

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

Lizandra Mistrello Funari

**MOTIVAÇÕES E MÉTODOS DE PREVISÃO DE
DEMANDA APLICADOS A EVENTOS**

Porto Alegre

2021

Lizandra Mistrello Funari

Motivações e métodos de previsão de demanda aplicados a eventos

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul como requisito parcial à obtenção do título de Mestre em Engenharia de Produção, modalidade Acadêmica, na área de concentração em Sistemas de Produção.

Orientador: Professor José Luis Duarte Ribeiro, Dr.

Porto Alegre

2021

Lizandra Mistrello Funari

Motivações e métodos de previsão de demanda aplicados a eventos

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Engenharia de Produção na modalidade Acadêmica e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul.

Prof. Orientador José Luis Duarte Ribeiro, Dr.
Orientador PPGEP/UFRGS

Prof. Alejandro Germán Frank
Coordenador PPGEP/UFRGS

Banca Examinadora:

Professor Michel Jose Anzanello, *Ph.D.* (PPGEP/UFRGS)

Professor Vinícius Andrade Brei, *Ph.D.* (Escola de Administração/UFRGS)

Professor Jonatas Ost Scherer, Dr.. (Campus Litoral Norte/UFRGS)

Dedico esta dissertação à minha família,
em especial, aos meus pais Luiz e Lisete,
sem eles nada disso seria possível.

AGRADECIMENTOS

Essa dissertação só foi possível devido ao apoio de muitos. Primeiramente, agradeço aos meus pais (Luiz e Lisete), sem os quais nada seria possível. Agradeço por sempre estarem do meu lado, me incentivarem a sempre buscar mais conhecimento e a crescer como pessoa e profissional. Vocês são a base de tudo e é por vocês que hoje concluo mais essa etapa.

Agradeço também a uma das pessoas que mais admiro, meu orientador Prof. Dr. José Luis Duarte Ribeiro. Agradeço por toda a compreensão durante o período de mestrado e toda a ajuda que me deste, mesmo quando eu não soube buscar. Tive a honra e o privilégio de ter por perto uma das pessoas mais humildes e mais inteligentes, que muito me ensinou e me guiou durante essa trajetória.

A todos os excelentes professores que tive o prazer de ter em minha vida acadêmica, mesmo que por um período curto. Principalmente àqueles que fazem parte deste programa de pós-graduação e que, de alguma forma, me incentivaram a continuar. Tenho certeza que sou uma profissional melhor por cursar meu mestrado aqui.

A todos os meus amigos (do trabalho, da UFRGS e da vida) e familiares, que dividiram comigo, não apenas os momentos felizes, mas também as preocupações, reclamações e angústias durante esse período. Sem vocês teria sido muito difícil chegar até aqui.

Por fim, à UFRGS e a todos aqueles que a tornam esse lugar incrível onde o conhecimento é a base de tudo e onde é possível sonhar em chegar mais longe.

RESUMO

Com a popularização do entretenimento, os eventos têm se tornado uma área de interesse crescente. Nesse contexto, a previsão de público é uma atividade importante para fins de planejamento e gestão do evento. Para compreender as motivações que conduzem ao comparecimento, esta dissertação, através de uma revisão sistemática da literatura, propõe-se a responder a duas questões de pesquisa: (i) quais são os fatores e motivações que interferem na decisão de participar de eventos; (ii) quais métodos de previsão podem ser utilizados para estimar o público em eventos. A partir de 56 artigos selecionados na literatura, foram identificados 47 fatores motivacionais relacionados ao comparecimento a eventos, organizados em 9 dimensões. Além disso, foram observados 24 métodos para prever e classificar a participação em eventos. Entre esses métodos, o mais utilizado foi a Regressão Logística e a Regressão Linear, geralmente associados à aplicação de *Machine Learning* em bases de dados obtidas através de redes sociais. Posteriormente, os fatores identificados foram organizados e apresentados em um *framework* que pode auxiliar os organizadores a prever o público em diferentes tipos eventos. O *framework* foi avaliado em duas simulações de eventos reais por uma profissional da área, onde foi possível avaliar a sua aplicabilidade e adaptabilidade. Por fim, essa dissertação aprofunda os conhecimentos de previsão de público com o uso de nove métodos de *Machine Learning* para a previsão de público total e pagante em jogos de futebol, proporcionando comparação de desempenho e identificação dos melhores modelos. Dados de jogos reais foram coletados em parceria com uma organização de futebol brasileira. Métodos de *ensemble* baseados em árvore de decisão foram avaliados e também redes neurais artificiais (ANN), modelo linear (LM) e *K-Nearest Neighbors* (KNN). Os resultados indicaram que os métodos de *ensemble* baseados em árvore de decisão são totalmente capazes de capturar os padrões de comportamento em relação ao comparecimento às partidas. Os modelos XGB e GBM foram os de melhor desempenho, apresentando os menores erros.

Palavras-chave: Métodos de previsão; Fatores motivacionais; Participação em eventos; Previsão de público; Gerenciamento de eventos; *Machine Learning*.

ABSTRACT

With the popularization of entertainment, events have become an area of growing interest. In this context, audience prediction is an important part of event planning and management. To understand the motivations that lead to attendance, this dissertation, through a systematic review of the literature, answers the following two research questions: (i) what are the motivational factors that interfere in the decision to participate in events and (ii) which forecasting methods can be used to estimate the audience at events. From 56 articles selected in the literature, 47 motivational factors related to attendance at events were identified, organized into 9 dimensions. In addition, 24 different methods were used to predict or classify participation in events. Among these methods, the most used was Logistic Regression and Linear Regression, generally associated with the application of *Machine Learning* in databases obtained through social networks. Subsequently, the identified factors were organized and presented in a framework that can help organizers to predict the public at different types of events. The framework was evaluated through two simulations of real events by a professional, where it was possible to evaluate its applicability and adaptability. Finally, this dissertation expands the knowledge of audience prediction with the use of nine *Machine Learning* algorithms for predicting the total and paying audience in football games, providing performance comparison and identification of best models. Real game data was collected in partnership with a Brazilian football team. Ensemble methods based on decision trees were evaluated and also artificial neural networks (ANN), linear model (LM) and K-Nearest Neighbors (KNN). The results indicated that the decision tree-based ensemble methods are fully capable of capturing behavior patterns in relation to attendance. The XGB and GBM models were the best performing, presenting the smallest forecasting errors.

Key words: Forecasting methods; Motivational factors; Event attendance; Audience prediction; Event management; *Machine Learning*.

LISTA DE FIGURAS

Figura 1 - Relação entre os artigos e os objetivos específicos da dissertação.....	18
Figura 2 - Combinação final de <i>query strings</i> para a busca nas bases de pesquisa.....	25
Figura 3 - Diagrama do processo de seleção de artigos	27
Figura 4 - Periódicos com as maiores quantidades de publicações.....	28
Figura 5 - Evolução temporal das publicações analisadas	28
Figura 6 - Tipos de eventos analisados encontrados na literatura.....	29
Figura 7 - Métodos de previsão de público encontrados na literatura.....	48
Figura 8 - Métodos de previsão associados ao tipo de evento	49
Figura 9 - <i>Framework</i> proposto para previsão de público em eventos.	63
Figura 10 - Avaliação dos fatores para edição do evento Gindulgence em Auckland.....	69
Figura 11 - Avaliação dos fatores para edição do evento Beervana em Auckland	70
Figura 12 - <i>Boxplot</i> e distribuição das variáveis resposta do banco de dados.....	83
Figura 13 - Relação entre o Público Total e as variáveis Campeonato e Atratividade do Adversário	84
Figura 14 - Relação entre (a) Público Total e as variáveis Campeonato e Fase e (b) Público Total e as variáveis Dia da Semana e Campeonato	85
Figura 15 - Diferentes curvas de aprendizagem: (a) bom ajuste do modelo (b) modelo de <i>overfitting</i> (c) modelo de <i>underfitting</i> (Adaptado de Géron, 2017)	89
Figura 16 - Procedimento experimental detalhado.....	90
Figura 17 - Comparação da distribuição dos erros de previsão de público total de cada algoritmo.....	92
Figura 18 - Comparação entre os valores previstos e valores reais dos diferentes métodos para público total.....	93
Figura 19 - Comparação da distribuição dos erros de previsão de público pagante de cada algoritmo.....	94
Figura 20 - Curvas de aprendizados dos diferentes modelos	95
Figura 21 - Comparação entre os valores previstos e valores reais dos diferentes métodos para público pagante.....	96

LISTA DE TABELAS

Tabela 1 – Resumo dos fatores encontrados na literatura e proposta de organização em perguntas para utilização de público em eventos	41
Tabela 2 – Resumo dos fatores encontrados na literatura e proposta de organização em perguntas para utilização de público em eventos (Continuação).	42
Tabela 3 - Métodos de previsão de público encontrados na literatura	47
Tabela 4 - Tabela resumo dos fatores motivacionais (Adaptado de Funari e Ribeiro, 2019).	62
Tabela 5 - Resumo dos resultados obtidos para o evento Gindulgence	67
Tabela 6 - Resumo dos resultados obtidos para o evento Beervana.....	68
Tabela 7 - Rótulos e descrição dos recursos do banco de dados obtido.	81
Tabela 8 - Performance dos modelos e métricas de erros para prever público total	92
Tabela 9 - Performance dos modelos e métricas de erros para prever público pagante.....	94
Tabela 10 - Variáveis mais importantes para os modelos XGB.....	98
Tabela 11 - Variáveis mais importantes para os modelos GBM	99

SUMÁRIO

1	INTRODUÇÃO	12
1.1	TEMA E OBJETIVOS	14
1.2	JUSTIFICATIVA DO TEMA E OBJETIVOS	14
1.3	MÉTODOS	15
1.4	LIMITAÇÕES DO ESTUDO	16
1.5	ESTRUTURA DA DISSERTAÇÃO	17
1.6	REFERÊNCIAS	18
2	ARTIGO 1 - FATORES MOTIVACIONAIS E MÉTODOS DE PREVISÃO APLICADOS A EVENTOS.....	20
2.1	INTRODUÇÃO	20
2.2	METODOLOGIA	22
2.3	SÍNTESE DOS RESULTADOS	27
2.3.1	MOTIVAÇÕES PARA O COMPARECIMENTO A EVENTOS	30
2.3.2	MÉTODOS DE PREVISÃO DE PÚBLICO EM EVENTOS.....	37
2.4	ANÁLISE DOS RESULTADOS OBTIDOS NA LITERATURA.....	39
2.4.1	FATORES QUE MOTIVAM A PARTICIPAÇÃO EM EVENTOS.....	39
2.4.2	MÉTODOS PARA PREVISÃO DE PÚBLICO EM EVENTOS	46
2.5	DISCUSSÃO.....	49
2.6	CONCLUSÃO	50
2.7	REFERÊNCIAS	52
3	ARTIGO 2 - MÉTODO DE PREVISÃO DE PÚBLICO EM EVENTOS BASEADO EM FATORES MOTIVACIONAIS.....	58
3.1	INTRODUÇÃO.....	58
3.2	REVISÃO DA LITERATURA	59
3.2.1	MÉTODOS UTILIZADOS PARA PREVISÃO DE PÚBLICO.....	60
3.2.2	FATORES MOTIVACIONAIS.....	61
3.3	RESULTADOS.....	63
3.3.1	ESTRUTURA DE PREVISÃO DE PÚBLICO EM EVENTOS	63
3.3.2	APLICAÇÃO DO MÉTODO DE PREVISÃO	65
3.4	DISCUSSÃO.....	71
3.5	CONSIDERAÇÕES FINAIS	72
3.6	REFERÊNCIAS	73
4	ARTIGO 3 - PREVISÃO DE PÚBLICO EM EVENTOS ESPORTIVOS COM O USO DE <i>MACHINE LEARNING</i>	75

4.1	INTRODUÇÃO	75
4.2	TRABALHOS RELACIONADOS	78
4.3	METODOLOGIA	79
4.3.1	FORMULAÇÃO DO PROBLEMA	80
4.3.2	COLETA DE DADOS	80
4.3.3	ANÁLISE DOS DADOS E PRÉ-PROCESSAMENTO	80
4.3.4	MÉTODOS DE REGRESSÃO	85
4.3.4.1	MODELO DE REGRESSÃO LINEAR	85
4.3.4.2	MODELO <i>K-NEAREST NEIGHBORS</i>	86
4.3.4.3	MODELOS BASEADOS EM ÁRVORE DE DECISÃO	86
4.3.4.4	MODELO DE REDES NEURAI	87
4.3.5	CRITÉRIOS DE AVALIAÇÃO E AJUSTES DOS HIPERPARÂMETROS	87
4.3.6	PROCEDIMENTO EXPERIMENTAL	89
4.4	TREINAMENTO E TESTE DOS MODELOS	91
4.5	DISCUSSÃO	97
4.6	CONCLUSÃO	99
4.7	MATERIAL COMPLEMENTAR	101
4.8	REFERÊNCIAS	101
5	CONSIDERAÇÕES FINAIS	103
5.1	CONCLUSÕES	103
5.2	SUGESTÕES PARA PESQUISAS FUTURAS	104

1 INTRODUÇÃO

Os métodos de previsão de demanda têm atraído atenção de diversas áreas do conhecimento, uma vez que permitem estimar valores futuros com base em valores passados, dentro de uma margem aceitável de erro (Tealab, 2018). Com os avanços nesta área, principalmente na manufatura, essas técnicas têm sido adaptadas para outras aplicações, como serviços, finanças e economia. Por exemplo, a demanda esportiva tem sido estudada por economistas, com implicações importantes na determinação de políticas de vendas e gerenciamento desses eventos. Além deles, Serwylo et al. (2014) aplicaram *Machine Learning* para previsão de ocorrências médicas em jogos esportivos e Jun et al. (2018) utilizaram métodos de *Machine Learning* para previsão de demanda em turismo.

O comportamento dinâmico da maioria das séries temporais observadas na prática apresenta o desafio de prever seus valores futuros com precisão e com metodologias lineares (Tealab, 2018). Sendo assim, vários métodos vêm sendo desenvolvidos com a intenção de capturar a realidade e os termos não lineares. As metodologias de inteligência computacional têm ganhado espaço dentro desse contexto, devido aos avanços tecnológicos dos últimos anos. Comparado com os métodos de regressão tradicionais, Tealab (2018) concluiu que o uso de métodos de inteligência computacional levou a uma maior precisão em seus resultados, sendo preferido aos métodos de regressão tradicionais.

O aprendizado de máquina, ou mais comumente conhecido como *Machine Learning*, nasceu na computação e é caracterizado pelo estudo de métodos e modelos estatísticos com o uso de sistemas computacionais, que aprendem uma tarefa de forma eficaz a partir de padrões (Bishop, 2006). Pela sua facilidade e praticidade, modelos com o uso de *Machine Learning* vem ganhando espaço em diferentes aplicações na literatura e na prática empresarial. Os modelos obtidos encontram uma combinação de parâmetros a fim de minimizar os erros de classificação ou regressão, aprendendo com os erros dos dados conforme os mesmos são apresentados ao computador (Géron, 2017).

Por outro lado, o interesse em estudar a demanda em festivais e eventos está crescendo, uma vez que possuem a habilidade de preservar, exibir e divulgar a cultura de um local. Existem vários motivos para o aumento desse interesse, relacionados à utilização da cultura como regeneração urbana e planos de desenvolvimento sociais (Devesa et al., 2015). A gestão de eventos, por exemplo, é um tema de crescente interesse no campo da pesquisa acadêmica sobre

turismo, devido aos benefícios significativos que o turismo de eventos pode trazer para a região (Kim e Kaewnuch, 2018). Os eventos e festivais são percebidos como fonte de riqueza e emprego, por atrair receitas relacionadas ao turismo cultural, que podem ajudar a desenvolver a economia de uma região. Eles também contribuem para melhorar a imagem de uma área, tornando-a competitiva através uma estratégia de diferenciação cultural. Além disso, Devesa et al. (2015) evidenciam que a cultura e os festivais também podem gerar benefícios sociais ligados à melhoria do bem-estar dos cidadãos e estabelecimento de relações entre pessoas.

O planejamento eficiente de eventos deve considerar traços inerentes de imprevisibilidade e os riscos do ambiente no qual estão inseridos, cujas características econômicas, ambientais, políticas, tecnológicas e sociais estão em constante mudança (Berridge, 2010; Adema e Roehl, 2010). Conhecer a quantidade de pessoas que participarão em eventos sociais é um tópico essencial para a sua organização, espera-se que as pessoas compareçam a eventos para socializar, trocar informações e expandir sua rede durante esses eventos (Li, 2018). O planejamento de grandes eventos é arriscado e demorado, com muitos detalhes a serem considerados, como dimensionamento de alimentação, pessoal e seleção do local.

Desta maneira, conhecer e entender as motivações que levam às pessoas a comparecerem aos eventos pode trazer benefícios para a eficiência da sua gestão, melhorando a produtividade e permitindo que os organizadores aumentem os resultados dos seus eventos (Zhang e Lv, 2018). Além disso, os mesmos autores afirmam que uma predição acurada da participação de pessoas nos eventos pode ser útil em termos de planejamento, publicidade e gerenciamento, assegurando uma melhor experiência aos participantes. Chiu, Won e Leng (2019) complementam que entender as motivações do comparecimento é fundamental para alavancar o público e manter a lealdade dos participantes. Sendo assim, o entendimento das motivações que levam às pessoas a comparecerem nos mais diferentes tipos de eventos é fundamental para a elaboração de um modelo de previsão de público.

Diante disso, a próxima seção deste capítulo apresenta o tema e os objetivos específicos desenvolvidos nesta dissertação. Após, tem-se a justificativa do tema e dos objetivos apresentados, seguido pelo método utilizado na pesquisa. Por fim, as limitações do estudo e a estrutura da dissertação são apresentados.

1.1 TEMA E OBJETIVOS

Diante do exposto anteriormente, esta dissertação tem como objetivo geral estender as aplicações dos métodos de previsão à predição de público em eventos. Especificamente, buscou-se:

- a) Identificar as motivações por trás da decisão de comparecer a diferentes tipos de eventos;
- b) Explorar como estas motivações podem ser compreendidas e estruturadas de forma a prever o público de eventos;
- c) Mapear e identificar quais métodos de previsão estão sendo utilizados para previsão de público em eventos;
- d) Propor um *framework* que auxilie na avaliação de eventos e na previsão de público;
- e) Demonstrar como o *framework* pode ser utilizado para previsão de público em eventos;
- f) Explorar métodos mais avançados citados na literatura para aplicação na previsão de público em eventos.

1.2 JUSTIFICATIVA DO TEMA E OBJETIVOS

Com base no exposto e dado que os eventos em geral podem desenvolver o turismo, criar novos empregos e impulsionar a economia local, esta dissertação proporciona conhecimento para o planejamento de eventos que podem contribuir para o desenvolvimento e gestão das cidades, como a alocação de recursos e pessoas. Aliado a isso, o conhecimento sobre as motivações para participar de eventos pode ser usado para entender porque as pessoas viajam para ir a eventos e também para impulsionar o turismo.

Assim, esse estudo visa coletar os achados sobre previsão de público e as motivações do comparecimento que se encontram dispersos na literatura, com uma abordagem voltada a diversos tipos de eventos. Os resultados podem contribuir para aprimorar o turismo de eventos por meio de um melhor planejamento e gestão dos mesmos, adquirindo conhecimento dos aspectos que os tornam atrativos e aqueles que precisam ser melhorados. Além do turismo de eventos, os achados dessa pesquisa podem ser utilizados para impulsionar o público dos eventos, aumentar os lucros e entender as motivações por trás do comportamento de seus

participantes. Esse conhecimento pode trazer benefícios para a gestão de *marketing* e estratégias para manter a lealdade de seus participantes para edições futuras. Essa compreensão aprofundada do evento e das motivações pode trazer grandes benefícios, melhorando a produtividade e os resultados dos eventos (Zhang e Lv, 2018).

Laing (2018) destaca, ainda, que faltam aplicações práticas nos estudos publicados sobre eventos. Além disso, sua pesquisa conclui que a maior parte dos estudos relacionados a eventos foca nas implicações turísticas que eventos podem trazer para a região, mas faltam estudos que os olhem para além do turismo, com foco na sua gestão. Desta forma, essa dissertação busca desenvolver essa área de estudo, com a elaboração de um método que irá ajudar organizações públicas e privadas a otimizar sua gestão, realizar um planejamento eficiente dos eventos e impulsionar a presença do público.

Por outro lado, os métodos de previsão têm se aperfeiçoado com o tempo e têm sido aplicados em diferentes áreas (Tealab, 2018). Sendo assim, essa dissertação busca ampliar o conhecimento desses métodos, aplicando-os de uma forma diferente às encontradas na literatura. Entre os métodos mais avançados e de crescente interesse e relevância, estão os métodos de *Machine Learning*. Claudino et al. (2019), por exemplo, mostraram em sua revisão sistemática que os métodos de *Machine Learning* já têm sido usados em aplicações esportivas, principalmente para prever lesões em jogadores e o resultado da partida. Porém, ainda não têm sido explorado para prever o público em eventos. Para isso, escolheu-se estudar como esses métodos podem ser aplicados à previsão de público em eventos esportivos e ampliar o conhecimento nesta área, trazendo uma alternativa aos organizadores de eventos em como obter resultados confiáveis referentes ao comparecimento.

1.3 MÉTODOS

Esta dissertação, do ponto de vista de sua natureza, pode ser classificada como uma pesquisa aplicada, uma vez que seus objetivos visam gerar conhecimento para uma aplicação, relacionada à previsão de público em eventos. Segundo Silva e Menezes (2001), uma pesquisa aplicada visa a geração de conhecimento para aplicação prática e solução de problemas. Além disso, pesquisas aplicadas também podem ser utilizadas para contribuição do conhecimento científico e sugerir novas questões a serem investigadas (Gil, 2019). Essa classificação pode ser dada uma vez que a identificação e estruturação das motivações que levam às pessoas a

comparecerem a eventos foram utilizadas no *framework* proposto para previsão de público e também na avaliação de eventos. Além disso, foi realizado um estudo prático utilizando outros métodos mais avançados encontrados na literatura.

Considerando os objetivos apresentados anteriormente, esta dissertação também se enquadra em uma pesquisa descritiva e exploratória. As pesquisas descritivas buscam descrever relações entre características de determinada população ou estabelecimento de relações entre variáveis. Já as pesquisas exploratórias, por sua vez, visam proporcionar maior familiaridade com o problema através de um levantamento bibliográfico ou entrevistas (Silva e Menezes, 2001). Sendo assim, esta dissertação fez uso de uma revisão sistemática para levantamento e estruturação das motivações por trás do comparecimento a eventos para, após, estabelecer relações entre essas motivações. Revisões sistemáticas fazem uso de um método estruturado e bem definido para responder questões de pesquisa previamente definidas, testar hipóteses ou teorias e descrever as contribuições sobre um assunto (Denyer e Tranfield, 2009).

Uma vez que essa dissertação tem como objetivo entender as motivações por trás do comparecimento a eventos e utilizar este conhecimento para prever o público, esta pesquisa pode ser classificada como qualitativa e quantitativa. Pesquisas qualitativas visam prover compreensões mais aprofundadas sobre determinado problema (Gil, 2019), o que essa dissertação faz ao compreender e estruturar os motivos que levam as pessoas a eventos. Já pesquisas quantitativas traduzem informações em números, de forma a classificá-las ou analisá-las (Silva e Menezes, 2001). Isto foi realizado durante o estudo, com o uso de métodos de regressão e o *framework* desenvolvido, esta dissertação utilizou as motivações levantadas para quantificar o número de pessoas que comparecerão em um evento.

1.4 LIMITAÇÕES DO ESTUDO

Como natural a qualquer pesquisa, este estudo apresenta limitações quanto a seus objetivos e resultados. Visto que utilizou uma revisão sistemática para levantamento das motivações, apenas aquelas contempladas nos estudos científicos selecionados pelos critérios arbitrados foram consideradas. Apesar disto, visto o grande número de estudos encontrados e de motivações identificadas, pode-se considerar que o assunto foi amplamente coberto. Além disso, o *framework* elaborado foi avaliado através de duas simulações de eventos, não sendo

validado em situações reais. Ainda assim, foi possível compreender sua funcionalidade e o mesmo foi aplicado por uma profissional do ramo da área de eventos.

Por fim, outra limitação desta pesquisa é decorrente dos dados coletados para aplicação com métodos de previsão levantados pela literatura. Os dados foram coletados em parceria com clube de futebol brasileiro e compreendem cinco anos de partidas de futebol. Porém, para parte dos métodos utilizados este tamanho de amostra é considerado pequeno, e é necessário tratamento especial dos dados. Além disso, torna o modelo de previsão específico para o evento em análise, como será discutido no Capítulo 4.

1.5 ESTRUTURA DA DISSERTAÇÃO

Esta dissertação está organizada na estrutura de três artigos, cada artigo está relacionado com um dos objetivos específicos deste estudo apresentados anteriormente conforme exposto na **Figura 1**. O Artigo 1, apresentado no capítulo 2, aborda a identificação e estruturação das motivações e métodos de previsão aplicados a eventos encontrados na literatura, por meio de uma revisão sistemática, intitulado: “Fatores motivacionais e métodos de previsão aplicados a eventos”. O capítulo 3 aborda o Artigo 2, no qual o framework desenvolvido no Artigo 1 é aplicado a dois eventos diferentes, demonstrando a maneira que organizadores de eventos podem utilizar as motivações identificadas na literatura, intitulado: “Método de previsão de público em eventos baseado em fatores motivacionais”. O Artigo 3, apresentado no capítulo 4, apresenta um método específico de previsão de público para eventos esportivos, utilizando de métodos de *Machine Learning*, intitulado: “Previsão de público em eventos esportivos com o uso de *Machine Learning*”. Por fim, o capítulo 5 apresenta as conclusões finais deste estudo, bem como sugestões de pesquisas futuras.

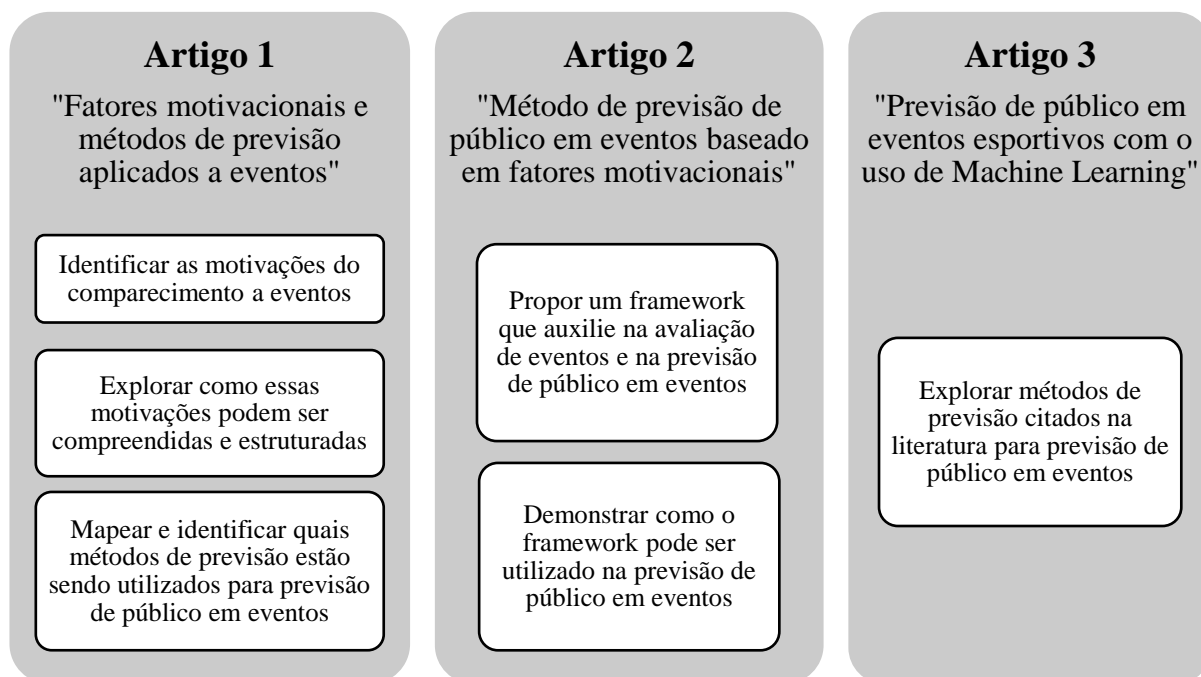


Figura 1 - Relação entre os artigos e os objetivos específicos da dissertação

1.6 REFERÊNCIAS

- ADEMA, K. L.; ROEHL, W. S.; Environmental scanning the future of event design. **International Journal of Hospitality Management**, v. 29, p. 199-207, 2010
- BERRIDGE, G.; Event pitching: The role of design and creativity. **International Journal of Hospitality Management**, v. 29, p. 208-215, 2010.
- BISCHOP, C., 2006. Pattern Recognition and Machine Learning. 1 ed. Springer. New York.
- CHIU, W.; WON, D.; LENG, H. K.. The relationship between sport involvement, perceived event prestige, and the intention to attend annual spectator events. **Asia Pacific Journal of Marketing and Logistics**, [s. l.], v. ahead-of-p, n. ahead-of-print, p. 1405–1421, 2019.
- CLAUDINO et al.; Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: a Systematic Review, **Sporting Medicine - Open**. 5 (2019) 28. <https://doi.org/10.1186/s40798-019-0202-3>.
- DENYER, D.; TRANFIELD, D.; **Producing a Systematic Review**. In: The Sage Handbook of Organizational Research Methods, edited by D. Buchanan and A. Bryman, 671-689. London: Sage. 2009.
- DEVESA, M. et al. Factors determining attendance at a film festival. **Event Management**, [s. l.], v. 19, n. 3, p. 317–330, 2015.

- GÉRON, A. **Hands-on Machine Learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems**. O'Reilly Media, Inc, 2017.
- GIL, A.C., 2019. **Como Elaborar Projetos de Pesquisa**. 6 ed. Atlas. São Paulo.
- KIM, Y.H., KAEWNUCH, K., 2018. Finding the gaps in event management research: A descriptive meta-analysis. **Event Management** 22, 453–467. doi:10.3727/152599518X15258922919408
- LAINING, Jennifer. Festival and event tourism research: Current and future perspectives. **Tourism Management Perspectives**, v.25, p. 165-168, 2018.
- LI, Libo. Predicting online invitation responses with a competing risk model using privacy-friendly social event data. **European Journal of Operational Research**, [s. l.], v. 270, n. 2, p. 698–708, 2018.
- SERWYLO et al.; Predicting patient presentation rates at mass gatherings using Machine Learning. **8th International ISCRAM Conference**, Lisboa, Portugal, maio 2014.
- SILVA, E.L., MENEZES, E.M., 2001. Metodologia de pesquisa e elaboração de dissertação. Laboratório de Ensino a Distância da UFSC, Florianópolis.
- TEALAB, A.; Forecasting using artificial neural network methodologies: A systematic review. **Future Computing and Informatics Journal**, v. 3, p. 334 – 340, 2018.
- W. JUN et al.; Modeling a combined forecast algorithm based on sequence patterns and near characteristics: An application for tourism demand. **Chaos, Solitons and Fractals**, v. 108, p. 136-147, 2018.
- ZHANG, Shuo; LV, Qin. Hybrid EGU-based group event participation prediction in event-based social networks. **Knowledge-Based Systems**, [s. l.], v. 143, p. 19–29, 2018.

2 ARTIGO 1 - FATORES MOTIVACIONAIS E MÉTODOS DE PREVISÃO APLICADOS A EVENTOS

RESUMO

Com a popularização do entretenimento, os eventos têm se tornado uma área de interesse crescente, pois conseguem aumentar a demanda turística local, trazendo benefícios econômicos e sociais para a região (Hall, O'Mahony e Vieceli, 2010). Neste contexto, o planejamento dos eventos é uma atividade valiosa que propicia otimização de recursos e aumento do lucro (Berridge, 2010). Além disso, uma predição acurada da participação de pessoas nos eventos pode ser útil em termos de planejamento, publicidade e gerenciamento (Zhang e Lv, 2018). Desta forma, para compreender as motivações por trás do comparecimento, este artigo, através de uma revisão sistemática da literatura, propõe-se a reunir os trabalhos de previsão aplicada a eventos e responder a duas questões de pesquisa: (i) Quais os fatores e motivações que interferem na decisão de participar de eventos; (ii) Quais métodos de previsão podem ser utilizados para estimar o público em eventos. Foram identificados a partir da literatura 47 fatores que motivam o comparecimento a eventos, que foram organizados e apresentados relacionados a perguntas que irão auxiliar os organizadores a prever o público nos eventos. Além disso, foram observados 24 métodos de previsão utilizados na literatura para prever e classificar a participação em eventos. Desses, o mais utilizado foi a Regressão Logística e a Regressão Linear, na sua maioria com aplicação de *Machine Learning* em bases de dados obtidas através de redes sociais.

2.1 INTRODUÇÃO

O ambiente no qual as organizações estão inseridas e as decisões ocorrem é caracterizado por riscos e incertezas. Neste cenário, ser capaz de estimar a demanda futura pela procura de produtos ou serviços é fundamental para a sobrevivência e eficiência das instituições. Além disso, os métodos de previsão têm atraído a atenção de diversas áreas do conhecimento, uma vez que eles podem possibilitar a estimativa da demanda futura a partir dos valores passados ou de outras informações com uma margem aceitável de erro (Tealab, 2018).

Com a popularização do entretenimento, como filmes, músicas e esportes, os eventos têm se tornado uma área de interesse global. Hall, O'Mahony e Vieceli (2010) evidenciam que os eventos conseguem aumentar a demanda turística local, trazendo grandes benefícios

econômicos e sociais. Yolal et al. (2016) também citam a geração de novos empregos, oportunidades de negócios e trocas culturais como benefícios de grandes eventos. Assim, a busca pelo gerenciamento eficiente dos eventos tem sido estudada de forma cada vez mais intensa (Getz, 2008).

Além disso, o planejamento dos eventos é uma atividade valiosa tanto para organizações públicas como privadas, propiciando a otimização do uso de recursos, o alcance de objetivos ou o aumento do lucro. Porém, o planejamento de eventos engloba traços inerentes de imprevisibilidade e riscos em um ambiente cujas características econômicas, ambientais, políticas, tecnológicas e sociais estão em constante mudança (Berridge, 2010; Adema e Roehl, 2010).

Desta maneira, conhecer e entender as motivações que levam às pessoas a comparecerem aos eventos pode trazer benefícios para a organização do mesmo, melhora a produtividade e permite que os organizadores aumentem os resultados dos seus eventos (Zhang e Lv, 2018). Além disso, os mesmos autores afirmam que uma predição acurada da participação de pessoas nos eventos pode ser útil em termos de planejamento, publicidade e gerenciamento, assegurando uma melhor experiência aos participantes. Chiu, Won e Leng (2019) complementam que entender as motivações do comparecimento é fundamental para alavancar o público e manter a lealdade dos participantes. Sung e Mills (2017) ainda, citam que a previsão da demanda esportiva, por exemplo, tem grandes implicações na determinação de políticas de vendas e gerenciamento dos eventos.

Getz (2008) exemplifica alguns eventos de maior importância mundial, como as Olimpíadas, a Copa do Mundo da FIFA, o automobilismo da Fórmula 1, as premiações Oscar e Grammy, estreias de cinema, festivais e o Carnaval brasileiro, que movimentam grandes volumes financeiros. Kim, Suh e Eves (2010) também trazem o conceito de eventos relacionados à comida e bebida, que atraem inúmeros participantes e podem ser incluídos nesta lista.

Embora seja um assunto de grande interesse e relevância crescente, a literatura em previsão de demanda aplicada a eventos encontra-se dispersa e segmentada. Por exemplo, Serwylo et al. (2014) aplicaram métodos de *Machine Learning* para previsão de ocorrências médicas em eventos; Sung e Mills (2017) elaboraram um modelo empírico de previsão de comparecimento a eventos esportivos baseado na incerteza do resultado da partida, preço e qualidade do time; enquanto Marshall, Dockendorff e Ibáñez (2013) elaboraram um sistema de

predição de comparecimento a estreias de filmes baseado em modelos de adesão a novos produtos. Além destes trabalhos, outros autores investigaram apenas as motivações para o comparecimento a diferentes tipos de eventos (Hall, O'Mahony e Vieceli, 2010; Kim, Suh e Eves, 2010).

Para compreender as contribuições já realizadas nesta área e as lacunas existentes, este artigo, através de uma revisão sistemática da literatura, propõe-se a reunir os trabalhos de previsão de público aplicada a eventos e responder a duas questões de pesquisa: (i) Quais os fatores ou motivações que interferem na decisão das pessoas a comparecerem a eventos; (ii) Quais métodos de previsão podem ser utilizados para estimar o público que deve estar presente em diferentes eventos.

Para responder estas questões, este artigo está dividido em seis seções. Após esta introdução, a Seção 2 detalha os procedimentos metodológicos que guiaram esta revisão sistemática, cujos resultados são descritos na Seção 3. Após, a Seção 4 apresenta a análise dos resultados seguido de uma discussão na Seção 5. Por fim, a Seção 6 apresenta as conclusões gerais obtidas da pesquisa e propostas para pesquisas futuras.

2.2 METODOLOGIA

Conforme mencionado, este artigo utiliza uma revisão sistemática da literatura para responder às questões de pesquisa. Uma revisão sistemática é um processo de revisão de literatura que adota uma série de passos rigorosos e bem definidos que permitem sintetizar de forma abrangente, transparente e reprodutível as contribuições de pesquisa em um determinado campo (Briner e Denyer, 2012). Além disso, uma revisão sistemática visa responder a alguma questão de pesquisa previamente definida, ou testar hipóteses e teorias, em vez de apenas descrever as contribuições sobre um assunto (Petticrew e Roberts, 2006; Denyer e Tranfield, 2009; Briner e Denyer, 2012). Segundo Briner e Denyer (2012), ainda, uma revisão sistemática permite tirar conclusões sobre o que é conhecido e o que ainda não se conhece sobre uma questão de pesquisa, com um certo nível de consistência e confiança.

Em função do uso de um protocolo definido, que deve ser descrito e comunicado, os resultados de revisões sistemáticas usualmente oferecem menor viés e são mais fáceis de serem reproduzidos, tornando-o mais vantajoso quando comparado ao método tradicional de revisão

da literatura (Petticrew e Roberts, 2006). Apesar de ser bem definido, o método é flexível aos critérios de busca definidos pelo pesquisador, desde que o mesmo deixe esses critérios documentados, justificando suas escolhas (Briner e Denyer, 2012). Além disso, o fato de ser reprodutível permite que o método possa ser revisto e avaliado criticamente ao longo do tempo pelos mesmos autores ou outros pesquisadores, possibilitando um estudo longitudinal do assunto.

Por estas razões, principalmente por ser um método que busca realizar uma pesquisa completa e não-enviesada do tema, optou-se por realizar uma revisão sistemática para reunir e analisar as contribuições acerca de previsão de público em eventos e verificar as lacunas existentes na área. Desta forma, a pesquisa seguiu seis passos-chaves no qual uma revisão sistemática pode ser resumida, apresentados por Briner e Denyer (2012): (i) planejamento da revisão de literatura; (ii) busca dos estudos relevantes nas bases de dados; (iii) avaliação das contribuições de cada estudo; (iv) síntese das informações coletadas; (v) análise e divulgação dos resultados obtidos; e (vi) discussão dos resultados.

O passo (i), planejamento da revisão de literatura, envolveu a definição do tema de estudo e uma breve análise exploratória do que está sendo realizado na área, para determinação, de forma clara e objetiva, das questões de pesquisa. Nesta fase, também, definiu-se as palavras-chave e as bases de dados que foram utilizadas na busca. Para isso, inicialmente procurou-se, em diversas bases, artigos relacionados ao tema de previsão de demanda e eventos e selecionou-se aqueles que respondiam a pelo menos uma das questões de pesquisa.

A definição das palavras-chave a serem utilizadas posteriormente foi feita a partir da recorrência das mesmas em diferentes artigos da primeira busca. Verificou-se que as palavras da língua inglesa “*prediction*”, “*forecasting*”, “*attendance*”, “*participation*”, “*event*” e “*crowd*” apresentavam maior relevância. Foram identificados três grupos de palavras que deveriam estar presentes para que a busca retornasse artigos relevantes para o estudo: o primeiro grupo é referente a predição (por exemplo: *prediction*); o segundo grupo é referente ao que se quer prever (por exemplo: *attendance*); e, o terceiro grupo é referente ao local de estudo (por exemplo: *event*). Sendo assim, definiu-se a seguinte combinação de termos (*query strings*) para a busca: (*predict* OR forecast* OR estimat**) AND (*participation OR attendance*) AND (*event OR crowd*). Sendo que cada parêntese representa um dos grupos identificados. As palavras de cada grupo foram ligadas pelo operador booleano OR, enquanto que os grupos foram ligados

pelo operador booleano AND. Desta forma, obrigatoriamente uma das palavras de cada grupo estará presente nos artigos resultantes da busca.

As bases de dados acadêmicas escolhidas para a realização das pesquisas foram a ScienceDirect® e a Scopus®. Essa escolha está baseada no fato das mesmas serem as duas bases de maior relevância no meio acadêmico. A ScienceDirect é uma plataforma da editora Elsevier que oferece acesso aos textos completos e compreende mais de 2.500 revistas revisadas por pares nas áreas de saúde, ciência e tecnologia (Elsevier, 2019). Já a Scopus é o maior banco de dados de resumos e citações de artigos com mais de 75 milhões de registros, cobrindo mais de 5.000 editoras internacionais e 23.500 revistas revisadas por pares (Scopus, 2019). Além disso, a Scopus oferece acesso a métricas de qualidade dos textos e revistas, como o seu Fator de Impacto, número de citações e o CiteScore™. Sendo assim, optou-se por utilizar ambas bases, garantindo o acesso às melhores editoras e revistas e uma cobertura abrangente do assunto.

Após a definição das *query strings* e das bases de dados, procedeu-se ao passo (ii), que envolve a busca dos estudos relevantes nas bases de dados. Nesta etapa, ocorreu a pesquisa e seleção dos artigos a serem avaliados a partir dos termos previamente definidos. Para isso, foi necessário também definir os critérios de inclusão e exclusão dos artigos para análise. Como critério de inclusão, optou-se por analisar apenas estudos publicados em periódicos (excluindo artigos de congressos, teses e dissertações, por exemplo). Além disso, também foram incluídos apenas artigos na língua inglesa publicados entre 2010 e o momento da busca. A pesquisa foi realizada nas duas bases de dados em setembro de 2019 e foram selecionados para análise apenas os artigos que continham os termos da *query strings* no Título, Resumo ou Palavras-chave, além de atender os critérios descritos anteriormente. No momento da busca, observou-se que a busca estava retornando vários artigos referentes a previsão do tempo, *weather forecasting*, e previsão de doenças, *disease prediction*. Sendo assim, adicionou-se à *query strings* “AND NOT (*weather OR disease*)”, de forma a evitar artigos que abordassem esses assuntos. A combinação final de termos pode ser vista na **Figura 2**, na qual os retângulos indicam os grupos e suas opções de termos.

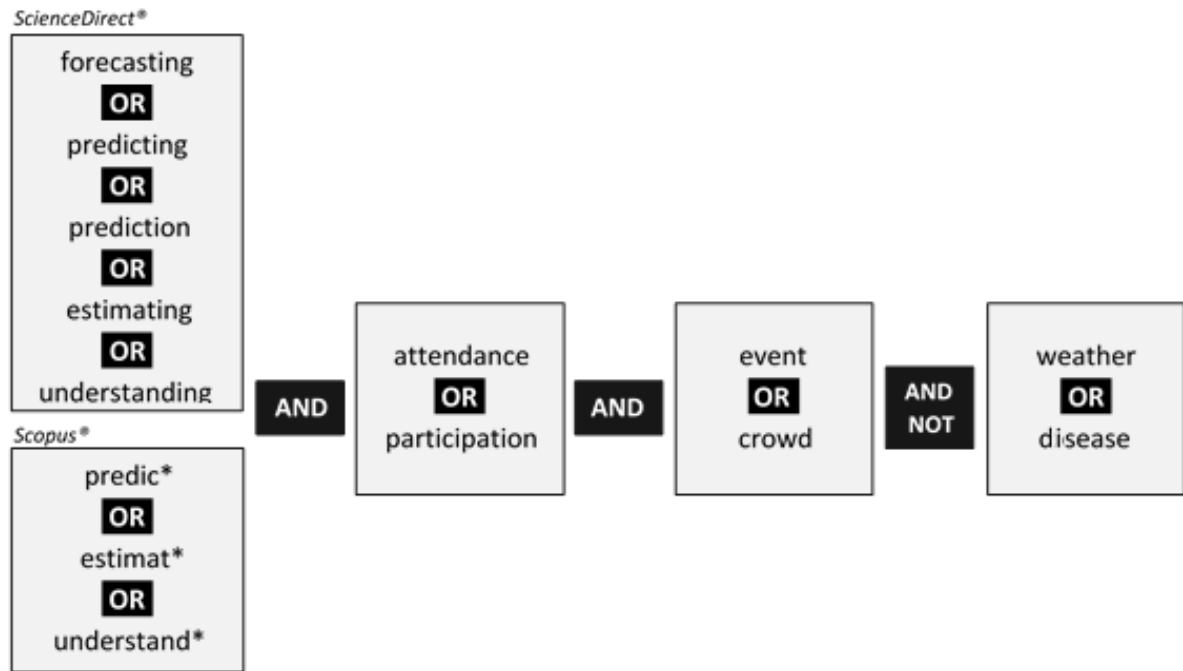


Figura 2 - Combinação final de *query strings* para a busca nas bases de pesquisa

A **Figura 2** também demonstra a diferença de *query strings* entre as bases de dados utilizadas, pois a base ScienceDirect não aceita o truncamento (*) no final das palavras. Após a busca, foram encontrados 545 artigos que atendiam aos critérios estabelecidos na base Scopus e 365 artigos na base ScienceDirect, resultando em um total de 910 artigos potencialmente relevantes. Foram extraídos das bases de dados apenas os títulos, DOI, palavras-chave, resumo, ano de publicação, revista e autores, que foram reunidos em uma planilha eletrônica única para facilitar a análise. Após remoção dos artigos que resultaram nas duas bases de dados, restaram 772 artigos potencialmente relevantes.

Como primeiro critério de exclusão, avaliaram-se os títulos desses 772 artigos e foram excluídos aqueles que estavam fora do escopo, como eventos médicos, doenças e eventos meteorológicos que não foram excluídos pela *string*. Após esta etapa, restaram 214 artigos que seguiram para análise dos resumos. Durante a análise dos resumos foram excluídos aqueles artigos que não citavam métodos de previsão de público em algum tipo de evento ou de possíveis motivações para o comparecimento, restando 80 artigos potencialmente relevantes. Na sequência, foi aplicado um critério de qualidade dos estudos, eliminando aqueles que foram publicados em revistas com CiteScore menor que 1 e fator de impacto menor que 0,8. Desta

forma, restaram 59 artigos para leitura do texto completo e análise da sua contribuição. Por fim, após leitura completa desses artigos e análise das suas referências, foram excluídos 8 artigos que não exemplificaram fatores de motivação e nem realizaram previsão de público. Além disso, foram incluídos 5 artigos do referencial que respondiam diretamente às questões de pesquisa, restando um total de 56 artigos relevantes para análise. A **Figura 3** demonstra a evolução do processo de seleção dos artigos.

Prosseguiu-se para o passo-chave (iii), contemplando a avaliação das contribuições de cada estudo. Nesta etapa, ocorreu a leitura completa dos artigos selecionados e extração dos dados relevantes. Para isso, usou-se a mesma planilha eletrônica que continha os dados básicos dos artigos para registrar, também, objetivos, métodos, resultados e contribuições dos estudos analisados durante a leitura completa. Assim, foram catalogados os dados de interesse que poderiam contribuir para responder às questões de pesquisa.

Os últimos passos da revisão sistemática, resumidos por Briner e Denyer (2012), envolvem (iv) síntese das informações coletadas, (v) análise e divulgação dos resultados obtidos e (vi) discussão dos resultados. Essas etapas são referentes à descrição e discussão dos resultados, incluindo a elaboração de quadros, diagramas e figuras que auxiliam a visualização das contribuições que já foram feitas na área de estudo. Assim, é possível extrair conclusões acerca das questões de pesquisa inicialmente impostas. Além disso, essas etapas permitem identificar aquelas questões que ainda necessitam de investigação na literatura, que serão apresentadas neste artigo nas seções 3, 4 e 5, respectivamente.

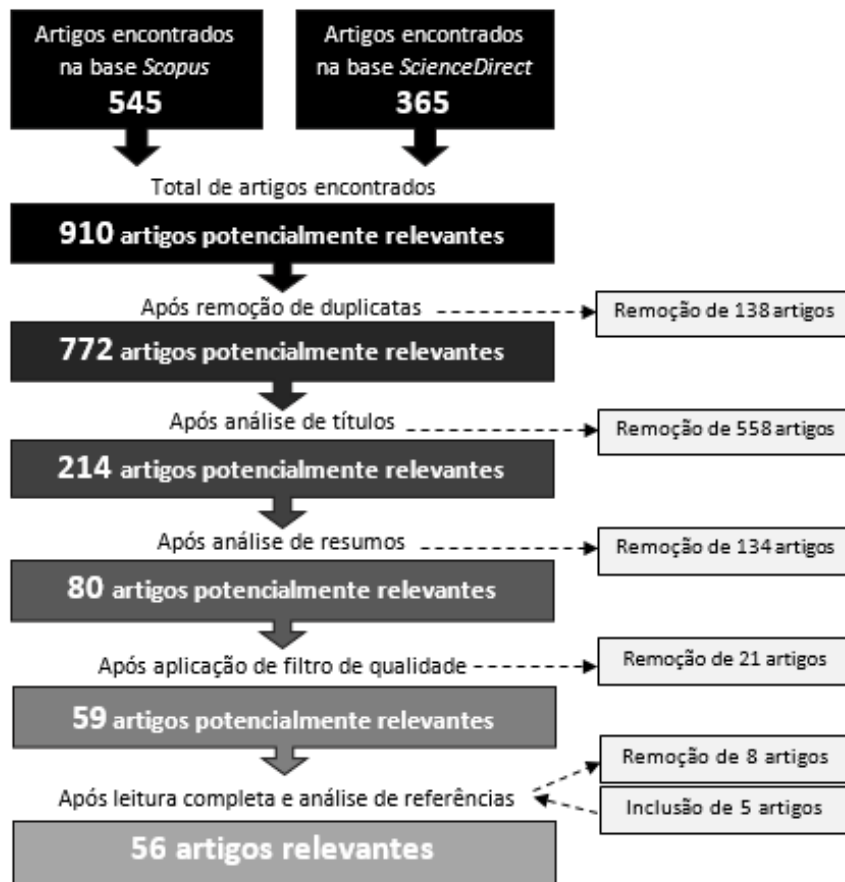


Figura 3 - Diagrama do processo de seleção de artigos

2.3 SÍNTESE DOS RESULTADOS

De forma a dar sequência à pesquisa com o passo (iv) da revisão sistemática, procurou-se organizar os artigos e extrair todas as informações relevantes sobre as motivações e métodos de previsão de público em eventos. Para uma melhor compreensão das informações extraídas, inicialmente foi avaliado o conteúdo bibliográfico das publicações. Dos 56 artigos selecionados após aplicação da metodologia apresentada, 51 artigos avaliaram motivações e fatores associados ao comparecimento e 17 artigos aplicaram algum método de previsão do público.

Com relação aos periódicos nos quais os artigos selecionados foram publicados, foi encontrada uma diversidade de 44 revistas. Esse número corrobora o fato de que as informações sobre previsão de público em eventos encontram-se dispersas na literatura, justificando a necessidade desta revisão sistemática. Dentre as revistas, destaca-se o periódico *Event Management* com cinco publicações no período. Seguido das revistas *International Journal of*

Event and Festival Management e *Journal of Sport and Tourism* com três publicações cada. A **Figura 4** traz a relação dos periódicos com mais de uma publicação no período analisado, totalizando sete revistas.

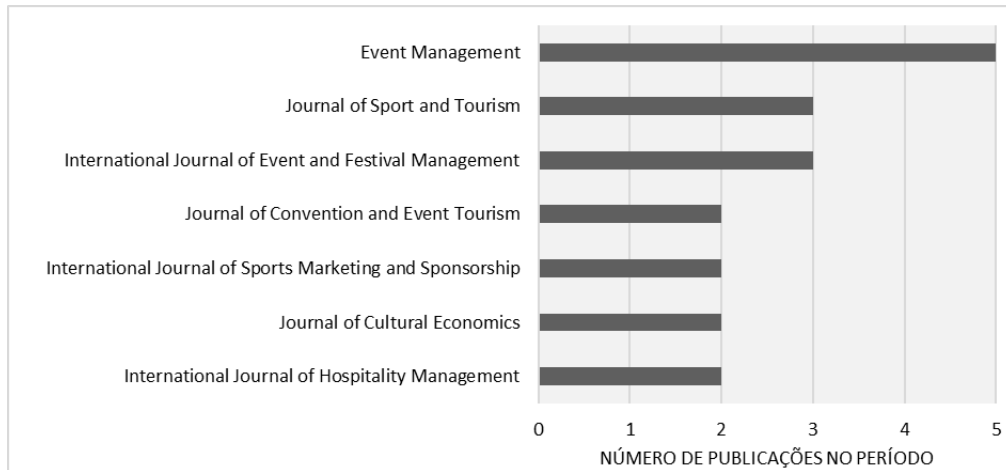


Figura 4 - Periódicos com as maiores quantidades de publicações

Quando comparada a evolução anual das publicações nos últimos dez anos, percebe-se que o interesse neste assunto vem crescendo, como pode ser visto na **Figura 5**. Nos últimos dois anos, foram publicados 17 artigos nesta área, enquanto nos dois primeiros anos da década foram publicados apenas 8. Isto representa um aumento de aproximadamente 100% nas contribuições a respeito de eventos.

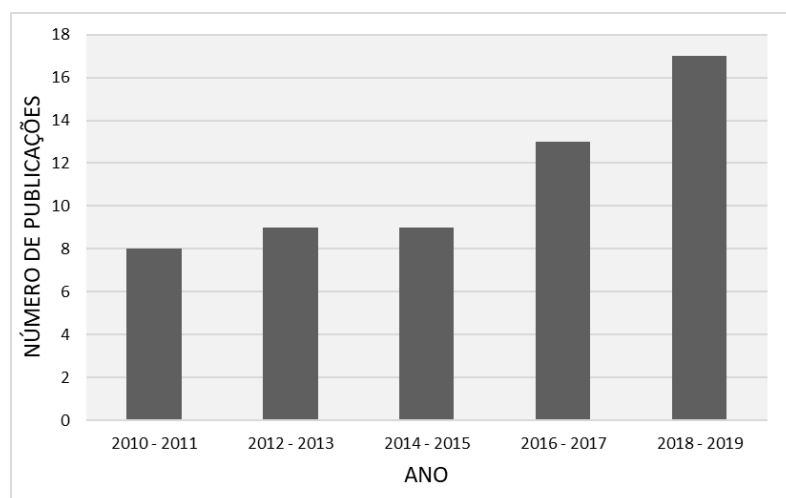


Figura 5 - Evolução temporal das publicações analisadas

Outro aspecto importante de ser analisado é os tipos de eventos que foram estudados pela literatura. Sabe-se que cada evento tem características inerentes ao seu tipo que podem influenciar as motivações. A **Figura 6** resume os tipos de eventos encontrados.

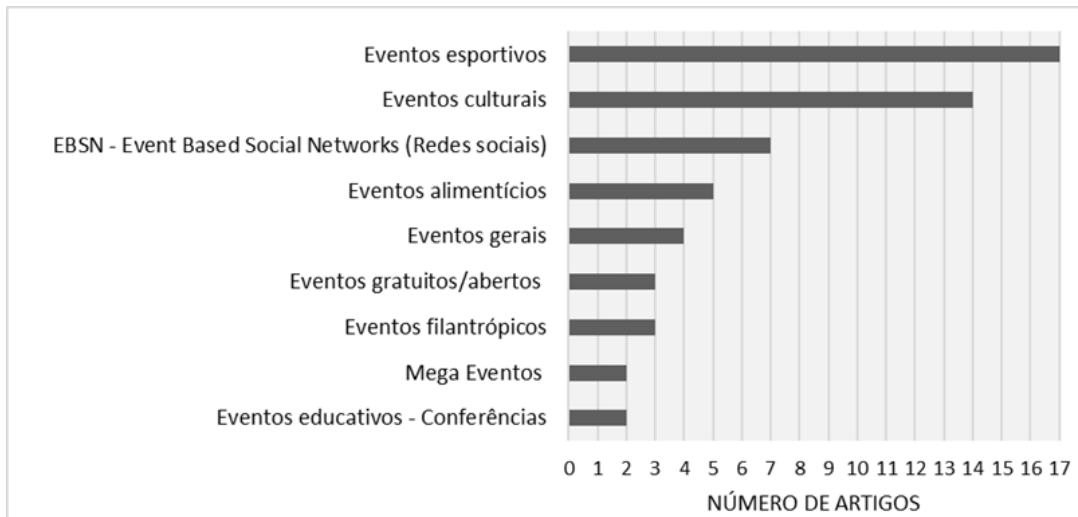


Figura 6 - Tipos de eventos analisados encontrados na literatura

A **Figura 6** revela que os eventos esportivos são o tipo de evento mais estudado. Os esportes encontrados na literatura foram futebol, basquete, baseball, futebol americano universitário, rugby, futebol americano profissional, hockey e maratonas, totalizando 17 publicações. Isto representa 30% de todos os artigos, refletindo a importância deste segmento. Além de eventos esportivos, os eventos culturais também totalizam um número considerável de publicações representando aproximadamente 25% do total de publicações. Foram considerados como eventos culturais aqueles relacionados a galerias de artes, museus, zoológicos, aquários, jardins botânicos, concertos musicais, teatro, dança, musicais, óperas e cinema.

Em seguida, os Eventos Baseados em Redes Sociais (EBSNs) são o terceiro grupo mais explorado, com oito artigos. É importante ressaltar o avanço da tecnologia dos últimos anos que pode ter inspirado o interesse nesse tipo de evento. EBSNs são serviços online nos quais os usuários podem iniciar diversos tipos de eventos sociais, como reuniões, grupos de estudo, caminhadas e jantares, especificando seu objetivo, interesses e geolocalização, além de

enviar convites para as mais diversas pessoas (Jiang e Li, 2019). Os eventos, então, são organizados pelas redes sociais e realizados em locais físicos. Algumas das plataformas mais conhecidas são o *MeetUp*, *Plancast* e o *Facebook Events*. Estes três tipos de eventos representam 68% das publicações selecionadas no estudo.

2.3.1 MOTIVAÇÕES PARA O COMPARECIMENTO A EVENTOS

Com base na leitura dos artigos selecionados, as considerações acerca das motivações aos eventos foram divididas em duas grandes dimensões, sendo elas uma dimensão de Fatores Internos e outra de Fatores Externos. Essa divisão é comum em análises motivacionais na literatura, sendo chamada também de Motivações Push ou Intrínsecas para os fatores internos e Motivações Pull ou Extrínsecas para os fatores externos (Priporas et al., 2018; Yahia e Al-Emadi, 2018; Getz, 2010).

As motivações intrínsecas são aquelas relacionadas ao próprio indivíduo, seus interesses e obstáculos internos, como suas atitudes e preferências (Priporas et al., 2018) e também suas emoções, paixões e crenças (Yahia e Al-Emadi, 2018). Já as motivações extrínsecas são relacionadas às características do evento, que podem atrair ou não a população. Podem ser consideradas fatores extrínsecos a percepção da cidade sede do evento, a segurança do local e as facilidades, por exemplo (Yahia e Al-Emadi, 2018). Percebe-se que os fatores externos são mais tangíveis e diretos, permitindo facilmente o seu uso em modelos de previsão e também a sua mensuração.

(i) Fatores Internos

Conforme definido anteriormente, fatores internos falam do comportamento e da percepção dos participantes do evento, são fatores pessoais e, às vezes, intangíveis. Muitos autores citam que os hábitos cultivados pela população, e que são relacionados com o tipo de evento, podem explicar a sua motivação. Para Palma, Palma e Aguado (2013), por exemplo, o consumo da cultura pela população aumenta o comparecimento a eventos. Ainda, o hábito de ir ao cinema (Devesa, 2015) e a prática regular de esportes (Priporas et al., 2018; Wicker e Hallmann, 2013; Snelgrove e Wood, 2010) também podem ser agentes motivacionais da participação.

Com relação ao consumo de eventos, ainda, a quantidade de vezes que a pessoa foi a um mesmo evento pode ser um forte preditor de comparecimentos futuros (Li, 2018; Klapanidou e Gibson, 2017; Popp, Jensen e Jackson, 2017; Boyle, 2016; Halpenny, Kulczycki e Moghimehfar, 2016; Trinh e Lam, 2016; Bruwer e Kelley, 2015; Palma, Palma e Aguado, 2013; Michalco, 2012). Além disso, comparecer a eventos semelhantes também pode explicar o comparecimento a um determinado evento, uma vez que isto captura as preferências pessoais do indivíduo (Lu, 2019; Jiang, 2019; Bogaert, Ballings e Van Den Poel, 2016; Boyle, 2016; Palma, Palma e Aguado, 2013).

Por outro lado, também, o quanto o indivíduo conhece e sabe sobre o assunto relacionado ao evento pode explicar as suas motivações. Por exemplo, o nível de conhecimento sobre música clássica (Dobson, 2010), vinhos e agricultura (Govindasamy e Kelly, 2014), dança de salão (Santos-Lewis e Moital, 2013) e esportes (Cheng, 2012) explicaram o comparecimento nos respectivos eventos analisados. Da mesma forma, a vontade de aprender mais sobre o assunto também leva às pessoas aos eventos (Cushing, 2018; Cheng, 2012; Tanford, Montgomery e Herzmand, 2012) e a falta de conhecimento sobre o tópico tem impacto negativo na atratividade geral do evento (Hamerman, 2016; Donson, 2010). Além disso, o conhecimento com o assunto do evento aumenta a intenção prévia de participação (Farrel et al., 2015).

A intenção de participar dos eventos foi utilizada como preditor do comparecimento por Lee et al. (2014), considerando o momento de formação da intenção e a força desta intenção. Além deste estudo, outros autores também afirmaram que o momento de formação da intenção é um grande fator explicativo, sendo um aspecto que deve ser melhor compreendido (Bruwer e Kelley, 2015; Chiu, Lee e Won, 2016; Yahia e Al-Emadi, 2018; Chiu, Won e Leng, 2019). Nesta linha, Hall, O'Mahony e Vieceli (2010) complementam que a emoção para com o evento está altamente relacionada com a intenção de participar. Compreender a ligação emoção de cada participante, portanto, torna-se essencial para as motivações do comparecimento.

Esta ligação emocional dos participantes associada ao evento foi citada por muitos autores como explicativa da participação, assim como ser fã de esportes e desejar compartilhar com os outros participantes a paixão pelo esporte (Chiu, Won e Leng, 2019; Yahia e Al-Emadi, 2018; Halpenny, Kulczycki e Moghimehfar, 2016; Devesa, 2015; Farrel et al., 2015; Wicker e Hallmann, 2013; Cheng, 2012; Mohan e Thomas, 2012; Tanford, Montgomery e Herzmand, 2012; Tanford, Montgomery e Nelson, 2012; Hall, O'Mahony e Vieceli, 2010; Snelgrove e Wood, 2010). Além disso, combinar o interesse de diversas pessoas e criar um sentimento de

grupo aumenta o comparecimento nos eventos (Biondi et al., 2019; Rodriguez-Campo, 2019; Hopkins, 2016). Segundo Snelgrove e Wood (2010), ainda, quanto maior for essa identidade de grupo, mais as pessoas tendem a retornar para um mesmo festival em edições posteriores.

Hopkins (2016) identificou que a identidade de grupo forte e inter-relações entre os participantes tiveram um efeito positivo e significativo na experiência durante o evento. A troca de experiência entre os participantes e a vontade de socializar e conhecer novas pessoas foram também fatores motivacionais levantados por alguns autores (Cushing, 2018; Priporas et al., 2018; Tanford, Montgomery e Hertzman; Snelgrove e Wood, 2010).

Além de conhecer novas pessoas, ter um tempo de qualidade com amigos e familiares também foi citado como motivação da participação em eventos por diversos autores (Rodríguez-Campo, 2019; Xu, 2019; Akgunduz e Cosar, 2018; Cushing, 2018; Zhang e Lv, 2018; Bogaert, Balling e Van Den Poel, 2016; Bruwer e Kelly, 2015; Farrel et al., 2015; Santos-Lewis e Moital, 2013; Michalco e Navrat, 2012; Hall, Griffith, 2010; O'Mahony e Vieceli, 2010). De maneira semelhante, a indicação de eventos por amigos e familiares tem efeito positivo em atrair participantes (Yazici, Koçak e Altunsoz, 2017). Por outro lado, Cheng et al. (2012) não achou efeito significativo de que estar com amigos e familiares tem efeito positivo no comportamento de fãs de baseball.

A influência de amigos e familiares vai além da companhia, Tanford, Montgomery e Hertzman (2012) ressaltam que ganhar popularidade e ser reconhecido pelos amigos e familiares também motivam os indivíduos. Ao encontro disso, Chiu, Ton e Len (2019) e Priporas et al. (2018) concordam que o prestígio percebido pelo indivíduo a ir ao evento é um fator determinante na decisão de comparecer a um evento. Sendo assim, a reputação do evento e a percepção social do seu comportamento são fatores importantes a serem considerados (Cheng et al., 2012; Tanford, Montgomery e Hertzman, 2012; Mohan e Thomas, 2012). Eventos que oferecem premiações aos participantes também trazem um sentimento de importância que atrai a população, sejam eles em dinheiro ou reconhecimento pela participação (Priporas et al., 2018; Yazici, Koçak e Altunsoz, 2017; Farrel et al., 2015);

Além disso, a sensação de sair da rotina e escapar das responsabilidades que os eventos propiciam é um forte agente motivacional para o comparecimento a eventos (Akgunduz e Coçar, 2018; Priporas et al., 2018; Yahia e Al-Emadi, 2018; Yazici, Koçak e Altunsoz, 2017; Cheng et al. 2012; Tanford, Montgomery e Hertzman, 2012). Sendo assim, cada indivíduo busca se sentir psicologicamente relaxado (Akgunduz e Coçar, 2018; Cheng et al., 2012) e obter

satisfação pessoal durante o evento (Akgunduz e Coçar, 2018; Yahia e Al-Emadi, 2018; Farrel et al., 2015; Hall, Basarin e Lockstone-Binney, 2010). Segundo Kaplanidou e Gibson (2010) e também Bruwer e Kelley (2015), a satisfação é um preditor do retorno do participante ao evento, estando altamente ligada às facilidades do evento, acesso e transporte. Vários autores complementam que a satisfação também tem efeito positivo na intenção de participar dos eventos (Rodríguez-Campo, 2019; Yazici, Koçak e Altunsöz, 2017; Tanford, Montgomery e Hertzman, 2012).

Outro sentimento importante que pode ser provocado pelos eventos nos seus participantes é o desejo de ajudar os outros e apoiar uma causa nobre. Os participantes podem se sentir motivados a ir aos eventos para levantar questões importantes para a sociedade como a luta contra o câncer, impactos ambientais e vida saudável (Snelgrove e Wood, 2010). Por outro lado, Born et al. (2014) não encontraram evidências de que ajudar os outros era a motivação por trás do comparecimento, mas sim a emoção pessoal associada à causa nobre, por exemplo ter vencido uma doença ou conhecer alguém que está com a doença associada à causa do evento. Trinh e Lam (2016) resumem que os benefícios percebidos do evento para si e para a população local que são importantes preditores do comparecimento.

Ainda sobre o comportamento dos participantes quanto indivíduo, quando analisada as redes sociais, o perfil do usuário, as publicações e o conteúdo do evento tornam-se importantes. As preferências e características do perfil como tags que são seguidas, amizade entre os usuários e o conteúdo de comentários realizados são fatores preditores do comparecimento (De Lira, 2019; Jiang, 2019; Shuo Zhang e LV, 2019; Xu, 2019; Zhang, 2018; Parra, 2016). Além disto, a diversidade do grupo organizador e quantidade de membros do grupo online podem ser também fatores explicativos do número de pessoas no evento (De Lira, 2019; Lu, 2019; Jiang, 2019; Shuo Zhang e LV, 2019; Xu, 2019; Zhang, 2018; Parra, 2016).

(ii) Fatores Externos

Em contraste aos fatores internos, os fatores externos são aqueles que caracterizam o evento em questão e discorrem sobre o contexto no qual ele está inserido. O ponto inicial de distinção é o tipo de evento (Born et al., 2014), uma vez que eventos maiores atraem mais público (Santos-Lewis e Moital, 2013) assim como a quantidade e a diversidade de atividades previstas (Priporas et al., 2018; Lee et al., 2014; Tanford, Montgomery e Nelson, 2012). Além disso, a penetração do evento da população pode ser um fator que auxilie na previsão do

público, capturando a quantidade de pessoas dispostas a ir no evento (Muñiz, Rodríguez e Soaréz, 2017; Trinh e Lam, 2016). Em eventos organizados na rede social, o conteúdo e a descrição do evento podem ser considerados o fator equivalente que pode auxiliar na previsão do público (Lu et al., 2019; Zhang e Lv, 2018; Michalco e Navrat, 2012).

Diversos autores citam que a cidade ou região do evento tem um impacto grande na decisão de comparecimento. A distância entre a cidade natal do participante, ou até mesmo entre a sua casa, e o evento tem um impacto negativo na participação (Jiang e Li, 2019; Priporas et al., 2018; Zhang e Lv, 2018; Devesa et al., 2015; Govindasamy e Kelley, 2014; Cheng et al., 2012; Michalco e Navrat, 2012; Griffith, 2010). A adequação do local ao tipo de evento também é um fator decisivo, influenciando na qualidade percebida do festival e na sua atratividade perante ao público (Biondi et al., 2019; Shuo Zhang, 2019; Lu et al., 2019; Akgunduz e Coçar, 2018; Li, 2018; Santos-Lewis e Moital, 2017; Farrel et al., 2015; Coates, Humphreys e Zhou, 2014; Palma, Palma e Aguado, 2013; Michalco e Navrat, 2012).

A adequação do local também se refere à atratividade da região do evento e a imagem que o local passa aos participantes (Yahia e Al-Emadi, 2018). Vinculado ao local, Akgunduz e Coçar (2018) concordam com Kaplanidou e Gibson (2010) ao afirmar que a imagem local além de motivar a participação, também é um preditor do retorno do participante ao mesmo evento. Um fator muito importante quando o local é analisado, é a sua capacidade, visto que lugares maiores comportam um maior número de pessoas (Popp, Jensen e Jackson, 2017). A atratividade do local também é examinada pela cultura da região e se esta disponibiliza atrações turísticas fora do evento que possam atrair público de cidades distantes (Yahia e Al-Emadi, 2018). Vários autores concordam que o desejo de aprender sobre a cultura da região do evento pode ser um motivador (Priporas et al., 2018; Sung e Mills, 2018; Snelgrove e Wood, 2010).

Além disso, quanto maior for a cidade do evento, mais a população participa de eventos, pelo aumento do consumo de cultura e escolaridade (Muñiz, Rodríguez e Soaréz, 2017; Trinh e Lam, 2016; Farrel et al., 2015; Govindasamy e Kelley, 2014; Palma, Palma e Aguado, 2013). Alguns estudos detectaram que pessoas com nível superior compareciam mais a eventos culturais, enquanto o nível de escolaridade não tinha efeito significativo em eventos esportivos, por ser um tipo de evento popular (Devesa et al., 2015; Montgomery e Robinson, 2010). De acordo com Yazaci e Li (2019), ainda, a educação em geral tem efeito positivo na participação em eventos.

Um fator muito associado à escolaridade da população da cidade do evento, a renda média também foi citada por alguns autores. Palma, Palma e Aguado (2013) avaliaram que quanto maior a renda familiar, maior a tendência de participação em diferentes eventos. Wicker e Hallmann (2013) acrescentam também que quanto maior a renda familiar, mais as pessoas viajam para cidades distantes para participar de eventos. Sendo assim, a renda pode ser relacionada ao comparecimento a diferentes tipos de evento (Muñiz, Rodríguez e Soaréz, 2017; Born et al., 2014; Montgomery e Robinson, 2010).

Referente à data do evento, diferentes considerações podem ser feitas. Muitos estudos identificaram que o dia da semana e o horário do evento interferem na quantidade de pessoas que comparecem ao evento (Zhuo Zhang e Lv, 2019; Akgunduz e Coçar, 2018; Li, 2018; Zhang e Lv, 2018; Chen et al., 2014; Coates, Humphreys e Zhou, 2014; Santos-Lewis, 2013; Beckamn et al., 2012; Cheng et al., 2012; Michalco e Navrat, 2012). Por exemplo, finais de semana tendem a atrair uma maior quantidade de pessoas, uma vez que a maioria da população possui tempo livre durante esses dias (Bogaert, Ballings e Van Den Poel, 2016; Born et al., 2014; Kim et al., 2014; Mohan e Thomas, 2012). Além disso, o mês para o qual o evento está previsto interfere na quantidade total de público. Por exemplo, meses que caem em estações do ano mais quentes como primavera e verão são mais atrativas ou meses que caem em períodos de férias podem impactar positiva ou negativamente o público (Bogaert, Ballings e Van Den Poel, 2016; Beckman et al., 2012).

Outro fator que pode impactar de diferentes formas o público é a previsão do tempo para o dia do evento, uma vez que dias muito quentes ou chuvosos podem ter um impacto negativo (Akgunduz e Coçar, 2018; Sung e Mills, 2018; Yahia e Al-Emadi, 2018; Muñiz, Rodríguez e Soaréz, 2017; Bogaert, Ballings e Van Den Poel, 2016; Michalco e Navrat, 2012; Hall, O'Mahony e Vieceli, 2010). Porém, Kim et al. (2014) destaca que a temperatura do dia do evento não teve significância estatística na influência do público uma vez que o evento analisado acontecia dentro de pavilhões fechados. Isto demonstra que o tipo de evento, mais uma vez, é um fator de grande importância na quantidade de participantes. Popp, Jensen e Jackson (2017) também não detectaram significância para a temperatura média do dia, pois estavam analisando o comportamento de fãs de football americano e os mesmos eram puramente motivados por fatores emocionais.

Ainda relacionado com o contexto do evento, as facilidades se mostraram um fator recorrente na literatura, aparecendo em uma grande quantidade de artigos. As facilidades

exemplificadas foram facilidade de estacionamento, acessibilidade às pessoas com algum tipo de deficiência física, acomodação na cidade no caso de eventos de mais um dia, presença de filas e limpeza do evento. Um evento com facilidades adequadas pode ser associado a um bom nível de satisfação e entretenimento dos participantes. (Shuo Zhanf e LV, 2019; Akgunduz e Coçar, 2018; Priporas et al., 2018; Yahia e Al-Emadi, 2018; Bruwer e Kelley, 2015; Chen et al., 2014; Santos-Lewis e Moital, 2013; Cheng et al., 2012; Tanford, Montgomery e Nelson, 2012; Hall, Basarin e Lockstone-Binney, 2010; Hall, O'Mahony e Viecegli, 2010; Kaplanidou e Gibson, 2010).

Hall, O'Mahony e Viecegli (2010) acrescentam que se deve incluir nas facilidades a oferta de alimentação e bebidas a um preço acessível. O gasto com transporte e alimentação é um dos fatores analisados pelo público e ponderado na decisão de ir ao evento (Priporas et al., 2018; Yahia e Al-Emadi, 2018; Boyle et al., 2016). Além desse, o preço do ingresso também influencia na decisão, muitas vezes associado ao valor emocional para com o evento (Griffith, 2010; Sung e Mills, 2018; Boyle et al., 2016; Farrel et al., 2015; Santos-Lewis, 2013; Cheng et al., 2012; Michalco e Navrat, 2012; Tanford, Montgomery e Nelson, 2012).

Considerando o caso particular de eventos esportivos, existem algumas características contextuais que podem ser consideradas na previsão de participantes. Coates, Humphreys e Zhou (2014) concordam com Sung e Mills (2018) de que a probabilidade do time da casa ganhar está fortemente relacionada com o número de pessoas que vão ao jogo. Porém discordam quanto a hipótese da incerteza do resultado (UOH), uma vez que Sung e Mills (2018) encontra evidências para contestar esta teoria. A hipótese da incerteza de resultado (UOH) postula que os fãs de esportes valorizam competições competitivas, preferindo jogos e temporadas incertas àquelas mais seguras.

Além desses, outros autores utilizam o percentual de vitória dos times que estão se enfrentando como uma métrica de qualidade e, também, outras características do time visitante, como o tamanho da sua cidade natal para prever o número de participantes (Popp, Jensen e Jackson, 2017; Coates, Humphreys e Zhou, 2014; Beckman et al., 2012; Mohan e Thomas, 2012). Sung e Mills (2018) acrescentam que a presença de um jogador ilustre em algum dos times aumenta a atratividade da partida e, por consequência, o público total. Como uma métrica de qualidade da partida, o reconhecimento e importância da série ou liga a qual o evento se refere também pode influenciar positivamente o público (Popp, Jensen e Jackson, 2017; Beckman et al., 2012).

Por fim, Kim et al. (2014) identificaram que divulgação nos meios de comunicação, como o jornal, pode atrair mais pessoas para o evento. Nos seus estudos, eles apontaram que o efeito referente a uma notícia sobre o evento tem um pico no quarto dia após a liberação da manchete e então começa a cair, perdendo o efeito após o oitavo dia. Outros autores também relataram que o acesso à informação sobre o evento atrai mais pessoas (Chen et al., 2014; Dobson, 2010) e que informações educativas sobre o evento também aumentam o comparecimento (Hamerman, 2016).

2.3.2 MÉTODOS DE PREVISÃO DE PÚBLICO EM EVENTOS

Dentre os 17 artigos selecionados e que utilizaram algum método de previsão, a maioria os empregou em eventos EBSNs, totalizando 9 artigos, ou em eventos esportivos, totalizando 6 artigos. Os eventos EBSNs analisados extraíram dados de eventos organizados pelo MeetUp (Xu et al., 2019; Li, 2018; Zhang e Lv, 3018), Facebook Events (Bogaert, Ballings e Van den Poel, 2016; Michalco e Navrat, 2012) e DoubanEvent (Lu, 2019) e também publicações e perfis de usuários do Facebook e Twitter (De Lira, 2019; Parra et al., 2016). Todos esses estudos utilizaram *Machine Learning* para obtenção de um modelo de classificação, no qual utilizavam características dos usuários e do grupo organizador para classificar cada indivíduo em um grupo de comparecimento ou não-comparecimento.

Os métodos de *Machine Learning* encontram uma combinação de parâmetros a fim de minimizar os erros de classificação ou regressão, aprendendo com os erros dos dados conforme os mesmos são apresentados ao computador. Para isso, os dados são divididos em um subconjunto de treinamento, no qual o modelo é concebido, e outro de teste, no qual este modelo é validado (Géron, 2017). Além disso, *Machine Learning* pode ser utilizada para obter resultados de regressão ou de classificação dependendo do problema e dados disponibilizados. Existem muitos métodos de classificação na literatura, sendo os mais comuns de Regressão Logística, Análise de Discriminantes Linear (LDA) e Quadrática (QDA), Classificador de Naive Bayes, Matriz de Fatoração, Árvores de Decisão, Redes Neurais, Random Forest, Suport Vector Machines, Bagging, Gradient Boosting e Adaptative Boosting (James et al., 2013).

De Lira (2019), Bogaert, Ballings e Van den Poel (2016) e Parra et al. (2016) testaram métodos de Regressão Logística, Random Forest, Gradient Boosting, Bagging, Boosting, Redes Neurais e Support Vector Machines nas suas bases de dados. De Lira (2019) e Parra et al. (2016)

analisaram o conteúdo e o sentimento de tweets para classificar os usuários, enquanto Bogaert, Ballings e Van den Poel (2016) utilizaram características do perfil do usuário como likes, amizades, preferências do usuário. Além das características do perfil e do grupo, Xu et al. (2019) analisaram a influência de certos usuários na decisão de outros em ir ou não ao evento, utilizando o algoritmo de Análise de Discriminante Linear.

Ainda sobre EBSNs, Lu et al. (2019) desenvolveram um algoritmo próprio de classificação baseado em Gradient Boosting, chamado MFDAU (Multi-features Event Participation Prediction Model), diferenciando usuários ativos e inativos da rede. Já Li (2018) utilizou dados do evento da plataforma MeetUp e as respostas passadas dos usuários para prever o comparecimento utilizando Modelos de Regressão de Cox. Os modelos de regressão de Cox permitem incorporar facilmente covariáveis dependentes do tempo e, no caso estudado, a probabilidade de responder positivamente ao convite do evento é dependente do tempo. E, finalmente, Michalco e Navrat (2012) utilizaram dados de um questionário com usuários do Facebook Events, assim como dados de eventos organizados nessa plataforma, para obter um modelo de predição. O modelo desenvolvido foi uma soma de probabilidades associada a fatores identificados no questionário, cujos pesos também foram obtidos através deste questionário.

Dentro dos métodos de *Machine Learning*, Mamei e Colonna (2016) utilizaram o método de regressão log-linear (de Poisson) para estimar a quantidade de participantes de diferentes eventos a partir de dados de celulares. Além disso, este foi o único estudo que utilizou um método de regressão com *Machine Learning*. O método de regressão de Poisson assume uma distribuição de Poisson para a variável resposta e modela o logaritmo do seu valor esperado como uma combinação linear das variáveis independentes. Desta maneira é conhecido também como modelo log-linear (Cameron e Trivedi, 1998).

Por outro lado, os artigos relacionados a eventos esportivos utilizaram modelos de regressão para estimar o público sem aplicar a metodologia de *Machine Learning*. Sendo assim, estes modelos foram desenvolvidos sem que houvesse um aprendizado sobre o erro durante a sua concepção. Griffith (2010), por exemplo, analisou dados de partidas de futebol americano universitário, obtidas através do site oficial da NCAA College Football, testando modelos de regressão linear e regressão log-linear (de Poisson). Popp, Jensen e Jackson (2017) também estudaram partidas de futebol americano universitário, utilizando um modelo de regressão

hierárquica, a fim de obter as variáveis mais importantes para maximizar o público vindo de outras cidades.

Sung e Mills (2018) utilizaram partidas de futebol com dados obtidos no site da Major League Soccer, obtendo um modelo quadrático. O termo quadrado foi utilizado para estimar a relação não linear proposta entre a incerteza do resultado e o comparecimento proposta. Além deles, Coates, Humphreys e Zhou (2014) e Beckman et al. (2012) analisaram partidas de baseball ao longo de diversos anos. Coates, Humphreys e Zhou (2014) utilizaram um modelo estrutural econométrico baseado na incerteza do resultado, enquanto Beckman et al. utilizaram um modelo de regressão tobit, uma vez que o estádio é limitado pela sua capacidade, violando as questões de linearidade. O modelo tobit é um modelo de regressão no qual a variável resposta é censurada de alguma forma. Além disso, os coeficientes do modelo são atualizados a cada cinco anos, pois a importância de cada variável se alterou no período de 25 anos analisado.

Por fim, Lee et al. (2014) avaliaram o comparecimento a um Mega Evento, utilizando dados de um questionário realizado antes do evento e desenvolvendo um modelo de classificação com regressão logística. Trinh (2018) e Trinh e Lam (2016) utilizaram dados obtidos na pesquisa Multi-purpose Household Survey (MPHS) em 2009 e 2010 pela Australian Bureau of Statistics para avaliar o comparecimento em diversos tipos de eventos. Esses autores utilizaram a Teoria da Distribuição Binomial Negativa para ajustar a frequência percentual de participação observada e avaliaram o modelo através do erro médio percentual absoluto (MAPE).

2.4 ANÁLISE DOS RESULTADOS OBTIDOS NA LITERATURA

Esta seção analisa os estudos selecionados e apresenta de forma estruturada os fatores e métodos encontrados na literatura, prosseguindo com o passo (v) desta revisão sistemática. Para responder às questões de pesquisa, a seção está dividida em (a) Fatores que motivam a participação em eventos e (b) Métodos de previsão utilizados para prever o público.

2.4.1 FATORES QUE MOTIVAM A PARTICIPAÇÃO EM EVENTOS

Após a leitura e extração das citações e achados da literatura, foram identificados 47 fatores relacionados a motivações para o comparecimento em diferentes tipos de eventos. Conforme apresentado anteriormente, as motivações podem ser divididas em internas e

externas. Desta forma, manteve-se este agrupamento identificado na literatura, dividindo os fatores nesses dois grupos. Além disso, os fatores foram organizados em nove dimensões, de acordo com a sua afinidade ao tópico.

Os comentários extraídos da literatura foram traduzidos na forma de fatores, e os mesmos foram classificados nos grupos e dimensões por afinidade. Os grupos, dimensões e fatores encontrados podem ser vistos na Tabela 1, onde podem ser observados os fatores, sua classificação e referências. A lista original com a rastreabilidade dos comentários realizados por cada autor, que originaram os fatores e a sua organização, pode ser conferida no <https://www.ufrgs.br/producao/wp-content/uploads/2020/05/appendix-article.xlsx>.

Além da classificação, a Tabela 1 relaciona cada fator a uma pergunta. A pergunta foi desenvolvida de forma a auxiliar o entendimento dos fatores e a sua utilização na previsão de público em eventos pelos organizadores.

Tabela 1 – Resumo dos fatores encontrados na literatura e proposta de organização em perguntas para utilização de público em eventos

Grupo	Dimensão	Fator	Pergunta relacionada	Referências
Externos (A)	Histórica	Tipo de Evento	Qual o tipo de evento?	[5], [32], [34], [41], [43], [56]
		Público médio	Qual o público médio neste tipo de evento?	[6], [7], [24], [38], [40], [51]
Internos (B)	Populacional	Cidade/Local	A cidade/local é propícia a participação no evento?	[1], [2], [3], [9], [12], [16], [19], [20], [21], [24], [27], [28], [31], [32], [34], [38], [41], [43], [44], [56]
		Tamanho da população	Qual a população do local/região sede do evento?	[20], [37], [40]
		Penetração	Qual o percentual da população que teria interesse nesse evento?	[6], [15], [32], [34], [37], [38], [51], [52], [53], [56]
		Hábitos culturais	Os hábitos da população local são favoráveis a participação?	[9], [13], [16], [17], [19], [20], [25], [38], [41], [43], [48], [52]
		Escolaridade	O nível de escolaridade da população local é favorável?	[6], [16], [19], [20], [30], [34], [36], [37], [55]
		Renda média	A renda média da população é favorável a participação?	[5], [20], [36], [37], [38], [52]
		Convites virtuais	A cultura local responde a convites virtuais?	[4], [27], [31], [34]
	Contextual	Cultura local	A região proporciona atrações turísticas fora do evento?	[1], [41], [45], [47], [49], [54]
		Benefícios para região	O evento oferece benefícios para o bairro/cidade/região?	[13], [51]
		Inclusão	O evento é amplamente inclusivo?	[3]
		Premiações	O evento oferece premiações pela participação?	[19], [41], [55]
		Reputação do evento	O evento possui boa reputação perante o público?	[11], [23], [35], [41], [49]
		Causa nobre	O evento está associado a causas nobres?	[3], [9], [10], [45], [48], [49]
	Emocional	Interações	O evento facilita interações entre participantes?	[1], [3], [4], [7], [9], [13], [16], [19], [21], [23], [26], [32], [34], [42], [41], [43], [45], [48], [49], [53], [56]
		Ligação emocional	O evento provoca emoções nos participantes?	[1], [3], [8], [9], [11], [16], [19], [23], [24], [25], [26], [28], [35], [42], [45], [48], [52], [54], [55]
		Sair da rotina	O evento oferece uma experiência tipicamente "fora da rotina"?	[1], [7], [9], [30], [41], [45], [48], [49], [54], [55]
		Satisfação	O evento gera satisfação que favorece a participação?	[1], [7], [19], [22], [28], [42], [48], [54], [55]
		Popularidade	O evento propicia popularidade àqueles que atendem?	[48], [52]
Admiradores		A atração possui um grande número de fãs?	[35], [42], [44]	

[1] Akgunduz e Cosar, 2018; [2] Beckman et al., 2012; [3] Biondi et al., 2019; [4] Bogaert, Ballings e Van den Poel, 2016; [5] Born et al., 2014; [6] Boyle et al., 2016; [7] Bruwer e Kelley, 2015; [8] Chen et al., 2014; [9] Cheng et al., 2012; [10] Chiu, Lee e Won, 2016; [11] Chiu, Won e Leng, 2019; [12] Coates, Humphreys e Zhou, 2014; [13] Cushing, 2018; [15] De Lira et al., 2019; [16] Devesa et al., 2015; [17] Dobson, 2010; [19] Farrel et al., 2015; [20] Govindasamy e Kelley, 2014; [21] Griffith, 2010; [22] Hall, Basarin e Lockstone-Binney, 2010; [23] Hall, O'Mahony e Vieceli, 2010; [24] Halpenny, Kulczycki e Moghimehfar, 2016; [25] Hamerman, 2016. [26] Hopkeins et al., 2016; [27] Jiang e Li, 2019; [28] Kaplanidou e Gibson, 2010; [29] Kim et al., 2014; [30] Lee et al., 2014; [31] Li, 2018; [32] Lu et al., 2019; [34] Michalco e Navrat, 2012; [35] Mohan e Thomas, 2012; [36] Montgomery e Robinson, 2010; [37] Muñoz, Rodríguez e Suárez, 2017; [38] Palma, Palma e Aguado, 2013; [39] Parra et al., 2016; [40] Popp, Jensen e Jackson, 2017; [41] Priporas et al., 2018; [43] Rodríguez-Campo et al., 2019; [43] Santos-Lewis e Moital, 2013; [44] Shuo Zhang e Lv, 2019; [45] Snelgrove e Wood, 2010; [47] Sung e Mills, 2018; [48] Tanford, Montgomery e Hertzman, 2012; [49] Tanford, Montgomery e Nelson, 2012; [51] Trinh e Lam, 2016; [52] Wicker e Hallmann, 2013; [53] Xu et al., 2019; [54] Yahia e Al-Emadi, 2018; [55] Yazici, Koçak e Altunsoz, 2017; [56] Zhang e Lv, 2018.

Tabela 2 – Resumo dos fatores encontrados na literatura e proposta de organização em perguntas para utilização de público em eventos (Continuação).

Grupo	Dimensão	Fator	Pergunta relacionada	Referências
Externos (C)	Infraestrutura	Acessibilidade	O evento oferece ampla acessibilidade a PCDs?	[1], [7], [3], [23], [43], [49]
		Transporte Público	O evento possui fácil acesso por transporte público?	[1], [7], [9], [22], [23], [28], [49]
		Estacionamento	Existe facilidade para estacionamento a preços acessíveis?	[1], [7], [8], [9], [22], [23], [28], [49]
		Existência de filas	Há previsão de filas que vão causar perda de tempo?	[7], [22]
		Alimentos e bebidas	O evento oferece alimentos/bebidas a preços acessíveis?	[1], [6], [7], [8], [20], [22], [44], [54]
		Amenidades	O local do evento é agradável e favorece a participação?	[1], [7], [8], [9], [22], [23], [28], [43], [44], [49], [54]
		Capacidade	Qual a capacidade do local onde irá ocorrer o evento?	[12], [40], [44], [47]
		Conforto	As pessoas podem ficar confortáveis durante o evento?	[1], [2], [7], [8], [22], [23], [49], [54]
	Temporal	Dia da semana	O dia da semana é favorável a participação?	[1], [2], [4], [5], [8], [9], [12], [29], [31], [34], [35], [43], [44], [56]
		Duração	A duração do evento é favorável a participação?	[4], [31]
		Hora do evento	O horário do evento é favorável a participação?	[1], [2], [4], [8], [9], [12], [31], [32], [34], [43], [44]
		Mês	O mês em que o evento irá ocorrer (estação do ano, férias) é favorável?	[2], [4], [29]
		Clima	A previsão do tempo para o dia ou dias do evento é favorável?	[1], [4], [23], [29], [34], [37], [40], [47], [54]
	Preço	Valor do ingresso	Qual o valor do ingresso comparado a outros eventos similares?	[6], [9], [19], [21], [34], [41], [43], [47], [49]
	Marketing	Propaganda de massa	Qual o nível de divulgação do evento em TV, Rádio, Jornal?	[8], [17], [25], [29]
		Propaganda virtual	Qual o nível divulgação do evento em mídias eletrônicas?	[8], [17], [25], [29]
		Interessados virtuais	Quantos usuários indicaram virtualmente interesse pelo evento?	[15], [39], [44]
		Tags do evento	Quantas tags (#) estão associadas ao evento?	[27]
		Número de convidados	Quantidade de pessoas convidadas para o evento?	[27], [53], [56]
		Presença Celebidades	Há influenciadores ou celebridades confirmadas para o evento?	[4], [34], [55]
	Esportiva	Jogos decisivos	O evento possui um caráter decisivo (no caso de jogos)?	[2], [47]
		Liga/Campeonato	O evento pertence a uma série ou campeonato importante?	[2], [40]
		Presença de Craques	O jogo terá a presença de jogadores estrelas?	[47]
		Resultado do jogo	O resultado esperado para o evento favorece a participação?	[12], [47]
		Time da casa	O time da casa está bem na série ou campeonato?	[2], [12], [21], [40], [47]
		Time visitante	O time visitante possui uma grande torcida ou reputação?	[2], [12], [47]

[1] Akgunduz e Cosar, 2018; [2] Beckman et al., 2012; [3] Biondi et al., 2019; [4] Bogaert, Ballings e Van den Poel, 2016; [5] Born et al., 2014; [6] Boyle et al., 2016; [7] Bruwer e Kelley, 2015; [8] Chen et al., 2014; [9] Cheng et al., 2012; [10] Chiu, Lee e Won, 2016; [11] Chiu, Won e Leng, 2019; [12] Coates, Humphreys e Zhou, 2014; [13] Cushing, 2018; [15] De Lira et al., 2019; [16] Devesa et al., 2015; [17] Dobson, 2010; [19] Farrel et al., 2015; [20] Govindasamy e Kelley, 2014; [21] Griffith, 2010; [22] Hall, Basarin e Lockstone-Binney, 2010; [23] Hall, O'Mahony e Vieceli, 2010; [24] Halpenny, Kulczycki e Moghimehfar, 2016; [25] Hamerman, 2016. [26] Hopkeins et al., 2016; [27] Jiang e Li, 2019; [28] Kaplanidou e Gibson, 2010; [29] Kim et al., 2014; [30] Lee et al., 2014; [31] Li, 2018; [32] Lu et al., 2019; [34] Michalco e Navrat, 2012; [35] Mohan e Thomas, 2012; [36] Montgomery e Robinson, 2010; [37] Muñiz, Rodríguez e Suárez, 2017; [38] Palma, Palma e Aguado, 2013; [39] Parra et al., 2016; [40] Popp, Jensen e Jackson, 2017; [41] Priporas et al., 2018; [43] Rodríguez-Campo et al., 2019; [43] Santos-Lewis e Moital, 2013; [44] Shuo Zhang e Lv, 2019; [45] Snelgrove e Wood, 2010; [47] Sung e Mills, 2018; [48] Tanford, Montgomery e Hertzman, 2012; [49] Tanford, Montgomery e Nelson, 2012; [51] Trinh e Lam, 2016; [52] Wicker e Hallmann, 2013; [53] Xu et al., 2019; [54] Yahia e Al-Emadi, 2018; [55] Yazici, Koçak e Altunsoz, 2017; [56] Zhang e Lv, 2018.

(i) Fatores Externos

Foram identificados 28 fatores externos, organizados em seis dimensões. A primeira dimensão, a Histórica, captura o Tipo de Evento que se está analisando, assim como o Público Médio a partir da existência de algum histórico de participação em edições passadas ou em eventos do mesmo tipo. Esta dimensão foi citada por 12 estudos diferentes e, em comparação com outras dimensões, não é a mais citada. Porém, é o ponto de partida para análise, visto que o tipo de evento condiciona as suas características e, conseqüentemente, os outros fatores (Born et al., 2014). Desta forma, esta dimensão foi relacionada ao grupo Fatores Externos (A), por ser o primeiro passo de uma abordagem de previsão de público.

As outras dimensões de Fatores Externos foram atribuídas ao grupo Externos (C). Essas dimensões capturam as particularidades de cada evento. A primeira delas é a Infraestrutura, que identifica a existência de acessibilidade para pessoas com deficiência física, facilidade de transporte até o evento ou estacionamento, capacidade do local, presença de filas e oferta de alimentos e bebidas. Além dessas, esta dimensão busca entender se o espaço é agradável e limpo e se as pessoas ficam confortáveis durante o evento, com local para descanso e sem aglomerações. Percebe-se na Tabela 1, que os fatores desta dimensão foram citados por diversos autores, evidenciando a importância de se ter dimensionada essa categoria. Ademais, todos esses fatores foram associados à satisfação dos participantes por diferentes autores (Hall, Basarin e Lockstone-Binney, 2010; Hall, O'Mahony e Vieceli, 2010). Sendo assim, esta dimensão também captura a qualidade do evento e possível retorno dos participantes para edições futuras, mensurando a lealdade dos seus participantes (Bruwer e Kelley, 2015; Kaplanidou e Gibson, 2010).

A segunda dimensão neste grupo é a Temporal, que identifica se o dia da semana, a duração do evento, a hora, o mês e o clima são favoráveis a participação no evento. Esta dimensão foi muito citada na literatura como influente na decisão de comparecimento. Contudo, foi observado que os fatores desta dimensão podem ser favoráveis ou desfavoráveis a participação. Por exemplo, Kim et al. (2014) não acharam significância na temperatura do dia, pois o evento foi realizado em pavilhões fechados. Além disso, a ocorrência do evento em mês de férias pode ou não atrair mais pessoas, de acordo com seu tipo, por exemplo.

Outra dimensão relacionada a fatores externos é o Preço, que foi citado por 9 autores e engloba o custo do ingresso do evento. O custo do ingresso deve ser analisado comparando-o com o valor de eventos similares ou de uma edição passada, visto que o público geralmente

realiza esta análise e se predispõe a pagar certo valor de acordo com o tipo de evento. Desta maneira, não se deve analisar o valor direto do ingresso, mas sim se este está mais caro, na média ou mais barato que os similares.

Além destas, outra dimensão que requer atenção é a de Marketing. Essa dimensão possui atributos que podem ser utilizados pelos organizadores para atrair mais público aos eventos. Os fatores de propaganda de massa e propaganda virtual, por exemplo, medem o nível de divulgação e facilidade de acesso a informações do evento pelos participantes em diferentes tipos de mídias. Já o fator número de convidados, mede o potencial de público já confirmado para o evento. Além destes, o fato da presença de celebridades no evento pode influenciar mais pessoas a participar. E, no caso de ser um evento organizado em redes virtuais, é interessante monitorar o número de indivíduos que manifestaram interesse no evento bem como as tags relacionadas.

Por fim, foi identificada a dimensão Esportiva, para agrupar alguns fatores que são específicos de jogos esportivos. Esta dimensão captura a importância e relevância dos jogos, definindo se o jogo possui um caráter decisivo, como jogos de finais ou contra adversários rivais, e se está relacionado a alguma liga ou campeonato importante. Também usa fatores como Time da casa e Time visitante para medir a qualidade dos times a partir das suas posições no campeonato, reputação e tamanho da torcida. Assim como na dimensão de Marketing, os fatores Esportivos utilizam a Presença de Craques no jogo para mensurar a atratividade do evento que pode aumentar o público. Por último, o fator Resultado do Jogo utiliza a Hipótese da Incerteza da Partida para verificar se o resultado esperado é favorável ao comparecimento. Esta hipótese foi citada na literatura por vários autores e diz que, quanto mais incerto for o resultado da partida, maior o público. Além disso, incorpora a ideia de que a participação em um evento esportivo depende do resultado real do jogo em relação a um ponto de referência que reflete a expectativa do participante (Coates, Humphreys e Zhou, 2014).

(ii) Fatores Internos

No âmbito de fatores internos foram identificados 19 fatores reunidos em três dimensões. Estas dimensões tentam capturar as preferências e emoções dos participantes de forma a serem mensuráveis. A primeira dimensão, a Populacional, identifica os hábitos e características da população alvo do evento. Por exemplo, a Cidade/Local mede o quanto a região é favorável à participação no evento em estudo. Dependendo do seu tamanho, deve-se

considerar como local a cidade ou apenas o bairro no qual o evento será realizado. O tamanho da população desta região também é um fator que influencia no público, pois delimita o alcance do evento. Já a Penetração analisa o percentual desta população que potencialmente tem interesse no tipo de evento analisado, restringindo ainda mais a sua abrangência. Esta dimensão também avalia se os Hábitos Culturais da população são favoráveis ao tipo de evento em estudo. Por exemplo, o evento pode ser melhor aceito em uma região do país, devido à cultura da região. Similarmente, os fatores de Escolaridade e Renda média da região podem ser indicativos do público. Além disso, é importante verificar se a população responde bem a Convites virtuais, que passam a ser o principal veículo de promoção de muitos tipos de eventos.

A segunda dimensão, a Contextual, trata dos atributos do evento que podem trazer benefícios para a região e também de como o evento é percebido pela população. Para isso, são levantados fatores como: ser um evento inclusivo, estar associado a alguma causa nobre e oferecer premiações aos participantes. Além disso, a reputação do evento perante o público é relevante por capturar o sentimento de prestígio que os participantes sentem a ir a eventos importantes. Quanto à região e contexto, é interessante verificar os benefícios que o evento pode trazer e quais os atributos e pontos turísticos que a região tem que podem atrair maior público.

Por último, a dimensão Emocional visa estimar os sentimentos individuais dos participantes, sendo os fatores mais citados na literatura em 35 estudos. O primeiro fator, o mais citado dentre os 47 fatores, é as Interações entre os participantes, que avalia o grau de interação e troca de experiências que o evento propicia. Por exemplo, feiras e convenções irão proporcionar mais interação do que concertos e jogos de futebol, nos quais os participantes são apenas espectadores. Outro fator relevante é a Ligação emocional que o evento pode proporcionar entre os participantes. Este fator compreende desde a paixão por determinado artista como ser um evento de caráter emotivo e social. A literatura indica que a ligação emocional dos participantes está correlacionada com a intenção de participação a eventos, tornando essa dimensão um aspecto chave para compreender as motivações do comparecimento. Aliado a isso, o número de Admiradores ou fãs de artistas ou bandas do evento é outro fator que contribui na predição do público. Outro fator considerado na dimensão Emocional é o fato de o evento ser capaz de proporcionar uma experiência fora da rotina para seus participantes. O nível de Satisfação em ir ao evento, proporcionada por boa organização, infraestrutura adequada e atividades que despertam emoções também são motivadores da

participação. Por fim, a Popularidade que os participantes podem ganhar com seus amigos e familiares por ter ido ao evento que consideram bem-conceituado, também, é um fator a ser considerado.

2.4.2 MÉTODOS PARA PREVISÃO DE PÚBLICO EM EVENTOS

Após leitura dos artigos selecionados, foram identificados 23 métodos diferentes na literatura, sendo que alguns estudos utilizaram mais de um método para previsão. Esses métodos foram distribuídos em três grupos, de acordo com a metodologia aplicada durante a pesquisa, são eles: *Machine Learning*, Métodos de Regressão e Outros Métodos. Na **Tabela 3** pode ser vista a classificação dos métodos de acordo com o grupo e também o tipo de dado utilizado na pesquisa.

Nota-se que o grupo de *Machine Learning* é o mais explorado dentre os três, totalizando nove artigos. Assim como também é o grupo de maior quantidade de métodos associados, com 14 métodos. Por outro lado, os métodos de *Machine Learning* foram utilizados apenas com eventos EBSNs, o que pode ser atribuído a necessidade de grandes bancos de dados para treinamento e teste do modelo. Sendo assim, esses estudos utilizaram dados de eventos organizados no *MeetUp*, *Facebook Events*, *DoubanEvent* e *Twitter*, como pode ser visto na **Tabela 3**. Além disso, os estudos que utilizaram *Machine Learning* aplicaram mais de um algoritmo em seus bancos de dados a fim de comparar qual seria o melhor para o problema proposto, uma prática cada vez mais empregada.

O segundo maior grupo é o de Métodos de Regressão, com seis estudos associados. Este grupo foi majoritariamente utilizado com eventos esportivos, cujos dados foram coletados em sites oficiais do esporte em análise. Além disso, a maior parte são métodos de regressão linear com alguma variação, como o modelo tobit e a regressão log-linear. Por fim, o terceiro grupo é o de Outros Métodos Estocásticos, que utilizam fundamentos da probabilidade para modelar seus dados e obter um modelo explicativo do comparecimento. Nesses casos, os dados foram obtidos através de questionários e *surveys*.

Tabela 3 - Métodos de previsão de público encontrados na literatura

Grupo	Método	Dados	Referências
<i>Machine Learning</i>	Regressão Logística	Facebook Events	[4]
		MeetUp	[31]
		Twitter	[15], [39]
	Adaptative Boosting	Facebook Events	[4]
		Twitter	[39]
	Gradient Boosting Decision Trees	DoubanEvent	[32]
		Twitter	[15]
	Naive Bayes	Facebook Events	[4]
		Twitter	[15]
	Random Forest	Facebook Events	[4]
		Twitter	[15], [39]
	Regressão Log-linear (de Poisson)	Dados de ligações de celular	[33]
	Matriz de Fatoração	MeetUp	[56]
	MFDAU	DoubanEvents	[32]
	Modelos de Regressão de Cox	MeetUp	[31]
Análise de Discriminante Linear - LDA	MeetUp	[53]	
Bagging	Twitter	[39]	
Bayesian Networks	MeetUp	[31]	
Redes Neurais	Facebook Events	[4]	
Suport Vector Machines	Twitter	[15], [39]	
Métodos de Regressão	Regressão Linear	Site oficial do Major League Baseball	[12]
		Site oficial do NCAA College Football	[21]
	Regressão Hierárquica	Site oficial do NCAA College Football	[40]
	Regressão Linear com Stepwise	Site oficial do NCAA College Football	[21]
	Regressão Linear Truncada (tobit)	Site oficial do Chicago Tribune	[2]
	Regressão Log-linear (de Poisson)	Site oficial do NCAA College Football	[21]
	Regressão Logística	Questionário antes do evento	[30]
Regressão Quadrática	Site oficial do Major League Soccer	[47]	
Outros Métodos	Distribuição Binomial Negativa (NBD)	Dados da pesquisa MPHS 2009-10	[50], [51]
	Soma de probabilidades de cada fator	Questionários e FacebookEvents	[34]

[2] Beckman et al., 2012; [4] Bogaert, Ballings e Van den Poel, 2016; [12] Coates, Humphreys e Zhou, 2014; [15] De Lira et al., 2019; [21] Griffith, 2010; [30] Lee et al., 2014; [31] Li, 2018; [32] Lu et al., 2019; [33] Mamei e Colonna, 2016; [34] Michalco e Navrat, 2012; [39] Parra et al., 2016; [40] Popp, Jensen e Jackson, 2017; [47] Sung e Mills, 2018; [50] Trinh, 2018; [51] Trinh e Lam, 2016; [53] Xu et al., 2019; [56] Zhang e Lv, 2018

Vale observar também que alguns métodos podem ser e foram utilizados com metodologias e aplicações diferentes, como pode ser visto na **Tabela 3** e também na **Figura 7**. Por exemplo, os métodos de Regressão Logística e Regressão Log-linear foram utilizados em métodos de *Machine Learning* e também apenas realizando uma modelagem direta dos dados.

A metodologia aplicada irá depender principalmente da quantidade e qualidade dos dados e os recursos computacionais disponíveis.

Na **Figura 7**, também, é possível visualizar o número de vezes que os métodos foram utilizados nos estudos bem como o grupo nos quais foram categorizados. Observa-se que os métodos de classificação associados a *Machine Learning* foram os mais presentes nos estudos. Ainda, o método de Regressão Logística foi o mais utilizado, aparecendo em cinco estudos e seguido de *Random Forest* em três artigos.

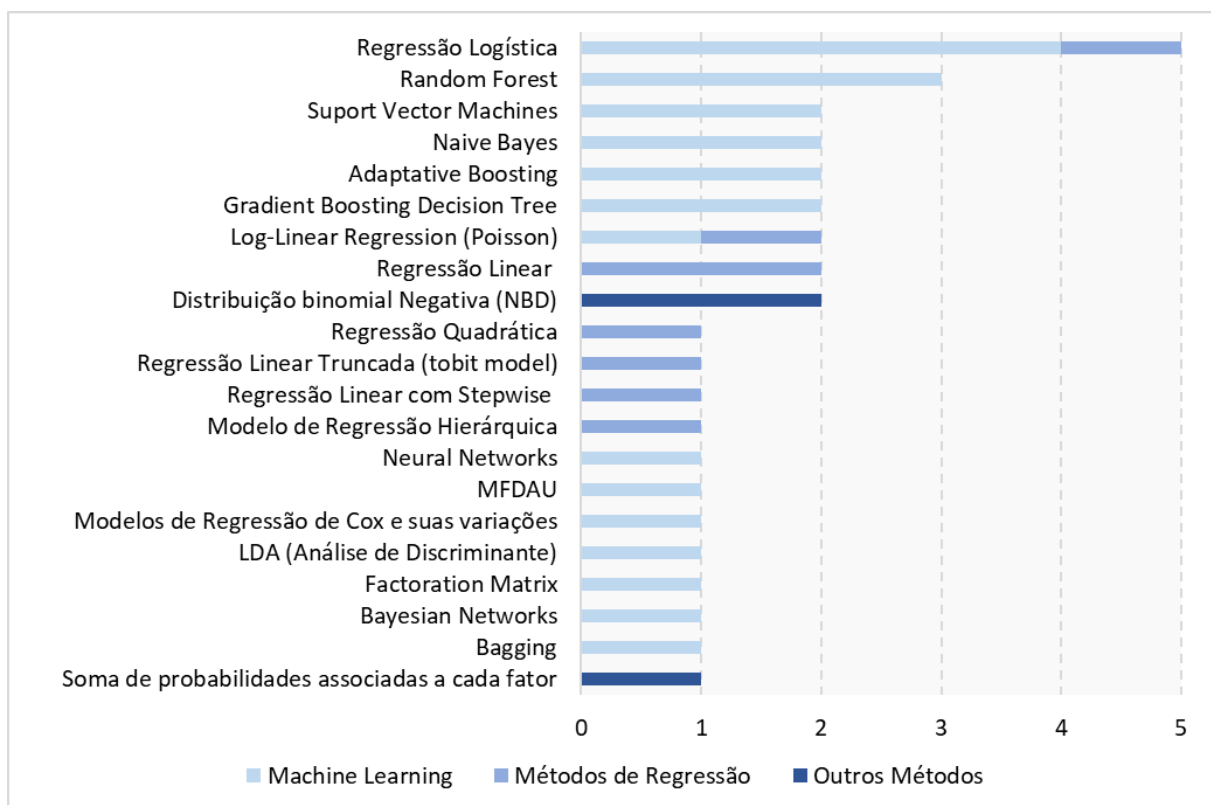


Figura 7 - Métodos de previsão de público encontrados na literatura

Por fim, ainda, a **Figura 8** apresenta os artigos analisados de acordo com os resultados obtidos, se foram de regressão ou de classificação e também ao tipo de evento analisado. Percebe-se que o número de estudos que utilizaram métodos de classificação é um pouco maior do que aqueles que utilizaram métodos de regressão, totalizando 53%. Dentre estes, a maioria utilizou métodos de classificação em eventos EBSN, com a exceção de um artigo, que classificou os participantes em relação a um Megaevento. Outro ponto importante é que os

eventos EBSN foram analisados apenas com métodos de classificação, enquanto os eventos esportivos foram em sua totalidade analisados com métodos de regressão.

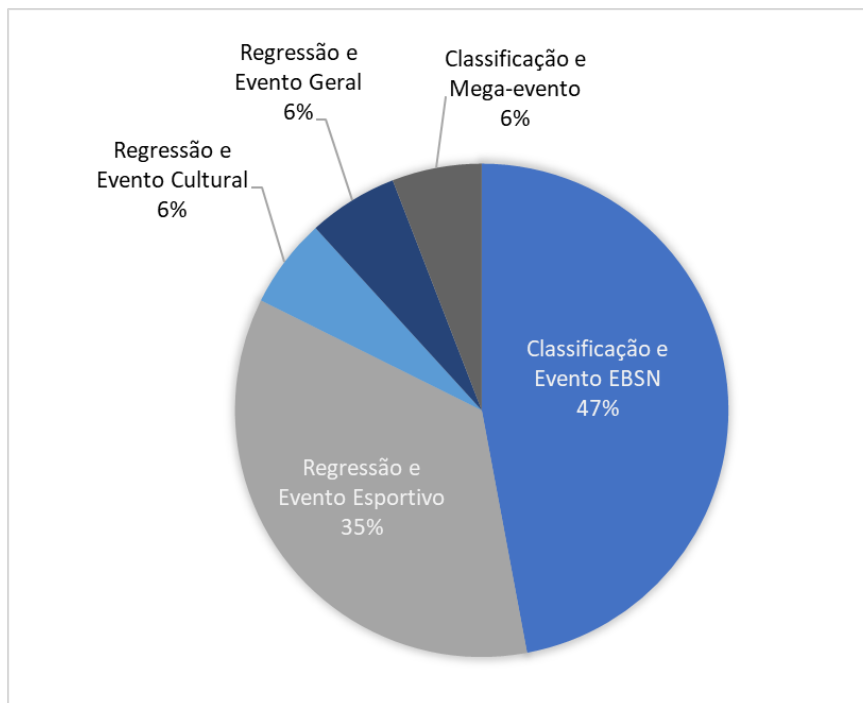


Figura 8 - Métodos de previsão associados ao tipo de evento

2.5 DISCUSSÃO

Entender as motivações para o comparecimento em eventos é fundamental para obtenção de uma estimativa confiável do público, o que facilita o gerenciamento e o monitoramento do evento (Davies, Coleman e Ramchandani, 2011). Diversos autores afirmam que o gerenciamento correto de eventos traz benefícios sociais e econômicos para a região, além de melhorar a experiência do usuário e aumentar a lealdade (Zhang e Lv, 2018; Lee et al., 2014). Porém, as informações acerca dessas motivações encontravam-se dispersas na literatura e segmentadas por tipo de evento. Este artigo propôs uma estruturação dessas motivações, identificando um total de 47 fatores em dois grupos de fatores internos e externos e classificados em nove dimensões.

Juntamente com a identificação dos fatores, foram levantadas perguntas associadas a cada um deles, conforme apresentado nas **Tabelas 1 e 2**. Essas perguntas podem ser utilizadas para a estimar a situação referente a cada fator a fim de obter uma previsão do público do evento.

O estudo sugere que a estimativa inicial de uma previsão de público seja realizada via o bloco de Fatores Externos (A), compreendido pela Dimensão Histórica, visto que este grupo é o principal balizador do evento, capturando o histórico de participação (Born et al., 2014). Paralelamente, também é indicado realizar uma avaliação do evento via os Fatores Internos e as demais dimensões de Fatores Externos. Conforme explicitado anteriormente, estas dimensões irão capturar as preferências e costumes da população local alvo do evento, bem como fatores externos que podem ser favoráveis ou desfavoráveis a participação.

Cada dimensão irá resultar em um coeficiente a partir da avaliação dos fatores que podem ser favoráveis ou desfavoráveis a participação em um evento específico, que poderá aumentar ou diminuir a previsão inicial. Por exemplo, um dia ensolarado pode trazer mais participantes do que um dia chuvoso, aumentando a estimativa inicial obtida através do histórico de participação. Da mesma forma, a participação de celebridades pode aumentar o público no evento específico que está em estudo.

Quanto ao método a ser utilizado, irá depender da quantidade de dados disponíveis. Foi verificado na literatura que a metodologia de *Machine Learning* está sendo explorada de forma crescente, porém ficando limitada a disponibilidade de grandes volumes de dados. Por esta razão, tem sido aplicada apenas em eventos EBSNs, onde grandes volumes de dados podem ser acessados mais facilmente. De forma geral, sugere-se a utilização de regressão linear e suas variantes para obtenção dos coeficientes relacionados a cada fator. Caso os dados sejam abundantes, é interessante utilizar *Machine Learning* em conjunto com métodos mais avançados, como *Random Forest*, *Gradient Boosting Decision Trees* e *Adaptative Boosting*.

2.6 CONCLUSÃO

Visando contribuir para o gerenciamento eficiente de eventos e melhorar a experiência dos participantes, através de uma revisão sistemática da literatura, este estudo investigou os fatores que impactam a participação em eventos e os métodos que têm sido utilizados para previsão de público.

Para responder a estas duas questões, foram buscados estudos nas duas bases de dados científicas de maior significância, a ScienceDirect e a Scopus. Após remoção das duplicatas, um total de 772 artigos foram para a etapa de análise de títulos, resumos e palavras-chaves. Após esta análise, restaram 80 artigos potencialmente relevantes nos quais foram aplicados um filtro de qualidade, removendo aqueles que foram publicados em periódicos de menor impacto.

Por fim, após inclusão de artigos obtidos das referências, chegou-se ao conjunto final de 56 artigos selecionados para o estudo.

A partir dos achados de cada estudo, foram identificados 47 fatores que afetam o comparecimento em eventos. Os fatores foram obtidos através dos comentários feitos nos 56 artigos acerca das motivações por trás da participação em diferentes tipos de eventos. Esses fatores foram organizados em nove dimensões e associados a perguntas que auxiliam os organizadores na gestão dos eventos e na previsão do público, apoiando a tomada de decisões.

Dentre as nove dimensões identificadas neste estudo, destacam-se a Populacional, a Emocional, a de Infraestrutura e a Temporal, sendo as mais recorrentes na literatura como influentes na decisão de comparecer a diferentes eventos. Além disso, os fatores Interação entre participantes e Ligação emocional foram os mais citados. Assim, avaliar a condição desses fatores é um aspecto importante para a previsão do público do evento. Já os fatores que integram a dimensão de Infraestrutura são fundamentais para entender o nível de satisfação dos participantes, que irá definir o retorno em novas edições do evento.

Como principal contribuição, este artigo propôs uma organização dos fatores em dimensões e, para cada fator, uma pergunta que pode auxiliar sua avaliação para previsão de público em eventos. Até este momento, nenhuma outra pesquisa reuniu os fatores e motivações para participação em eventos de forma completa, sem restringir a um tipo de evento específico. Assim, este estudo oferece organização aos conhecimentos que se encontravam dispersos na literatura.

Além de profissionais que trabalham com eventos, este estudo beneficia a literatura acadêmica, avançando a pesquisa na área e reunindo os avanços feitos até agora. Os achados da literatura também indicaram que há poucos avanços quanto a previsão de público em eventos, uma vez que apenas 17 estudos dentre os selecionados utilizaram algum método de previsão. Além disso, na sua maioria, as previsões foram aplicadas apenas a eventos EBSNs e com a utilização apenas de métodos de classificação. Observa-se ainda que a maior parte dos métodos de regressão foram aplicados a eventos esportivos. Desta forma, há espaço para trabalhos que envolvam métodos de regressão em diferentes tipos de eventos.

Neste contexto, a estrutura apresentada neste estudo pode ser utilizada para estabelecer um procedimento de previsão para diferentes tipos de eventos, facilitando a obtenção de uma estimativa quantitativa para público total. Ademais, as perguntas associadas a cada fator podem auxiliar os organizadores a mensurar a qualidade do seu evento e avaliar os pontos que devem

ser trabalhados para alavancar o público e manter a lealdade e eventual retorno dos participantes.

2.7 REFERÊNCIAS

- ADEMA, K. L.; ROEHL, W. S.; Environmental scanning the future of event design. **International Journal of Hospitality Management**, v. 29, p. 199-207, 2010
- AKGUNDUZ, Yilmaz; COŞAR, Yeşim. Motivations of event tourism participants and behavioural intentions. **Tourism and Hospitality Management**, [s. l.], v. 24, n. 2, p. 341–358, 2018.
- BECKMAN, Elise M. et al. Explaining Game-to-Game Ticket Sales for Major League Baseball Games Over Time. **Journal of Sports Economics**, [s. l.], v. 13, n. 5, p. 536–553, 2012.
- BERRIDGE, G.; Event pitching: The role of design and creativity. **International Journal of Hospitality Management**, v. 29, p. 208-215, 2010.
- BIONDI, L. et al. Understanding orchestrated participatory cultural initiatives: Mapping the dynamics of governance and participation. **Cities**, [s. l.], v. 96, n. April 2019, p. 102459, 2019.
- BOGAERT, Matthias; BALLINGS, Michel; VAN DEN POEL, Dirk. The added value of Facebook friend’s data in event attendance prediction. **Decision Support Systems**, [s. l.], v. 82, p. 26–34, 2016.
- BORN, Kathryn et al. Joint model of weekend discretionary activity participation and episode duration. **Transportation Research Record**, [s. l.], n. 2413, p. 34–44, 2014.
- BOYLE, Stephen et al. Predicting international visitors’ interest in Hong Kong arts festival. **Event Management**, [s. l.], v. 20, n. 4, p. 593–605, 2016.
- BRINER, B.; DENYER, D. **Systematic Review and Evidence Synthesis as a Practice and Scholarship Tool**. Newgen, pp.112-129, 2012.
- BRUWER, Johan; KELLEY, Kathleen. Service performance quality evaluation and satisfaction in a USA wine festivalscape buying behavioural effects. **International Journal of Event and Festival Management**, [s. l.], v. 6, n. 1, p. 18–38, 2015.
- CAMERON, A.C.; TRIVEDI, P.K. **Regression analysis of count data**, Cambridge University Press, 1998.
- CHEN, Po Ju et al. Can fundraising be fun? An event management study of unique experiences, performance and quality. **Tourism Review**, [s. l.], v. 69, n. 4, p. 310–328, 2014.

- CHENG, Chih Fu et al. Fan participation behaviour in baseball: An application of the theory of planned behaviour. **International Journal of Sports Marketing and Sponsorship**, [s. l.], v. 14, n. 1, p. 22–33, 2012.
- CHIU, Weisheng; LEE, Young joo; WON, Doyeon. Bifactor analysis of motivation for charity sport event participation. **International Review on Public and Nonprofit Marketing**, [s. l.], v. 13, n. 1, p. 1–13, 2016.
- CHIU, Weisheng; WON, Doyeon; LENG, Ho Keat. The relationship between sport involvement, perceived event prestige, and the intention to attend annual spectator events. **Asia Pacific Journal of Marketing and Logistics**, [s. l.], v. ahead-of-p, n. ahead-of-print, p. 1405–1421, 2019.
- COATES, Dennis; HUMPHREYS, Brad R.; ZHOU, Li. Reference-dependent preferences, loss aversion, and live game attendance. **Economic Inquiry**, [s. l.], v. 52, n. 3, p. 959–973, 2014.
- CUSHING, Amber L. “We’ve no problem inheriting that knowledge on to other people”: Exploring the characteristics of motivation for attending a participatory archives event. **Library and Information Science Research**, [s. l.], v. 40, n. 2, p. 135–143, 2018.
- DAVIES, Larissa; COLEMAN, Richard; RAMCHANDANI, Girish. Measuring attendance: Issues and implications for estimating the impact of free-to-view sports events. **International Journal of Sports Marketing and Sponsorship**, [s. l.], v. 12, n. 1, p. 11–23, 2011.
- DE LIRA, Vinicius Monteiro et al. Event attendance classification in social media. **Information Processing and Management**, [s. l.], v. 56, n. 3, p. 687–703, 2019.
- DENYER, D.; TRANFIELD, D.; **Producing a Systematic Review**. In: *The Sage Handbook of Organizational Research Methods*, edited by D. Buchanan and A. Bryman, 671-689. London: Sage. 2009.
- DEVESA, María et al. Factors determining attendance at a film festival. **Event Management**, [s. l.], v. 19, n. 3, p. 317–330, 2015.
- DOBSON, Melissa C. New audiences for classical music: The experiences of non-attenders at live orchestral concerts. **Journal of New Music Research**, [s. l.], v. 39, n. 2, p. 111–124, 2010.
- DŽUPKA, Peter; ŠEBOVÁ, Miriam. Local economic impact of the white night festival in Košice. **E a M: Ekonomie a Management**, [s. l.], v. 19, n. 2, p. 132–141, 2016.
- Elsevier, 2019. Disponível em: <<https://www.elsevier.com/solutions/scopus>>. Acesso em: 23 set. 2019.

- Elsevier, 2019. Disponível em: <<https://www.elsevier.com/solutions/sciencedirect>>. Acesso em: 23 set. 2019.
- FARRELL, Jocelyn M. et al. Dog-sport competitors: What motivates people to participate with their dogs in sporting events? *Anthrozoos*, [s. l.], v. 28, n. 1, p. 61–71, 2015.
- GÉRON, A. **Hands-on Machine Learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems**. O'Reilly Media, Inc, 2017.
- GETZ, D.; Event Tourism definition, evolution and research. *Tourism Management*, v. 29, p. 403-428, 2008.
- GETZ, D.; *Event Studies: Theory, Research and Policy for Planned Events*. Elsevier, Oxford, 2008.
- GETZ, D.; The nature and scope of festival studies. *International Journal of Event Management Research*, v. 5, n. 1, 2010.
- GOVINDASAMY, Ramu; KELLEY, Kathleen. Agritourism consumers' participation in wine tasting events: An econometric analysis. *International Journal of Wine Business Research*, [s. l.], v. 26, n. 2, p. 120–138, 2014.
- GRIFFITH, Daniel A. An analytical perspective on sporting events attendance: The 2007-2008 US NCAA college bowl games. *Applied Geography*, [s. l.], v. 30, n. 2, p. 203–209, 2010.
- HALL, John; BASARIN, V. John; LOCKSTONE-BINNEY, Leonie. An empirical analysis of attendance at a commemorative event: Anzac Day at Gallipoli. *International Journal of Hospitality Management*, [s. l.], v. 29, n. 2, p. 245–253, 2010.
- HALL, John; O'MAHONY, Barry; VIECELI, Julian. An empirical model of attendance factors at major sporting events. *International Journal of Hospitality Management*, [s. l.], v. 29, n. 2, p. 328–334, 2010.
- HALPENNY, Elizabeth A.; KULCZYCKI, Cory; MOGHIMEHFAR, Farhad. Factors effecting destination and event loyalty: examining the sustainability of a recurrent small-scale running event at Banff National Park. *Journal of Sport and Tourism*, [s. l.], v. 20, n. 3–4, p. 233–262, 2016.
- HAMERMAN, Eric J. Cooking and disgust sensitivity influence preference for attending insect-based food events. *Appetite*, [s. l.], v. 96, p. 319–326, 2016.
- HOPKINS, Nick et al. Explaining effervescence: Investigating the relationship between shared social identity and positive experience in crowds. *Cognition and Emotion*, [s. l.], v. 30, n. 1, p. 20–32, 2016.
- JAMES, G. et al. **An Introduction to Statistical Learning: with Applications in R**. Springer, 2013.

- JIANG, Jyun Yu; LI, Cheng Te. Who should I invite: predicting event participants for a host user. **Knowledge and Information Systems**, [s. l.], v. 59, n. 3, p. 629–650, 2019.
- KAPLANIDOU, Kyriaki(Kiki); GIBSON, Heather J. Predicting behavioral intentions of active event sport tourists: The case of a small-scale recurring sports event. **Journal of Sport and Tourism**, [s. l.], v. 15, n. 2, p. 163–179, 2010.
- KIM, Tae Kyun et al. The Carryover Effect of Newspaper Reports on a Mega Event: Ex Post Analysis of the 2012 Expo Yeosu Korea. **Asia Pacific Journal of Tourism Research**, [s. l.], v. 19, n. 9, p. 1009–1022, 2014.
- KIM, Y. G.; SUH, B. W.; EVES, A.; The relationships between food-related personality traits, satisfaction and loyalty among visitors attending food events and festivals. **International Journal of Hospitality Management**, v. 29, p. 216–226, 2010.
- LEE, Choong Ki et al. Estimating the intention-behavior gap associated with a mega event: The case of the Expo 2012 Yeosu Korea. **Tourism Management**, [s. l.], v. 41, p. 168–177, 2014.
- LI, Libo. Predicting online invitation responses with a competing risk model using privacy-friendly social event data. **European Journal of Operational Research**, [s. l.], v. 270, n. 2, p. 698–708, 2018.
- LU, Meilian et al. MFDAU: A multi-features event participation prediction model for users of different activity levels. **Physica A: Statistical Mechanics and its Applications**, [s. l.], v. 534, p. 122244, 2019.
- MAMEI, Marco; COLONNA, Massimo. Estimating attendance from cellular network data. **International Journal of Geographical Information Science**, [s. l.], v. 30, n. 7, p. 1281–1301, 2016.
- MARSHALL, P.; DOCKENDORFF, M.; IBÁÑEZ, S.; A forecasting system for movie attendance. **Journal of Business Research**, v. 66, p. 1800–1806, 2013.
- MICHALCO, Jaroslav; NAVRAT, Pavol. Arrangement of face-to-face meetings using social media. **Studies in Informatics and Control**, [s. l.], v. 21, n. 4, p. 383–392, 2012.
- MOHAN, Leon J.; THOMAS, Lionel. Effect of identification on attendance at team sporting events. **Event Management**, [s. l.], v. 16, n. 4, p. 341–349, 2012.
- MONTGOMERY, Sarah S.; ROBINSON, Michael D. Empirical evidence of the effects of marriage on male and female attendance at sports and arts. **Social Science Quarterly**, [s. l.], v. 91, n. 1, p. 99–116, 2010.
- MUÑIZ, Cristina; RODRÍGUEZ, Plácido; SUÁREZ, María José. Participation in cultural activities: specification issues. **Journal of Cultural Economics**, [s. l.], v. 41, n. 1, p. 71–93, 2017.

- PALMA, María Luisa; PALMA, Luis; AGUADO, Luis Fernando. Determinants of cultural and popular celebration attendance: The case study of Seville Spring Fiestas. **Journal of Cultural Economics**, [s. l.], v. 37, n. 1, p. 87–107, 2013.
- PARRA, Denis et al. Twitter in academic events: A study of temporal usage, communication, sentimental and topical patterns in 16 Computer Science conferences. **Computer Communications**, [s. l.], v. 73, p. 301–314, 2016.
- PETTICREW, R.; ROBERTS, H.; **Systematic Reviews in the Social Sciences. A Practical Guide**. Oxford, Englad: Blackwell Publishing, 2006.
- POPP, Nels; JENSEN, Jonathan; JACKSON, Rhett. Maximizing visitors at college football bowl games. **International Journal of Event and Festival Management**, [s. l.], v. 8, n. 3, p. 261–273, 2017.
- PRIPORAS, Constantinos Vasilios et al. The effect of sport tourists' travel style, destination and event choices, and motivation on their involvement in small-scale sports events. **Event Management**, [s. l.], v. 22, n. 5, p. 745–765, 2018.
- RODRÍGUEZ-CAMPO, Lorena et al. A holistic understanding of the emotional experience of festival attendees. **Leisure Sciences**, [s. l.], v. 0, n. 0, p. 1–19, 2019.
- SANTOS-LEWIS, Rhiannon; MOITAL, Miguel. Constraints to attend events across specialization levels. **International Journal of Event and Festival Management**, [s. l.], v. 4, n. 2, p. 107–124, 2013.
- SERWYLO et al.; Predicting patient presentation rates at mass gatherings using Machine Learning. **8th International ISCRAM Conference**, Lisboa, Portugal, maio 2011.
- SHUO ZHANG, Jason; LV, Qin. Understanding event organization at scale in event-based social networks. **ACM Transactions on Intelligent Systems and Technology**, [s. l.], v. 10, n. 2, p. 1–23, 2019.
- SNELGROVE, Ryan; WOOD, Laura. Attracting and leveraging visitors at a charity cycling event. **Journal of Sport and Tourism**, [s. l.], v. 15, n. 4, p. 269–285, 2010.
- SUMATHI, V. P. et al. Crowd estimation at a social event using call data records. **International Journal of Business Information Systems**, [s. l.], v. 28, n. 2, p. 246–261, 2018.
- SUNG, Hojun; MILLS, Brian M. Estimation of game-level attendance in major league soccer: Outcome uncertainty and absolute quality considerations. **Sport Management Review**, [s. l.], v. 21, n. 5, p. 519–532, 2018.
- TANFORD, Sarah; MONTGOMERY, Rhonda; HERTZMAN, Jean. Towards a Model of Wine Event Loyalty. **Journal of Convention and Event Tourism**, [s. l.], v. 13, n. 2, p. 77–99, 2012.

- TANFORD, Sarah; MONTGOMERY, Rhonda; NELSON, Kathleen B. Factors that Influence Attendance, Satisfaction, and Loyalty for Conventions. **Journal of Convention and Event Tourism**, [s. l.], v. 13, n. 4, p. 290–318, 2012.
- TEALAB, A.; Forecasting using artificial neural network methodologies: A systematic review. **Future Computing and Informatics Journal**, v. 3, p. 334 – 340, 2018.
- TRINH, Giang Tue. The attendance at sporting events: A generalized theory and its implications. **International Journal of Market Research**, [s. l.], v. 60, n. 3, p. 232–237, 2018.
- TRINH, Giang; LAM, Desmond. Understanding the attendance at cultural venues and events with stochastic preference models. **Journal of Business Research**, [s. l.], v. 69, n. 9, p. 3538–3544, 2016.
- W. JUN et al.; Modeling a combined forecast algorithm based on sequence patterns and near characteristics: An application for tourism demand. **Chaos, Solitons and Fractals**, v. 108, p. 136-147, 2018.
- WICKER, Pamela; HALLMANN, Kirstin. Estimating consumer’s willingness-to-pay for participation in and traveling to marathon events. **Event Management**, [s. l.], v. 17, n. 3, p. 271–282, 2013.
- XU, Tong et al. Exploiting the dynamic mutual influence for predicting social event participation. **IEEE Transactions on Knowledge and Data Engineering**, [s. l.], v. 31, n. 6, p. 1122–1135, 2019.
- YAHIA, Imene Ben; AL-EMADI, Mohammed. Exploring the determinants of 2022 FIFA World Cup attendance in Qatar. **International Journal of Sport Management and Marketing**, [s. l.], v. 18, n. 1–2, p. 116–129, 2018.
- YAZICI, Tuba; KOÇAK, Settar; ALTUNSÖZ, Irmak Hürmeriç. Examining the effect of experiential marketing on behavioral intentions in a festival with a specific sport event. **European Sport Management Quarterly**, [s. l.], v. 17, n. 2, p. 171–192, 2017.
- YOLAL et al.; Impacts of festivals and events on residents’ well-being. **Annals of Tourism Research**, v. 61, p. 1-18, 2016.
- ZHANG, Shuo; LV, Qin. Hybrid EGU-based group event participation prediction in event-based social networks. **Knowledge-Based Systems**, [s. l.], v. 143, p. 19–29, 2018.

3 ARTIGO 2 - MÉTODO DE PREVISÃO DE PÚBLICO EM EVENTOS BASEADO EM FATORES MOTIVACIONAIS

RESUMO

Os eventos têm se tornado uma área de crescente interesse, pois são uma forma de impulsionar o turismo e dinamizar a economia local, criando novos empregos e atrações. Nesse contexto, a previsão do público é uma parte importante do planejamento e execução de eventos. Para entender as motivações que levam as pessoas ao comparecimento com o objetivo de prever o público do evento, este artigo, se propõe a elaborar uma estrutura de avaliação e previsão de público em eventos, baseados em fatores motivacionais encontrados na literatura. Foram utilizados 47 fatores relacionados ao comparecimento aos eventos para obtenção do *framework* de previsão. O uso do *framework* foi ilustrado por meio de duas simulações de previsão de público para edições futuras de eventos de alimentos e bebidas. O *framework* elaborado foi capaz de avaliar os eventos e estimar a audiência nos dois casos com sucesso.

3.1 INTRODUÇÃO

A busca por um gerenciamento eficiente de eventos tem sido estudada intensamente nos últimos anos (Getz, 2008). Muitos autores afirmam que a gestão correta e eficiente de eventos pode trazer grandes benefícios, tanto econômicos como sociais, para a região na qual estão inseridos e proporciona otimização no uso de recursos e aumento do lucro provenientes dos mesmos (Lee et al., 2014; Zhang e Lv, 2018). Além disso, Laing (2018) sugere que a pesquisa sobre gestão e operação de eventos e festivais foram limitados até agora. Neste contexto, conseguir prever o número correto de pessoas que irão comparecer a um evento pode ser crucial para uma gestão eficiente, assim como entender os motivos que levam às pessoas aos eventos de forma a atrair um maior público.

Essa compreensão aprofundada do evento e das motivações por trás do comparecimento pode trazer benefícios, melhorando a produtividade e permitindo que os organizadores alcancem melhores resultados (Zhang e Lv, 2018). Uma previsão precisa dos participantes também pode ser útil em termos de publicidade e gestão, garantindo uma melhor experiência aos participantes, além de alavancar a economia local. Por exemplo, Sung e Mills

(2018) concluem que a previsão da demanda esportiva tem implicações críticas na política de vendas e gerenciamento desses tipos de eventos.

Apesar de ser uma ferramenta importante no gerenciamento de eventos, a previsão de público aplicada a eventos tem sido pouco explorada. Funari e Ribeiro (2019), em um estudo de revisão sistemática, acharam poucos estudos abordando previsão de público em eventos, apenas 17 dentre 59 artigos selecionados. Por exemplo, Sung e Mills (2018) desenvolveram um modelo empírico de comportamento de atendimento a eventos esportivos enquanto Marshall et al. (2013) desenvolveram um modelo de previsão de frequência ao cinema, com base em modelos de comportamento do consumidor. Uma breve revisão da literatura sugere que os modelos desenvolvidos até o momento são específicos por tipo de evento e pouco adaptáveis.

Sendo assim, esse artigo se propõe a desenvolver um método de previsão de público em eventos que pode ser aplicado a diferentes tipos de eventos, baseados em fatores motivacionais que levam às pessoas a comparecerem aos mesmos. Por ser baseado em fatores motivacionais, o método também pode servir como orientação para realizar uma análise e avaliação do evento, apontando pontos fortes e pontos a melhorar para tornar o evento mais atrativo.

Para atingir este objetivo, este artigo está dividido em seis seções. Após esta introdução, a seção 2 apresenta uma breve revisão da literatura com os métodos de previsão de público desenvolvidos até o momento. Após, a Seção 3 detalha o procedimento metodológico que guiou essa pesquisa, cujos resultados são descritos na Seção 4. Na sequência, é apresentada uma discussão dos resultados obtidos na Seção 5. Por fim, a Seção 6 apresenta as conclusões e contribuições gerais do estudo.

3.2 REVISÃO DA LITERATURA

Esta seção apresenta uma breve discussão da literatura, buscando quais métodos e eventos que estão sendo analisado de forma a prever o público e também quais fatores motivacionais podem ser encontrados na literatura. Para isso, a seção está dividida em (a) Métodos utilizados para previsão de público e (b) Fatores motivacionais.

3.2.1 MÉTODOS UTILIZADOS PARA PREVISÃO DE PÚBLICO

Os eventos mais estudados quanto a previsão de público na literatura são eventos esportivos e eventos baseados em redes sociais (EBSNs). Os EBSNs são plataformas online, tais como o MeetUp e Facebook Events, nos quais os usuários podem iniciar e realizar a gestão de diversos tipos de eventos, enviar convites, atribuir locais e reunir pessoas com interesses em comum (Jiang e Li, 2019). Esses tipos de eventos são caracterizados por serem organizados em redes sociais e realizados em locais físicos.

Entre os métodos de previsão mais explorados estão aqueles que utilizam as técnicas de *Machine Learning*, para análise de eventos EBSNs. Os métodos com *Machine Learning* encontram uma combinação de parâmetros a fim de minimizar os erros de previsão, aprendendo com os erros anteriores conforme os dados são apresentados ao computador (Géron, 2017). Sendo assim, esses métodos geralmente necessitam de grandes bases de dados para modelos mais eficazes, o que justifica seu uso apenas com EBSNs.

Por exemplo, De Lira (2019) e Parra et al. (2016) analisaram o conteúdo de *tweets* extraídos do Twitter para classificar os usuários em dois grupos, um que iria ao evento e outro que não iria, utilizando as técnicas de *Machine Learning*. Sob outra perspectiva, Bogaert, et al. (2016) utilizaram características do perfil do usuário como likes, amizades e preferências para realizar suas previsões. Além das características do perfil e do grupo, Xu et al. (2019) ainda analisaram a influência de certos usuários na decisão de outros em ir ou não ao evento. Lu et al. (2019) desenvolveram um algoritmo próprio de classificação chamado MFDAU (Multi-features Event Participation Prediction Model), diferenciando usuários ativos e inativos da rede.

Dentro dos métodos de *Machine Learning* encontrados na literatura, Mamei e Colonna (2016) utilizaram o método de regressão log-linear (de Poisson) para estimar a quantidade de participantes de diferentes eventos a partir de dados de celulares. Além disso, este foi o único estudo encontrado que utilizou um método de regressão com técnicas de *Machine Learning*. A maioria dos estudos que utilizaram *Machine Learning* obtiveram um modelo de classificação, nos quais diferentes fatores e características são avaliadas para classificar cada indivíduo em um grupo que irá ou não irá comparecer ao evento.

Já os estudos encontrados relacionados a eventos esportivos usaram diferentes modelos de regressão para estimar o público sem aplicar nenhuma metodologia de *Machine Learning* (Coates et al., 2014; Sung e Mills, 2018). Griffith (2010), por exemplo, analisou dados de partidas de futebol americano universitário, obtidas através do site oficial da NCAA College

Football, para obtenção de modelos de regressão linear e regressão log-linear (de Poisson). Em contrapartida, Popp, Jensen e Jackson (2017) também estudaram partidas de futebol americano universitário, porém, utilizando um modelo de regressão hierárquica, a fim de obter as variáveis mais importantes para maximizar o público vindo de outras cidades. Sung e Mills (2018) utilizaram partidas de futebol com dados obtidos no site da Major League Soccer, obtendo um modelo quadrático de previsão. Além deles, Coates et al. (2014) e Beckman et al. (2012) analisaram partidas de baseball ao longo de diversos anos.

3.2.2 FATORES MOTIVACIONAIS

Para identificação dos fatores motivacionais, Funari e Ribeiro (2019) realizaram uma revisão sistemática da literatura, na qual encontraram 47 Fatores Motivacionais em diversos estudos realizados em eventos. Uma revisão sistemática adota uma série de passos que permitem sintetizar de forma abrangente, transparente e reproduzível as contribuições de pesquisa em um determinado campo (Briner e Denyer, 2012). Apesar de bem definidos, os passos são flexíveis aos critérios do pesquisador, desde que documentados e justificados (Briner e Denyer, 2012). Desta forma, Funari e Ribeiro (2019) reuniram estudos sobre motivações e previsão de público em eventos publicados na língua inglesa e em periódicos de impacto entre os anos de 2010 a 2019. A busca ocorreu nas duas bases de dados científicas de maior significância, a ScienceDirect® e a Scopus® para uma maior abrangência do assunto, resultando em 56 artigos relevantes para o tema.

A revisão sistemática completa está descrita no artigo “Fatores motivacionais e métodos de previsão aplicados a eventos” (Funari e Ribeiro, 2019). Após leitura dos artigos e extração dos achados da literatura, foram identificados 47 fatores motivacionais que podem influenciar a decisão de comparecer a diferentes tipos de eventos. Os fatores foram categorizados em nove dimensões, de acordo com a sua similaridade e afinidade, conforme apresentado na **Tabela 4**. Como contribuição da revisão sistemática, cada fator está associado a uma pergunta para facilitar a sua compreensão e avaliação para posterior previsão de público (Funari e Ribeiro, 2019).

Tabela 4 - Tabela resumo dos fatores motivacionais (Adaptado de Funari e Ribeiro, 2019).

Dimensão	Fator Motivacional	Pergunta Relacionada
Histórica	Tipo do evento	Qual o tipo de evento?
	Público médio	Qual o público médio neste tipo de evento?
		Qual o público mínimo no pior cenário razoável?
		Qual o público máximo devido à capacidade ou outras restrições?
Populacional	Cidade/Local	A cidade/local é propícia a participação no evento?
	População	Qual a população do local/região sede do evento?
	Penetração	Qual o percentual da população que teria interesse nesse evento?
	Hábitos culturais	Os hábitos da população local são favoráveis a participação?
	Escolaridade	O nível de escolaridade da população local é favorável?
	Renda média	A renda média da população é favorável a participação?
	Convites virtuais	A cultura local responde a convites virtuais?
Contextual	Cultura local	A região proporciona atrações turísticas fora do evento?
	Benefícios para região	O evento oferece benefícios para o bairro/cidade/região?
	Inclusão	O evento é amplamente inclusivo?
	Premiações	O evento oferece premiações pela participação?
	Reputação do evento	O evento possui boa reputação perante o público?
	Causa nobre	O evento está associado a causas nobres?
Emocional	Interações	O evento facilita interações entre participantes?
	Ligação emocional	O evento provoca emoções nos participantes?
	Sair da rotina	O evento oferece uma experiência tipicamente "fora da rotina"?
	Satisfação	O evento gera satisfação que favorece a participação?
	Popularidade	O evento propicia popularidade àqueles que atendem?
	Admiradores	A atração possui um grande número de fãs?
Infraestrutura	Acessibilidade	O evento oferece ampla acessibilidade a PCDs?
	Transporte Público	O evento possui fácil acesso por transporte público?
	Estacionamento	Existe facilidade para estacionamento a preços acessíveis?
	Existência de filas	Há previsão de filas que vão causar perda de tempo?
	Alimentos e bebidas	O evento oferece alimentos/bebidas a preços acessíveis?
	Amenidades	O local do evento é agradável e favorece a participação?
	Capacidade	Qual a capacidade do local onde irá ocorrer o evento?
	Conforto	As pessoas podem ficar confortáveis durante o evento?
Temporal	Dia da semana	O dia da semana é favorável a participação?
	Duração	A duração do evento é favorável a participação?
	Hora do evento	O horário do evento é favorável a participação?
	Mês	O mês em que o evento irá ocorrer é favorável?
	Clima	A previsão do tempo para o dia ou dias do evento é favorável?
Preço	Valor do ingresso	Qual o valor do ingresso comparado a outros eventos similares?
Marketing	Propaganda de massa	Qual o nível de divulgação do evento em TV, Rádio, Jornal?
	Propaganda virtual	Qual o nível de divulgação do evento em mídias eletrônicas?
	Interessados virtuais	Quantos usuários indicaram virtualmente interesse pelo evento?
	Tags do evento	Quantas tags (#) estão associadas ao evento?
	Número de convidados	Quantidade de pessoas convidadas para o evento?
	Presença Celebridades	Há influenciadores ou celebridades confirmadas para o evento?
Esportiva	Jogos decisivos	O evento possui um caráter decisivo (no caso de jogos)?
	Liga/Campeonato	O evento pertence a uma série ou campeonato importante?
	Presença de Craques	O jogo terá a presença de jogadores estrelas?
	Resultado do jogo	O resultado esperado para o evento favorece a participação?
	Time da casa	O time da casa está bem na série ou campeonato?
	Time visitante	O time visitante possui uma grande torcida ou reputação?

3.3 RESULTADOS

Conforme dito anteriormente, este artigo buscou a elaboração de um método de previsão de público em eventos baseado em fatores motivacionais. Esta seção apresenta os resultados obtidos nas três etapas de pesquisa apresentadas na metodologia, de forma a atingir este objetivo.

3.3.1 ESTRUTURA DE PREVISÃO DE PÚBLICO EM EVENTOS

De posse dos fatores motivacionais e suas categorias, foi desenvolvido um *framework* para orientar a previsão do público do evento. Essa estrutura é mostrada na **Figura 10** e envolve a execução de duas etapas principais. Dentro de cada dimensão apresentada no *framework*, encontram-se os fatores motivacionais identificados na literatura, classificados de acordo com o seu assunto e similaridade.

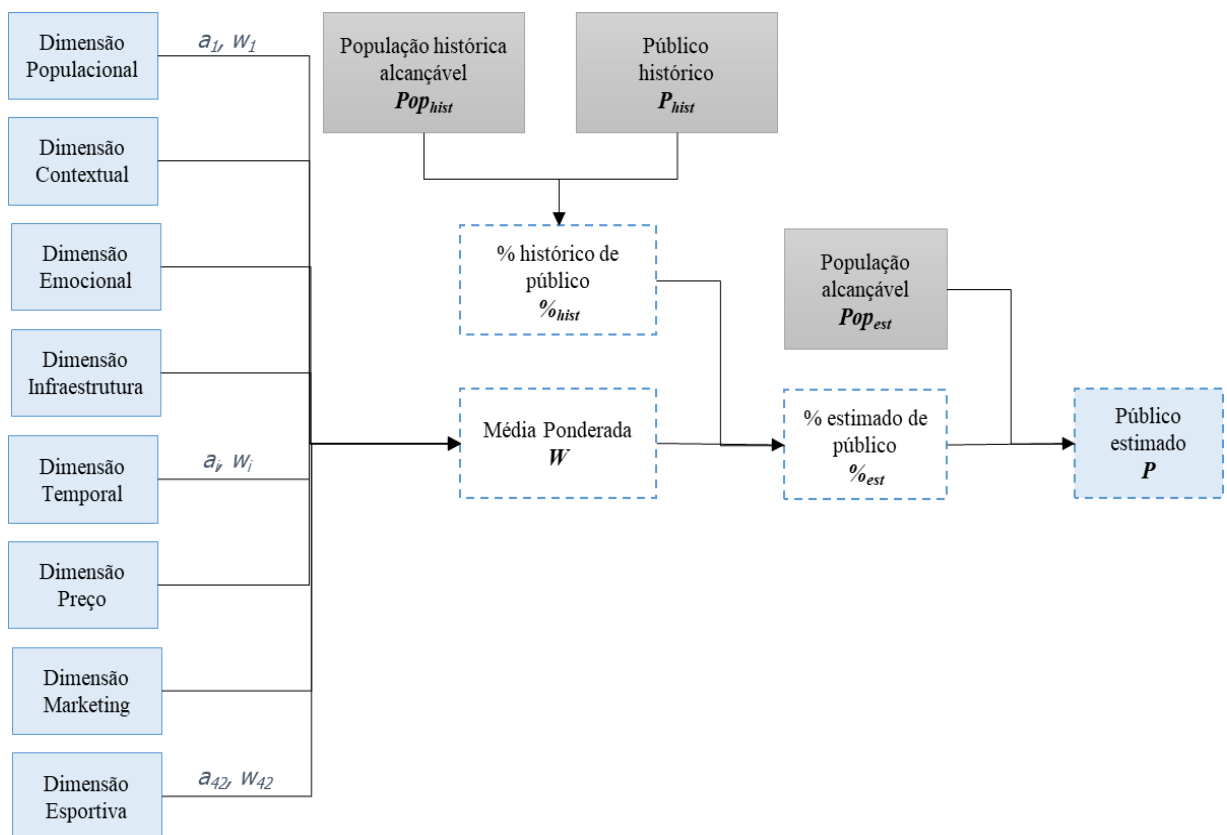


Figura 9 - Framework proposto para previsão de público em eventos.

A primeira etapa proposta está relacionada à dimensão histórica, abrangendo o público médio histórico (P_{hist}) e a população da região alcançável da edição anterior (Pop_{hist}). O público médio histórico leva em considerações edições passadas do evento a ser analisado, ou até, eventos similares que possam ser comparados ao evento a ser analisado caso não se tenha um histórico. Já a população da região alcançável da edição anterior, leva em consideração a população que pode ser impactada pelo evento, sendo influenciada a ir ou não. Esta região alcançável pode ser uma cidade, região metropolitana ou apenas um bairro, a depender do tipo de evento e tamanho do mesmo. Com esses dados, é possível calcular o percentual de audiência histórica alcançável, de acordo com a Equação (1). Este percentual irá indicar a quantidade de pessoas que, historicamente, tendem a ir a eventos do tipo sendo analisado. Este valor será usado posteriormente para estimar o público.

$$\%_{hist} = \frac{P_{hist}}{Pop_{hist}} \quad (1)$$

Onde P_{hist} é o público médio histórico e Pop_{hist} é a população na região alcançável. Portanto, $\%_{hist}$ representa a porcentagem histórica da população que tem interesse no evento.

A segunda etapa é avaliar os demais fatores levantados na literatura relacionados às demais dimensões, representados pelos blocos azuis da **Figura 10**. Essa avaliação deve ser realizada comparando o evento em análise com as edições anteriores ou eventos anteriores similares, informando se o fator atualmente é mais ou menos favorável à participação. No método proposto, esta avaliação é feita em uma escala que varia de 0 a 2, onde 0 significa que o fator está muito abaixo das edições anteriores, 1 significa semelhante e 2 significa muito acima. Caso algum fator não seja aplicável ao evento analisado, por exemplo aqueles dentro da dimensão esportiva para eventos gastronômicos, estes fatores podem ser desprezados e não são incluídos nos cálculos. As avaliações de cada fator (a_i) serão ponderadas por um peso atribuído a cada dimensão (w_i).

Para as simulações realizadas, esses pesos foram calculados dividindo 1 pelo número de fatores correspondentes à dimensão e, portanto, a mesma importância foi atribuída a todas as dimensões. Isso fornece um ponto de partida para a tarefa de atribuição de peso, porém, considerando as condições específicas de um determinado evento, os responsáveis pela previsão do público podem atribuir pesos diferentes para algumas dimensões ou fatores cuja influência deve ser maior ou menor na avaliação em andamento.

Uma vez que todos os fatores forem avaliados, faz-se uma média ponderada das avaliações, resultando no coeficiente W de acordo com a Equação 2. Este coeficiente foi utilizado para calcular a porcentagem de público estimado ($\%_{est}$) ajustando os dados históricos com a média ponderada da avaliação dos fatores, conforme mostrado na Equação 3.

$$W = \frac{\sum a_i w_i}{\sum w_i} \quad (2)$$

Onde a_i é a avaliação dada ao fator i enquanto w_i é o peso atribuído ao fator i .

$$\%_{est} = W \times \%_{hist} \quad (3)$$

A porcentagem de audiência potencial será maior na medida em que o coeficiente W for maior que um. Se W for maior que um, significa que fatores atrativos são dominantes para o evento em análise e espera-se uma maior penetração na população, aumentando o público previsto. Por outro lado, se W for menor que um, a porcentagem de audiência potencial, estimada, será menor que a histórica.

Por fim, é possível estimar a audiência (P_{est}) multiplicando a porcentagem de audiência potencial ($\%_{est}$) e a população alcançável potencial (Pop_{est}), conforme a Equação 4. Este cálculo ajustará a avaliação dos fatores para a região do novo evento. Caso o evento seja realizado no mesmo local, apenas a avaliação dos fatores motivacionais irá influenciar a previsão do público. Além disso, deve-se considerar qualquer capacidade ou restrições locais. Por exemplo, número máximo de pessoas em um local permitidas por espaço físico ou alvarás. Desta forma, a estimativa é limitada pela capacidade do local, conforme indicado pela literatura (Priporas et al., 2018; Sung and Mills, 2018; Popp et al., 2017).

$$P_{est} = \%_{est} \times Pop_{est} \quad (4)$$

3.3.2 APLICAÇÃO DO MÉTODO DE PREVISÃO

Para avaliação e um melhor entendimento do *framework* proposto, foi realizada uma simulação e previsão do público do evento. Para esta etapa, uma entrevista aberta com profissional da área de eventos foi conduzida de modo a aplicar o método de previsão. Durante a entrevista, os fatores motivacionais foram apresentados junto às perguntas associadas para

facilitar o entendimento. Foram analisados dois eventos de grande reconhecimento na Nova Zelândia, o Gindulgence Festival e o Beervana, com os quais a profissional entrevistada está familiarizada com a organização.

(i) Gildungence Festival

O primeiro evento analisado foi o Gindulgence Festival, o evento é um festival de gin, que inclui degustações, palestras, comidas e entretenimento ao vivo (Gindulgence, 2020). O festival geralmente ocorre em parques cercados nas cidades de Wellington e Christchurch, durante a estação do verão. O novo cenário proposto foi estimar o público para uma nova edição em Auckland, no parque Smales Farm. Por ser uma região mais populosa, Auckland teria um potencial de público maior. Por outro lado, o perfil da população de Auckland difere bastante da população de Wellington e Christchurch, que já estão acostumados com o evento. Sendo assim, o *framework* se propõe a avaliar se a mudança de local do evento será favorável à participação.

Conforme dito anteriormente, foi utilizado o *framework*, contendo os fatores motivacionais e as perguntas associadas, para auxiliar na avaliação do especialista. As respostas estão apresentadas na **Figura 11**, assim como as avaliações (a_j) e os pesos dos fatores (w_i) nas colunas à direita da figura. A avaliação de cada fator dada pela especialista está destacada em preto na imagem. Para aqueles fatores que não eram aplicáveis ao evento, a resposta foi NA e não foram considerados para estimar a audiência e, por essa razão, não possuem avaliação nem peso associado. Cada fator foi avaliado como sendo mais ou menos favorável à participação em comparação com a edição anterior do evento realizada em Christchurch.

Como primeiro passo, foi realizada a identificação dos fatores da dimensão histórica. A população histórica alcançável em Christchurch é a área urbana na costa leste da Ilha Sul, logo ao norte da Península de Banks, pois o evento reúne pessoas de toda esta região. A área urbana abriga 377.200 residentes e a autoridade territorial tem 385.500 habitantes, o que a torna a segunda cidade mais populosa da Nova Zelândia depois de Auckland. Ademais, a audiência histórica do evento é de 2.000 pessoas, resultando em uma porcentagem de audiência histórica de 0,52%. Já Auckland, cidade onde será prevista uma nova edição do evento, tem a área urbana mais populosa do país, com uma população total de 1.642.800 habitantes e que inclui áreas rurais periféricas e as ilhas do Golfo Hauraki.

Prosseguindo com o método, após a análise de todos os fatores, a média ponderada W resultou 1,18. Uma vez que W é maior do que 1, já há indicação de que o evento em Auckland tem um ambiente mais favorável, atraindo um percentual da população alcançável maior. Este percentual, de acordo com as equações apresentadas anteriormente é de 0,61%. Isso indicaria um público total de 10.058 pessoas.

Porém, durante a entrevista com a especialista foi levantado que o evento está limitado por uma capacidade de 10.000 pessoas, devido ao alvará de autorização de venda de bebidas, necessário na Nova Zelândia. Desta forma, o novo cenário iria atingir a capacidade do evento, resultando em um público estimado (P_{est}) de 10.000 pessoas. Esses cálculos podem ser vistos na **Tabela 5**.

Tabela 5 - Resumo dos resultados obtidos para o evento Gindulgence

HISTÓRICO: Christchurch	
<i>População histórica alcançável</i>	385.500
<i>Público médio histórico</i>	2.000
<i>% histórico de público</i>	0,52%
NOVO CENÁRIO: Auckland	
<i>População alcançável</i>	1.642.800
<i>Capacidade / Restrições</i>	10.000
Média ponderada de fatores W	1,18
% estimado de público	0,61%
Estimativa sem restrições	10.058
Estimativa final	10.000

(ii) Beervana

O segundo evento analisado foi o Beervana. Este evento é um festival anual de cerveja artesanal, contendo centenas de diferentes cervejas para aprender, beber e celebrar as pessoas que as produzem (Beervana, 2020). O festival conta com mais de 60 cervejarias e ocorre na cidade de Wellington na Nova Zelândia, no Estádio Regional de Wellington. Os cervejeiros vêm de toda a Nova Zelândia, Austrália e algumas cervejarias de outros países também são escolhidas para participar. O evento ocorre em quatro sessões, cada qual respeitando o limite de 5.000 pessoas decorrente da licença para venda de bebidas alcoólicas. Considerando todas as seções, o público médio histórico conta com a participação de 14.000 pessoas. O evento, por

ser de grande porte e reconhecido na região, possui uma população histórica alcançável de 506.800 pessoas. O que resulta em um percentual histórico de público de 2,76%.

Assim como no caso anterior, o cenário proposto foi mudar o local do evento para a cidade de Auckland, visto que é a mais populosa da Nova Zelândia. Neste novo cenário, o evento também ocorreria dentro de um estádio para manter o formato conhecido do festival. Como os perfis de população são diferentes, o *framework* pode ser utilizado para avaliar se a mudança de local é favorável a participação e deve ser realizada.

Novamente, os fatores foram avaliados comparando-os com a última edição realizada e ponderando se os mesmos eram mais ou menos favoráveis à participação. As avaliações e seus pesos associados podem ser vistas na **Figura 12**. Após esta análise, a média ponderada dos fatores (W) foi de 0,97. Isto indica que o novo cenário é um pouco menos favorável à participação do que a última edição, diminuindo o percentual estimado de público para 2,67%. Ainda assim, como a população alcançável na região de Auckland é de 1.642.800, o público estimado sem restrições é de 43.837 pessoas. Considerando as informações coletadas na entrevista com a profissional da área de eventos, o evento é realizado em quatro sessões com limitação de 5.000 pessoas. Assim, o público fica restringido ao total de 20.000 ingressos. De acordo com a previsão, o evento esgotaria ao limite estabelecido. Os resultados para a simulação do Beervana em Auckland podem ser vistos na **Tabela 6**.

Tabela 6 - Resumo dos resultados obtidos para o evento Beervana

HISTÓRICO: Wellington	
<i>População histórica alcançável</i>	506.800
<i>Público médio histórico</i>	14.000
<i>% histórico de público</i>	2,76%
NOVO CENÁRIO: Auckland	
<i>População alcançável</i>	1.642.800
<i>Capacidade / Restrições</i>	20.000
Média ponderada de fatores W	0,97
% estimado de público	2,67%
Estimativa sem restrições	43.837
Estimativa final	20.000

	NA	0	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2	aj	wi
Cidade/Local							1						1.0	0.14
Hábitos culturais								1.2					1.2	0.14
Escolaridade								1.2					1.2	0.14
Renda média									1.4				1.4	0.14
Convites virtuais							1						1.0	0.14
Cultura local							1						1.0	0.17
Benefícios para região							1						1.0	0.17
Inclusão											1.8		1.8	0.17
Premiações	1													
Reputação do evento							1						1.0	0.17
Causa nobre							1						1.0	0.17
Interações											1.8		1.8	0.33
Ligação emocional							1						1.0	0.33
Sair da rotina					0.8		1						0.8	0.33
Satisfação								1.4					1.4	0.33
Popularidade							1						1.0	0.33
Admiradores							1						1.2	0.33
Acessibilidade											1.8		1.8	0.33
Transporte Público								1.2					1.2	0.33
Estacionamento								1.2					1.2	0.33
Existência de filas				1.6									1.6	0.33
Alimentos e bebidas									1.4				1.4	0.33
Amenidades											1.8		1.8	0.33
Conforto									1.6				1.6	0.33
Dia da semana							1						1.0	0.33
Duração								1.2					1.2	0.33
Hora do evento							1						1.0	0.33
Mês				0.4									0.4	0.33
Clima								1.2					1.2	0.33
Valor do ingresso								1.2					1.2	1.00
Propaganda de massa						0.8							0.8	0.17
Propaganda virtual								1.2					1.2	0.17
Interessados virtuais							1						1.0	0.17
Tags do evento							1						1.0	0.17
Número de convidados									1.4				1.4	0.17
Presença Celebidades							1						1.0	0.17
Jogos decisivos	1													
Liga/Campeonato														
Presença de Craques														
Resultado do jogo														
Time da casa														
Time visitante														

Figura 10 - Avaliação dos fatores para edição do evento Gindulgence em Auckland

	NA	0	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2	aj	wi
Cidade/Local						0.8							0.8	0.20
Hábitos culturais				0.4									0.4	0.20
Escolaridade									0.4				1.4	0.20
Renda média								0.4					1.2	0.20
Convites virtuais							1.0						1.0	0.20
Cultura local							1.0						1.0	0.20
Benefícios para região									0.4				1.4	0.20
Inclusão							1.0						1.0	0.20
Premiações	0.4													
Reputação do evento				0.4									0.4	0.20
Causa nobre					0.6								0.6	0.20
Interações								0.8					1.2	0.17
Ligação emocional					0.6								0.6	0.17
Sair da rotina								0.8					1.2	0.17
Satisfação							1.0						1.0	0.17
Popularidade						0.8							0.8	0.17
Admiradores					0.6								0.6	0.17
Acessibilidade							1.0						1.0	0.14
Transporte Público							1.0						1.0	0.14
Estacionamento			0.2										0.2	0.14
Existência de filas										0.4			0.4	0.14
Alimentos e bebidas								1.0					1.0	0.14
Amenidades								1.2					1.2	0.14
Conforto							1.0						1.0	0.14
Dia da semana							1.0						1.0	0.20
Duração							1.0						1.0	0.20
Hora do evento							1.0						1.0	0.20
Mês							1.0						1.0	0.20
Clima	0.4													
Valor do ingresso						1.2							1.2	1.00
Propaganda de massa							1.0						1.0	0.17
Propaganda virtual							1.0						1.0	0.17
Interessados virtuais							1.0						1.0	0.17
Tags do evento							1.0						1.0	0.17
Número de convidados							1.0						1.0	0.17
Presença Celebidades							1.0						1.0	0.17
Jogos decisivos	0.4													
Liga/Campeonato	0.4													
Presença de Craques	0.4													
Resultado do jogo	0.4													
Time da casa	0.4													
Time visitante	0.4													

Figura 11 – Avaliação dos fatores para edição do evento Beervana em Auckland

3.4 DISCUSSÃO

Os resultados demonstram que o *framework* elaborado é de fácil compreensão e aplicação para diferentes tipos e tamanhos de eventos. As perguntas associadas aos fatores auxiliam os organizadores de eventos na avaliação dos mesmos, orientando sobre o que cada fator se refere. Quanto à primeira simulação, do evento Gindulgence Festival em Auckland, a avaliação mostra que o novo evento é promissor devido à atratividade de alguns fatores. É possível reparar na **Figura 11**, que a avaliação da especialista está concentrada à direita da imagem, indicando que a maior parte dos fatores são mais favoráveis no cenário proposto do que na edição anterior. Portanto, a média ponderada dos fatores W resultou em um número maior que 1.

Para este cenário, os fatores “Inclusão”, “Interações”, “Acessibilidade” e “Amenidades” foram os mais favoráveis à nova edição, o que aumenta o número de pessoas que podem se interessar pelo evento. Nesta nova edição no parque Smales Farm, é esperado a existência uma infraestrutura de mais fácil acesso e que possibilita uma interação entre os participantes maior, sendo assim, estes fatores foram os melhores avaliados.

O *framework*, ainda, pode fornecer informações sobre os pontos fracos que podem ser trabalhados neste novo cenário para alavancar o público. Nesse caso, os fatores “Mês” e “Existência de Filas” foram os fatores mais críticos. O mês é um fator crítico pois, na época no ano em que é realizado, existem muitos outros eventos ocorrendo na região de Auckland que podem oferecer concorrência para o Gindulgence Festival. Os fatores “Divulgação em massa” e “Fora da rotina” também poderiam ser analisados e aprimorados, pois estão abaixo da condição observada no cenário anterior. Ainda assim, a previsão de público final indica que a capacidade máxima do evento será alcançada, devido a atratividade do evento em uma região receptiva a ele.

Em relação à segunda simulação, do evento Beervana em Auckland, pode-se concluir que também é favorável a realização de uma edição nesta região, apesar da média ponderada dos fatores ser inferior a 1. Isto é, o evento em Auckland é menos atrativo nesta região quando avaliado o conjunto de fatores e terá uma penetração menor na população. Por ser um evento de porte maior que o Gindulgence e já consolidado na região de Wellington, muitos fatores motivacionais acabam sendo avaliados como abaixo da condição observada na última edição.

Os fatores mais críticos são “Hábitos culturais”, “Reputação do evento” e “Estacionamento”, visto que a população de Auckland não está acostumada com este tipo de

evento e o mesmo já ser bem consolidado na região atual. Por outro lado, como a região de Auckland é mais populosa, ainda pode ser atrativo realizar uma nova edição do evento neste local. Uma vez que a previsão indica um número de pessoas interessadas substancialmente maior que a restrição, mesmo que o percentual estimado de público seja menor que o histórico. Sendo assim, os organizadores podem avaliar a ampliação da duração do evento, abrindo um número maior de sessões do que o usual.

3.5 CONSIDERAÇÕES FINAIS

Compreender as motivações da participação em eventos é fundamental para uma estimativa confiável de público além de melhorar a gestão e organização de eventos (Davies et al., 2011). No entanto, os métodos de previsão encontrados na literatura até este momento não levam em consideração todos os fatores que podem influenciar a decisão em ir a eventos e são restritos ao tipo de evento analisado.

Com o objetivo de contribuir para o gerenciamento eficiente de eventos e melhorar a experiência dos participantes, este estudo propôs um *framework* de avaliação e previsão de público em eventos utilizando os fatores motivacionais encontrados na literatura. Esses fatores foram extraídos dos achados e comentários apresentados ao longo dos artigos, discutindo as motivações para a participação em diferentes tipos de eventos. Os fatores, ainda, foram organizados em nove dimensões e associados a questões que podem auxiliar os organizadores do evento no processo de gestão do evento e na utilização do *framework*.

Como principal contribuição, este artigo propôs um *framework* de previsão do público em eventos com base nos fatores motivacionais identificados na literatura. Até o momento, nenhuma outra pesquisa abrangeu de forma ampla os fatores e motivações para a participação em eventos em um método de previsão, sem restringi-los a um tipo específico de evento. Além disso, foi demonstrado como o *framework* pode ser utilizado para estimar o público do evento por meio de duas simulações com o auxílio de uma profissional da área de eventos.

O principal benefício da estrutura proposta é que ela é facilmente adaptada a diferentes tipos de eventos. Se algum fator ou dimensão é considerado mais ou menos importante para um determinado evento, os responsáveis pela previsão podem ajustar os pesos de previsão para lidar com as especificidades de acordo com a sua experiência ou também eventos passados. A estrutura também pode ajudar os organizadores a medir a qualidade de seu evento e avaliar os pontos fracos a serem melhorados, a fim de atrair público e manter a lealdade dos participantes.

Além disso, essas descobertas ajudam a melhorar a qualidade geral do evento e seus benefícios sociais e econômicos para as cidades e regiões.

3.6 REFERÊNCIAS

BECKMAN, Elise M. et al. Explaining Game-to-Game Ticket Sales for Major League Baseball Games Over Time. **Journal of Sports Economics**, [s. l.], v. 13, n. 5, p. 536–553, 2012.

Beervana. Beervana. Disponível em: <<https://beervana.co.nz/>>. Acesso em: 01 dez. 2020.

BOGAERT, Matthias; BALLINGS, Michel; VAN DEN POEL, Dirk. The added value of Facebook friend’s data in event attendance prediction. **Decision Support Systems**, [s. l.], v. 82, p. 26–34, 2016.

COATES, Dennis; HUMPHREYS, Brad R.; ZHOU, Li. Reference-dependent preferences, loss aversion, and live game attendance. **Economic Inquiry**, [s. l.], v. 52, n. 3, p. 959–973, 2014.

DE LIRA, Vinicius Monteiro et al. Event attendance classification in social media. **Information Processing and Management**, [s. l.], v. 56, n. 3, p. 687–703, 2019.

FUNARI, L.M.; RIBEIRO, J.L.D. (em preparação). Fatores motivacionais e métodos de precisão aplicados a eventos, 2019.

GÉRON, A. **Hands-on Machine Learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems**. O’Reilly Media, Inc, 2017.

GETZ, D.; Event Tourism definition, evolution and research. **Tourism Management**, v. 29, p. 403-428, 2008.

Gindulgence. Gindulgence. Disponível em: <<https://www.gindulgence.co.nz/>>. Acesso em: 01 dez. 2020.

GRIFFITH, Daniel A. An analytical perspective on sporting events attendance: The 2007-2008 US NCAA college bowl games. **Applied Geography**, [s. l.], v. 30, n. 2, p. 203–209, 2010.

JIANG, Jyun Yu; LI, Cheng Te. Who should I invite: predicting event participants for a host user. **Knowledge and Information Systems**, [s. l.], v. 59, n. 3, p. 629–650, 2019.

LAINING, Jennifer. Festival and event tourism research: Current and future perspectives. **Tourism Management Perspectives**, v.25, p. 165-168, 2018.

LU, Meilian et al. MFDAU: A multi-features event participation prediction model for users of different activity levels. **Physica A: Statistical Mechanics and its Applications**, [s. l.], v. 534, p. 122244, 2019.

- MAMEI, Marco; COLONNA, Massimo. Estimating attendance from cellular network data. **International Journal of Geographical Information Science**, [s. l.], v. 30, n. 7, p. 1281–1301, 2016.
- MARCONI M.A.; LAKATOS, E.M. Técnicas de pesquisa. 8. ed. São Paulo: Atlas, 2017.
- MARSHALL, P.; DOCKENDORFF, M.; IBÁÑEZ, S.; A forecasting system for movie attendance. **Journal of Business Research**, v. 66, p. 1800-1806, 2013.
- PARRA, Denis et al. Twitter in academic events: A study of temporal usage, communication, sentimental and topical patterns in 16 Computer Science conferences. **Computer Communications**, [s. l.], v. 73, p. 301–314, 2016.
- POPP, Nels; JENSEN, Jonathan; JACKSON, Rhett. Maximizing visitors at college football bowl games. **International Journal of Event and Festival Management**, [s. l.], v. 8, n. 3, p. 261–273, 2017.
- PRIPORAS, Constantinos Vasilios et al. The effect of sport tourists’ travel style, destination and event choices, and motivation on their involvement in small-scale sports events. **Event Management**, [s. l.], v. 22, n. 5, p. 745–765, 2018.
- SUNG, Hojun; MILLS, Brian M. Estimation of game-level attendance in major league soccer: Outcome uncertainty and absolute quality considerations. **Sport Management Review**, [s. l.], v. 21, n. 5, p. 519–532, 2018.
- SUNG, Hojun; MILLS, Brian M. Estimation of game-level attendance in major league soccer: Outcome uncertainty and absolute quality considerations. **Sport Management Review**, [s. l.], v. 21, n. 5, p. 519–532, 2018.
- XU, Tong et al. Exploiting the dynamic mutual influence for predicting social event participation. **IEEE Transactions on Knowledge and Data Engineering**, [s. l.], v. 31, n. 6, p. 1122–1135, 2019.
- ZHANG, Shuo; LV, Qin. Hybrid EGU-based group event participation prediction in event-based social networks. **Knowledge-Based Systems**, [s. l.], v. 143, p. 19–29, 2018.

4 ARTIGO 3 - PREVISÃO DE PÚBLICO EM EVENTOS ESPORTIVOS COM O USO DE *MACHINE LEARNING*

RESUMO

A participação em eventos esportivos representa uma receita importante para as associações esportivas, com benefícios econômicos e sociais para sua cidade e regiões de origem (Hall et al., 2010). O presente estudo aplicou nove métodos de *Machine Learning* para prever o público total e pagante em jogos de futebol, visando obter o melhor modelo. Os dados dos jogos foram coletados em parceria com um time brasileiro de futebol e um conjunto de dados de 210 jogos foi reunido, contendo as seguintes variáveis: Data e hora do jogo, atratividade do adversário, campeonato, fase do campeonato, escalação do time da casa, temperatura, chuva, nível de marketing e evolução da venda de ingressos nos seis dias anteriores à partida. Foram avaliados métodos *ensemble* baseado em árvore de decisão, redes neurais artificiais (ANN), modelo linear (LM) e *K-nearest neighbors* (KNN). Os resultados indicaram que os métodos baseados em árvore de decisão são totalmente capazes de capturar os padrões de comportamento relativos ao comparecimento à partida. Os modelos XGB e GBM foram os de melhor desempenho, com os erros absolutos mais baixos.

4.1 INTRODUÇÃO

Os eventos esportivos são uma classe de eventos mundiais que movimentam grandes quantias de dinheiro, impulsionando o turismo e a economia local (Getz, 2008). Dentre eles, o futebol é um dos esportes mais atraentes da atualidade, conquistando cada vez mais espaço no cenário mundial (Sung e Mills, 2018). No Brasil, além de ser reconhecido como “a paixão nacional”, o futebol é um poderoso mecanismo de integração social, solidificação de uma identidade nacional, revelação de características sociais e também de forte importância política (Helal e Gordon, 2002).

A participação em eventos esportivos representa uma importante receita para as organizações esportivas, trazendo também benefícios econômicos e sociais para as cidades e regiões envolvidas (Hall et al., 2010). Além disso, entender as motivações que levam ao comparecimento a esse tipo de evento é de grande valor para os gestores esportivos na tomada de decisões para as operações locais e a precificação de ingressos. Hall et al. (2010) concluem que essas motivações incluem características inerentes ao evento, ambiente do estádio, experiências sociais e emoções pessoais relacionadas ao esporte.

Por outro lado, os métodos de previsão de demanda têm atraído atenção de diversas áreas do conhecimento (Tealab, 2018). A demanda esportiva já está sendo estudada por economistas, com implicações importantes para a determinação de políticas de vendas e estratégias de gestão. No entanto, o foco principal tem sido o uso da teoria da decisão sob incerteza (Sung e Mills, 2018). Dentre as probabilidades estudadas por essa teoria, Sung e Mills (2018) citam a incerteza da partida, a incerteza do campeonato e a incerteza do resultado como atrativos para assistir a esse tipo de evento.

Com os principais avanços nos métodos de previsão de demanda na manufatura, essas técnicas foram adaptadas para aplicações também em serviços. Por exemplo, Serwylo et al. (2011) aplicou métodos de *Machine Learning* (ML) para prever ocorrências médicas em partidas esportivas. Além disso, Jun et al. (2018) usou técnicas de ML para prever a demanda turística. Segundo eles, o método de Redes Neurais Artificiais (ANN) foi o mais utilizado nessa área. Comparado com os métodos de regressão tradicionais, Tealab (2018) concluiu que o uso de ML levou a uma maior precisão, sendo preferido aos métodos tradicionais.

Além disso, Claudino et al. (2019) mostraram em sua revisão sistemática que os métodos de ML já têm sido usados em aplicações esportivas, principalmente para prever lesões em jogadores e o resultado da partida. Os métodos mais comumente utilizados encontrados por eles foram ANN, *Tree Based Models* e *Support Vector Machine*. No entanto, não há muitos estudos realizados com o objetivo de prever o público da partida. Assim, o presente trabalho propõe a aplicação de diferentes métodos de ML para previsão de público total e pagante em jogos de futebol, proporcionando uma comparação e identificação dos melhores modelos de previsão.

Os métodos de ML foram aplicados a partir de dados coletados em jogos reais de um time de futebol brasileiro da primeira divisão, compreendendo seis temporadas completas. Portanto, foi possível comparar os resultados da previsão com os dados reais anteriores e avaliar a eficiência do modelo. Nove métodos diferentes foram testados, incluindo regressão linear e redes neurais artificiais, conforme apresentado na próxima sessão. Em seguida, esses métodos foram comparados entre si para detectar o mais preciso para prever o atendimento com as características consideradas. As variáveis preditoras foram determinadas após análise da literatura e validadas por especialistas em futebol.

Os modelos foram avaliados considerando sua capacidade de prever o público dois dias antes da partida, com tempo suficiente para ajustar fatores gerenciais importantes do

evento, como tamanho da equipe de assistência, policiais e ambulâncias. Esse tempo de antecedência é fundamental para os gestores da equipe de futebol parceira do estudo, pois o número de funcionários que atuam na organização do evento, bem como o número de portões e lanchonetes a serem abertos são diretamente proporcionais ao total de pessoas presentes. Além disso, essa previsão é importante para a compra de material de bar e definição da quantidade de policiais e ambulâncias que estarão presentes no local.

As contribuições da pesquisa e suas diferenças com outros estudos encontrados na literatura são resumidas a seguir.

- (i) O estudo amplia a literatura aplicando *Machine Learning* (ML) a um problema resolvido até o momento com a teoria da decisão sob incerteza, demonstrando modelos novos e mais avançados para resolver esse problema. Além disso, diferentes métodos foram testados e avaliados para analisar o problema. Assim, o estudo fornece conhecimento sobre as deficiências e vantagens de cada método para prever o comparecimento do público, que pode ser extrapolado para eventos semelhantes.
- (ii) Esta pesquisa utiliza dados reais de jogos obtidos em parceria com uma organização de futebol e incorpora a evolução da venda de ingressos até dois dias antes da partida nas modelos. A incorporação dessa variável é importante, pois carrega informações sólidas sobre os potenciais espectadores, contribuindo para o desempenho das modelos.
- (iii) A pesquisa avalia os métodos selecionados para prever a audiência total de uma partida e também o público pagante, que irá adquirir ingressos. Essa diferenciação é importante para os gerentes do evento, ao obterem informações sobre a receita do jogo, além de apenas o número total de pessoas presentes.
- (iv) Os gerentes de eventos podem se beneficiar das previsões para tomar decisões de gerenciamento, como número de funcionários, policiais e atendimento médico no local, bem como comprar produtos para lanchonetes e facilidades. Além disso, os modelos podem fornecer uma visão sobre as motivações por trás da participação em eventos de jogos que ajuda os gestores a criar um sentimento de lealdade e também impulsionar aqueles jogos que não são tão atraentes perante o público.

Este artigo está dividido em seis seções. Após esta introdução, a Seção 2 fornece uma breve revisão da literatura sobre trabalhos relacionados com a previsão de frequência. Em seguida, a Seção 3 detalha os procedimentos metodológicos que nortearam esta pesquisa, bem

como a coleta e visualização dos dados. A Seção 4 descreve os principais resultados obtidos e sua análise é apresentada na Seção 5. Por fim, a Seção 6 apresenta as conclusões gerais da pesquisa e sugestões de trabalhos futuros.

4.2 TRABALHOS RELACIONADOS

Para obter um maior conhecimento sobre o problema e os desafios enfrentados pela previsão de demanda esportiva, foi realizada uma breve revisão da literatura sobre estudos sobre previsão de comparecimento a eventos esportivos. Além disso, a revisão da literatura forneceu informações sobre quais variáveis são eficazes na previsão do público. Vários estudos na área foram conduzidos utilizando métodos estocásticos, como regressão linear múltipla, regressão hierárquica e regressão log-linear. Por exemplo, Coates et al. (2014) desenvolveram um modelo de comparecimento baseado nas preferências de comportamento do consumidor, usando um modelo estrutural econométrico e a hipótese de incerteza de resultado para prever o comparecimento em jogos de beisebol. Seu estudo mostra a utilização de fatores que captam a qualidade das equipes envolvidas na partida e também a probabilidade de vitória da equipe da casa como preditores do comparecimento.

Beckman et al. (2012) também investigaram o comparecimento a jogos de beisebol, fazendo uso de 25 anos de dados históricos para desenvolver um modelo de regressão linear múltipla, que previu a venda de ingressos para os jogos. Bem como Coates et al. (2014), seu modelo capturou as características de qualidade do time da casa e também a escalação do time da casa, presença de jogadores famosos, idade do estádio, mês e dia do jogo e a força da rivalidade entre os times. Em concordância, Sung e Mills (2018) também consideraram a escalação do time da casa e a presença de jogadores famosos em seu modelo. Eles desenvolveram um modelo de regressão linear que, além desses fatores, incluiu o preço do ingresso, a incerteza do resultado e a qualidade da equipe adversária.

Outro exemplo de modelo estocástico que captura as preferências de comportamento do consumidor é o modelo de distribuição binomial negativa (NBD), como um dos modelos mais simples. Este modelo foi utilizado para ajustar o comparecimento a diversos eventos esportivos no trabalho de Trinh (2018). O modelo desenvolvido ajustou a probabilidade de um percentual da população comparecer ao evento utilizando a frequência média de comparecimento ao longo do tempo, coletada por meio de pesquisas.

Além deles, Griffith (2010) analisou dados de jogos de futebol americano universitário, testando modelos de regressão lineares e log-lineares. Suas descobertas mostram que o modelo log-linear é melhor para prever o comparecimento a eventos esportivos, uma vez que o estádio é limitado por sua capacidade, o que viola a restrição de linearidade dos modelos lineares. Popp et al. (2017) também testou uma regressão linear para os seus dados e introduziu outros fatores explicativos do comparecimento, como o reconhecimento da liga ou campeonato, porcentagem de vitórias da equipe da casa na temporada, capacidade do estádio e audiência média da equipe na temporada.

Embora haja algum progresso no campo da previsão de público em jogos esportivos, foi difícil encontrar trabalhos utilizando técnicas de ML. Claudino et al. (2019) realizaram uma revisão sistemática sobre o uso de ML no esporte e apenas um estudo surgiu que previa frequência do comparecimento a jogos. Este estudo usou ANN para prever o público de três times diferentes participantes do Campeonato da Liga de Futebol da Inglaterra (Strand et al, 2017). Os fatores utilizados no estudo incluíram a posição das equipes no campeonato, o dia da partida, se foi feriado ou não, as condições climáticas, a distância entre a cidade natal das equipes competidoras como medida de rivalidade, a média de público fora de casa da equipe para os jogos fora de casa e o comparecimento da última partida como um sinal de série temporal. Como conclusão do seu estudo, os modelos de ANN captaram com sucesso os padrões do comparecimento em eventos de futebol.

Claudino et al. (2019) também verificaram que a variável mais significativa na previsão do público foi a média de público visitante para jogos fora de casa, seguido pela posição da equipe da casa no campeonato e pela distância entre as duas cidades dos times do jogo. Por outro lado, a temperatura e o dia do jogo foram as variáveis menos influenciáveis para o comparecimento. Sahin e Erol (2018) também avaliaram modelos de ANN para prever o comparecimento em jogos de clubes de futebol europeus com dados obtidos de duas temporadas para três times diferentes. Neste estudo, foi utilizado como variáveis as medições de desempenho das equipes em casa e fora, o dia do jogo, a distância das cidades natais das equipes e a incerteza do resultado.

4.3 METODOLOGIA

Nesta seção, inicialmente são apresentados a formulação do problema e os detalhes da coleta de dados e pré-processamento dos dados. Em seguida, são apresentados os métodos de

predições selecionados para este estudo e os critérios de avaliação utilizados para comparar os modelos. Por fim, o procedimento experimental foi detalhado para facilitar a compreensão e reprodutibilidade.

4.3.1 FORMULAÇÃO DO PROBLEMA

Conforme afirmado anteriormente, o objetivo desta pesquisa é prever o público total e pagante em jogos de futebol com dois dias de antecedência. No Brasil, o público total da partida costuma ser composto por um número de espectadores que são associados a equipe da casa e não precisam comprar ingressos, os associados, e quem precisa comprar ingressos, o público pagante. Para efeito desta pesquisa, estamos interessados em prever o público total e também o número pessoas do público pagante. Assim, permite que os dirigentes das equipes estejam preparados para o número total de pessoas presentes e também fornece uma estimativa do faturamento da partida.

4.3.2 COLETA DE DADOS

Os dados foram obtidos em parceria com um time de futebol do sul do Brasil e por meio de buscas na internet. Após análise da literatura, foram considerados como variáveis relevantes a data e hora do jogo, se é feriado ou não, a reputação do adversário, o campeonato, a fase do campeonato, a escalação da equipe da casa, atratividade do próximo jogo em casa, resultado da partida anterior e a previsão de temperatura e chuva obtidas em sites da internet. Além disso, foi incluída no conjunto de dados a evolução da venda de ingressos nos seis dias anteriores ao jogo em seis temporadas completas e o nível de marketing realizado para a partida, obtido diretamente do time de futebol parceiro. O conjunto de dados continha um total de 210 jogos e cobriu os anos de 2014 a 2019, sem dados faltantes. Todas as partidas ocorreram no mesmo estádio; portanto, este fator não foi considerado como uma variável do modelo. A capacidade total do estádio é de 55.400 pessoas sentadas.

4.3.3 ANÁLISE DOS DADOS E PRÉ-PROCESSAMENTO

As variáveis relevantes para a obtenção do modelo foram discutidas e validadas em reunião com três especialistas que atuavam na organização parceira de futebol. Além disso, as reputações da equipe adversária foram convertidas em uma escala de atratividade do adversário, de muito pouco atraente a extremamente atraente por dois torcedores do time parceiro e

validada pelos especialistas. Essa variável captura a qualidade do convidado e também a rivalidade entre as equipes como fatores atrativos para aumentar a audiência, conforme indicado pela literatura. Todas as variáveis consideradas e suas descrições são apresentados na **Tabela 7**, incluindo o intervalo e as categorias de cada variável.

Tabela 7 - Rótulos e descrição dos recursos do banco de dados obtido.

<i>Variável</i>	<i>Descrição</i>	<i>Intervalo / Categorias</i>	<i>Tipo</i>
Hora	A hora da partida	Formato: HH:MM	Categórica
DiaSem	O dia da semana em que a partida ocorreu	Segunda-feira Terça-feira Quarta-feira Quinta-feira Sexta-feira Sábado Domingo	Categórica
Feriado	Se o dia no qual ocorreu a partida foi feriado	Sim Não	Booleana
Camp	O campeonato correspondente à partida	Campeonato Brasileiro Campeonato Regional Primeira Liga Copa Sul-americana Copa do Brasil Copa Libertadores	Categórica
Fase	Qual fase do campeonato a partida ocorreu	Eliminatórias Grupo Oitavas de final Quartas de final Semifinal Final	Categórica
Escal	A escalação do time da casa	Reservas Maioria reservas Misto Maioria titulares Titulares Titular + estreia	Categórica

<i>Variável</i>	<i>Descrição</i>	<i>Intervalo / Categorias</i>	<i>Tipo</i>
Atrat	Medida de atratividade do adversário	Excepcionalmente atrativo Muito atrativo Atrativo Médio Pouco atrativo Muito pouco atrativo	Categórica
OpenTV	Se a partida teve transmissão na TV aberta ou não	Sim Não	Booleana
ResultAnt	Saldo de gol da partida anterior calculado pela diferença entre os gols feitos e os gols levados	-4 – 4 gols	Numérica
AtratProx	Atratividade da próxima partida em casa (no mesmo estádio analisado)	Muito baixa Baixa Normal Alta Muito alta	Categórica
Marketing	Nível de publicidade realizada para a partida	Normal Acima do normal Muito acima do normal	Categórica
Chuva	Milímetros de chuva previstos no dia da partida	0 – 200 mm	Numérica
Temp	Temperatura prevista para o dia da partida	0 – 32 °C	Numérica
D-6, D-5, D-4, D-3, D-2	Número de ingressos acumulados vendidos 6, 5, 4, 3 e 2 dias antes do jogo	0 – 27.000 ingressos	Numérica

O conjunto de dados também contém todas as três variáveis de resposta possíveis: público total, pagante e associado para cada jogo. As distribuições dessas variáveis de resposta são apresentadas na **Figura 13**. É possível observar que o Público total e o Público pagante têm distribuições com formas semelhantes e são delimitadas à direita (devido a restrições de capacidade). Por outro lado, os Associados apresentaram uma distribuição ligeiramente diferente e alguns valores extremos (*outliers*). Além disso, a distribuição do público total e do

público pagante apresenta ser uma distribuição bimodal, com dois picos distintos. Isso indica que a maioria dos jogos ou são muito atraentes, com capacidade de lotar o estádio, ou pouco atraentes e com um público menor. Sendo que, os jogos com público total em torno de 20.000 pessoas são maioria. Como o foco é a previsão do público total e pagante, nenhum tratamento de dados foi aplicado aos *outliers* vistos no público associado.

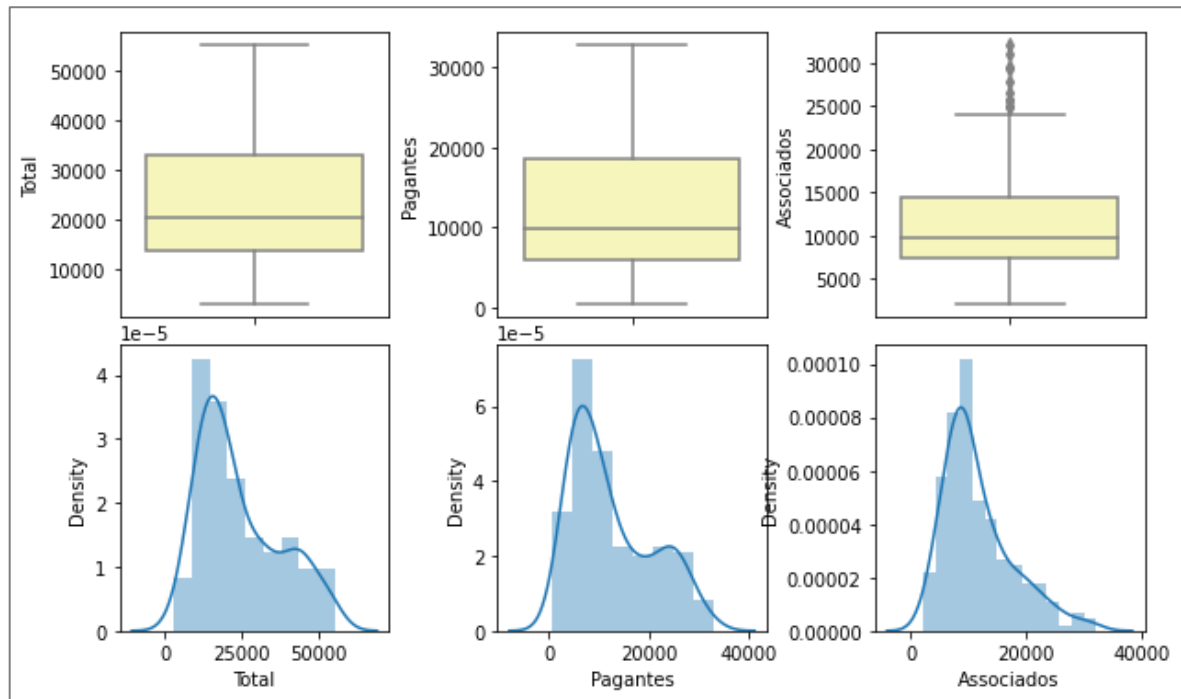


Figura 12 - Boxplot e distribuição das variáveis resposta do banco de dados

Além da distribuição dos dados, é interessante obter uma análise inicial das variáveis e como elas interferem na variável resposta antes de iniciar o estudo. Para isso, foram plotadas várias combinações das variáveis predictoras para obter informações sobre o banco de dados. É possível ver na **Figura 14** a relação entre o público total e as variáveis campeonato e atratividade do adversário. A figura indica que existe uma correlação entre a atratividade do adversário e o campeonato referente à partida. Por exemplo, a maior parte dos adversários muito pouco atrativos estão relacionados ao campeonato regional, ambos resultando em um público total baixo. Além disso, é possível notar também que, quanto maior a atratividade do adversário, maior o público total independentemente do campeonato ao qual a partida se refere.

Outras relações interessantes podem ser vistas na **Figura 15**. Por exemplo, é possível notar na Figura 15(a) que a medida que o campeonato se encaminha para as fases finais (quartas

de final, semi-final e final) o público aumenta para todos os campeonatos, como é esperado. Porém, a intensidade do aumento desse público depende do campeonato ao qual a partida faz parte. Já a Figura 15(b) demonstra que existe uma relação entre o dia da semana em que a partida ocorre e o campeonato ao qual ela pertence. Jogos do campeonato Libertadores da América, por exemplo, tendem a ocorrer nas terças, quartas e quintas-feiras e possuem um público sempre maior que os outros campeonatos. Além disso, jogos do campeonato Brasileiro tendem a atrair um público maior nos finais de semana do que quando ocorrem em dias da semana. Outras análises podem ser vistas no *link* disponibilizado no material suplementar.

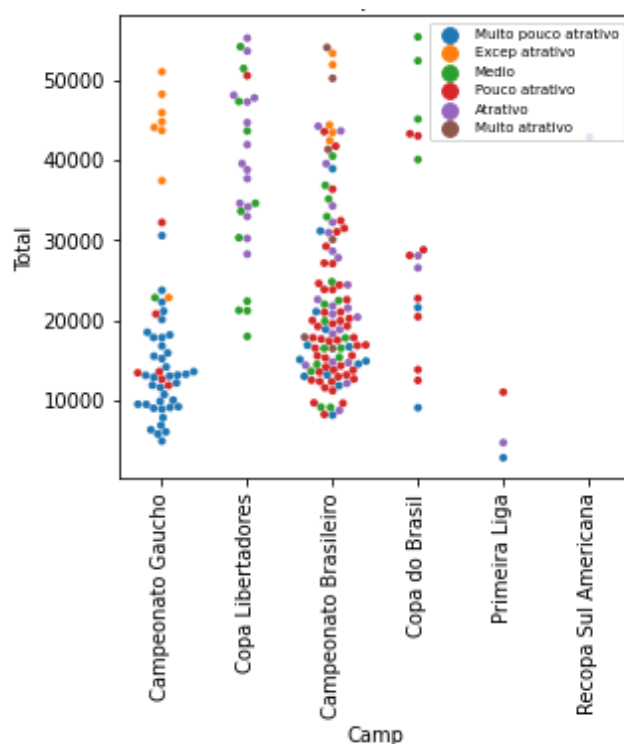


Figura 13 - Relação entre o Público Total e as variáveis Campeonato e Atratividade do Adversário

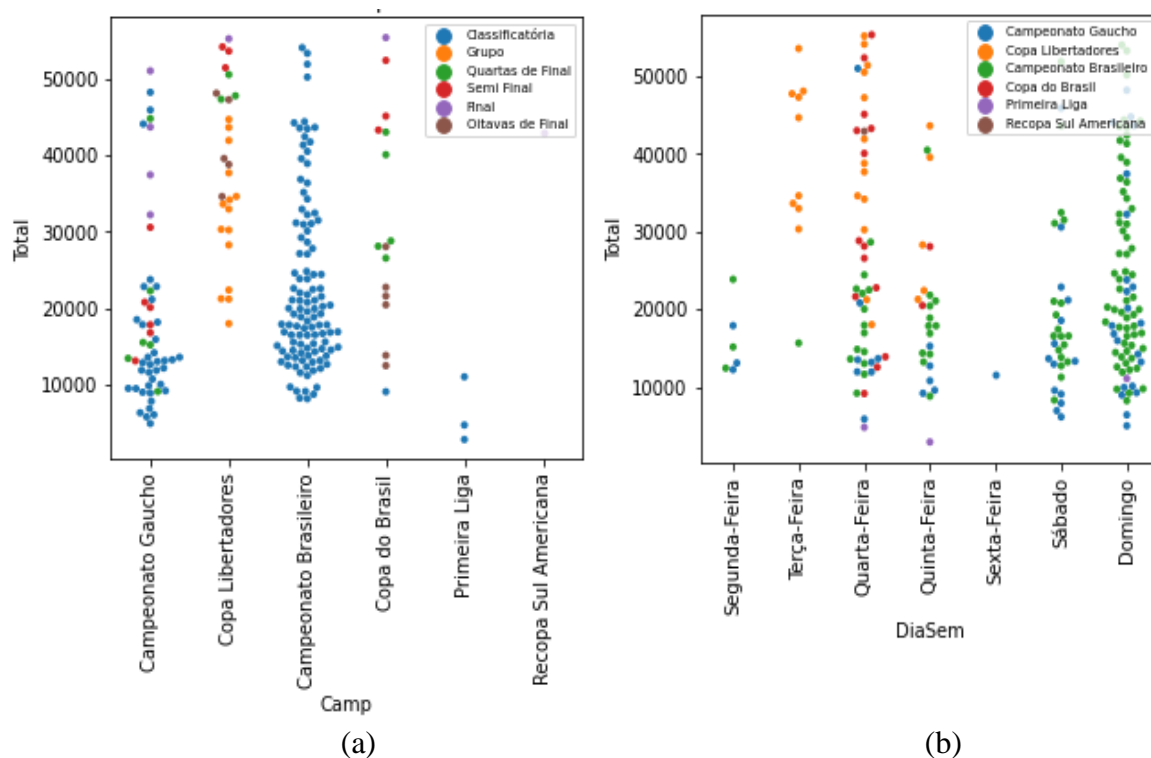


Figura 14 - Relação entre (a) Público Total e as variáveis Campeonato e Fase e (b) Público Total e as variáveis Dia da Semana e Campeonato

4.3.4 MÉTODOS DE REGRESSÃO

Uma vez que não foi encontrado nenhum estudo semelhante na literatura, este artigo utilizou uma variedade de métodos de previsão para avaliar o desempenho e obter o melhor modelo. Foram testados métodos simples de regressão como Modelo de Regressão Linear (LM) e *K-Nearest Neighbors* (KNN) e métodos mais avançados, como modelos baseados em árvore de decisão, métodos de *ensemble* e redes neurais artificiais. Os métodos testados neste estudo são apresentados nas próximas subseções.

4.3.4.1 MODELO DE REGRESSÃO LINEAR

LM é um dos métodos mais antigos conhecidos e tenta descrever uma variável de resposta em função de uma ou mais variáveis independentes. Um importante algoritmo de otimização implementado em modelos lineares para ML é o gradiente descendente ou *gradient descent*. O gradiente descendente utiliza uma abordagem de otimização iterativa que ajusta gradativamente os hiperparâmetros do modelo para minimizar uma função de custo durante o treinamento do modelo, tentando alcançar um custo mínimo. Normalmente, este custo é uma

métrica de erro de previsão (por exemplo, função do erro quadrático). Existem algumas variações que podem ser aplicadas ao gradiente descendente. Neste estudo, o modelo linear foi implementado utilizando o *Stochastic Gradient Descent* (SGD) (Gerón, 2017).

4.3.4.2 MODELO K-NEAREST NEIGHBORS

O modelo KNN é um dos métodos de aprendizagem mais comuns e pode ser usado tanto para problemas de classificação e também para regressão. Um algoritmo de regressão KNN tenta primeiramente identificar as observações de treinamento K que estão mais próximas da observação a ser prevista em termos das variáveis explicativas. E então, a estimativa é feita usando a média de todas as K respostas de treinamento mais próximas. O número de vizinhos K é dependente do conjunto de dados (James et al., 2013).

4.3.4.3 MODELOS BASEADOS EM ÁRVORE DE DECISÃO

Os modelos baseados em árvore de decisão são uma família de métodos que utilizam como base as árvores de decisão. Os modelos simples de árvores de decisão (DT) são um algoritmo de ML não linear que podem ser utilizados para problemas de classificação e regressão. Eles são muito poderosos e capazes de ajustar conjuntos de dados muito complexos e, além disso, podem ser facilmente interpretados (Gerón, 2017). O algoritmo tenta dividir o espaço variáveis em regiões distintas, cada região é conhecida como nós terminais ou folhas. O valor previsto é estimado como a resposta média para a região que a observação a ser prevista (James et al., 2013).

O algoritmo *Random Forest* (RF) é um método *ensemble* baseado na construção de múltiplas árvores de decisão e o valor previsto é a média de todas as respostas individuais das diferentes árvores de decisão elaboradas (Jiang et al., 2018). Cada árvore é construída sem considerar todas as variáveis preditoras disponíveis, a fim de diminuir a variância das respostas, e o mais profundo possível, para diminuir o *bias* (Jiang et al, 2018; James et al., 2013). Portanto, os modelos de RF geralmente funcionam bem mesmo na presença de *outliers*.

Outro algoritmo da família de árvores de decisão é o algoritmo *Extremely Randomized Forest* (XRF), outro método *ensemble* baseado em árvores de decisão. Ele difere de RF por randomizar ainda mais as variáveis e a escolha do ponto de corte ao dividir um nó da árvore. Portanto, ele cria árvores totalmente aleatórias com estruturas independentes que geralmente diminuem a variância dos resultados (Geurts et al., 2006).

Gradient Boosting (GBM) é outro método *ensemble* que visa melhorar as previsões de uma árvore de decisão. O método consiste em construir as árvores sequencialmente; portanto, cada árvore recebe informações da árvore anteriormente desenvolvida, pois sua entrada é a saída da árvore anterior (James et al., 2013). Os métodos de *boosting* visam aprender com os erros anteriores e melhorar a previsão final. No algoritmo GBM, a melhoria é feita ajustando a próxima árvore de decisão aos erros residuais cometidos pela árvore anterior (Gerón, 2017). Outros métodos de *boosting* escolhidos para esta pesquisa foram o *Adaptive Boosting* (ADA) e o *Extreme Gradient Boosting* (XGB).

O algoritmo ADA, em contraste com o GBM, corrige a previsão aprendendo com as observações de treinamento que foram mal ajustadas e ajustando os pesos das variáveis ao desenvolver a nova árvore (Gerón, 2017). Já o *Extreme Gradient Boosting* (XGB) é uma implementação do GBM que melhora a velocidade e o desempenho do algoritmo, desenvolvido por Chen e Guestrin (Chen e Guestrin, 2016). O XGB é capaz de resolver problemas complexos e, por isso, tem sido amplamente utilizado por cientistas de dados e também em competições de ML.

4.3.4.4 MODELO DE REDES NEURAIAS

O modelo de Redes Neurais (ANN) é um algoritmo de *deep learning* que consiste em uma rede de neurônios ou nós conectados em camadas. Cada neurônio nas camadas pode ser ativado por uma função de ativação e transmitir um sinal através da rede para outros neurônios. As conexões entre neurônios contêm um parâmetro de peso associado que é atualizado quando o modelo é treinado, pela técnica de retropropagação (Chollet, 2017). Este algoritmo tem sido amplamente utilizado, uma vez que pode facilmente aprender com padrões de comportamentos, tentando replicar o que acontece dentro do cérebro humano. No entanto, ao lidar com dados limitados, o principal problema do ANN é o *overfitting*. Para contornar esse problema, é utilizada a técnica de *dropout*, na qual a rede esquece e omite aleatoriamente um percentual de neurônios em cada etapa (Jiang et al., 2018). O método de *dropout* foi usado nesta pesquisa, uma vez que o conjunto de dados contém apenas 210 observações, número considerado baixo para aplicações em redes neurais.

4.3.5 CRITÉRIOS DE AVALIAÇÃO E AJUSTES DOS HIPERPARÂMETROS

Os resultados de cada algoritmo foram avaliados principalmente pelo erro absoluto médio (MAE), uma vez que é importante saber quantos indivíduos o algoritmo está prevendo errado. O número absoluto de pessoas se torna importante para o problema em questão pois as decisões de gerenciamento dependem desse número. Por exemplo, para cada 10.000 pessoas dentro do estádio é necessária uma ambulância extra, além de um aumento no número de policiais. Portanto, quanto menor o MAE, menos erros os gestores cometerão. O MAE pode ser calculado pela equação (1), e é simplesmente a média aritmética de todas as diferenças absolutas entre os valores previstos e reais (Gerón, 2017).

Além disso, os modelos foram comparados por sua raiz do erro quadrático médio (RMSE) e o coeficiente de determinação (R^2) entre o valor real e o previsto. O RMSE é uma medida de desempenho típica ao lidar com problemas de regressão. Ele representa quanto erro o modelo geralmente comete e atribui maior peso a erros maiores. O RMSE pode ser calculado pela equação (2). Além dele, o R^2 também é muito utilizado para problemas de regressão pois é uma medida de quão bom foi o ajuste dos dados, calculado pela equação (3).

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_i - y_i| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - y_i)^2} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (Y_i - y_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2} \quad (3)$$

Onde, Y_i são os valores reais, \bar{Y} é a média dos valores reais, y_i são os valores previstos e N é o número de observações.

Por fim, os métodos foram posteriormente comparados por meio de suas curvas de aprendizado, para ajudar a avaliar se o modelo obtido não está com problemas de *underfitting* ou de *overfitting*. Portanto, foram plotados os escores de treinamento e validação durante o treinamento do modelo para ver como o algoritmo estava aprendendo. A **Figura 16** demonstra diferentes curvas de aprendizado para modelos com *underfitting*, *overfitting* e bom ajuste. Observe que para um bom modelo de ajuste, o erro de treinamento começa em um determinado

valor e diminui até atingir um valor constante menor que o valor alcançado no conjunto de validação. Quando o modelo está *overfitting*, o erro de treinamento começa em zero e permanece neste valor até um ponto em que começa a aumentar. Por outro lado, quando o modelo está *underfitting*, o erro aumenta no início e se estabiliza em um valor mais alto.

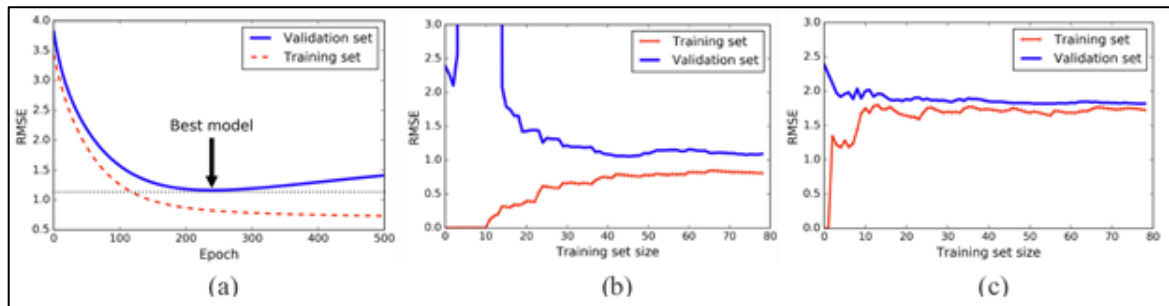


Figura 15 - Diferentes curvas de aprendizagem: (a) bom ajuste do modelo (b) modelo de *overfitting* (c) modelo de *underfitting* (Adaptado de Géron, 2017)

4.3.6 PROCEDIMENTO EXPERIMENTAL

A primeira etapa desta pesquisa foi coletar todos os dados e obter o banco de dados. A coleta foi realizada de acordo com o especificado no item anterior, por meio de busca na internet e reuniões com especialistas da equipe parceira. Assim que o banco de dados estava completo, a próxima etapa de análise de dados e pré-processamento pode ser executada. Nesta etapa, os dados foram analisados quanto à sua distribuição, foram obtidas estatísticas e correlações entre as variáveis consideradas. Alguns desses resultados também foram apresentados anteriormente nesta seção. O procedimento experimental detalhado pode ser observado na **Figura 17**.

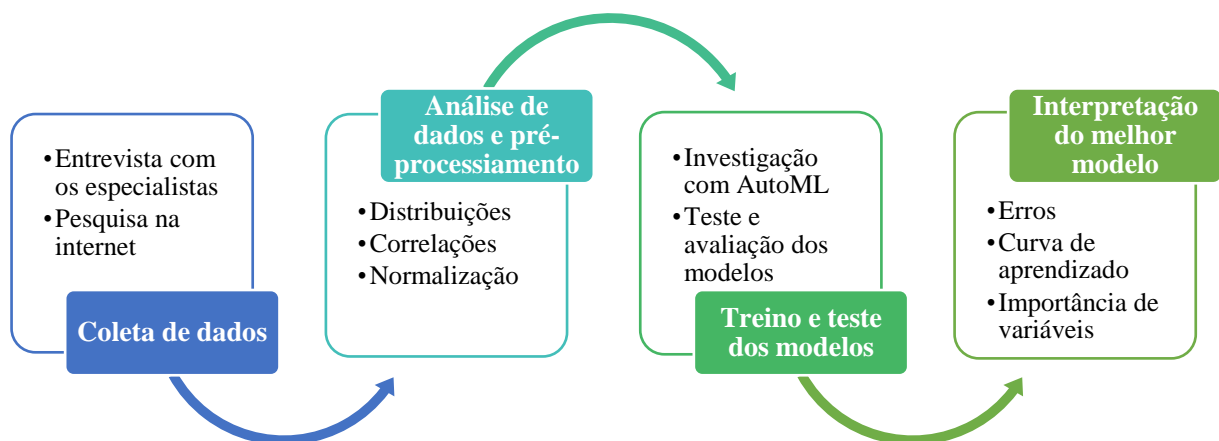


Figura 16 - Procedimento experimental detalhado

Em seguida, o treinamento e o teste dos modelos foram realizados. O estudo foi desenvolvido em um ambiente Python 3 usando as seguintes bibliotecas de código aberto: Numpy, Pandas, ScikitLearn, Keras, Tensorflow e AutoML, da plataforma h2o. O código completo, a análise e os resultados também podem ser encontrados no material suplementar em forma de *notebook* Python.

Todos os métodos foram testados e os modelos desenvolvidos com a aplicação do método de validação cruzada *Leave-One-Out Cross Validation* (LOOCV), uma vez que os dados eram limitados. Este método divide o conjunto de dados em duas partes, uma para treino e outra para teste do modelo. Porém, uma vez que a base de dados é pequena, o conjunto de teste contém apenas uma observação. Em seguida, é aplicado um *loop* ao algoritmo para treinar e testar o modelo em todas as observações. A medida de erro calculada para o modelo é a média de todos os *loops* de validação cruzada (James et al., 2013).

Para este estudo, a primeira etapa do desenvolvimento dos modelos foi investigar os possíveis métodos e hiperparâmetros com a ajuda do aprendizado de máquina automatizado. Para tanto, foi utilizado o pacote AutoML desenvolvido pela h2o. Este pacote testa uma variedade de métodos, desde os mais simples aos mais complexos, e busca os melhores hiperparâmetros que minimizam os erros de previsão dos modelos. Sendo assim, primeiramente obteve-se os melhores hiperparâmetros dos métodos selecionados com o AutoML para serem utilizados posteriormente nos testes com a base de dados. Esse procedimento foi aplicado para a previsão tanto do público total quanto do público pagante.

Por fim, todos os métodos selecionados foram analisados com ScikitLearn e Keras para obtenção dos melhores modelos de previsão de público total. Os dados foram padronizados para avaliação, uma vez que alguns métodos requerem essa etapa. Para padronização, as variáveis foram escalonadas para uma variância unitária e média zero. O único algoritmo que foi avaliado sem padronização dos dados foi o KNN. Após a obtenção dos modelos, os resultados foram convertidos de volta para a escala original para melhor compreensão dos resultados. As métricas para avaliação também foram descritas anteriormente. Em seguida, os quatro modelos de melhor desempenho para prever o público total foram avaliados para prever o público pagante. Finalmente, o melhor modelo foi interpretado para encontrar as variáveis mais significativas usando o algoritmo de importância da biblioteca do ScikitLearn.

4.4 TREINAMENTO E TESTE DOS MODELOS

Conforme detalhado anteriormente, a primeira etapa foi aplicar o aprendizado de máquina automatizado a ambos os conjuntos de dados. Os resultados do AutoML indicaram que os melhores métodos para prever o comparecimento ao jogo eram os métodos *ensemble* baseados em árvore de decisão para público total e também para o público pagante. O algoritmo XGB apresentou o melhor desempenho, seguido de GBM, XRT, RF e ANN para ambos os casos. Os hiperparâmetros indicados pelo AutoML diferiram ligeiramente para os métodos de *ensemble* baseados em árvore de decisão quando aplicados para público total ou pagante, enquanto que para a ANN foram exatamente os mesmos. Esses hiperparâmetros foram registrados e utilizados para a análise posterior.

Uma vez que as distribuições das variáveis e os métodos encontrados eram semelhantes, análises mais profundas foram conduzidas utilizando apenas o banco de dados para prever apenas o público total. Além disso, considerando o desempenho dos métodos de *ensemble* baseado em árvore de decisão, foram incluídos para análise os métodos ADA e DT. Além deles, métodos simples como LM e KNN também foram avaliados para comparação. Nesta seção serão apresentados os resultados obtidos nessa análise, que serão discutidos na seção seguinte. A **Tabela 8** resume as métricas avaliadas para os nove métodos.

Na **Tabela 8**, pode-se notar que o modelo XGB apresentou melhor desempenho nas três métricas. Desta forma, teve, em média, o menor erro absoluto e foi o menos penalizado por erros grandes por apresentar também o menor RMSE. O RF, GBM e ADA foram, respectivamente, os seguintes melhores métodos. Ao comparar os modelos RF e GBM, eles

obtiveram o mesmo MAE. No entanto, o modelo RF foi menos penalizado por grandes erros, uma vez que seu RMSE é menor. A distribuição dos erros absolutos dos valores preditos para cada algoritmo é apresentada na **Figura 6**, por meio de gráficos de boxplots.

Tabela 8 - Performance dos modelos e métricas de erros para prever público total

	LM	KNN	DT	RF	XRF	GBM	ADA	XGB	ANN
MAE	4.537	4.311	4.500	3.576	3.830	3.508	3.520	3.353	4.060
RMSE	5.863	5.806	6.049	4.767	5.157	4.598	4.688	4.468	5.971
R²	0,803	0,807	0,790	0,870	0,848	0,879	0,874	0,886	0,796

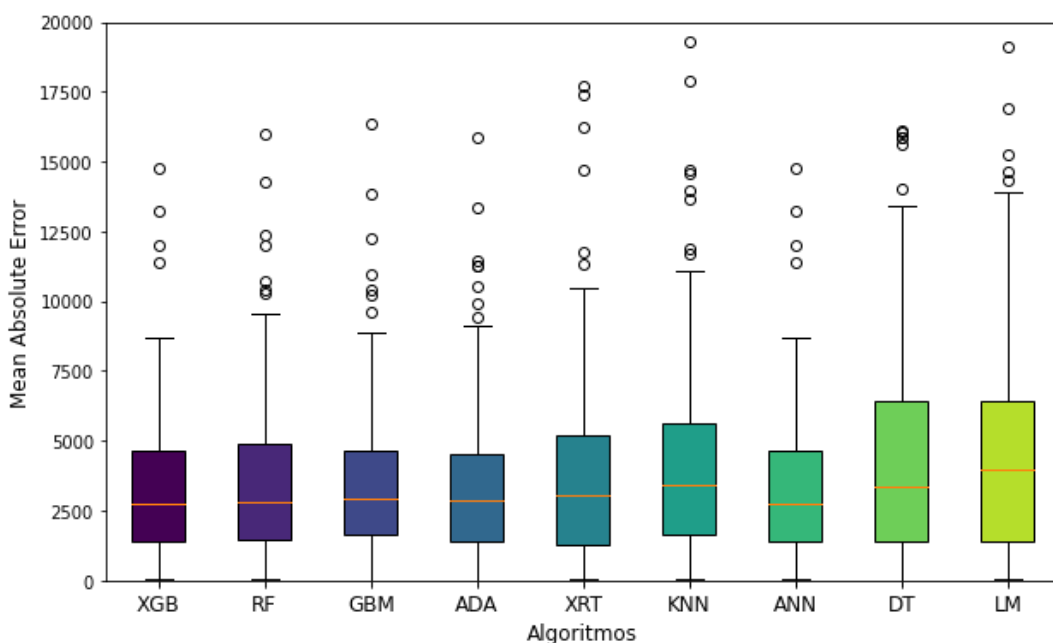


Figura 17 - Comparação da distribuição dos erros de previsão de público total de cada algoritmo

A **Figura 18** revela que todos os métodos tiveram alguns *outliers* em sua distribuição de erros, apresentando alguns erros acima de 7.500 pessoas. No entanto, os métodos XGB e RF apresentaram menos *outliers* e seus boxplots eram mais estreitos em comparação com os outros, com a maioria dos erros em torno de 2.000 a 5.000 pessoas. Além disso, a **Figura 19** exibe uma representação gráfica dos valores previstos em relação aos valores reais para os nove métodos. Quanto mais próximas as observações estiverem da linha vermelha, melhor será o algoritmo.

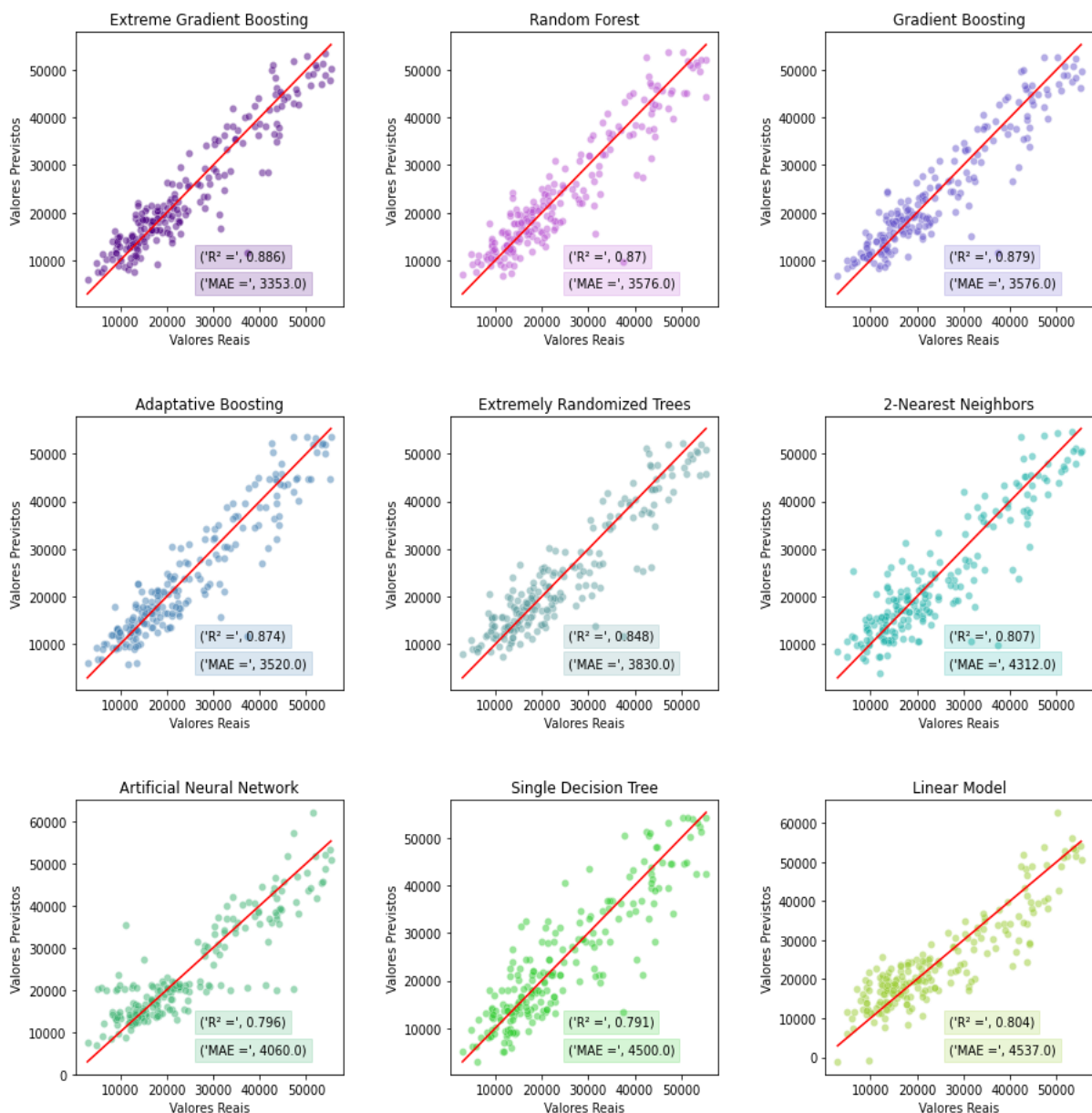


Figura 18 - Comparação entre os valores previstos e valores reais dos diferentes métodos para público total

Por fim, para completar a avaliação inicial de desempenho dos modelos, as curvas de aprendizagem foram geradas, como pode ser visto na **Figura 21**. Depois de analisar o desempenho de todos os modelos, os quatro melhores métodos foram XGB, RF, GB e ADA. Em seguida, apenas esses quatro métodos foram avaliados no conjunto de dados público pagante para confirmar sua capacidade de previsão.

Os resultados de performance dos modelos para prever o público pagante são mostrados na **Tabela 9**. Mais uma vez, o modelo XGB teve o melhor desempenho. No geral, todos os quatro métodos analisados obtiveram desempenhos semelhantes. O modelo XGB

mostrou-se mais eficiente nas três métricas de erros, com os menores MAE e RMSE, assim como melhor aproximação dos dados reais com R^2 maior.

Tabela 9 - Performance dos modelos e métricas de erros para prever público pagante

	RF	GBM	ADA	XGB
MAE	2.617	2.636	2.617	2.456
RMSE	3.661	3.646	3.718	3.358
R²	0,797	0,798	0,791	0,829

Assim como foi feito para avaliação de previsão do público total, os *boxplots* de distribuição dos erros de cada algoritmo para previsão do público pagante foram plotados para comparação, de acordo com a **Figura 20**.

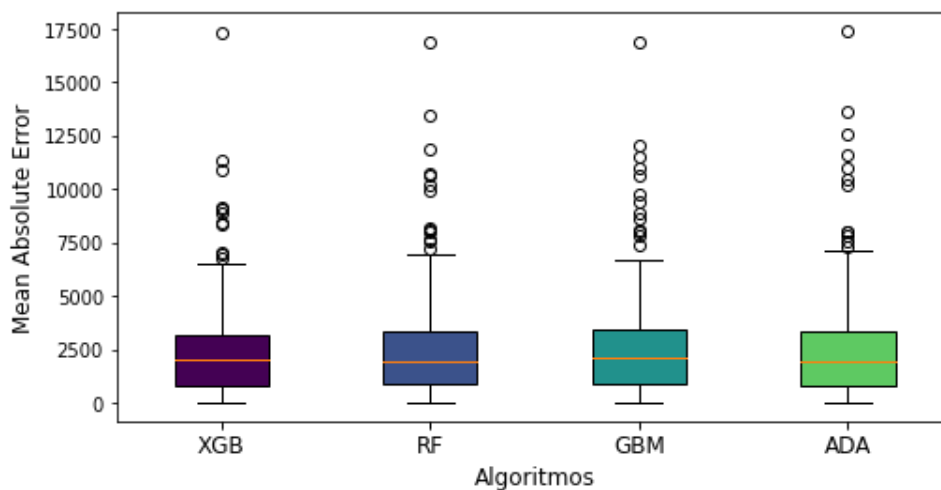


Figura 19 - Comparação da distribuição dos erros de previsão de público pagante de cada algoritmo

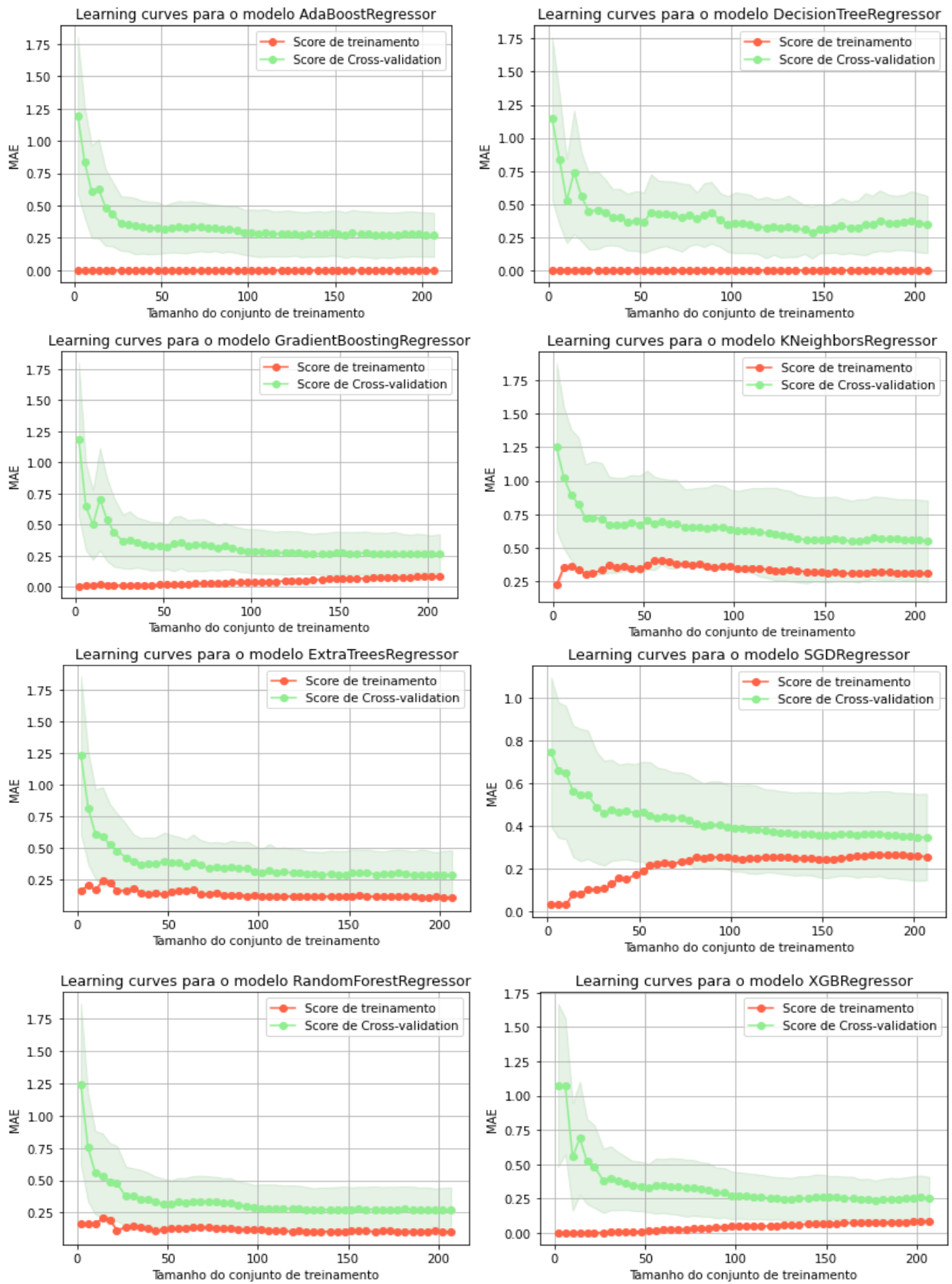


Figura 20 - Curvas de aprendizados dos diferentes modelos

A **Figura 20** mostra que a previsão do público pagante também apresentou alguns *outliers* na sua distribuição de erros, com uma observação específica muito distante das outras, em torno de 17.000 pessoas para os quatro métodos. Além disso, todos os métodos foram muito semelhantes em sua distribuição de erros, com a maioria dos erros ocorrendo até cerca de 3.000. Os métodos XGB e GBM apresentaram os boxplots mais estreitos, com os *outliers* mais baixos. Os valores previstos para este cenário foram plotados contra os valores reais e mostrados na **Figura 22**.

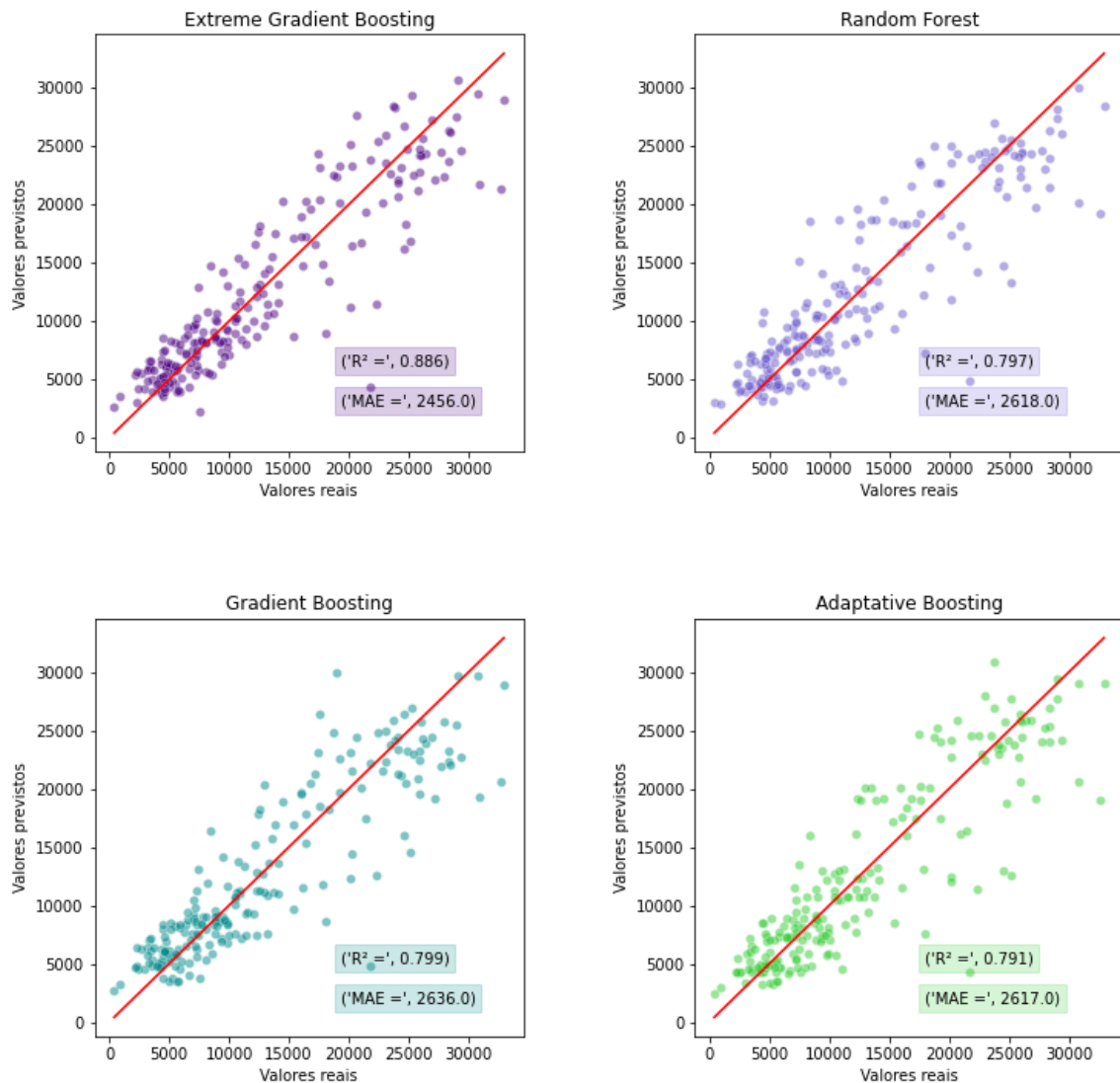


Figura 21 - Comparação entre os valores previstos e valores reais dos diferentes métodos para público pagante

4.5 DISCUSSÃO

Os resultados obtidos a partir do conjunto de dados de futebol total e público pagante demonstram que os métodos de conjunto baseados em árvore são um poderoso conjunto de métodos para prever o público do jogo. Mesmo RF, como uma técnica de conjunto mais simples, teve um bom desempenho com os recursos considerados. O XGB, GBM, ADA e XRT tiveram eficiências semelhantes e também curvas de aprendizado de acordo com uma curva de modelo de bom ajuste (Géron, 2017).

Ao analisar os métodos mais simples LM e KNN, é possível perceber que eles geraram erros maiores e também as distribuições de erros mais dispersas. Além disso, eles foram os únicos métodos que tiveram uma curva de aprendizado típica para um modelo de *underfitting* (Géron, 2017). A indicação de *underfitting* é o aumento do erro de treinamento no início da curva e posterior estabilização em um valor maior que o inicial.

O modelo de ANN apresentou grandes erros quando comparado aos modelos baseados em árvore de decisão. Isso poderia ser explicado pelo número reduzido de observações ou a necessidade de maior exploração da arquitetura da rede neural. Considerando a distribuição do erro, esse método é promissor por ter um formato estreito, semelhante à distribuição XGB, com poucos *outliers*. Além disso, houve algum progresso na literatura usando este algoritmo para aplicações semelhantes, corroborando seu potencial.

Desse modo, pode-se concluir que XGB e GBM foram os melhores modelos para a previsão do total e do público pagante em uma partida de futebol, uma vez que obtiveram os menores erros. O modelo XGB obteve os menores erros absolutos MAE para a previsão de público total e pagante, confirmando sua eficácia. Usando a função de importância das variáveis disponível no pacote ScikitLearn, foi possível analisar quais as variáveis preditoras foram mais decisivas na previsão para esses dois métodos. Além disso, podem ser comparadas as diferenças entre o público total e o público pagante. O resultado de importância dado pela função é um valor entre 0 e 1, quanto mais próximo de um, mais importante é esse preditor para o modelo. Além disso, todos os pesos de importância somam 1 e a função considera cada categoria das variáveis como uma variável diferente.

A **Tabela 10** descreve as variáveis mais importantes dos modelos obtidos com o algoritmo XGB. Como pode ser visto, o recurso D-3 foi o mais importante para a previsão de público total e pagante. No entanto, o peso do associado para prever o público total é ligeiramente maior do que para o público pagante. Portanto, essa variável pode explicar melhor

a variância do público total. Os times cuja atratividade é Excepcionalmente Atrativo também possuem uma grande importância para ambos os públicos total e pagante. Por outro lado, podemos notar que o público total é influenciado pelo resultado da partida anterior, enquanto o público pagante é mais influenciado pelo campeonato e pela fase a qual a partida pertence.

Tabela 10 - Variáveis mais importantes para os modelos XGB

Público Total		Público Pagante	
Nome da variável	Peso de importância	Nome da variável	Peso de importância
<i>D-3</i>	<i>0,584</i>	<i>D-3</i>	<i>0,425</i>
<i>Atrat_Excep atrativo</i>	<i>0,060</i>	<i>Atrat_Excep atrativo</i>	<i>0,057</i>
<i>D-5</i>	<i>0,033</i>	<i>D-5</i>	<i>0,047</i>
<i>ResultAnt</i>	<i>0,020</i>	<i>Fase_Classificatória</i>	<i>0,044</i>
<i>Hora_16:00</i>	<i>0,019</i>	<i>D-4</i>	<i>0,039</i>
<i>Escal_Reservas</i>	<i>0,017</i>	<i>Camp_Copa Libertadores</i>	<i>0,036</i>
<i>AtratProx_Muito baixa</i>	<i>0,017</i>	<i>Fase_Grupo</i>	<i>0,036</i>
<i>Camp_Campeonato Brasileiro</i>	<i>0,015</i>	<i>D-6</i>	<i>0,034</i>
<i>Fase_Classificatória</i>	<i>0,015</i>	<i>Marketing_Acima do normal</i>	<i>0,029</i>
<i>Marketing_Acima do normal</i>	<i>0,014</i>	<i>Temp</i>	<i>0,017</i>

Além destas, as características mais importantes dos modelos GBM podem ser vistas na **Tabela 11**. Da mesma forma, para o modelo XGB, o recurso D-3 foi o mais importante para prever tanto o público total quanto o pagante. No entanto, os pesos associados foram muito maiores para ambos os casos em comparação com o modelo XGB. Além disso, os modelos GBM deram uma importância muito grande para a evolução de venda de ingressos, ficando dependente dessa variável. Ainda assim é possível perceber algumas diferenças nas influências das variáveis entre os dois públicos. O público total é mais influenciado pela temperatura, chuva do que o público pagante. Isso é razoável, uma vez que o público comprador de ingressos escolherá as melhores partidas para comprar um ingresso, sendo mais influenciados pela atratividade do adversário e campeonato e fase da partida. Por fim, o resultado da partida anterior também parece ter importância relevante para os dois públicos.

Assim, pode-se concluir que o público pagante é sim mais influenciado pelo campeonato e fase enquanto o público total é influenciado pela temperatura, chuva e data do

jogo. Além disso, esses achados estão de acordo com o encontrado na literatura. Isso confirma que a importância da liga é um forte preditor do comparecimento, conforme Popp et al. (2017) concluiu em seu trabalho. Além disso, as condições meteorológicas, dia e hora do jogo foram considerados importantes em outros estudos da literatura (Sahin e Erol, 2018; Popp et al, 2017; Coates et al., 2014; Griffith 2010).

Tabela 11 - Variáveis mais importantes para os modelos GBM

Público Total		Público Pagante	
Nome da variável	Peso de importância	Nome da variável	Peso de importância
<i>D-3</i>	0,886	<i>D-3</i>	0,835
<i>D-5</i>	0,033	<i>D-5</i>	0,047
<i>Temp</i>	0,015	<i>D-4</i>	0,026
<i>Chuva</i>	0,012	<i>D-6</i>	0,013
<i>D-4</i>	0,009	<i>Atrat_Except atrativo</i>	0,012
<i>Atrat_Except atrativo</i>	0,008	<i>Atrat_Atrativo</i>	0,008
<i>ResultAnt</i>	0,007	<i>Temp</i>	0,007
<i>Fase_Final</i>	0,007	<i>Fase_Final</i>	0,007
<i>D-6</i>	0,006	<i>Chuva</i>	0,005
<i>Arear_muito pouco atrativo</i>	0,002	<i>ResultAnt</i>	0,004

4.6 CONCLUSÃO

Este estudo analisou com sucesso diferentes métodos de *Machine Learning* para encontrar o melhor modelo de previsão de público total e de público pagante em jogos de futebol, com dois dias antes do dia do jogo. Para atingir os objetivos da pesquisa, dados de jogos foram coletados em parceria com um time brasileiro de futebol e um banco de dados contendo 210 jogos foi coletado junto às seguintes características: data e hora do jogo, atratividade do adversário, campeonato, fase do campeonato, escalação da equipe da casa, temperatura, chuva, nível de *marketing*, resultado da partida anterior, transmissão em TV aberta e evolução da venda de ingressos de seis a dois dias antes da partida.

Foram avaliados nove métodos diferentes, utilizando a técnica LOOCV e encontrando os melhores hiperparâmetros dos modelos com o auxílio do aprendizado de máquina automatizado AutoML. Os resultados indicaram que os métodos de *ensemble* baseados em árvore de decisão são capazes de capturar os padrões de comportamento em relação ao

comparecimento à essas partidas. Ademais, os modelos XGB e GBM obtiveram os melhores desempenhos, com os erros mais baixos e os *bloxplots* de distribuição de erros mais estreitos. Além disso, o modelo de ANN mostrou resultados promissores, indicando que poderia ter um desempenho melhor com um conjunto de dados maior. Por outro lado, os modelos LM e KNN tiveram problemas de *underfitting*, comprovando a não linearidade do problema.

O estudo, ainda, mostrou uma forma diferente de prever o comparecimento do público no jogo, com a utilização de métodos mais avançados, mas simples, e que não são caros computacionalmente. Além disso, difere da literatura por utilizar as informações de venda de ingressos nos dias anteriores à partida. Os resultados mostraram a importância desse recurso e os gestores podem se beneficiar disso, uma vez que é um dado facilmente obtido por eles.

Outra importante diferença e contribuição da pesquisa foi a previsão tanto do público total quanto do público pagante. Essa separação pode beneficiar fortemente os gerentes de eventos, fornecendo duas informações diferentes: a quantidade total de pessoas presentes e a receita esperada com a venda de ingressos. Além disso, essas informações foram fornecidas com o tempo necessário para a tomada de decisões gerenciais relevantes. Assim, os gerentes podem dimensionar de forma confiável a equipe, o policiamento e os cuidados médicos necessários para o evento.

Do mesmo modo, os modelos desenvolvidos puderam ser interpretados, fornecendo *insights* sobre as motivações do comparecimento ao jogo. De forma geral, o número acumulado de ingressos vendidos dois dias antes do jogo foi a característica mais importante. Também mostrou que o público pagante, aqueles que compram o ingresso, é mais influenciado pela qualidade do jogo, sendo motivado pela atratividade do campeonato, fase e adversário. Na análise do público total, variáveis mais diretas estão envolvidas, como dia do jogo, temperatura e hora. Essa interpretação também pode ser usada por gestores para entender o nível de público e tentar impulsioná-lo. Os resultados deste estudo podem ser extrapolados para eventos semelhantes (por exemplo, outros esportes ou shows) e os modelos podem ser ajustados seguindo a mesma metodologia apresentada anteriormente.

Uma limitação importante deste estudo foi o número restrito de observações no conjunto de dados. Uma melhoria futura seria expandi-lo e adicionar diferentes times da casa à análise. Além disso, a presença de *outliers* nas distribuições dos erros pode indicar que há outra variável explicativa que poderia ser adicionada à análise. Portanto, esta é outra melhoria futura que pode ser aplicada. No entanto, os valores de MAE obtidos com o modelo XGB foram

considerados satisfatórios pelos especialistas e foram capazes de fazer previsões confiáveis, visto que suas decisões gerenciais são impactadas a cada 5.000 pessoas no estádio.

4.7 MATERIAL COMPLEMENTAR

O código completo das análises, bem como comentários adicionais, pode ser visto no Notebook do Google Colab através do link abaixo:

https://colab.research.google.com/drive/18cOVGpfdZ1n6WBRk5bT8CnjIB4LauOUP?usp=s_haring

4.8 REFERÊNCIAS

- A. GÉRON, **Hands-On Machine Learning with Scikit-Learn and TensorFlow**, 1st ed., O'Reilly Media, United States of America, 2017.
- A. TEALAB, Time series forecasting using artificial neural networks methodologies: A systematic review, **Future of Computing Informatics J.** 3 (2018) 334–340. <https://doi.org/10.1016/j.fcij.2018.10.003>.
- D. COATES, B.R. HUMPHREYS, L. ZHOU, Reference-dependent preferences, loss aversion, and live game attendance, **Economics Inquires** 52 (2014) 959–973. <https://doi.org/10.1111/ecin.12061>.
- D. GETZ, **Event Studies: Theory, Research and Policy for Planned Events**, Elsevier, Oxford, 2008.
- D. STRNAD, A. NERAT, Š. KOHEK, Neural network models for group behavior prediction: a case of soccer match attendance, **Neural Computer Applied.** 28 (2017) 287–300. <https://doi.org/10.1007/s00521-015-2056-z>.
- D.A. GRIFFITH, An analytical perspective on sporting events attendance: The 2007-2008 US NCAA college bowl games, **Applied Geography.** 30 (2010) 203–209. <https://doi.org/10.1016/j.apgeog.2009.01.005>.
- E.M. BECKMAN, W. CAI, R.M. ESROCK, R.J. LEMKE, Explaining Game-to-Game Ticket Sales for Major League Baseball Games Over Time, **Journal of Sports Economy.** 13 (2012) 536–553. <https://doi.org/10.1177/1527002511410980>.
- F. CHOLLET, **Deep Learning with Python**, Manning Publications Company, Berkeley, CA, 2017.
- G. JAMES, D. WITTEN, T. HASTIE, R. TIBSHIRANI, **An Introduction to Statistical Learning**, Springer New York, New York, NY, 2013. <https://doi.org/10.1007/978-1-4614-7138-7>.

- G.T. TRINH, The attendance at sporting events: A generalized theory and its implications, **International Journal of Marketing Research**. 60 (2018) 232–237. <https://doi.org/10.1177/1470785318774677>.
- H. SUNG, B.M. MILLS, Estimation of game-level attendance in major league soccer: Outcome uncertainty and absolute quality considerations, **Sport Management**. Rev. 21 (2018) 519–532. <https://doi.org/10.1016/j.smr.2017.12.002>.
- J. HALL, B. O'MAHONY, J. VIECELI, An empirical model of attendance factors at major sporting events, **International Journal of Hospitality Management**. 29 (2010) 328–334. <https://doi.org/10.1016/j.ijhm.2009.10.011>.
- J.G. CLAUDINO, D. DE O. CAPANEMA, T.V. DE SOUZA, J.C. SERRÃO, A.C. MACHADO PEREIRA, G.P. NASSIS, Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: a Systematic Review, **Sporting Medicine - Open**. 5 (2019) 28. <https://doi.org/10.1186/s40798-019-0202-3>.
- M. ŞAHİN, R. EROL, Prediction of Attendance Demand in European Football Games: Comparison of ANFIS, Fuzzy Logic, and ANN, **Computational Intelligence and Neuroscience**. 2018 (2018). <https://doi.org/10.1155/2018/5714872>.
- N. POPP, J. JENSEN, R. JACKSON, Maximizing visitors at college football bowl games, **International Journal of Event and Festival Management**. 8 (2017) 261–273. <https://doi.org/10.1108/IJEFM-02-2017-0014>.
- P. GEURTS, D. Ernst, L. Wehenkel, Extremely randomized trees, **Mach. Learn**. 63 (2006) 3–42. <https://doi.org/10.1007/s10994-006-6226-1>.
- P.E. SERWYLO, P. ARBON, G.W. RUMANTIR, Predicting patient presentation rates at mass gatherings using Machine Learning, in: **ICIS**, 2011.
- R. HELAL, C. GORDON, A crise do futebol brasileiro: perspectivas para o século XXI., Rev. **Eco-Pós**. 5 (2002) 37–55. <https://doi.org/https://doi.org/10.29146/eco-pos.v5i1.1155>.
- S. JIANG, K.-S. CHIN, G. QU, K.L. TSUI, An integrated Machine Learning framework for hospital readmission prediction, **Knowledge-Based Systems**. 146 (2018) 73–90. <https://doi.org/10.1016/j.knosys.2018.01.027>.
- T. CHEN, C. GUESTRIN, XGBoost, in: Proc. 22nd ACM SIGKDD **International conference on knowledge discovery & data mining - KDD '16**, ACM Press, New York, New York, USA, 2016: pp. 785–794. <https://doi.org/10.1145/2939672.2939785>.
- W. JUN, L. YUYAN, T. LINGYU, G. PENG, Modeling a combined forecast algorithm based on sequence patterns and near characteristics: An application for tourism demand forecasting, **Chaos, Solitons & Fractals**. 108 (2018) 136–147. <https://doi.org/10.1016/j.chaos.2018.01.028>.

5 CONSIDERAÇÕES FINAIS

Este capítulo resume os principais achados decorrentes do estudo e apresenta sugestões para pesquisas futuras.

5.1 CONCLUSÕES

A importância dos métodos de previsão de demanda para uma gestão eficiente de empresas e corporações, em vários cenários, já é amplamente conhecida. Por esta razão, esses métodos vêm sendo adaptados e aplicados para diferentes setores, como serviços, turismo e esportes. Em vista disso, esta dissertação de mestrado buscou ampliar o conhecimento nesta área, investigando como os métodos de previsão podem ser aplicados para prever o número de pessoas em um evento e auxiliar na gestão dos mesmos. Neste contexto, foram delimitados os objetivos específicos apresentados no capítulo 1, atingidos por meio dos três estudos expostos nos capítulos seguintes.

Primeiramente, no Artigo 1, buscou-se identificar quais fatores motivacionais levam as pessoas a comparecerem em eventos e quais métodos de previsão estavam sendo utilizados nesta área. Para isso, foi realizada uma revisão sistemática da literatura que, após aplicar os critérios expostos no capítulo 2, encontrou 56 artigos relevantes que contribuiriam para o estudo. Com base nos achados da literatura, foram identificados 47 fatores motivacionais afetam o comparecimento em eventos. Esses fatores foram organizados em nove dimensões e associados a perguntas que auxiliam os organizadores na gestão dos eventos, apoiando a tomada de decisões. Até este momento, nenhuma outra pesquisa reuniu os fatores e motivações para participação em eventos de forma completa, sem restringir a um tipo de evento específico.

Além disso, a revisão sistemática da literatura revelou poucos artigos que, de alguma forma, faziam previsão de público em eventos. Foram encontrados apenas 17 artigos entre os estudos selecionados que tentavam prever o público. As previsões, ainda, na sua maioria, foram aplicadas a eventos EBSNs e com a utilização de métodos de classificação associados a *Machine Learning*. Aqueles estudos que utilizaram algum método de regressão foram aplicados a eventos esportivos, porém sem fazer uso de metodologias mais avançadas, como *Machine Learning*. Desta forma, o artigo revelou espaço para trabalhos que envolvam métodos de regressão em diferentes tipos de eventos e aplicação de métodos mais avançados.

Com isto em mente, o Artigo 2 utilizou os fatores motivacionais levantados, bem como as perguntas associadas, para desenvolver um *framework* de previsão de público que pode ser

aplicado a diferentes tipos de eventos. Este *framework* foi avaliado por uma profissional da área de eventos ao realizar duas simulações de previsão de público. O principal benefício da estrutura proposta é que ela é facilmente aplicada a diferentes tipos de eventos, ao contrário dos métodos encontrados na literatura. Aliado a isso, o *framework* é adaptável às especificidades de cada evento pois os responsáveis pela previsão podem ajustar os pesos de previsão de acordo com a sua experiência e o histórico de participação. Por fim, a estrutura também pode ajudar os organizadores a medir a qualidade de seu evento e avaliar os pontos fracos (fatores motivacionais com avaliação mais baixa) a serem melhorados, a fim de atrair maior público.

Por fim, o Artigo 3 se propôs a ampliar a literatura em previsão de público em eventos esportivos, ao aplicar métodos mais avançados de *Machine Learning*. Até o momento, a literatura indicou que o uso de metodologias de *Machine Learning* na área de eventos esportivos foi utilizada apenas para previsão do resultado da partida e ocorrências médicas. Assim, o estudo demonstrou uma forma diferente de prever o comparecimento do público no jogo, com a utilização de métodos mais avançados, mais simples, e que não são caros computacionalmente. Além disso, difere da literatura por utilizar as informações de venda de ingressos nos dias anteriores à partida. Os resultados mostraram a importância desse recurso e os gestores podem se beneficiar com ele, uma vez que é um dado à disposição e facilmente obtido.

Em síntese, os objetivos específicos desta dissertação foram atingidos ao longo dos três artigos apresentados, estendendo o conhecimento sobre os métodos de previsão de demanda em outros contextos, aplicando-os à previsão de público em eventos. Além disso, os estudos demonstraram a importância desses métodos para uma gestão eficiente dos eventos. É possível, através da previsão de público, entender os pontos fortes e fracos dos eventos, estabelecer ações para tornar os eventos mais atrativos e dimensionar corretamente a infraestrutura de suporte ao evento.

5.2 SUGESTÕES PARA PESQUISAS FUTURAS

Com base no desenvolvimento e nas contribuições dos estudos apresentados, considerou-se relevantes alguns pontos para que pesquisas futuras avancem o assunto abordado nesta dissertação. Primeiramente, sabe-se que cada evento possui características intrínsecas que os diferenciam dos demais. Sendo assim, sugere-se que o *framework* elaborado no Artigo 2 seja validado para outros tipos de eventos, ampliando a sua aplicação e demonstrando a sua

adaptabilidade. Além disso, é possível ainda realizar um estudo da importância de cada fator motivacional para o evento em análise, alterando-se os pesos associados a partir de um histórico de comparecimento ao evento. Sendo assim, é possível aprofundar o estudo iniciado, obtendo mais resultados práticos sobre a gestão de eventos conforme requerido pela literatura. Ademais, é interessante validar os resultados obtidos com as simulações com a realização dos cenários propostos. Além de utilizar o *framework* para prever o público em outros tipos de eventos, pesquisas futuras poderiam explorar o seu potencial para identificar pontos fracos, planejar ações de melhoria e ampliar público. Sugere-se a realização de um estudo de caso, no qual os fatores motivacionais que se encontram abaixo do esperado sejam objeto de ações de desenvolvimento e melhoria do evento.

Por fim, quanto a utilização de métodos de *Machine Learning* para a previsão de público em eventos, sugere-se que pesquisas futuras aprofundem os métodos propostos. Foi visto que o algoritmo de redes neurais é promissor para a previsão do público em jogos de futebol. Sendo assim, é interessante que mais dados sejam coletados e que se obtenha um banco de dados compatível para a utilização de redes neurais. Para isso, pode ser incluído outros times de futebol na análise e, desta maneira, o fator “estádio” tornar-se-ia relevante. Além disso, é interessante a aplicação desses métodos para outros tipos de eventos uma vez que são de fácil aplicação. Para isso, pesquisas futuras devem levantar uma base de dados com diferentes eventos para realizar uma análise semelhante a apresentada.