

CIÊNCIA DE DADOS APLICADA A CINEMA DE PEQUENO PORTE

Bruno Belluzzo-Bacharelado em Ciência de Dados
Orientador: João Pedro Albino

RESUMO

Apesar do grande impacto na economia mundial e do sucesso que são os cinemas ao redor do mundo, com o crescimento de serviços de vídeo online as pessoas, cada vez mais, estão deixando de ir ao cinema, pois, eventualmente, o filme estará disponível online. Em meio a toda essa revolução cinematográfica, cinemas de pequeno porte, que possuem uma única sala para exibição de filmes, sofrem com a disponibilidade de cópias, pois os estúdios e suas distribuidoras dão preferência às grandes redes de cinemas, pois estas geram maior receita. Procurando entender as limitações desses pequenos cinemas, este projeto busca utilizar técnicas de ciência de dados e *machine learning* para levantar dados por meio de um *survey* sobre os frequentadores do cinema com o intuito de compreender o *perfil desses clientes* e buscar sugerir soluções inteligentes com o intuito de maximizar os lucros do estabelecimento, e melhorando o atendimento e a satisfação deste público. A pesquisa (*survey*) será realizada utilizando um formulário de pesquisa disponível na internet com os usuários do *Cine Belluzzo*, um cinema de pequeno porte situado na cidade de Bariri, interior do Estado de São Paulo, com cerca de 35 mil habitantes.

1. INTRODUÇÃO

Data Science ou **Ciência de Dados** é uma área que já existe há mais de 30 anos, mas vem ganhando destaque nos últimos anos, devido ao **Big Data**. Segundo Chen, Mao e Liu (2014), o termo *big data* refere-se a *um grande conjunto de dados gerados e armazenados* com os quais os aplicativos de processamento de dados tradicionais não conseguem lidar em um tempo tolerável. O surgimento do termo está relacionado com o aumento exponencial da quantidade de dados gerados a cada minuto no mundo. Ainda segundo os autores, o *big data* representou uma nova era na sociedade moderna, em que *os dados se tornaram cada vez mais valiosos*, mudando a forma como a economia e a ciência observam os processos, extraem e geram valor desse **caos** de dados.

De acordo com Silveira (2016), ciência de dados (em inglês: *data science*) pode ser definida como uma área interdisciplinar voltada para o **estudo e análise de dados**, estejam estes dados estruturados ou não, e que busca a extração de conhecimento ou *insights*¹ para possíveis tomadas de decisão. Ainda de acordo com Silveira (2016), a ciência de dados pode transformar *uma grande quantidade de dados brutos (big data)* em inteligência de negócios, e com isso, auxiliar organizações em tomadas de decisões para que as mesmas obtenham melhores resultados.

O desenvolvimento de áreas como *machine learning* - uma área da ciência da computação que significa "aprendizado da máquina" e que por sua vez embute o conceito de

¹ Insight é um substantivo do idioma inglês e que significa compreensão súbita de alguma coisa ou determinada situação. Está também está relacionado com a capacidade de discernimento, indicando um momento único de esclarecimento. Pode também ser entendido como um acontecimento cognitivo associado a vários fenômenos podendo ser sinônimo de compreensão, conhecimento, intuição ou a capacidade de apreender alguma coisa e acontece quando uma solução surge de forma repentina. Fonte: <https://www.significados.com.br/insight/>.

inteligência artificial, que estuda meios para que os computadores possam desenvolver tarefas que seriam executadas por pessoas - reforçam o crescimento e a importância da Ciência de Dados sendo que é cada vez mais comum as organizações de qualquer ramo da atividade econômica se beneficiarem do estudo para a tomada de decisões, de forma a alavancar os seus crescimentos (SILVEIRA, 2016).

Trabalhos publicados entre 2010 e 2018 atestam, de acordo com o INSTITUTO ATLANTICO (2018), que *dados são o novo petróleo e o grande recurso desta década*. Nos últimos dez anos (de 1997 a 2017), o mundo gerou uma quantidade muito grande de informação, e as organizações têm interesse nesse recurso como forma de aprimorar os seus serviços (INSTITUTO ATLANTICO, 2018). Os mais diferentes tipos de negócios podem utilizar os benefícios da Ciência de Dados para alavancar seus negócios, e isto também não seria diferente com as empresas do ramo do entretenimento, mais especificamente, como discutido neste trabalho, do ramo cinematográfico.

A indústria cinematográfica mundial gera uma receita bilionária todos os anos. Segundo IstoÉ (2018), os filmes registraram, em 2018, um lucro recorde de 96,8 bilhões de dólares em todo o mundo, incluindo mais de 40 bilhões apenas em bilheterias, de acordo com números publicados pela *Motion Picture Association of America* (MPAA).

No Brasil, números levantados pela *Motion Picture Association na América Latina* (MPA-AL) em parceria com o *Sindicato Interestadual da Indústria do Audiovisual*, mostram que a indústria cinematográfica injetou em 2018 mais de R\$19 bilhões na economia brasileira. (ISTOÉ, 2018).

Apesar do grande impacto na economia mundial e do sucesso dos cinemas ao redor do mundo, com o crescimento de serviços de vídeo online as pessoas estão cada vez mais deixando de ir ao cinema, pois "uma hora ou outra" o filme estará disponível online. Segundo Exame (2017), na sétima edição da CinemaCon, o principal evento de exibição cinematográfica dos Estados Unidos, em 2017, teve como tema informal o futuro do negócio da exibição cinematográfica na era do streaming².

Muitos estúdios já começaram versões *online* de seus canais, afirma IstoÉ (2018), contudo é preciso balancear as coisas com os donos das salas de cinema, que sempre foram fiéis aliados dos grandes estúdios cinematográficos.

Em meio a toda essa revolução cinematográfica, cinemas de pequeno porte, que possuem uma única sala para exibição de filmes, sofrem com a disponibilidade de cópias, pois os estúdios e suas distribuidoras dão preferência aos grandes cinemas, que irão gerar uma maior receita.

Com isso em mente e buscando entender as limitações desses pequenos cinemas, este projeto pretende discutir e apresentar como técnicas de ciência de dados e *machine learning* podem ser utilizadas para maximizar os lucros e melhorar a satisfação dos clientes.

A pesquisa será realizada utilizando um formulário de pesquisa (*survey*) disponível na internet com os usuários do *Cine Belluzzo*, um cinema de pequeno porte situado na cidade de Bariri, interior do Estado de São Paulo, com cerca de 35 mil habitantes.

² **Streaming** significa ouvir música ou assistir a um vídeo em 'tempo real', em vez de baixar um arquivo no computador e vê-lo mais tarde. Com vídeos na Internet e webcasts de eventos ao vivo, não há nenhum arquivo para download, apenas um fluxo contínuo de dados. Fonte: <https://www.palpitedigital.com/que-streaming-quais-melhores-audio-video/>.

2. PROBLEMA

As limitações físicas e financeiras de um cinema de pequeno porte, assim como o número de habitantes da cidade em que normalmente estas salas estão localizadas, dificultam a possibilidade de se obter maiores ganhos. As distribuidoras de filmes ficam com uma parcela da renda da venda de ingressos, logo é mais vantajoso para estas organizações que utilizem redes de cinema de grande porte, que possuem um público muito maior e que trarão maior faturamento.

Além disso, pelo fato de as distribuidoras possuírem cópias limitadas dos filmes, os pequenos cinemas sofrem também uma defasagem em relação aos cinemas de grande porte, exibindo filmes com atraso de uma ou duas semanas, além de, por geralmente possuírem apenas uma sala para exibição, terem um limite de dois e no máximo três filmes sendo exibidos simultaneamente.

Com essas restrições em ênfase, a utilização da ciência de dados pode ser extremamente benéfica para compreender as preferências e os gostos das pessoas que frequentam os cinemas de pequeno porte, e assim apoiá-los a tomar melhores decisões no que tange a escolha de filmes, criação de promoções, combos na compra de produtos, entre outras escolhas. Os algoritmos de ciência de dados podem também auxiliar na busca do entendimento dos motivos pelos quais as pessoas não irem ao cinema, elaborando soluções inteligentes para atrair essas pessoas para frente da "grande tela".

3. JUSTIFICATIVA

O conhecimento obtido a partir do uso de Data Science é fundamental para apoiar o processo de tomada de decisão. A informação gerada pelas aplicações informáticas disponibiliza aos gestores um conjunto de indicadores sobre o negócio, que lhe dão indicações do que aconteceu no passado e lhe permitem traçar cenários para o futuro (SANTOS; RAMOS, 2006).

“As empresas que possuem ferramentas de *business intelligence* e a utilizam em seus processos durante uma tomada de decisão, apresentam vantagem, para se posicionarem a frente de uma nova oportunidade no mercado” (SILVA; TERRA, 2015, p. 11).

O termo *inteligência de negócios* (ou *Business Intelligence*, em inglês) refere-se ao processo de coleta, organização, análise, compartilhamento e monitoramento de informações que oferecem suporte a gestão de negócios, afirma Rud (2009). É um conjunto de técnicas e ferramentas para auxiliar na transformação de dados brutos em informações significativas e úteis a fim de analisar o negócio. Identificando novas oportunidades e implementando uma estratégia efetiva baseada nos dados, também pode promover negócios com vantagem competitiva no mercado e estabilidade a longo prazo. No ramo do entretenimento esse conhecimento pode ser de extrema importância para atrair a maior quantidade possível de pessoas ao estabelecimento, assim como elaborar as melhores estratégias para garantir o maior lucro possível.

Com as limitações dos cinemas de pequeno porte citadas anteriormente, as tomadas de decisões devem ser muito bem pensadas antes de serem executadas, e possuindo as informações coletadas por meio de levantamento dos dados, bem como experiências do passado, é possível realizar as melhores escolhas para evitar ao máximo a perda de patrimônio.

4. OBJETIVOS

4.1. Objetivo geral

Utilizar ciência de dados e seus ferramentais, tais como *machine learning* e *estatística*, para compreender os padrões dos clientes de um cinema de pequeno porte e, com base

nos dados levantados, buscar as melhores decisões e criar soluções inteligentes com o intuito de maximizar os lucros do estabelecimento.

4.2. Objetivos específicos

- Levantar dados por meio de uma pesquisa sobre os frequentadores do cinema
- Estudar os dados levantados e gerar gráficos para compreender o perfil dos clientes
- Elaborar soluções inteligente com base nos dados coletados
- Aplicar as soluções em um cinema de pequeno porte
- Avaliar o desempenho das receitas do cinema com as soluções aplicadas.

5. MÉTODO DE PESQUISA

Para a pesquisa e desenvolvimento do projeto serão necessárias várias etapas. A primeira etapa consiste em estudar as bibliotecas disponíveis para a linguagem Python e suas ferramentas para o uso em aplicações de Ciência de Dados, tais como: *Matplotlib* e *Seaborn* para visualização dos dados e *Pandas* para manipulação e análise dos dados. Posteriormente será necessário elaborar um formulário e coletar as respostas das pessoas da cidade de Bariri, interior do Estado de São Paulo, e usuários do cinema usado como base para este **estudo de caso**.

Segundo Yin (1993), **caso de estudo** ou **estudo de caso** são expressões sinônimas que designam um método da abordagem de pesquisa e que consiste na utilização de um ou mais métodos **qualitativos** de coleta de informação. O estudo de caso caracteriza-se por descrever um evento ou caso de uma forma longitudinal. O **caso** consiste geralmente no estudo aprofundado de uma **unidade individual**, tal como: uma pessoa, um grupo de pessoas, **uma instituição**, um evento cultural, etc.

O caso de estudo deste trabalho é o *Cine Belluzzo*, um cinema de pequeno porte situado na cidade de Bariri, no estado de São Paulo, com cerca de 35 mil habitantes³.

O método de pesquisa utilizado neste trabalho será o **survey**. Segundo Babbie (2001) e Freitas et. ali (2000), *survey* é um método de pesquisa social que utiliza técnicas estatísticas e que normalmente implica na construção de investigações por meio de questionários. Os pesquisadores podem utilizar a informação por este método para desenhar estratégias e estabelecer planos de ação.

Segundo Babbie (2001), as principais características da pesquisa do tipo survey são:

- Sempre que o pesquisador procura explicar as razões para e as fontes de eventos, características e correlações observados, a pesquisa deve assumir uma postura *determinística* baseada nas *relações de causa e efeito*;
- Surveys amostrais são realizados para entender-se a população maior da qual a amostra foi inicialmente selecionada. Análises explicativas em pesquisas de survey visam a desenvolver proposições gerais sobre o comportamento humano;
- Replicar um achado entre subgrupos diferentes (por sexo, por idade, etc.) fortalece a certeza de que ele representa um fenômeno geral na sociedade;
- Com a pesquisa, busca-se o máximo de compreensão com o menor número de variáveis possível;
- A conceituação e a medição de variáveis estão no âmago da prática científica. Se as variáveis não são adequadamente conceituadas e medidas, correlações observadas entre elas podem não fazer sentido. Assim, se observadores independentes discordam sobre como variáveis foram medidas, podem discordar logicamente da conclusão geral;

³ IBGE Censo [2010]. Fonte: <https://cidades.ibge.gov.br/brasil/sp/bariri/panorama>.

Para a criação e distribuição do formulário de pesquisa será utilizado o aplicativo *Google Forms*. O *Google Forms* é uma ferramenta que permite coletar informações dos usuários por meio de um *survey* (pesquisa) ou questionário personalizado. As informações são coletadas e automaticamente armazenadas em uma planilha. A planilha é automaticamente preenchida com as respostas da pesquisa ou do questionário.

Com as respostas em mãos, serão utilizadas as bibliotecas Python citadas anteriormente para fazer uma análise detalhada dos dados e seus padrões, para enfim inferir as melhores estratégias e soluções para a organização e em seguida, com as estratégias estabelecidas, aplica-las na prática.

Caso sejam constatadas situações em que inteligência artificial possa realizar o trabalho de tomada de decisões, como, por exemplo, a escolha do melhor filme entre os disponíveis pelas distribuidoras, será realizado também um estudo sobre *Machine Learning* (aprendizado de máquina) e alguns de seus algoritmos, como árvores de decisão e classificação *Naive Bayes*.

Para a implementação de IA serão utilizadas as bibliotecas *TensorFlow*, *SciKit Learn* e *Keras*, disponíveis em Python para *Machine Learning*.

Por fim será realizada outra pesquisa para verificarmos se, com as modificações e soluções aplicadas, a satisfação dos clientes e o lucro do cinema tiveram melhora, comparando com os dados anteriores à realização da primeira pesquisa.

Python foi escolhida por ser uma das linguagens mais utilizadas para pesquisa, conter ferramentas que facilitam o estudo e a implementação de Ciência de Dados e Machine Learning, e possuir uma comunidade muito ativa (BABU L, 2019).

O equipamento usado para a pesquisa será um notebook com sistema operacional Microsoft Windows e placa de vídeo NVIDIA.

Para o estudo e o desenvolvimento dos modelos estatísticos, a planilha do Google Docs deverá ser lida pela linguagem Python.

5.1. Cronograma Proposto

Identificação da Etapa	Início (dd/mm/aaaa)	Término (dd/mm/aaaa)
Análise dos dados obtidos e levantamento de soluções.	01/09/2019	30/09/2019
Verificação e análise do teste piloto. Revisão do survey.	08/07/2019	10/07/2019
Elaboração do formulário	06/05/2019	28/06/2019
Elaboração do formulário web final e coleta eletrônica das respostas	10/07/2019	30/08/2019
Revisão bibliográfica	06/05/2019	31/05/2019
Estudo das bibliotecas e ferramentas disponíveis em Python	06/05/2019	31/05/2019
Teste piloto	28/06/2019	05/07/2019
Realização da segunda pesquisa e comparação com os dados anteriores	01/11/2019	22/11/2019
Redação do TCC.	16/07/2019	01/11/2019
Aplicação das soluções no cinema.	01/10/2019	31/10/2019

REFERÊNCIAS

BABBIE, E. Métodos de Pesquisas de Survey, Belo Horizonte, Editora UFMG, 2001.

BABU L, A. Top 8 programming languages every data scientist should master in 2019, Big Data Made Simple, Jan 24, 2019. Disponível em: <<https://bigdata->

madesimple.com/top-8-programming-languages-every-data-scientist-should-master-in-2019/>. Acesso em: 03 de fevereiro de 2019.

CHEN, M.; MAO, S.; LIU, Y. Big Data: A Survey, Mobile Networks and Applications, 19 (2). 2014. pp: 171–209.

EXAME, Cinema vs. Netflix, o combate. Disponível em: <<https://exame.abril.com.br/estilo-de-vida/cinema-vs-netflix-o-combate/>>. Acesso em 01 de maio de 2019.

FREITAS, H., OLIVEIRA, M., SACCOL, A. Z., & MOSCAROLA, J. O método de pesquisa Survey. Revista de Administração da Universidade de São Paulo, 35(3), 2000.

ISTOÉ, Indústria cinematográfica registra lucro recorde em 2018. Disponível em: <<https://www.istoedinheiro.com.br/industria-cinematografica-registra-lucro-recorde-em-2018/>>. Acesso em 02 de abril de 2019.

INSTITUTO ATLÂNTICO. Data Science: entenda a importância dos dados para sua empresa. 2018. Disponível em: <<https://blog.atlantico.com.br/data-science-entenda-a-importancia-dos-dados-para-sua-empresa/>>. Acesso em: 17 de março de 2019.

RUD, O. P. Business Intelligence Success Factors: Tools for Aligning Your Business in the Global Economy. Hoboken, N.J: Wiley & Sons, 2009.

SANTOS, M. Y.; RAMOS, I. Business Intelligence: Tecnologias da informação na gestão de conhecimento. FCA Editora de Informática, Ltda., 2006.

SILVA, V. C. L.; TERRA, L. A. A. Business intelligence como fator decisivo na competitividade empresarial: Uma análise a partir de multi casos. Revista Inteligência Competitiva, v. 5, n. 1, p. 1-13, 2015.

SILVEIRA, D. P. O que é Data Science? 2016. Disponível em: <<https://www.oficinadanet.com.br/post/16919-o-que-e-data-science>>. Acesso em: 17 de março de 2019.

YIN, R. K. Pesquisa Estudo de Caso - Desenho e Métodos, 2 ed., Porto Alegre, Bookman, 1994.