICIONANTES FINANCEIROS

<b>Gastos com propaganda</b>	CLEMENT; WU; FISCHER (2014)	MCKENZIE; WALLS (2013) CLEMENT; WU; FISCHER (2014)
------------------------------	--------------------------------	---

<b>Orçamento</b>		BASUROY; CHATTERJEE; RAVID (2003) CHANG; KI (2005) BREWER; KELLEY; JOZEFOWICZ (2009) TREME (2010) NELSON; GLOTFELTY (2012) KIM; PARK; PARK (2013) MCKENZIE; WALLS (2013) CRAIG; GREENE; VERSACI (2015)
------------------	--	---

<b>Custo Negativo</b>		WALLS (2005)
-----------------------	--	--------------

#### CONDIÇÕES DA DISTRIBUIÇÃO E DO LANÇAMENTO

<b>Data de lançamento</b>		BASUROY; CHATTERJEE; RAVID (2003) CHANG; KI (2005) DELMESTRI; MONTANARI; USAI (2005) WALLS (2005) BREWER; KELLEY; JOZEFOWICZ (2009) TREME (2010) KIM; PARK; PARK (2013)
---------------------------	--	---

<b>Distribuidor</b>		KIM; PARK; PARK (2013) DERRICK; WILLIAMS; SCOTT (2014) PACKARD et al. (2016)
---------------------	--	--

---

**REDE E REPUTAÇÃO DO ELENCO  
E EQUIPE TÉCNICA**


---

<b>Diretor</b>	CLEMENT; WU; FISCHER (2014)	CHANG; KI (2005) DELMESTRI; MONTANARI; USAI (2005) KIM; PARK; PARK (2013) NELSON; GLOTFELTY (2012) CLEMENT; WU; FISCHER (2014)
----------------	--------------------------------	--

---

<b>Elenco</b>	CLEMENT; WU; FISCHER (2014)	SAWHNEY; ELIASHBERG (1996) BREWER; KELLEY; JOZEFOWICZ, (2009) CHANG; KI (2005) COLLINS; HAND; SNELL (2002) BASUROY; CHATTERJEE; RAVID (2003) DELMESTRI; MONTANARI; USAI (2005) TREME (2010) KARNIOUCHINA (2011) KIM; PARK; PARK (2013) MCKENZIE; WALLS (2013) CLEMENT; WU; FISCHER (2014) DERRICK; WILLIAMS; SCOTT (2014) NELSON; GLOTFELTY (2012) PACKARD et al. (2016) SHARDA; DELEN (2006) WALLS (2005) CRAIG; GREENE; VERSACI (2015)
---------------	--------------------------------	--

---

**CRÍTICAS E BOCA A BOCA**


---

<b>Redes sociais</b>	CLEMENT; WU; FISCHER (2014)	KARNIOUCHINA (2011) KIM; PARK; PARK (2013) CLEMENT; WU; FISCHER (2014) CRAIG; GREENE; VERSACI (2015) CHEN; XU; ZHANG (2016) KIM; HONG; KANG (2015, 2017)
----------------------	--------------------------------	---

---

<b>Review da audiência</b>		CHANG; KI (2005) NELSON; GLOTFELTY (2012) PACKARD et al. (2016)
----------------------------	--	---

---

---

**Review da crítica**

CLEMENT; WU;  
FISCHER (2014)

SAWHNEY; ELIASHBERG (1996)  
BREWER; KELLEY; JOZEFOWICZ (2009)  
BASUROY; CHATTERJEE; RAVID (2003)  
CHANG; KI (2005)  
COLLINS; HAND; SNELL (2002)  
TREME (2010)  
KIM; PARK; PARK (2013)  
MCKENZIE; WALLS (2013)  
CLEMENT; WU; FISCHER (2014)  
PACKARD et al. (2016)  
TERRY; BUTLER; ARMOND (2003)

---

**Nomeações e prêmios**

TERRY; BUTLER; ARMOND (2003)  
PACKARD et al. (2016)

---



## 2.1 Modelo preditivo de demanda

A base deste estudo é composta de 403 filmes lançados e exibidos entre 2009 e 2016. Há um total de nove atributos que são utilizados como variáveis preditoras, selecionadas a partir da larga adoção em estudos empíricos, como mostra o **Quadro 1**, e da disponibilidade de informação. São elas: gênero, classificação indicativa (CI), tipo de produtora, sazonalidade, sequência, salas no lançamento e grau de fama do elenco (star power). Tempo de projeção e tempo de produção, apesar de não terem sido empregadas em estudos, foram aqui adicionadas e mostraram-se significativas, como se discutirá mais à frente.

A variável a ser predita foi originada a partir da renda bruta do filme, divulgada pela Ancine, corrigida pelo IPCA – Cinema Mensal e posteriormente convertida em oito classes divididas por quartilhes de 12,5%. Os intervalos são representados na **Tabela 1**. Como se pode notar, para os fins deste modelo, um filme é considerado um sucesso de bilheteria (*blockbuster*) quando supera uma renda de R\$ 5,8 milhões.

**Tabela I: Distribuição de filmes nas classes de desempenho de bilheteria (R\$)**

INTERVALO	CLASSE	QUANTIDADE DE FILMES
[5.800.000,00)	A	51
[920.000,5.800.000)	B	50
[170.000,920.000)	C	49
[62.000,170.000)	D	52
[29.000,62.000)	E	50
[13.000,29.000)	F	50
[5.500,13.000)	G	51
[0,5.500)	H	50



### 2.1.1. Variáveis preditoras

#### *Gênero*

O impacto de categorias de gênero específicas, isto é, se o filme é um romance ou drama, por exemplo, varia de país para país, em razão de condicionantes culturais que acabam por afetar o gosto e a preferência dos consumidores (NELSON; GLOTFELTY, 2012). Como a classificação de gênero não segue padrões nos estudos empíricos, é impossível comparar os resultados dessa variável nos distintos modelos existentes na literatura.

Para o presente estudo, buscou-se superar tal adversidade por meio do cruzamento das classificações das diferentes fontes. Assim, a construção da variável gênero foi obtida por meio de uma análise de frequência entre as classificações do IMDb, do Ministério da Justiça e do AdoroCinema.

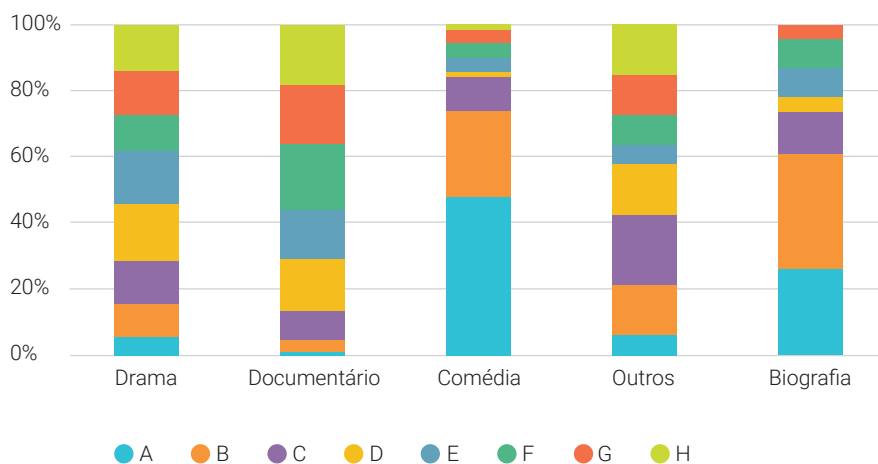
Algumas classificações de gênero possuíam poucos exemplares, ou mesmo eram únicas na amostra, como é possível observar na **Tabela 2**, o que traz um problema para as técnicas de aprendizado de máquinas. Primeiramente, porque não há exemplos suficientes para treinar e testar o algoritmo; depois, porque cada classificação se transforma em uma dimensão distinta do modelo. Dessa forma, optou-se por agrupar todos os gêneros que possuíam menos de 20 filmes na classe “outros”. A classificação final utilizada com a respectiva frequência ficou como se mostra na **Tabela 2**.

**Tabela II: Classificação de Gênero**

CLASSIFICAÇÃO INICIAL		CLASSIFICAÇÃO FINAL	
GÊNERO	QUANTIDADE	GÊNERO	QUANTIDADE
Drama	151	Drama	151
Documentário	127	Documentário	127
Comédia	69	Comédia	69
Biografia	23	Outros	33
Suspense	8	Biografia	23
Aventura	5		
Ação	4		
Animação	4		
Musical	3		
Terror	2		
Crime	1		
Histórico	1		
Família	1		
Romance	1		
Faroeste	1		
Fantasia	1		
Policial	1		

A **Figura 1** traz o percentual de filmes por gênero e classe. Pode-se observar que 48% das comédias e 26% das biografias da amostra são *blockbusters*. Em outras palavras, esses, somados, são os gêneros que alcançam melhor desempenho de bilheteria.

**Figura 1: Percentual de filmes por gênero e classe**



### *Classificação indicativa (CI)*

A classificação indicativa visa sinalizar o conteúdo do filme, informando pais e responsáveis sobre possíveis cenas de sexo e violência. Em termos de mercado, a classificação indicativa indica o tamanho da audiência potencial (CHANG; KI, 2005). Nesse sentido, espera-se que, quanto maior a restrição, menor o público potencial e, portanto, menores as chances de ter um bom desempenho de bilheteria. Empiricamente, tal hipótese foi corroborada por alguns estudos (CHANG; KI, 2005; NELSON; GLOTFELTY, 2012; WALLS, 2005), enquanto outros não puderam constatar sua influência (SHARDA; DELEN, 2006; BREWER; KELLEY; JOZEFOWICZ, 2009; KIM; PARK; PARK, 2013).

No Brasil, a classificação indicativa das obras audiovisuais é feita pela Coordenação de Classificação Indicativa (Cocind), do Departamento de Promoção de Políticas de Justiça (DPJUS) filiado à Secretaria Nacional de Justiça (SNJ) do Ministério da Justiça. A quantidade de filmes por CI da amostra é representada na **Tabela 3**.

O desempenho dos filmes em cada categoria é apresentado na **Figura 2**. Enquanto a participação de *blockbusters* é mais bem distribuída, observa-se que, em filmes para maiores de 18 anos, as chances de ser um verdadeiro fracasso de bilheteria aumentam significativamente.

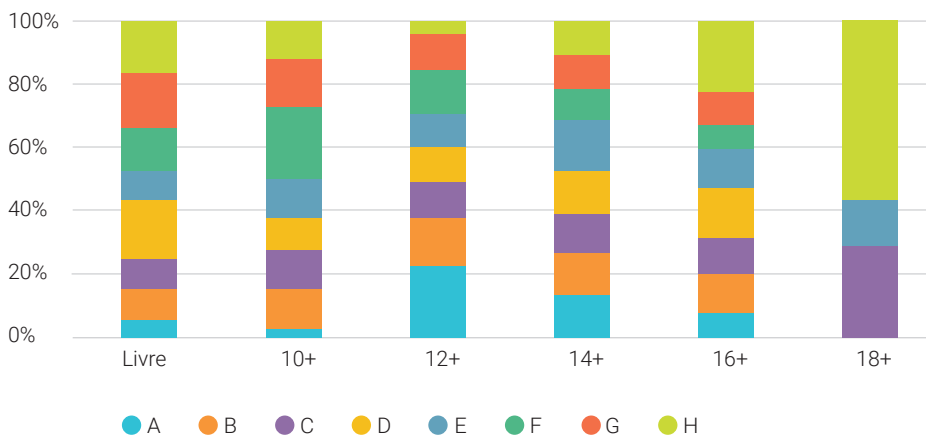
**Tabela III - Distribuição da Classificação Indicativa na Amostra**

CLASSIFICAÇÃO INDICATIVA	QUANTIDADE
Livre	73
10+	40
12+	120
14+	112
16+	51
18+	7
<b>Total</b>	<b>403</b>





**Figura 2: Porcentual de filmes por classe e por classificação indicativa**



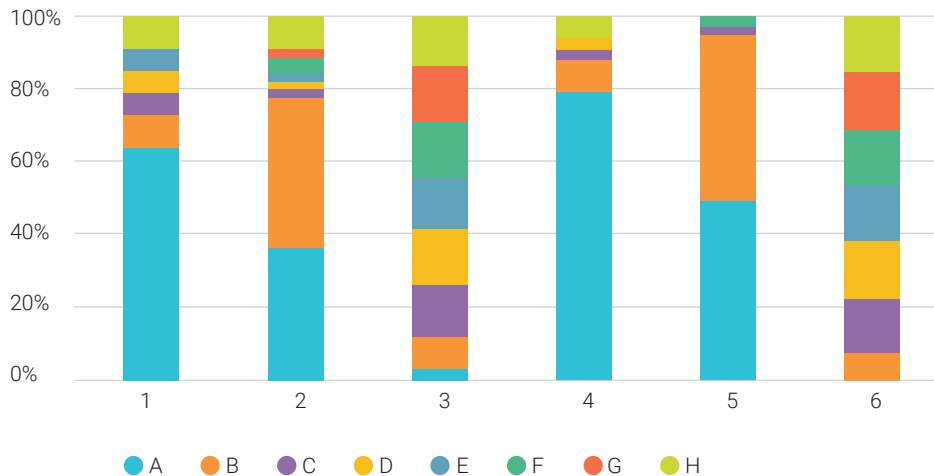
### *Tipo de produtora*

Grande parte dos estudos busca compreender a força dos estúdios no direcionamento do sucesso de filmes. Como os estúdios atuam tanto na produção quanto na distribuição do filme, eles costumam avaliar o “poder do distribuidor”. No Brasil, as atividades de produção e distribuição de filmes não são verticalmente integradas. Nesse sentido, buscou-se avaliar de que maneira uma produtora tende a influenciar o sucesso de bilheteria. A análise do agente produtor se deu pois a capacidade de influência do distribuidor deve estar refletida no número de telas em que o filme é exibido no lançamento, outra variável do modelo.

A operacionalização da influência de agentes da cadeia ocorre, usualmente, por meio da determinação de classes de poder econômico, sendo *majors* aqueles agentes com maior poder econômico. Neste estudo, buscou-se criar três classes de influência: *major*, *medium* e *minor*. *Major* refere-se àquelas produtoras que acumulam aproximadamente 85% do total arrecadado com bilheteria para o período analisado; as *medium* acumulam 10% do total da bilheteria; e as *minor* são aquelas que abocanham a “franja”, isto é, os 5% restantes da bilheteria.

Dois períodos de determinação da influência do produtor foram considerados, sendo adotado aquele que apresentou melhores resultados. No primeiro, buscou-se categorizar a produtora em função do desempenho do ano anterior. No segundo, em função do desempenho médio dos três últimos anos por filme. Os resultados são apresentados na **Figura 3**.

**Figura 3: Comparativo dos métodos de classificação da produtora**



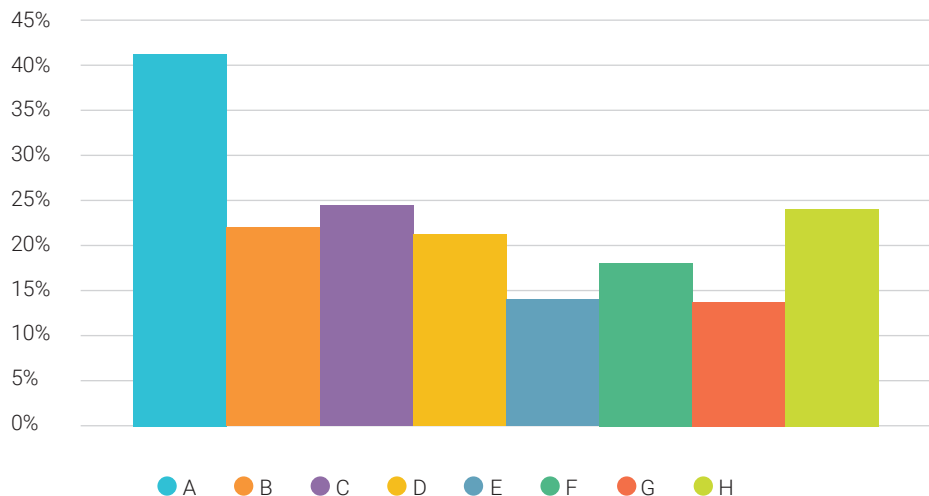
Observa-se que, ao utilizar a média dos últimos três anos, é possível obter uma separação melhor entre as classes. Ao comparar *Medium* na média dos três últimos anos, a maioria dos filmes está nas classes A e B (96% dos filmes), enquanto que, no método do ano anterior, *Medium* classifica nas classes A e B apenas 78% dos filmes. Na classe *Minor*, no método dos três últimos anos, não há filme algum da classe A, enquanto que no método do ano anterior há 4% dos filmes, o que gera mais ruído para o algoritmo de aprendizado de máquina, incrementando o erro em suas classificações.

### Sazonalidade

A indústria cinematográfica é caracterizada por uma elevada sazonalidade, com os picos de demanda ocorrendo nas férias e em feriados (DELMESTRI; MONTANARI; USAI, 2005; EINA, 2007). Nesse sentido, os filmes que são lançados nesse período tendem a ter melhor desempenho de bilheteria.

Considerou-se como período de férias escolares os meses completos de janeiro, julho e dezembro. A **Figura 4** mostra o porcentual de filmes lançados nas férias ou em feriados.

**Figura 4: Porcentual de filmes lançados nas férias ou em feriados por classe**



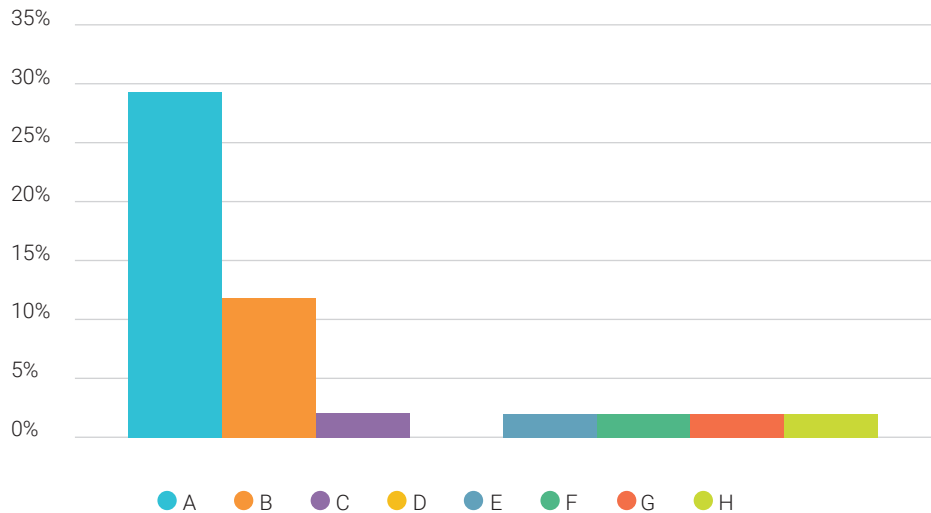
É possível observar que há certa correlação entre a sazonalidade e as classes. Entre os *blockbusters* (classe A), 41% dos filmes foram lançados nessas datas especiais. Nas demais classes, esse percentual vai caindo quase que linearmente, não fosse pela classe H, que inverte a tendência e traz ruído para a variável. Suspeita-se que, como nessas datas costuma ocorrer os lançamentos dos grandes estúdios, como Disney e Fox, pode haver uma tendência de filmes nacionais serem preteridos.

### Sequência

Sequências são filmes cujos materiais e conteúdo baseiam-se em outro filme, programa de televisão, videogame ou livro (BREWER; KELLEY; JOZEFOWICZ, 2009). Por já terem uma base pré-existente de fãs, supõe-se menor nível de risco e menos chance de fracasso (CRAIG; GREENE; VERSACI, 2015).

Os filmes da amostra foram classificados manualmente e os resultados por classe de desempenho são apresentados na **Figura 5**. À primeira vista, a hipótese de que as sequências diminuem o risco é comprovada. No entanto, a baixa quantidade de filmes classificada como sequência, 26 filmes ou 6,5% da base, faz com que o número de exemplos para aprendizado seja insuficiente, mesmo nas classes em que há maior quantidade de sequências (A e B).

Figura 5: Participação de seqüências por classe

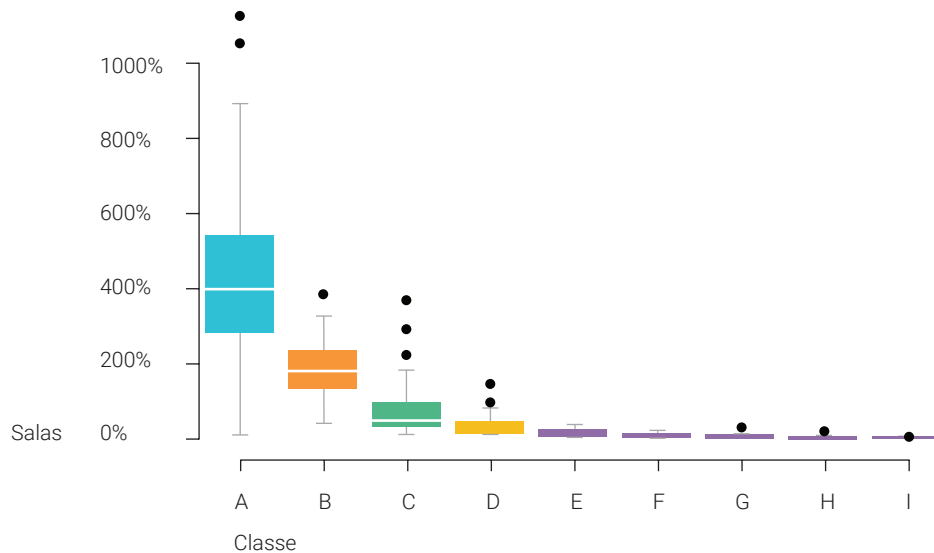


*Salas no lançamento*

O número de salas no lançamento reflete as expectativas do distribuidor quanto ao desempenho futuro do filme (CHANG; KI, 2005; CLEMENT; WU; FISCHER, 2014; MCKENZIE; WALLS, 2013). Como o total de bilheteria é, em grande medida, determinado pelo desempenho do filme na primeira semana, quanto maior o número de salas em que for exibido no lançamento, maior o alcance ao público e, conseqüentemente, maior a chance de sucesso.

Como mostra a **Figura 6**, o número de salas no lançamento está positivamente relacionado com a renda que o filme atinge.

Figura 6: Número de salas no lançamento por classe de desempenho



#### Grau de fama do elenco (star power)

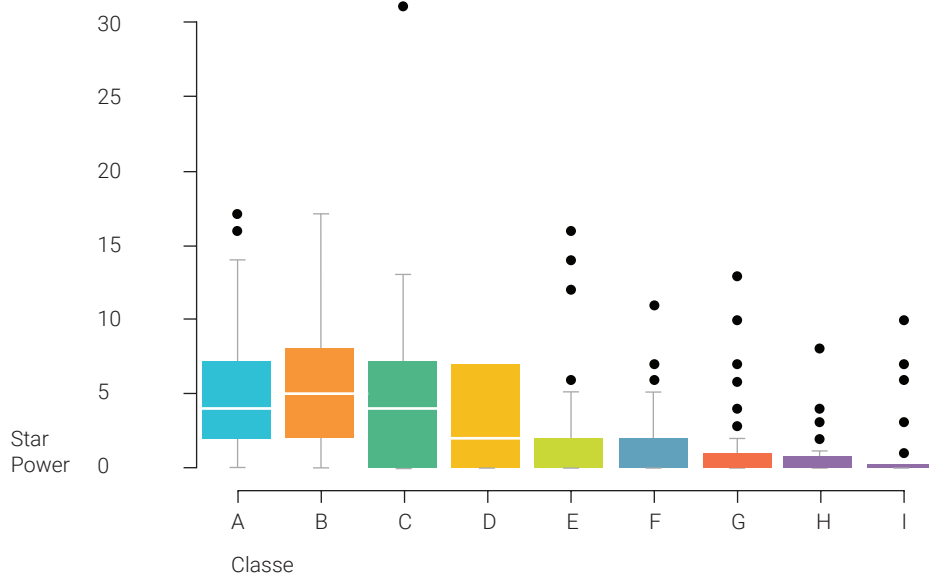
Como é possível observar na **Tabela 1**, a capacidade do ator famoso de atrair público foi largamente abordada em estudos empíricos. Mais uma vez, a comparação dos resultados é difícil de ser feita, em razão da falta de padronização na forma como se constrói a variável nos diferentes modelos. Enquanto alguns estudos definiram como ator famoso aquele que já tivesse alcançado algum desempenho expressivo anteriormente, em termos seja de atração de público (HUR; KANG; CHO, 2016), seja de bilheteria (PACKARD et al., 2016), outros definiram em função de aparições em ranques e revistas populares (BREWER; KELLEY; JOZEFOWICZ, 2009; TREME, 2010), ou ainda nomeações e/ou premiações (BASUROY; CHATTERJEE; RAVID, 2003).

No Brasil, dada a grande popularidade das novelas da Rede Globo, a operacionalização da variável (I) levou em consideração aparição de atores do elenco nesse tipo de produção nos cinco anos antecessores ao lançamento.

$$(I) \text{ Star Power} = \sum_1^n \alpha_n, \text{ onde } \alpha \text{ representa a quantidade de novelas em que o ator } n \text{ apareceu nos últimos cinco anos.}$$

O valor final da variável é a somatória, portanto, de cada ator. **A Figura 7** apresenta a relação entre a fama dos atores do filme e o desempenho de bilheteria. Apesar de as classes com melhor desempenho econômico apresentarem médias superiores em relação ao grau de fama dos atores, é possível observar muitos *outliers* nas classes de pior desempenho econômico. Dessa forma, a presença de ator famoso não diminui o risco de o filme ser um fracasso de bilheteria.

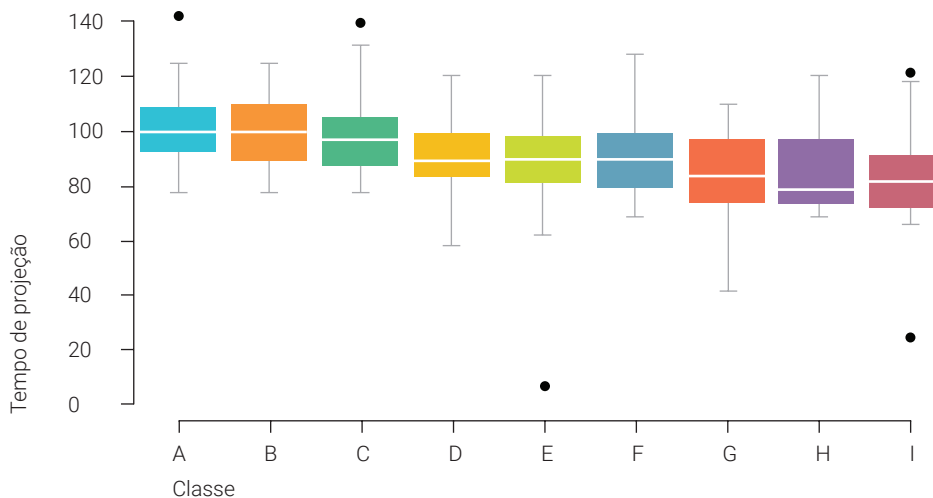
**Figura 7: Grau de fama do elenco por classe de desempenho**



### *Tempo de projeção*

A influência do tempo de projeção no desempenho de bilheteria é pouco estabelecida (GHIASSI; LIO; MOON, 2015). Neste modelo, a variável é um número inteiro que indica qual a duração do filme em minutos. Os filmes com melhor desempenho de bilheteria têm, em média, cem minutos. Na medida em que o tempo de projeção cai, o seu desempenho de bilheteria também tende a cair, como é possível notar na **Figura 8**.

**Figura 8: Tempo de projeção por classe de desempenho**



### *Tempo de produção*

A variável “tempo de produção” ainda não foi explorada na literatura. Representa o tempo decorrido entre o ano de produção e seu lançamento<sup>5</sup>. A racionalidade para inclusão da variável é que seja uma *proxy* para capacidade empreendedora.

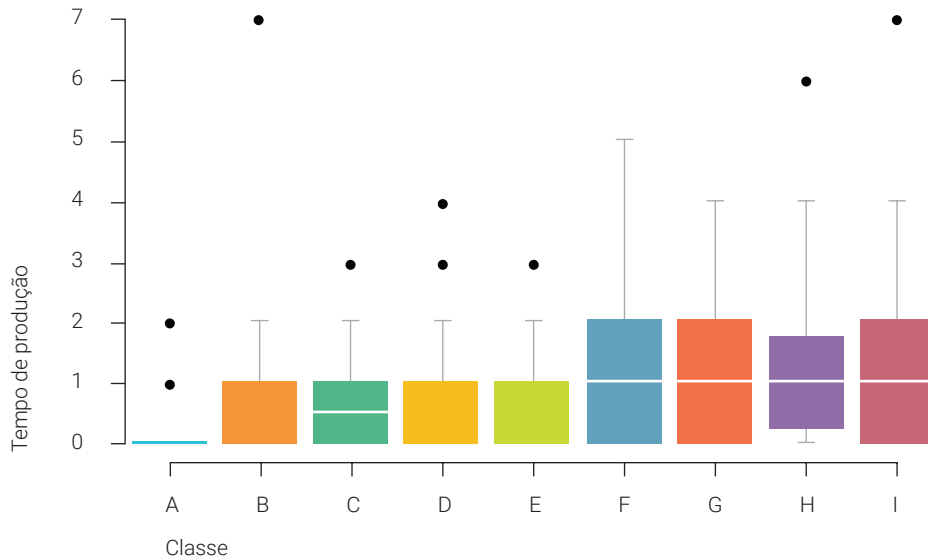
Nesse sentido, quanto menor o tempo decorrido entre o início da produção e o seu lançamento, maior a capacidade do empreendedor cultural de mobilizar recursos. Essa capacidade empreendedora, por sua vez, aumenta as chances de o público perceber a qualidade do filme.

Tal hipótese só faz sentido para o contexto nacional, largamente baseado na produção de dramas, documentários e comédias que se utilizam muito pouco de efeitos especiais. Dessa forma, a etapa de pós-produção é naturalmente reduzida.

A **Figura 9** traz a relação entre o tempo de produção e as classes de desempenho de bilheteria. Observa-se que, quanto maior o tempo de produção, pior, em média, é o seu desempenho.

<sup>5</sup> O ano de produção foi retirado do site Filmow por meio de Scraper, enquanto o ano de lançamento foi retirado da base da Ancine.

Figura 9: Relação entre tempo de produção e desempenho de bilheteria



### 3. Metodologia

Dos 403 filmes, 80% foram utilizados para treinamento (322 filmes) e 20% para teste (81 filmes). Para dividir a base, foi utilizado o parâmetro *random\_state* = 1, importante para reprodutibilidade. Dessa forma, em todas as vezes que o experimento for repetido a divisão será sempre a mesma. A construção da base de treinamento e teste foi feita de forma estratificada, ou seja, as proporções entre as classes foram mantidas tanto na base de treinamento quanto na de teste, conforme observado na **Tabela 4**.

Para treinar e validar o modelo, bem como otimizar seus parâmetros, foi utilizada a validação cruzada "*k-fold*", em que a base de treinamento foi dividida cinco vezes (5-*folds*) entre treinamento e validação, e o resultado final é a média de cada *fold*. A **Figura 10** ilustra o conceito.

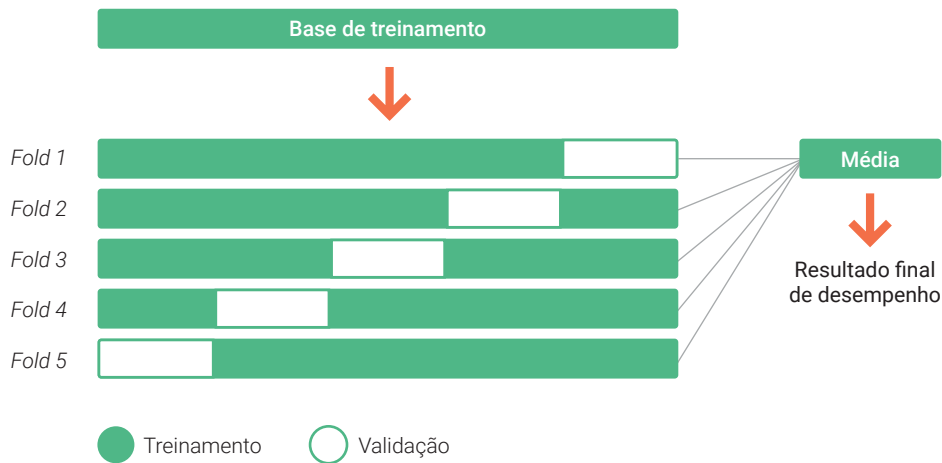




Tabela IV – Proporções das classes das bases de treinamento e teste

CLASSE	TREINAMENTO	TESTE
A	41	10
B	40	10
C	39	10
D	41	11
E	40	10
F	40	10
G	41	10
H	40	10
<b>Total</b>	<b>322</b>	<b>81</b>

Figura 10: Validação cruzada “k-fold”



Quando treinamos um modelo, esse pode sofrer ou por ficar muito simples (*high bias*) e não se ajustar aos dados (*underfitting*) ou por ficar muito complexo (*high variance*) e se ajustar demais aos dados (*overfitting*). Nos dois casos, quando apresentados novos dados ao modelo, este não terá capacidade suficiente para generalizar suas recomendações. Esse problema é conhecido como *bias-variance trade-off* (RASCHKA, 2015). A validação cruzada ajuda a estimar o erro de generalização do modelo em dados que o modelo nunca viu e foi utilizada para calibrar os parâmetros do modelo.

Para calibrar o modelo foi utilizada uma técnica chamada *grid search*, recomendada quando o espaço de parâmetros não for muito grande, ou seja, foi realizada uma busca exaustiva entre todas as combinações de parâmetros. O **Quadro 2** apresenta os parâmetros, os valores a serem otimizados e as métricas empregados no modelo de regressão logística:

**Quadro II – Parâmetros, valores a serem otimizados e métricas do modelo de regressão logística**

CLASSIFICADOR	HIPERPARÂMETRO	VALORES	MÉTRICA
Logistic regression (LR)	C	10%, $x \in \{-7, -6.82, \dots, 2\}$	Accuracy
	solver	lbfgs	
	multi_class	multinomial	

## 4. Resultados

Para poder comparar os resultados, foi treinado um *dummy classifier* para ser usado como *baseline*. O *dummy classifier* é um classificador que faz previsões baseado em regras simples. Utilizamos a estratégia da classe majoritária. O algoritmo usa a classe com mais casos da base de treinamento e prediz essa classe para todos os casos da base de teste. Por exemplo, como as classes A, D e G possuem 41 casos na base de treinamento, o algoritmo escolhe a primeira classe (A) e prediz na base de teste que todas as observações são "A", o que gera um acerto de 10/81 (12,35%). Essa quantidade de acertos é a acurácia de *baseline* do modelo.

Inicialmente, o modelo foi treinado sem otimizar seus hiperparâmetros, para serem comparados com o *baseline*. A acurácia média foi de 39,42 ( $\pm 3,35\%$ ) na validação cruzada com *5-folds* e de 48,15% na base de teste. Comparado ao *baseline* de 12,35%, foi quase quatro vezes superior.

Após a otimização dos hiperparâmetros, a acurácia média foi de 42,47 ( $\pm 4,65\%$ ) na base de treinamento e de 46,91% na base de teste. Apesar da melhora na base de treinamento, não houve melhoria na base de teste. Nesse sentido, o modelo sem otimizar os parâmetros teve um resultado melhor que o otimizado.

Na **Tabela 5** é apresentado o coeficiente de cada variável do modelo para cada classe. A partir dela, foi construída a **Tabela 6**, que apresenta indícios da importância de cada variável para cada classe. Para tanto, foi utilizado o valor absoluto de cada coeficiente e analisado se sua influência é negativa ou positiva para definir cada classe. Em seguida, visando entender a influência de cada variável para definir as classes, foi desenvolvida a **Tabela 7**. Nessa tabela, para cada classe, é apresentado em qual posição a variável ficou e então calcula-se a mediana e a média da posição de cada variável. As variáveis foram ordenadas primeiramente pela mediana e, como critério de desempate, foi utilizada a média.

A partir da **Tabela 7** é possível entender um pouco melhor como cada variável influenciou de forma geral a classificação das classes. Observa-se que a variável mais importante é a quantidade de salas no lançamento, e que surpreendentemente a segunda variável é o tempo de produção, que não foi utilizado em trabalhos anteriores, trazendo uma grande contribuição para a área.

Tabela V – Coeficientes da Regressão Logística

VARIABLE	CLASS							
	A	B	C	D	E	F	G	H
Tempo_de_Projeção	-0.05461	-0.03466	-0.01590	0.00301	0.01725	0.01854	0.02251	0.04387
Tempo_producao	-0.00274	-0.04151	-0.07159	-0.12149	-0.15833	0.17523	0.17653	0.04389
Sazonalidade	0.03471	0.00802	0.08062	-0.06886	-0.06327	-0.00862	-0.05029	0.06768
Sequencia	-0.00784	0.00653	0.01145	-0.01759	0.00629	0.00975	-0.00575	-0.00285
Salas_lancamento	0.26543	0.25686	0.22728	0.16278	0.04478	-0.03485	-0.11161	-0.81069
Star_Power	-0.02274	0.03620	0.07242	0.02940	-0.00166	-0.09813	0.03948	-0.05497
Gênero_Biografia	0.00820	0.07231	-0.03177	-0.01798	-0.00200	0.00183	-0.01744	-0.01314
Gênero_Comédia	-0.00304	0.01365	0.03166	-0.02266	-0.01451	0.00908	0.00468	-0.01887
Gênero_Documentário	-0.01156	-0.02605	0.02153	-0.03572	-0.04546	0.03220	0.00977	0.05528
Gênero_Drama	0.02954	-0.08244	-0.01898	0.06655	0.06455	-0.02585	-0.00259	-0.03079
Gênero_Outros	-0.04474	0.01103	0.01494	-0.00874	-0.01943	0.00751	0.00633	0.03310
Livre	-0.02680	0.02479	0.02932	0.06272	-0.07037	-0.03487	-0.00139	0.01661
14+	0.02526	-0.04289	-0.02233	-0.04641	0.10019	-0.01112	0.03856	-0.04128
10+	-0.02021	0.00193	0.04046	-0.04551	-0.01020	0.03432	0.01694	-0.01774
16+	-0.00142	0.01795	-0.01791	0.02902	-0.03442	-0.03276	-0.05408	0.09361
18+	-0.00020	-0.00088	0.03815	-0.01352	-0.00080	-0.02333	-0.03072	0.03130
12+	0.00176	-0.01239	-0.05031	-0.00485	-0.00127	0.09253	0.03143	-0.05691
Tipo_produtores_Major	0.05833	-0.05146	-0.00816	0.00561	-0.01029	-0.00878	-0.01214	0.02689
Tipo_produtores_Medium	0.08457	0.03785	-0.08434	-0.02940	-0.00478	-0.00202	-0.00102	-0.00086
Tipo_produtores_Minor	-0.16451	0.00212	0.10988	0.00525	-0.00178	0.03557	0.01391	-0.00044

## 5. Discussão

O modelo alcançou resultados significativos, com cerca de 48% de acurácia, sendo que cada classe de desempenho foi impactada de maneira diferente pelos parâmetros.

O número de salas no lançamento, como esperado, é a variável que mais impacta positivamente os sucessos de bilheteria. Na medida em que reflete as expectativas do distribuidor e é resultado das negociações entre exibidores e distribuidores, sugere-se uma análise mais aprofundada dessa relação em estudos futuros.



Entre os tipos de produtoras, as *minors* se relacionam negativamente com os *blockbusters*. Assim, é possível inferir que dificilmente uma pequena produtora conseguirá produzir um grande sucesso de bilheteria.

Apesar da relação positiva do tempo de projeção com o desempenho de bilheteria, como visto na **Figura 6**, observa-se que, entre as classes de melhor desempenho, a relação é inversa, ou seja, quanto maior o tempo de exibição, menor a bilheteria alcançada.

O tempo de produção, ou seja, a capacidade empreendedora, apresentava uma relação negativa com as classes: quanto maior o tempo de produção, pior a classe. Como consequência, os coeficientes dessa variável se mostraram negativos para as melhores classes, penalizando o filme com maior tempo de produção. Para as piores classes, os coeficientes se tornam positivos, reforçando que, quanto maior o tempo, pior a classe.

Para algumas classes, como D, E, F e G, o valor desse coeficiente é alto, e em alguns casos é o maior coeficiente (E, F, G), mostrando-se mais importante que a quantidade de salas no lançamento. O tempo foi dado em anos, o que trouxe baixa variação entre os *blockbusters* produzidos e lançados no mesmo ano. Nesse sentido, se os valores fossem dados em meses, os resultados poderiam ser ainda mais significativos.

Para a sazonalidade, é possível observar como os valores dos coeficientes vão diminuindo a cada classe, como verificado na análise descritiva dessa variável.

Vale destacar que o trabalho apresenta algumas limitações, como o tamanho da base e o fato de a definição de algumas características não ter sido feita com base em uma avaliação de especialistas. No entanto, a introdução da variável “tempo de produção” é uma importante contribuição à literatura, visto que não foi utilizada em trabalhos passados e mostrou-se uma variável importante na separação das classes. Outras variáveis precisam ser mais bem estudadas, como gênero, classificação indicativa e sequência, que em outros estudos mostraram-se mais importantes do que neste, deixando oportunidades abertas para pesquisas futuras.

VI – Variáveis da Regressão Logística ordenadas por valor absoluto e classe

VARIÁVEL	A	INFLU- ÊNCIA	VARIÁVEL	B	INFLU- ÊNCIA	VARIÁVEL	C	INFLU- ÊNCIA	VARIÁVEL	D	INFLU- ÊNCIA
Salas_lancamento	0.26543	+	Salas_lancamento	0.25866	+	Salas_lancamento	0.22728	+	Salas_lancamento	0.16278	+
Tipo_produtores_Minor	0.16451	-	Gênero_Drama	0.08244	-	Tipo_produtores_Minor	0.10988	+	Tempo_producao	0.12149	-
Tipo_produtores_Medium	0.08457	+	Gênero_Biografia	0.07231	+	Tipo_produtores_Medium	0.08434	-	Sazonalidade	0.06886	-
Tipo_produtores_Major	0.05833	+	Tipo_produtores_Major	0.05146	-	Sazonalidade	0.08062	+	Gênero_Drama	0.06655	+
Tempo_de_Projeção	0.05461	-	14+	0.04289	-	Star_Power	0.07242	+	Livre	0.06272	+
Gênero_Outros	0.04474	-	Tempo_producao	0.04151	-	Tempo_producao	0.07159	-	14+	0.04641	-
Sazonalidade	0.03471	+	Tipo_produtores_Medium	0.03785	+	12+	0.05031	-	10+	0.04551	-
Gênero_Drama	0.02954	+	Star_Power	0.0362	+	10+	0.04046	+	Gênero_Documentário	0.03572	-
Livre	0.0268	-	Tempo_de_Projeção	0.03466	-	18+	0.03815	+	Star_Power	0.0294	+
14+	0.02526	+	Gênero_Documentário	0.02605	-	Gênero_Biografia	0.03177	-	Tipo_produtores_Medium	0.0294	-
Star_Power	0.02274	-	Livre	0.02479	+	Gênero_Comédia	0.03166	+	16+	0.02902	+
10+	0.02021	-	16+	0.01795	+	Livre	0.02932	+	Gênero_Comédia	0.02266	-
Gênero_Documentário	0.01156	-	Gênero_Comédia	0.01365	+	14+	0.02233	-	Gênero_Biografia	0.01798	-
Gênero_Biografia	0.0082	+	12+	0.01239	-	Gênero_Documentário	0.02153	+	Sequencia	0.01759	-
Sequencia	0.00784	-	Gênero_Outros	0.01103	+	Gênero_Drama	0.01898	-	18+	0.01352	-
Gênero_Comédia	0.00304	-	Sazonalidade	0.00802	+	16+	0.01791	-	Gênero_Outros	0.00874	-
Tempo_producao	0.00274	-	Sequencia	0.00653	+	Tempo_de_Projeção	0.0159	-	Tipo_produtores_Major	0.00561	+
12+	0.00176	+	Tipo_produtores_Minor	0.00212	+	Gênero_Outros	0.01494	+	Tipo_produtores_Minor	0.00525	+
16+	0.00142	-	10+	0.00193	+	Sequencia	0.01145	+	12+	0.00485	-
18+	0.0002	-	18+	0.00088	-	Tipo_produtores_Major	0.00816	-	Tempo_de_Projeção	0.00301	+

Tabela VI – Variáveis da Regressão Logística ordenadas por valor absoluto e classe

VARIÁVEL	E	INFLU- ÊNCIA	VARIÁVEL	F	INFLU- ÊNCIA	VARIÁVEL	G	INFLU- ÊNCIA	VARIÁVEL	H	INFLU- ÊNCIA
Tempo_producao	0.15833	-	Tempo_producao	0.17523	+	Tempo_producao	0.17653	+	Salas_lancamento	0.81069	-
14+	0.10019	+	Star_Power	0.09813	-	Salas_lancamento	0.11161	-	16+	0.09361	+
Livre	0.07037	-	12+	0.09253	+	16+	0.05408	-	Sazonalidade	0.06768	+
Gênero_Drama	0.06455	+	Tipo_produtores_Minor	0.03557	+	Sazonalidade	0.05029	-	12+	0.05691	-
Sazonalidade	0.06327	-	Livre	0.03487	-	Star_Power	0.03948	+	Gênero_Documentário	0.05528	+
Gênero_Documentário	0.04546	-	Salas_lancamento	0.03485	-	14+	0.03856	+	Star_Power	0.05497	-
Salas_lancamento	0.04478	+	10+	0.03432	+	12+	0.03143	+	Tempo_producao	0.04389	+
16+	0.03442	-	16+	0.03276	-	18+	0.03072	-	Tempo_de_Projeção	0.04387	+
Gênero_Outros	0.01943	-	Gênero_Documentário	0.0322	+	Tempo_de_Projeção	0.02251	+	14+	0.04128	-
Tempo_de_Projeção	0.01725	+	Gênero_Drama	0.02585	-	Gênero_Biografia	0.01744	-	Gênero_Outros	0.0331	+
Gênero_Comédia	0.01451	-	18+	0.02333	-	10+	0.01694	+	18+	0.0313	+
Tipo_produtores_Maior	0.01029	-	Tempo_de_Projeção	0.01854	+	Tipo_produtores_Minor	0.01391	+	Gênero_Drama	0.03079	-
10+	0.0102	-	14+	0.01112	-	Tipo_produtores_Maior	0.01214	-	Tipo_produtores_Maior	0.02689	+
Sequencia	0.00629	+	Sequencia	0.00975	+	Gênero_Documentário	0.00977	+	Gênero_Comédia	0.01887	-
Tipo_produtores_Medium	0.00478	-	Gênero_Comédia	0.00908	+	Gênero_Outros	0.00633	+	10+	0.01774	-
Gênero_Biografia	0.002	-	Tipo_produtores_Maior	0.00878	-	Sequencia	0.00575	-	Livre	0.01661	+
Tipo_produtores_Minor	0.00178	-	Sazonalidade	0.00862	-	Gênero_Comédia	0.00468	+	Gênero_Biografia	0.01314	-
Star_Power	0.00166	-	Gênero_Outros	0.00751	+	Gênero_Drama	0.00259	-	Sequencia	0.00285	-
12+	0.00127	-	Tipo_produtores_Medium	0.00202	-	Livre	0.00139	-	Tipo_produtores_Medium	0.00086	-
18+	0.0008	-	Gênero_Biografia	0.00183	+	Tipo_produtores_Medium	0.00102	-	Tipo_produtores_Minor	0.00044	-

Tabela VII – Classificação final das variáveis

VARIÁVEL	A	B	C	D	E	F	G	H	MEDIANA	MÉDIA	CLASSIFICAÇÃO FINAL
Salas_lancamento	1	1	1	1	7	6	2	1	1.0	2.5	1
Tempo_producao	17	6	6	2	1	1	1	7	4.0	5.1	2
Sazonalidade	7	16	4	3	5	17	4	3	4.5	7.4	3
Star_Power	11	8	5	9	18	2	5	6	7.0	8.0	4
14+	10	5	13	6	2	13	6	9	7.5	8.0	5
Gênero_Drama	8	2	15	4	4	10	18	12	9.0	9.1	6
Gênero_Documentário	13	10	14	8	6	9	14	5	9.5	9.9	7
16+	19	12	16	11	8	8	3	2	9.5	9.9	8
Tempo_de_Projeção	5	9	17	20	10	12	9	8	9.5	11.3	9
Livre	9	11	12	5	3	5	19	16	10.0	10.0	10
12+	18	14	7	19	19	3	7	4	10.5	11.4	11
10+	12	19	8	7	13	7	11	15	11.5	11.5	12
Tipo_produtoira_Medium	3	7	3	10	15	19	20	19	12.5	12.0	13
Tipo_produtoira_Major	4	4	20	17	12	16	13	13	13.0	12.4	14
18+	20	20	9	15	20	11	8	11	13.0	14.3	15
Gênero_Biografia	14	3	10	13	16	20	10	17	13.5	12.9	16
Gênero_Comédia	16	13	11	12	11	15	17	14	13.5	13.6	17
Tipo_produtoira_Minor	2	18	2	18	17	4	12	20	14.5	11.6	18
Gênero_Outros	6	15	18	16	9	18	15	10	15.0	13.4	19
Sequência	15	17	19	14	14	14	16	18	15.5	15.9	20



## Referências bibliográficas

**ANCINE.** *Anuário estatístico do cinema nacional brasileiro*. 2017. Disponível em: <[https://oca.ancine.gov.br/sites/default/files/repositorio/pdf/anuario\\_2017.pdf](https://oca.ancine.gov.br/sites/default/files/repositorio/pdf/anuario_2017.pdf)>. Acesso em: 4 jan. 2018.

**BASUROY, S.; CHATTERJEE, S.; RAVID, S. A.** How critical are critical reviews? The box office effects of film critics, star power, and budgets. *Journal of Marketing*, v. 67, n. 4, 2003, p. 103-117.

**BREWER, S. M.; KELLEY, J. M.; JOZEFOWICZ, J. J.** A blueprint for success in the US film industry. *Applied Economics*, v. 41, n. 5, 2009, p. 589-606.

**CHANG, B. H.; KI, E. J.** Devising a practical model for predicting theatrical movie success: focusing on the experience good property. *Journal of Media Economics*, v. 18, n. 4, 2005, p. 247-269.

**CHEN, R.; XU, W.; ZHANG, X.** Dynamic box office forecasting based on microblog data. *Filomat*, v. 30, n. 15, 2016, p. 4.111-4.124.

**CHISHOLM, D. C.; FERNÁNDEZ-BLANCO, V.; RAVID, S. A.; WALLS, W. D.** Economics of motion pictures: the state of the art. *Journal of Cultural Economics*, v. 39, n. 1, 2015, p. 1-13.

**CLEMENT, M.; WU, S.; FISCHER, M.** Empirical generalizations of demand and supply dynamics for movies. *International Journal of Research in Marketing*, v. 31, n. 2, 2014, p. 207-223.

**COLLINS, A.; HAND, C.; SNELL, M. C.** What makes a blockbuster? Economic analysis of film success in the United Kingdom. *Managerial and Decision Economics*, v. 23, n. 6, 2002, p. 343-354.

**CRAIG, C. S.; GREENE, W. H.; VERSACI, A.** E-word of mouth: early predictor of audience engagement: How pre-release “E-WOM” drives box-office outcomes of movies. *Journal of Advertising Research*, v. 55, n. 1, 2015, p. 62-72.

**DELEN, D.; SHARDA, R.** Predicting the financial success of Hollywood movies using an information fusion approach. *EndüstriMühendisliğiDergisi*, v. 21, n. 1, 2010, p. 30-37.

**DELMESTRI, G.; MONTANARI, F.; USAI, A.** Reputation and strength of ties in predicting commercial success and artistic merit of independents in the Italian feature film industry. *Journal of Management Studies*, v. 42, n. 5, 2005, p. 975-1.002.

**EINAV, L.** Seasonality in the U. S. motion picture industry. *The Rand Journal of Economics*, v. 38, n. 1, 2007, p. 127-145.

**GHIASSI, M.; LIO, D.; MOON, B.** Pre-production forecasting of movie revenues with a dynamic artificial neural network. *Expert Systems with Applications*, v. 42, n. 6, 2015, p. 3.176-3.193.

**GUO, Z.; ZHANG, X.; HOU, Y.** Predicting box office receipts of movies with pruned random forest. International Conference on Neural Information Processing. 22. 2015, Istanbul. Springer. 758p. Disponível em: <<https://www.springer.com/gp/book/9783319265544>>. Acesso em: 5 dez. 2018.

**HUR, M.; KANG, P.; CHO, S.** Box-office forecasting based on sentiments of movie reviews and Independent subspace method. *Information Sciences*, v. 372, 2016, p. 608-624.

**KANZLER, M.; MILLA, J.** Focus 2016 – world film market trends. [s.l.: s.n.].

**KARNIOUCHINA, E. V.** Impact of star and movie buzz on motion picture distribution and box office revenue. *International Journal of Research in Marketing*, v. 28, n. 1, 2011, p. 62-74.

**KIM, S. H.; PARK, N.; PARK, S. H.** Exploring the effects of online word of mouth and expert reviews on theatrical movies' box office success. *Journal of Media Economics*, v. 26, n. 2, 2013, p. 98-114.

**KIM, T.; HONG, J.; KANG, P.** Box office forecasting using machine learning algorithms based on SNS data. *International Journal of Forecasting*, v. 31, n. 2, 2015, p. 364-390.

**KIM, T.; HONG, J.; KANG, P.** Box office forecasting considering competitive environment and word-of-mouth in social networks: a case study of Korean film market. *Computational Intelligence and Neuroscience*, 2017.

**LEE, K.; PARK, J.; KIM, I.; CHOI, Y.** Predicting movie success with machine learning techniques: ways to improve accuracy. *Information Systems Frontiers*, v. 20, n. 3, 2018, p. 577-588.



**LEGOUX, R.; LAROCQUE, D.; LAPORTE, S.; BELMATI, S.; BOQUET, T.** The effect of critical reviews on exhibitors' decisions: do reviews affect the survival of a movie on screen? *International Journal of Research in Marketing*, v. 33, n. 2, 2016, p. 357-374.

**LIU, L.; ZHAO, Y.** Research of box-office prediction based on rough set and support vector machine. *International Journal of Hybrid Information Technology*, v. 9, n. 2, 2016, p. 417-426.

**MCKENZIE, J.; WALLS, W. D.** Australian films at the Australian box office: performance, distribution, and subsidies. *Journal of Cultural Economics*, v. 37, 2013, p. 247-269.

**MESSERLIN, P.; PARC, J.** The effect of screen quotas and subsidy regime on cultural industry: a case study of French and Korean film industries the effect of screen quotas and subsidy regime on cultural industry. *Journal of International Business and Economy*, v. 1.517, n. 2, 2014, p. 57-73.

**MICHEL, R. C.; AVELLAR, A. P.** Indústria cinematográfica brasileira de 1995 a 2012: Estrutura de mercado e políticas públicas. *Nova Economia*, v. 24, n. 3, 2014, p. 491-516, 2014.

**NELSON, R. A.; GLOTFELTY, R.** Movie stars and box office revenues: an empirical analysis. *Journal of Cultural Economics*, v. 36, 2012, p. 141-166.

**PACKARD, G.; ARIBARG, A.; ELIASHBERG, J.; FOUTZ, N. Z.** The role of network embeddedness in film success. *International Journal of Research in Marketing*, v. 33, n. 2, 2016, p. 328-342.

**RASCHKA, S.; MIRJALILI, V.** Python Machine Learning. 2. ed., 2017.

**SAWHNEY, M. S.; ELIASHBERG, J.** A parsimonious model for forecasting gross box-office revenues of motion pictures. *Marketing Science*, v. 15, n. 2, 1996, p. 113.

**SHARDA, R.; DELEN, D.** Predicting box-office success of motion pictures with neural networks. *Expert Systems with Applications*, v. 30, n. 2, 2006, p. 243-254.

**TERRY, N.; BUTLER, M.; ARMOND, D. A. DE.** Determinants of the box office performance of motion pictures. *Proceedings of the Academy of Marketing Studies*, v. 8, n. 2, 2003, p. 23-28.

**TREME, J.** Effects of celebrity media exposure on box-office performance. *Journal of Media Economics*, v. 23, n. 1, 2010, p. 5-16.

**WALLS, W. D.** Modeling movie success when “nobody knows anything”: conditional stable-distribution analysis of film returns. *Journal of Cultural Economics*, v. 29, n. 3, 2005, p. 177-190.

**ZHANG, L.; LUO, J.; YANG, S.** Forecasting box office revenue of movies with BP neural network. *Expert Systems with Applications*, v. 36, n. 3 (II), 2009, p. 6.580-6.587.



