

Uso de Inteligência Artificial para a Classificação de Comentários Turísticos do Jalapão nos Domínios da Experiência

Jhonathan de Souza Lima¹
João Eugenio Marynowski²
José Elmar Feger³
Rafaela Mantovani Fontana⁴

Resumo

O turismo é uma área fundamental para o bem-estar das pessoas e para a economia de uma localidade. Com vistas à melhoria e ao desenvolvimento da atividade turística, a análise da experiência dos visitantes pode fornecer dados importantes sobre a qualidade da oferta de uma região. As opiniões sobre a experiência de turistas são coletadas por meio de comentários em redes sociais ou questionários. Interpreta-se o comentário do turista e verifica-se se ele se refere a uma experiência de turismo relacionada ao entretenimento, à educação/aprendizagem, à evasão/escapismo ou à estética/contemplação. Este trabalho contribui com essa função, apresentando a utilização de algoritmos de Inteligência Artificial (IA) como estratégia para a identificação automática do domínio de experiência expressa pelo turista. Para tal, toma-se uma base de dados já constituída de comentários de turistas classificados em domínios de experiência sobre o Parque Estadual do Jalapão (PEJ). Realizou-se o teste de diferentes técnicas de IA: Gaussian Naive Bayes, Support Vector Machine (SVM) e Long Short Term Memory (LSTM); demonstrando que a análise automática desses dados é possível. O método envolveu cinco etapas: 1) preparação dos dados; 2) partição do conjunto de dados; 3) treinamento; 4) ajuste de hiperparâmetros; e 5) predição. Como resultado, observou-se que as técnicas apresentaram desempenhos diferentes, dependendo do preparo e da partição do conjunto de dados para o emprego das técnicas. Os melhores resultados foram apresentados pela técnica SVM, com F1-Score de 88%.

Palavras-chave: Turismo de experiência; Redes sociais; Inteligência artificial; Comunidade virtual.

1. Especialista em Inteligência Artificial e Gestão de Projetos, Bacharel em Engenharia Elétrica, experiência em pesquisa e desenvolvimento de produtos eletrônicos. E-mail: jhonathanslima93@gmail.com. ORCID: <https://orcid.org/0009-0009-6545-009X>
2. Doutor, Mestre e Bacharel em Ciência da Computação, Professor e Pesquisador do Setor de Educação Profissional e Tecnológica da Universidade Federal do Paraná (UFPR). E-mail: jeugenio@ufpr.br. ORCID <https://orcid.org/0000-0002-0168-7217>
3. Doutor em Desenvolvimento Regional, Bacharel em Administração, Professor e Pesquisador do Programa de Pós-graduação em Turismo da Universidade Federal do Paraná (UFPR). E-mail: elmar@ufpr.br. ORCID <https://orcid.org/0000-0002-1982-4179>.
4. Doutora em Informática, Professora Titular no Setor de Educação Profissional e Tecnológica da Universidade Federal do Paraná (UFPR). E-mail: rafaela.fontana@ufpr.br. ORCID <https://orcid.org/0000-0001-6350-4167>.

Abstract

Classification of Jalapão Tourist Comments in the Domains of Experience Based on Artificial Intelligence

Tourism is an essential activity for people's well-being and for the economy surrounding touristic areas. When one is interested in improving and developing touristic activities, analyzing tourist experiences may provide useful insights. One way to understand tourist experiences is by collecting their comments on social media. This kind of analysis is made by interpreting people's comments and classifying their experience as being related to entertainment, education/learning, evasion/escapism, or aesthetics/contemplation. These categories are called experience domains. This paper presents a solution for automatically classifying touristic comments in experience domains by using artificial intelligence (AI). To conduct the study, we applied a database with tourists' comments about the Jalapão State Park (Parque Estadual do Jalapão, Brazil) that had been manually classified in experience domains. We tested different AI techniques, i.e., Gaussian Naive Bayes, Support Vector Machine (SVM), and LSTM, and demonstrated that the automatic analysis of comments is feasible. Our research method comprised five steps: 1) data preparation, 2) dataset partition, 3) training, 4) hyperparameters adjustment, and 5) prediction. Our results show that IA techniques present different performances, depending on how we prepare and partition data. The best results were obtained with SVM technique, with an 88% F1 Score.

Keywords: Experience tourism; Social networks; Artificial intelligence; Virtual community.

Resumen

Clasificación de los comentarios turísticos del Jalapão en los dominios de experiencia basados en Inteligencia Artificial

El turismo es un ámbito fundamental para el bienestar de las personas y la economía de una localidad. Con vistas a mejorar y desarrollar la actividad turística, el análisis de la experiencia de los visitantes puede proporcionar datos importantes sobre la calidad de la oferta de una región. Se recogen opiniones sobre la experiencia turística a través de comentarios en redes sociales o cuestionarios. Se interpreta el comentario del turista y se verifica si se refiere a una experiencia turística relacionada con el entretenimiento, la educación/aprendizaje, la evasión/escapismo o la estética/contemplación. Este trabajo contribuye a esta función, presentando el uso de algoritmos de Inteligencia Artificial (IA) como estrategia para identificar automáticamente el dominio de experiencia expresada por el turista. Para ello, se utiliza una base de datos ya formada por comentarios de turistas clasificados en dominios de experiencia sobre el Parque Estadual Jalapão (PEJ). Se probaron diferentes técnicas de IA: Gaussian Naive Bayes, Support Vector Machine (SVM) y LSTM; demostrando que el análisis automático de estos datos es posible. El método implicó cinco pasos: 1) preparación de datos; 2) partición del conjunto de datos; 3) formación; 4) hiperajuste de parámetros; y 5) predicción. Como resultado, se observó que las técnicas presentaron diferentes desempeños, dependiendo de la preparación y partición del conjunto de datos para el uso de las técnicas. Los mejores resultados los presentó la técnica SVM, con un F1-Score del 88%.

Palabras-clave: Turismo de experiencias; Redes sociales; Inteligencia artificial; Comunidad virtual.

INTRODUÇÃO

O turismo é um fenômeno social caracterizado pelo deslocamento temporário de pessoas até outro local, gerando, com isso, relações econômicas, sociais e cul-

turais (Torre, 1992). Assim, a atividade de turismo expõe o patrimônio natural e cultural com o objetivo de despertar a consciência ambientalista e promover o bem-estar da comunidade autóctone (Brasil, 2010).

Dentro desse contexto, o ecoturismo é a prática turística em ambientes naturais, com o objetivo de relacionar-se de forma mais próxima com o ecossistema e o meio social (Beni, 2004). Para essa prática, é necessário se dirigir a um lugar que reúna elementos paisagísticos e naturais que conectem o turista ao espaço (Beni, 2004). Dentre os pontos turísticos nacionais e internacionais que são proeminentes para a prática de turismo em ambiente natural, pode-se citar o Parque Estadual do Jalapão (PEJ), situado ao leste do estado do Tocantins, reunindo ricos elementos para o ecoturismo (Tocantins, 2022).

Por ser o turismo uma atividade que necessita estar sempre em evolução, torna-se relevante entender as experiências dos visitantes de uma localidade turística. Destaca-se que os indivíduos, ao decidirem viajar, utilizam comentários na internet tanto no que concerne ao atrativo a ser visitado quanto em relação às opções da infraestrutura turística (alimentação e meios de hospedagem), como já tratado por Law, Buhalis e Cobanoglu (2014), Boaria e Santos (2018) e Bandeira e Menezes (2022). Por essa razão, torna-se relevante aos gestores realizar o monitoramento da reputação online dos empreendimentos, a fim de refinar a prestação dos serviços e melhorar a competitividade dos estabelecimentos hoteleiros (Bandeira & Menezes, 2022), o que pode ser extrapolado para os demais elementos do sistema turístico.

A partir desse entendimento, podem ser formuladas estratégias empresariais ou políticas públicas de melhoria do turismo e, conseqüentemente, de desenvolvimento para a região (Kaizer et al., 2021). Estudos alinhados com essa proposta são apresentados por Caracristi et al. (2020) e Kaizer et al. (2021), momento em que os autores se utilizam de comentários escritos no website de viagens TripAdvisor (TripAdvisor, 2022) para identificar qual foi a experiência de turistas em pontos turísticos do PEJ. A análise dos autores foi realizada conforme a proposta de Pine e Gilmore (1999), que descrevem quatro domínios da experiência turística, classificados da seguinte forma: entretenimento, educação/aprendizagem, evasão/escapismo e estética/contemplação.

Embora os trabalhos de Caracristi et al. (2020) e Kaizer et al. (2021) tenham efetivamente encontrado os domínios de experiência relevantes para os turistas do PEJ, a análise manual de comentários de websites de viagem pode ser lenta e custosa. A classificação automatizada de texto tem sido considerada vital para gerenciar e processar uma grande quantidade de documentos em formato digital que são difundidos e estão em constante crescimento. Normalmente, a maioria dos dados para classificação de gêneros é coletada da web, através de grupos de notícias, boletins e notícias transmitidas ou impressas (Kennedy et al., 2015). Assim, a partir do acesso ao grande volume de dados disponíveis na internet, pode-se disponibilizar conhecimento por meio de pesquisas acadêmicas que subsidiem decisões no campo do turismo. Todavia, há necessidade de conseguir lidar com um grande volume de dados textuais e, nesse sentido, o emprego da Inteligência Artificial (IA) torna-se uma alternativa interessante.

No que tange ao emprego da IA para a pesquisa científica e à compreensão de sua implicação no turismo, há ainda espaço para aprofundamentos, visto que se trata de processo recente. Ao abordar o tema da experiência do cliente possibi-

litada pela IA, Ghesh, Alexander e Davis (2024) indicam que há, ainda, possibilidade para futuras pesquisas que abordem a identificação de ferramentas mais eficazes para monitorar, medir e otimizar a experiência do usuário de serviços turísticos. Nessa mesma linha, após amplo estudo visando obter insights relacionados às implicações da IA no turismo, Kim et al. (2024) asseveram que a IA pode fornecer feedback imediato, mostrando as reações dos clientes a partir de diferentes abordagens e, assim, contribuir para criar produtos turísticos que atendam melhor aos requisitos dos clientes. Tem aumentado o uso de mídias sociais para a divulgação e busca por informações sobre destinos turísticos, disponibilizados em forma de textos, fotografias e vídeos, conforme demonstram Hu e Chen (2024). Nesse sentido, há possibilidades de ampliar pesquisas para verificar o efeito do emprego de análises por meio de IA em canais de trocas de mensagens e impressões na internet, para coletar dados sobre hóspedes e desenvolver estratégias que alcancem maior satisfação e lealdade dos clientes, conforme foi apontado por Al-Hyari, Al-Smadi e Weshah (2024).

Mediante a isso, de forma a construir uma solução automatizada para o processo de avaliação de atrativos turísticos, este estudo propõe a utilização de técnicas de IA para a classificação automática dos domínios de experiência presentes nos comentários do website TripAdvisor para o PEJ. Foram avaliados os comentários públicos sobre os pontos turísticos Cachoeira da Velha, Cachoeira da Formiga, Fervedouro, Dunas do Jalapão, Povoado de Mumbuca e Serra do Espírito Santo, dentro dos quatro domínios de experiência de Pine e Gilmore (1999). Dentre as técnicas de IA mais populares utilizadas para classificação de textos, citadas por Singh et al. (2019), são usadas neste trabalho: SVM, Gaussian Naive Bayes e Redes Neurais Recorrentes (RNR) com memória de longo prazo de curto alcance, conhecida como LSTM.

Este documento apresenta o desenvolvimento do estudo e está organizado conforme segue. A seção a seguir apresenta uma revisão de literatura sobre os domínios da experiência e trabalhos relacionados também com a classificação de textos utilizando técnicas de IA. A metodologia empregada é apresentada na Seção 3. Depois, são apresentados os resultados obtidos e, por fim, o artigo discute os resultados e apresenta considerações finais.

REVISÃO DE LITERATURA

No final dos anos 1990, concebeu-se que a evolução da economia seria marcada pela demanda por experiências memoráveis, indicando que a procura caminharia para além dos aspectos utilitários dos bens ou serviços, migrando da busca por elementos externos ao consumidor para a vazão por elementos internos a ele (Pine & Gilmore, 1998). Uma experiência extraordinária consiste em um processo ativo, dinâmico e com forte dimensão social, possuindo significados e sentimentos alegres de forma integral, gerando envolvimento por meio de absorção e controle pessoal; assim como deve depender de um contexto imprevisível e novo (Hanefors & Mossberg, 2003). Essa perspectiva corrobora Pine e Gilmore (1998), para os quais a experiência é uma construção cognitiva que abrange aspectos de caráter físico, emocional, intelectual e espiritual decorrentes do contato entre o destino (considerado pelos autores como um “teatro”) e o estado men-

tal do visitante (definido como um “ator”), o que gera, por sua vez, um resultado diferente para cada pessoa.

Dessa forma, os “produtos” na economia da experiência advêm da oferta de recursos encenados, pessoais e que despertam sensações. Consome-se uma experiência a partir da imersão ou absorção do visitante no ambiente de consumo como um todo. A primeira ocorre quando o consumidor se envolve como parte do evento ou do seu desempenho, enquanto a segunda compreende captar a sua atenção. Já em relação às características passiva ou ativa, na atuação passiva não há influência direta do consumidor sobre a experiência, enquanto na ativa essa influência ocorre (Pine & Gilmore, 1998).

A interseção entre a participação passiva e a absorção em uma atividade proporciona uma experiência de entretenimento (como, por exemplo, um espetáculo musical). Segundo Pine e Gilmore (1999), visitantes que se encaixam na dimensão entretenimento possuem uma característica mais passiva na experiência, estando em estado de relaxamento e procurando absorver sensações a partir dos eventos ou atividades que praticam, e se caracterizam pela busca por emoções e excelência nos serviços consumidos, com destaque à hospitalidade. Isso foi corroborado nos achados de Loureiro (2014), que observou que o entretenimento no meio rural surge quando o visitante observa de forma passiva as atividades alheias. Nesta dimensão, Gosling et al. (2016) afirmam que a possibilidade de vivências culturais nos museus está geralmente associada à dimensão entretenimento. Achados com essa perspectiva foram encontrados por Pezzi e Vianna (2015); Gândara, Brea e Manosso (2013); Añaña et al. (2016); Alencar et al. (2019) para diferentes tipos de atrativos turísticos.

Já no caso das vivências advindas de participação ativa e absorção, proporcionam experiências educativas, como a participação em um congresso ou práticas desportivas. Na dimensão aprendizagem, considerada a mais ativa, insere-se aquilo que o empreendimento quer que os clientes aprendam com os serviços prestados (Gândara, Brea & Manosso, 2013). Para Pezzi e Vianna (2015), compõem essa dimensão a aquisição de novos conhecimentos, a aprendizagem de coisas novas (incluindo sobre si mesmo) e a melhoria nas habilidades. Essa mesma perspectiva foi corroborada nos estudos de Gosling et al. (2016), em museus, e Silva e Trentin (2018), em uma região de turismo étnico. A aquisição de novos conhecimentos, assim como a melhoria das habilidades, pôde ser observada por Loureiro (2014), Mateiro (2018) e Guissoni et al. (2019), destacando que a aprendizagem pode ocorrer através da interação com outros hóspedes e com os funcionários que atuam no destino.

Para as atividades que envolvem participação ativa e a imersão do visitante, ocorre uma experiência de escapismo (como, por exemplo, as atividades de cunho religioso ou social). A evasão inclui atividades que fogem ao cotidiano do visitante (Mondo et al., 2016), fazendo-o sentir-se em outro papel, fora de sua realidade (Pezzi & Vianna, 2015), o que pode incluir a vivência da cultura e do estilo de vida local (Loureiro, 2014), vivências espirituais (Silva & Trentin, 2018) e atividades aventureiras em atrativos naturais (pesca, passeio de canoa no período de chuva, banho de rio, acompanhamento da rotina dos pantaneiros) que permitam uma fuga da rotina (Benetti et al., 2018), confirmando as afirmações de Santos et al. (2018).

E, por fim, aquelas resultantes do cruzamento entre participação passiva e imersão proporcionam uma experiência estética (como no caso de passeios panorâmicos) (Pine & Gilmore, 1998). A dimensão estética envolve o “estar lá”, permitindo usufruir do momento e envolver-se na esfera local (Silva & Trentin, 2018). A estética engloba o visual da cidade, a arquitetura, o cenário harmonioso, o prazer (Pezzi & Vianna, 2015), a beleza dos atrativos (Mondo et al., 2016), a paisagem natural (parques, jardins, locais para atividades ao ar livre) (Añaña et al., 2016), assim como a localização/entorno, a manutenção e a decoração (Gândara, Brea & Manosso, 2013), aspectos corroborados por Loureiro (2014) e Alencar et al. (2019), os quais apontam que a estética pode ser contemplada através da atmosfera da hospedagem, da sua paisagem, logo, utilizando os sentidos.

A investigação sobre os domínios da experiência em comentários da internet, porém, de forma manual, tem sido recorrente. No estudo de Alencar et al. (2019), foram avaliadas as características dos turistas que demandam entretenimento no Paraná. Um estudo comparando o marketing de dois destinos turísticos na Espanha foi elaborado por Barreto e Martinez (2016). Os investigadores Benetti et al. (2018) analisaram manualmente comentários do website TripAdvisor para estudar a experiência turística em destinos culturais e naturais na Transpantaneira-MT. Nos trabalhos de Caracristi et al. (2020) e Kaizer et al. (2021), utiliza-se dos comentários escritos no TripAdvisor para se determinar manualmente a experiência de turistas em pontos turísticos do PEJ. Em Caracristi et al. (2021), ainda é apresentada a caracterização da demanda turística para o PEJ baseada nos comentários do TripAdvisor.

A experiência é avaliada mediante a aplicação de técnicas de IA em algumas áreas, como a análise da satisfação de usuários em ambientes hospitalares, como no estudo de Li et al. (2022), que identificou os impactos positivos e negativos de diferentes fatores do atendimento hospitalar. Por exemplo, o serviço de alimentação influencia o entendimento do paciente quanto à qualidade dos serviços recebidos, indicando que a inovação pode melhorar a eficiência na produção dos serviços médicos. No campo do turismo, García-Madurga e Grilló-Méndez (2023) fazem uma síntese abrangente da literatura sobre aplicações da IA no turismo, enfatizando suas aplicações, bem como novas vias de investigação. Diante do aumento do uso das mídias sociais com o intuito de disseminar informações sobre o turismo, há necessidade de aprimorar a classificação de textos, conforme indicam Hu e Chen (2024).

A classificação de texto, com o auxílio de técnicas de IA, tem sido utilizada para estimar a probabilidade de um texto pertencer a uma classe (Ogada et al., 2015; Violos et al., 2018; Singh et al., 2019). A classificação é uma tarefa da Mineração de Dados, principal fase do processo de extração de conhecimento em bases de dados (Goldschmidt et al., 2015). O processo de extração inicia-se com o pré-processamento, que compreende a obtenção e a preparação dos dados para aplicação das técnicas; segue-se a mineração de dados, que busca efetivamente informações e conhecimentos úteis no contexto da aplicação, utilizando as técnicas de IA; e, por fim, a etapa de pós-processamento, que tem como objetivo viabilizar a avaliação da utilidade do conhecimento descoberto.

A maioria das técnicas usadas para a classificação de texto lida com um pacote de palavras em combinação com modelos probabilísticos, redes neurais e técnicas de hiperplano multidimensional para estimar a probabilidade de um texto per-

tencer a uma classe. Ogada et al. (2015) avaliam a categorização de texto usando classificação e grupos de palavras com diversas quantidades. Violos et al. (2018) utilizam a classificação textual de sequências de palavras com grafos direcionados, empregando análise de fluxos de dados contínuos (*streaming*). A avaliação foi feita utilizando as técnicas de SVM e Bayes. Singh et al. (2019) comparam a aplicação das técnicas de Bayes para classificação textual de sentimentos.

METODOLOGIA

O objetivo deste estudo foi desenvolver uma solução computacional que aplique técnicas de IA para classificar comentários de turistas em domínios da experiência (Pine & Gilmore, 1998). Os comentários utilizados são oriundos da base de dados empregada em Caracristi et al. (2020), Caracristi et al. (2021) e Kaizer et al. (2021). Os autores extraíram os comentários da plataforma TripAdvisor por meio de uma técnica de captura de dados da internet chamada *Web Scraping* (Munzert et al., 2014). Os comentários considerados estavam em português e foram postados entre os anos de 2012 e 2019, referentes às visitas realizadas entre 2011 e 2019. Detalhes sobre os dados e o processamento realizado para tratamento dos dados e classificação são apresentados em Kaizer et al. (2021).

Os comentários foram agrupados em quadrigramas. Um quadrigrama é um grupo de quatro palavras que, segundo Hyland (2008), carrega maior significação e pode ser encontrado com maior frequência em um texto. Tal estrutura foi empregada pois foi identificada como eficiente para o entendimento e a identificação dos domínios da experiência (Caracristi et al., 2020; Kaizer et al., 2021) e é considerada uma peça fundamental para a aprendizagem de uma linguagem (Hyland, 2008). No Quadro 1 são exemplificados quadrigramas dos quatro domínios da experiência que foram classificados manualmente.

Quadro 1. Quadrigramas classificados por domínio de experiência

Domínios da experiência	Exemplo de quadrigrama	Ponto turístico
Educação/Aprendizado	“amante natureza biodiversidade pais”	Cachoeira da Formiga
Educação/Aprendizado	“muita pedra solta areia”	Serra do Espírito Santo
Educação/Aprendizado	“duna formada erosao serra”	Dunas do Jalapão
Entretenimento	“artesanato feito capim dourado”	Povoado de Mumbuca
Entretenimento	“Água nao deixa afundar”	Fervedouro
Entretenimento	“nao possível entrar agua”	Cachoeira da Velha
Estética/contemplação	“cachoeira maravilhosa agua azul”	Cachoeira da Formiga
Estética/contemplação	“madrugada ver nascer sol”	Serra do Espírito Santo
Estética/contemplação	“serra espirito santo fundo”	Dunas do Jalapão
Evasão/Escapismo	“Nao da vontade sair”	Fervedouro
Evasão/Escapismo	“nao quer ir embora”	Cachoeira da Formiga
Evasão/Escapismo	“agua sentir grandeza natureza”	Cachoeira da Velha

Fonte: baseado em Kaizer et al., 2021.

O objetivo deste trabalho é avaliar o emprego de técnicas de aprendizado de máquina e IA para a classificação dos quadrigramas de forma automatizada. A fim de identificar a técnica com melhor desempenho, este estudo envolve testes com as técnicas Gaussian Naive Bayes, SVM e RNR com memória longa de curto prazo, conhecida como LSTM. As técnicas foram aplicadas para a classificação automática de quadrigramas de comentários de turistas nos domínios da experiência.

A execução do trabalho foi dividida em cinco etapas: 1) preparação dos dados; 2) partição do conjunto de dados; 3) treinamento; 4) ajuste de hiperparâmetros; e 5) predição. As etapas são detalhadas a seguir, nas subseções seguintes.

PREPARAÇÃO DOS DADOS

A coleta de dados foi feita a partir de uma base oriunda do trabalho realizado pelos autores Caracristi et al. (2020) e Kaizer et al. (2021), que realizaram a classificação manual dos textos em seus respectivos domínios de experiência. Essa classificação foi utilizada nas etapas seguintes para o treinamento dos modelos de IA.

Os quadrigramas que ocorreram no mínimo duas vezes foram reunidos a partir dos comentários sobre os pontos turísticos Cachoeira da Velha, Cachoeira da Formiga, Fervedouro, Dunas do Jalapão, Povoado de Mumbuca e Serra do Espírito Santo, de acordo com as quantidades apresentadas na Tabela 1.

Tabela 1. Quantidade de quadrigramas por ponto turístico

Ponto turístico	Número de quadrigramas
Povoado de Mumbuca	172
Cachoeira da Velha	173
Dunas do Jalapão	184
Serra do Espírito Santo	183
Fervedouro	179
Cachoeira da Formiga	175

Fonte: os autores.

A base de dados utilizada totaliza 1.066 registros de quadrigramas referentes aos comentários sobre o PEJ, na seguinte ordem: educação/aprendizagem (282 quadrigramas), entretenimento (451 quadrigramas), estética/contemplação (274 quadrigramas) e evasão/escapismo (59 quadrigramas), conforme mostrado na Tabela 2.

Tabela 2. Quantidade de quadrigramas por domínio de experiência

Domínios da experiência	Quadrigramas	Porcentagem
Educação/Aprendizado	282	26%
Entretenimento	451	42%
Estética/contemplação	274	26%
Evasão	59	6%
Total	1.066	100%

Fonte: os autores.

Neste trabalho, o primeiro passo para a extração de características de texto foi um processo chamado *tokenização*. A *tokenização*, também conhecida como segmentação de palavras, quebra a sequência de caracteres em um texto, localizando o limite de cada palavra, ou seja, os pontos onde uma palavra termina e outra começa. Para fins de linguística computacional, as palavras assim identificadas são frequentemente chamadas de *tokens* (Barbosa et al., 2017). Ou seja, a *tokenização* é a transformação de texto legível por humanos em *tokens* legíveis por máquina (Mullen et al., 2018).

Em seguida, aplicou-se uma técnica de vetorização de palavras para transformar os dados do texto original em dados numéricos, processáveis pelas técnicas de IA (Yang, 2022). Dentre as técnicas de vetorização de palavras, destaca-se o *Term Frequency-Inverse Document Frequency* (TF-IDF) (Yang, 2022). O TF-IDF pode ser formulado pela multiplicação entre a frequência do termo e a frequência inversa de documentos. A Frequência do Termo (TF) é a medição de quão frequentemente um termo ocorre em um documento (Croft et al., 2010). A Frequência Inversa de Documentos (IDF) mede a importância de um termo dentro de um conjunto de documentos (Rahman et al., 2018).

Na Tabela 3, são mostrados, a título de exemplo, valores de TF-IDF de alguns quadrigramas extraídos da base de dados.

Tabela 3. Valores de TF-IDF para cada domínio de experiência e quadrigrama específico

Domínio de experiência	Quadrigrama	Valor TF-IDF
Educação/Aprendizado	“sorvete fruta cerrado tem”	0.516395958135
Entretenimento	“visita trilha suspensa facil”	0.492845919043
Evasão/Escapismo	“jalapao bruto todo repetem”	0.476863555460
Estética/Contemplação	“linda beleza lembra mini”	0.461324794431

Fonte: os autores.

Após o processo de *tokenização*, também foi utilizado um conceito chamado *Word Embedding* para o emprego do método Bi-LSTM. Os *Word Embedding* são modelos matemáticos que codificam relações de palavras dentro de um espaço vetorial, por meio de um processo de treinamento baseado em informações de coocorrência entre palavras (Heimerl & Gleicher, 2018).

PARTIÇÃO DO CONJUNTO DE DADOS

O treinamento é o momento em que se “ensina” às técnicas de IA a qual categoria cada comentário pertence. Por isso, utilizaram-se as classificações manuais realizadas por Kaizer et al. (2021) como referência. Nesta etapa, os dados foram divididos em dois grupos: um que foi usado para treinar o modelo e outro que foi utilizado para testá-lo. É o teste que, mais tarde, identificará quão eficiente é a técnica para prever o domínio de experiência de um determinado quadrigrama.

Neste estudo, utilizaram-se três cenários distintos de análise, nos quais diferentes estratégias foram empregadas para a separação dos grupos de dados de teste e treinamento (Quadro 2). No Cenário 1, utilizou-se a divisão de 67% da base de dados para o treinamento dos modelos e 33% da base de dados para o

teste do modelo. No Cenário 2, empregou-se a técnica de validação cruzada k -fold (Yadav & Shukla, 2016), com parâmetro $k = 10$; e, no Cenário 3, foi utilizada a mesma técnica, mas com parâmetro $k = 50$.

Na técnica de validação cruzada k -fold, os dados são divididos em k partições iguais. O treinamento do modelo é feito em k partes, excetuando-se uma, que é usada para teste (Yadav & Shukla, 2016). A escolha ideal para o treinamento de um modelo é a validação cruzada k -fold com grande valor de k , porém menor que o número de instâncias (Yadav & Shukla, 2016).

Neste trabalho, também foi necessário o uso de uma técnica de correção de prevalência de dados chamada Synthetic Minority Oversampling Technique (SMOTE), para balancear a quantidade de elementos em cada categoria (domínio de experiência). Essa técnica foi utilizada no Cenário 3 para a análise dos dados (Quadro 2). Segundo Chawla (2002), o SMOTE é uma técnica popular de balanceamento de dados, que gera dados sintéticos para a classe minoritária a fim de equilibrar os dados de treinamento. De forma geral, a técnica utiliza a distância entre os dados vizinhos para gerar novos valores aleatórios (Bernardo, 2020).

Quadro 2. Cenários de análise dos dados

	Cenário 1	Cenário 2	Cenário 3
Separação dos dados de teste e treinamento	67% treinamento 33% teste	Validação cruzada k-fold com k=10.	Validação cruzada k-fold com k=50.
Quantidade de quadrigramas por categoria	Desbalanceada	Desbalanceada	Balanceada com a técnica SMOTE

Fonte: os autores.

TREINAMENTO

Os modelos de classificação de textos foram implementados por meio de uma biblioteca da linguagem de programação Python chamada Scikit-Learn ou de um *framework* de IA (conjunto de códigos genéricos) chamado TensorFlow, juntamente com a Application Programming Interface (API). Nessa API, existem funções previamente construídas para o treinamento e teste de um modelo de classificação.

Os modelos de classificação utilizados no trabalho foram o Gaussian Naive Bayes, SVM e Bi-LSTM. O modelo Gaussian Naive Bayes considera os pontos de dados com n características, nas quais as características podem ser quaisquer dados de texto. Nesse contexto de classificação de texto, a probabilidade de um texto dado Y pertencer a uma classe X é calculada usando o Teorema de Bayes (Venzatesh & Ranjitha, 2018).

O modelo SVM é um dos algoritmos de aprendizado de máquina mais utilizados para a classificação de pequenos conjuntos de dados (Hyland, 2008). O objetivo do modelo é construir um hiperplano entre as amostras utilizadas para o treinamento da rede (Hyland, 2008). Além disso, o modelo visa maximizar a margem entre os vetores de suporte; esses dados são relevantes para descobrir o

melhor separador — o hiperplano — encontrando o plano ótimo entre as amostras de diferentes classes (Hyland, 2008).

O modelo LSTM é um tipo especial de Redes Neurais Recorrentes. Essas redes possuem ciclos de realimentação em sua arquitetura, permitindo que as informações anteriores influenciem o processamento atual (Köksal, 2022). As Redes Neurais Recorrentes possuem memórias de curto prazo, portanto sua computação pode ser lenta, além de essas memórias sofrerem com o problema do gradiente de fuga (Köksal, 2022). A técnica foi melhorada por Hochreiter e Schmidhuber (1997) e Jang (2020), dando origem à rede neural LSTM Bidirecional (Bi-LSTM), utilizada neste trabalho. Essa é composta por unidades LSTM que operam juntas para aprender problemas de modelagem sequencial de dependências de longo prazo e é amplamente usada para a classificação de texto (Jang, 2020).

AJUSTE DE HIPERPARÂMETROS

Em geral, construir um modelo de aprendizado de máquina ou de aprendizado profundo eficaz é um processo complexo e demorado, que envolve a determinação de um algoritmo apropriado e a obtenção de uma arquitetura de modelo ideal por meio do ajuste de seus hiperparâmetros (Yang & Shami, 2020).

Neste trabalho, é utilizada uma técnica para ajuste de hiperparâmetros chamada Pesquisa em Grade (*Grid Search*), que é um dos métodos mais usados para explorar o espaço de configuração dos hiperparâmetros. A pesquisa em grade pode ser considerada uma busca ativa ou um método de força bruta que avalia todas as combinações de hiperparâmetros dadas a grade de configurações. Essa técnica trabalha avaliando o produto cartesiano de um conjunto finito de valores especificados pelo usuário (Yang & Shami, 2020). Os valores utilizados e os resultados obtidos são apresentados na Seção Resultados.

PREDIÇÃO

O objetivo da etapa de predição é verificar quão eficientes os modelos obtidos são para prever o domínio de experiência dos quadrigramas. Inicialmente, contabiliza-se a quantidade de quadrigramas classificados de maneira correta e incorreta pelo modelo, gerando uma tabela conhecida como matriz de confusão. A partir desses valores, têm-se diversas métricas de avaliação da classificação, dentre as quais foram utilizadas: precisão, *recall* e F1-score. Tanto a precisão quanto o *recall* indicam quantos acertos o modelo fez considerando as classificações positivas para uma classe, mas a precisão desconta as classificações positivas erradas (falsos positivos), e o *recall* desconta as classificações negativas erradas (falsos negativos). Por fim, a F1-score é uma média harmônica entre a precisão e o *recall*.

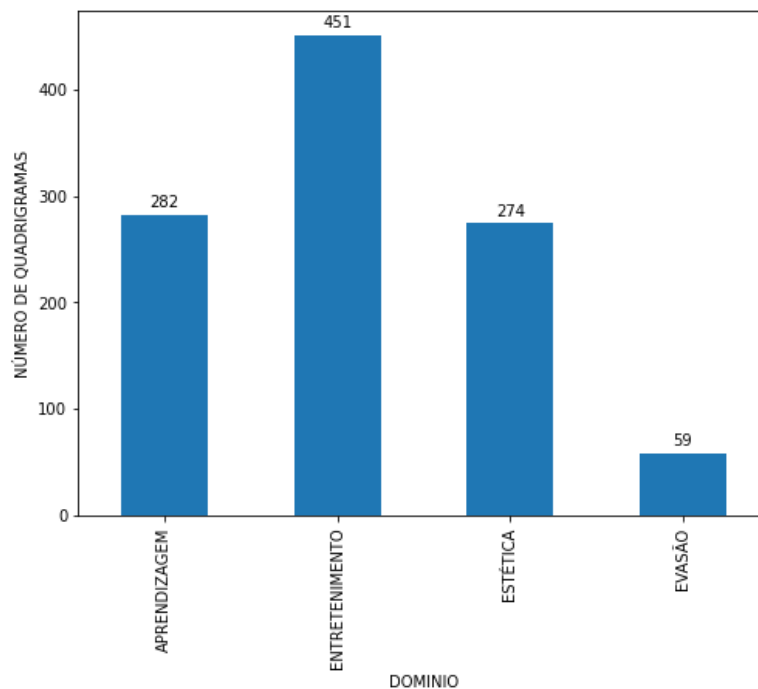
A próxima seção mostra os resultados obtidos nos testes dos modelos como preditores dos domínios de experiência dos quadrigramas.

RESULTADOS

A fim de avaliar as melhores configurações dos modelos de IA para a predição dos domínios de experiência dos quadrigramas gerados a partir dos comentários de turistas, definiram-se três cenários de análise (conforme mostrado na Seção 3, Quadro 2).

O Cenário 1 compreende os testes realizados com os dados desbalanceados, ou seja, nos quais a quantidade de quadrigramas para cada classificação de domínio de experiência era diferente. A Figura 1 mostra a quantidade de quadrigramas em cada domínio de experiência. Aqui, foram utilizados 67% da base de dados para o treinamento dos modelos e 33% da base de dados para o teste dos modelos.

Figura 1. Demonstração da quantidade desbalanceada de quadrigramas por categoria

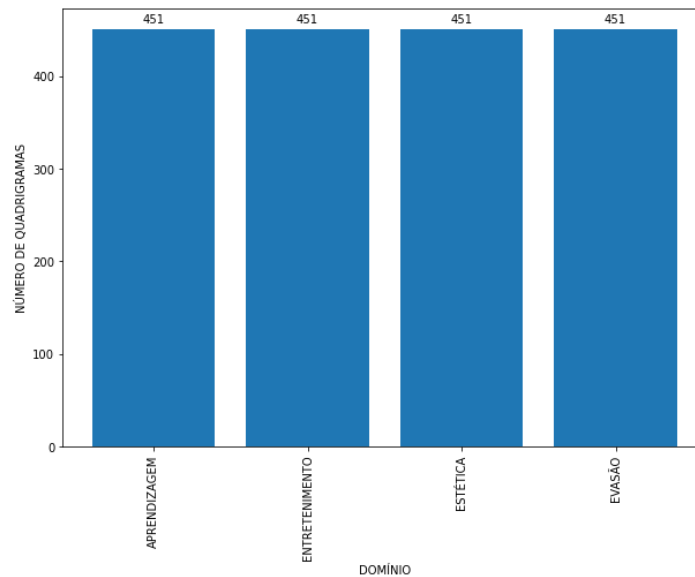


Fonte: os autores.

O Cenário 2 foi realizado também com os dados desbalanceados. No entanto, nessa segunda configuração, alterou-se a técnica para dividir os dados em treinamento e teste. Aqui, utilizou-se a técnica de validação cruzada k -fold, com parâmetro $k=10$.

No último cenário, o Cenário 3, realizou-se o balanceamento dos dados com o emprego da técnica SMOTE. Assim, a quantidade de dados de todos os domínios de experiência ficou em 451 quadrigramas, como demonstra a Figura 2. Além disso, foi utilizada a técnica de validação cruzada k -fold, com parâmetro $k=50$.

Figura 2. Demonstração da quantidade balanceada (por meio da técnica SMOTE) de quadrigramas por categoria



Fonte: os autores.

A Tabela 4 demonstra, em detalhes, os resultados obtidos com os três modelos, nos três cenários descritos. Verificam-se, também, os resultados obtidos em cada um dos domínios de experiência para as métricas: precisão, *recall* e F1-score. A linha “média” resume, para cada modelo, o resultado consolidado de todos os domínios para cada métrica.

Tabela 4. Valores de precisão, recall e F1-Score para as três técnicas aplicadas, nos três cenários de análise

	Domínios da experiência	Cenário 1 Precisão	Cenário 2 Precisão	Cenário 3 Precisão	Cenário 1 <i>Recall</i>	Cenário 2 <i>Recall</i>	Cenário 3 <i>Recall</i>	Cenário 1 F1-Score	Cenário 2 F1-Score	Cenário 3 F1-Score
Gaussian Naive Bayes	Educação/Aprendizado	63%	53%	75%	46%	28%	88%	53%	36%	81%
	Entretenimento	63%	55%	85%	76%	80%	46%	69%	65%	60%
	Estética/contemplação	70%	69%	75%	78%	66%	90%	74%	67%	82%
	Evasão/escapismo	100%	100%	92%	9%	2%	100%	17%	3%	96%
	MÉDIA	74%	69%	82%	52%	58%	81%	53%	55%	80%

	Domínios da experiência	Cenário 1 Precisão	Cenário 2 Precisão	Cenário 3 Precisão	Cenário 1 Recall	Cenário 2 Recall	Cenário 3 Recall	Cenário 1 F1-Score	Cenário 2 F1-Score	Cenário 3 F1-Score
SVM	Educação/ Aprendizado	85%	50%	84%	27%	42%	82%	41%	46%	83%
	Entretenimento	46%	60%	76%	93%	78%	83%	62%	68%	79%
	Estética/ contemplação	53%	80%	94%	12%	64%	88%	19%	71%	91%
	Evasão/ escapismo	0%	74%	100%	0%	29%	100%	0%	41%	100%
	MÉDIA	46%	66%	88%	33%	53%	88%	30%	56%	88%
Bi-LSTM	Educação/ Aprendizado	54%	54%	79%	65%	58%	90%	59%	56%	84%
	Entretenimento	72%	69%	75%	68%	73%	70%	70%	71%	72%
	Estética/ contemplação	75%	74%	88%	64%	67%	93%	69%	70%	91%
	Evasão/ escapismo	28%	50%	93%	38%	27%	82%	32%	35%	87%
	MÉDIA	57%	62%	84%	59%	56%	84%	58%	58%	84%

Fonte: os autores.

Percebe-se como as técnicas de IA diferem em performance, dependendo da técnica e da base de dados utilizada. Destaca-se, por exemplo, na técnica Gaussian Naive Bayes, como os valores das métricas *recall* e F1-Score para a categoria Evasão/Escapismo variam significativamente entre os três cenários (9%, 2% e 100%; 17%, 3% e 96%, respectivamente). O mesmo pode ser observado para a técnica SVM, na mesma categoria, quanto às métricas precisão, *recall* e F1-Score (0%, 74% e 100%; 0%, 29% e 100%; 0%, 41% e 100%). Isso demonstra o impacto do balanceamento dos dados e da mudança dos hiperparâmetros das técnicas, que foram as alterações realizadas entre os diferentes cenários.

Observa-se também como, num mesmo cenário, técnicas diferentes apresentam desempenhos notavelmente melhores. Por exemplo, os valores da métrica precisão, no Cenário 1, apresentam os valores de 63%, 85% e 54% para a categoria Educação/Aprendizado nas técnicas Gaussian Naive Bayes, SVM e Bi-LSTM, respectivamente. Por outro lado, em alguns testes realizados, houve pouca diferença nos resultados obtidos. A categoria Entretenimento, por exemplo, no Cenário 2, na métrica *recall*, obteve os resultados de 80%, 78% e 73% para Gaussian Naive Bayes, SVM e Bi-LSTM, respectivamente.

A conclusão sobre a técnica com melhor potencial de aplicação pode ser retirada dos valores das médias apresentados na Tabela 5. Observa-se que o melhor desempenho foi obtido pelo modelo do tipo SVM para classificação de quadrigramas, com F1-Score de 88%, quando comparado com os modelos de aprendizado Gaussian Naive Bayes e Bi-LSTM. Esses resultados foram obtidos notavelmente no Cenário 3, em que a técnica de validação cruzada *k*-fold foi aplicada com *k*=50 sobre os dados balanceados.

Tabela 5. Melhores resultados obtidos cada modelo de classificação de quadrigramas de acordo com as métricas de avaliação precisão, *recall* e *F1-score*

Modelo de classificação	Precisão	Recall	F1-Score
Gaussian Naive Bayes	82%	81%	80%
SVM	88%	88%	88%
Bi-LSTM	84%	84%	84%

Fonte: os autores.

CONSIDERAÇÕES

Entre as técnicas avaliadas, obteve-se o melhor resultado com a SVM, o que está de acordo com outros estudos, como os de Ying (2021) e Joachims (1998). Duas constatações foram importantes para a obtenção desse resultado: a necessidade de balanceamento da quantidade de dados para cada classe na base e a técnica para partição do conjunto de dados em teste e treinamento.

Neste estudo, utilizou-se a técnica SMOTE (Chawla, 2002) para balancear a quantidade de elementos em cada categoria (domínio de experiência), ou seja, os dados foram gerados de forma artificial porque não houve a possibilidade de coletá-los de forma balanceada na fonte. A expansão da base de dados de treinamento é uma estratégia largamente utilizada e é tratada por autores como Ying (2019).

Quanto à técnica de partição do conjunto de dados em teste e treinamento, verificou-se que a validação cruzada k -fold com $k=50$ obteve o melhor resultado. No entanto, foi notável o aumento do tempo para treinamento do modelo, e isso deve ser considerado.

Para implementar um classificador automatizado, é necessário formar conjuntos de termos semelhantes e individualizados. A partir desses conjuntos, é possível automatizar a classificação para novos termos. Assim, os comentários, que normalmente possuem mais de um domínio da experiência, dificultariam ou até inviabilizariam essa distinção e a utilização direta deles para a classificação automatizada.

Os resultados obtidos indicam o potencial de aplicação da classificação automatizada de quadrigramas nos domínios da experiência para um atrativo turístico. Os quadrigramas obtidos dos comentários apresentam maior representatividade e possibilidade de classificação nos domínios da experiência, mesmo podendo, ainda assim, indicar mais de um domínio (Caracristi et al., 2020; Kaizer et al., 2021). Este trabalho contribui nesse sentido, pois indica que a classificação utilizando os quadrigramas possibilitou sua automatização com o uso de técnicas de IA.

O esforço empreendido nesta pesquisa foi identificar formas automatizadas, por meio de IA, para a classificação de textos retirados de comentários espontâneos postados por turistas em redes sociais. Como apontado por Ghesh, Alexander e Davis (2024), há necessidade de identificar ferramentas que auxiliem os gestores a utilizarem o potencial da IA na gestão. Nesse sentido, o resultado desta pesquisa apresenta a melhor alternativa em termos de método que pode ser utilizado para organizar os dados nos domínios da experiência, evitando que se perca tempo testando diferentes opções disponíveis no mercado. Ao desenvolver e testar os procedimentos, pretende-se que os dados sejam organizados de

forma padronizada e ordenada. Nesse sentido, ao se classificar os dados, pode-se realizar as classificações de diferentes períodos e, assim, comparar os resultados após mudanças realizadas pela gestão. Por exemplo, no contexto analisado, verificou-se que há pouca identificação de impressões dos turistas quanto aos aspectos ecológicos do local analisado. Visto que os parques ecológicos têm a função de disseminar conhecimentos sobre os biomas, a gestão pode tomar medidas para incrementar essa experiência. Após um tempo, poderiam ser colhidos novamente os dados e classificados, a fim de verificar possíveis mudanças nos volumes de quadrigramas relacionados à aprendizagem. Essa possibilidade pode contribuir para aprimorar a coleta e o tratamento de dados, que, na maioria das vezes, são relativos a um momento específico no tempo, havendo necessidade de realizar estudos longitudinais. Essa necessidade foi apontada por Lozano e Gutiérrez (2018) como uma limitação da maioria dos estudos realizados com dados da Internet.

Além das questões relacionadas a um atrativo ou destino particular, pela facilidade e velocidade que a IA proporciona no que tange ao tratamento de dados, a metodologia poderia ser aplicada para estudar diferentes atrativos ou destinos turísticos de uma área ou região, e ainda comparar seus resultados. Isso ajudaria os gestores do turismo na identificação de diferenças entre os objetos analisados e na análise de características distintas de diferentes competidores do setor para, com base nisso, adotar estratégias no sentido de melhorar o seu posicionamento.

Uma contribuição do trabalho que ultrapassa os limites da coleta de dados e de seu tratamento diz respeito à forma como os procedimentos adotados para a realização do estudo foram descritos. Isso permite aos não desenvolvedores de programação, como gestores ou turismólogos, compreender os bastidores do desenvolvimento de programas de IA, contribuindo, dessa forma, para disseminar conhecimentos da área de tecnologia da informação para outros domínios do conhecimento, como a administração e o turismo.

DISCUSSÕES E CONSIDERAÇÕES FINAIS

O objetivo deste trabalho foi aplicar técnicas de IA para a classificação automática dos domínios de experiência presentes nos comentários do website TripAdvisor para o PEJ. Os resultados aqui obtidos demonstram o potencial da aplicação de tais técnicas na automatização da avaliação de experiências de turismo e, conseqüentemente, na facilitação de sua aplicação em políticas públicas.

O compartilhamento de experiências de turistas em plataformas online está em expansão. Esse conjunto de dados é utilizado pelos próprios usuários no momento de planejar, realizar e avaliar uma viagem. Por isso, a interpretação adequada dos fatores que motivam a viagem é importante para a gestão do turismo. Ocorre que a capacidade humana de interpretar os dados é limitada e, por isso, geralmente se recorre a técnicas de amostragem. Todavia, por melhor que seja a metodologia adotada, pode-se deixar de observar fatores que são relevantes para a escolha do destino pelo turista.

Nesse sentido, este trabalho contribui com a aplicação de fundamentos teóricos que abordam aspectos da economia, considerando fatores subjetivos do comportamento humano. Além disso, propõe instrumentos que possibilitam tra-

tar um grande volume de dados, a fim de compreender a experiência dos turistas — metodologia que pode ser replicada para outros tipos de atrativos e permite, inclusive, efetuar comparações com outros destinos com características semelhantes de forma mais rápida e econômica. Contribui ainda para que a gestão de destinos turísticos adote estratégias mais assertivas. Por se caracterizar como um processo automatizado, permite o acompanhamento histórico dos dados, possibilitando avaliar a efetividade das estratégias adotadas pelos gestores do destino, especialmente verificar se sua atuação no sentido de melhorar o desempenho em algum domínio da experiência trouxe melhorias em termos de experiência na visão do turista.

Entende-se que o estudo traz contribuições no sentido de aprimorar métodos de captura e análise de textos; todavia, não esgota o assunto. O desenvolvimento da IA está em evolução e ainda apresenta limitações. Assim, há espaço para replicar o estudo em outros tipos de atrativos, a fim de verificar a efetividade do processo. Sugere-se aplicar a metodologia a outros tipos de destinos, com distintos atrativos, a fim de verificar a capacidade do processo em reconhecer adequadamente os comportamentos dos turistas.

Vale ressaltar que o presente estudo procurou atender apenas a um dos itens apontados por Ghesh, Alexander e Davis (2024) como possibilidade de avanço nas pesquisas envolvendo a IA e o Turismo — no caso, a análise de dados que facilitem a gestão. Há ainda um espectro de outras possibilidades de estudos indicados no trabalho de Kim et al. (2024), os quais realizaram ampla pesquisa com diferentes atores, tanto acadêmicos como empresariais e especialistas em tecnologia, o que permite ampliar os estudos na área.

Uma limitação deste estudo decorre da coleta de comentários espontâneos dos turistas, que muitas vezes não expressam sensações relacionadas às variáveis necessárias para compreender a situação, muito menos ao balanceamento desses dados. Esse fato pode contribuir para que haja restrições quanto à extrapolação estatística dos achados, e demanda aprimoramentos na forma de captura e mineração dos textos, a fim de que os resultados sejam mais próximos da realidade.

REFERÊNCIAS

- Alencar, D. G., Santos, M. L. dos, Souza, A. A. e, & Gândara, J. M. G. (2019). Produtos turísticos para demandantes de experiências da dimensão entretenimento de Pine & Gilmore: Novas características e tendências para o Paraná. *Turismo Visão e Ação*, 21(2), 46–67.
- AlHyari, H. S., AlSmadi, I. M., & Weshah, S. R. (2024). The impact of artificial intelligence (AI) on guest satisfaction in hotel management: An empirical study of luxury hotels. *GeoJournal of Tourism and Geosites*, 48(2spl), 810–819. <https://doi.org/10.30892/gtg.482spl15-1081>
- Añaña, E. da S., Anjos, F. A. dos, & Pereira, M. de L. (2016). Imagem de destinos turísticos: Avaliação à luz da teoria da experiência na economia baseada em serviços. *Revista Brasileira de Pesquisa Em Turismo*, 10(2). <https://doi.org/10.7784/rbtur.v10i2.1093>
- Bandeira, L. C. L., & Menezes, P. D. L. (2022). A experiência dos hóspedes na hotelaria de João Pessoa: Uma análise a partir do conteúdo gerado no website Booking.com.

- Revista Turismo em Análise*, 33(1), 191–212. <https://doi.org/10.11606/issn.1984-4867>
- Barbosa, J. L. N. et al. (2017). Introdução ao processamento de linguagem natural usando Python. In *III Escola Regional de Informática do Piauí: Anais – Artigos e minicursos* (pp. 336–360).
- Benetti, A. C., Carrión Carracedo Ozelame, Â. M., Acosta Pereira, L., & Torres Tricárico, L. (2018). Turismo de experiência em áreas patrimoniais: Uma análise das emoções a partir dos comentários do TripAdvisor sobre a Estrada Parque Transpantaneira–MT–Brasil. *PASOS. Revista de Turismo y Patrimonio Cultural*, 18(3), 565–581. <https://doi.org/10.25145/j.pasos.2018.16.042>
- Beni, M. C. (2004). Turismo: Da economia de serviços à economia da experiência. *Revista Turismo – Visão e Ação*, 6(3), 296–306.
- Bernardo, A. et al. (2020). *CSMOTE: Continuous Synthetic Minority Oversampling for Evolving Data Streams*. Politecnico di Milano. Dipartimento di Elettronica Informatica e Bioingegneria.
- Boaria, F., & dos Santos, C. A. F. (2018). Análise do impacto do conteúdo gerado pelos usuários nas mídias sociais e agências de viagens online na gestão hoteleira. *Marketing & Tourism Review*, 3(3). <https://doi.org/10.29149/mtr.v3i3.4594>
- Brasil. (2010). *Ecoturismo: Orientações básicas* (2ª ed.). Ministério do Turismo.
- Caracristi, M. de F. de A., Feger, J. E., Silva, T. M. da, & Marynowski, J. E. (2020). Uma viagem pelo Jalapão, Brasil: análise das experiências turísticas. *Revista Paranaense de Desenvolvimento (RPD)*, 41(138), 89–110.
- Caracristi, M. F. A., Feger, J. E., Marynowski, J. E., & Minasi, S. M. (2021). A demanda turística do Parque Estadual do Jalapão (PEJ, TO, Brasil) baseada em comentários de redes sociais. *Revista Brasileira de Ecoturismo*, 14(3), 291–314. <https://periodicos.unifesp.br/index.php/ecoturismo/article/view/11406/8604>
- Chawla, N. V. (2002). SMOTE: Synthetic minority oversampling technique. *Journal of Artificial Intelligence Research*, 16(1), 321–357.
- Croft, W. B., Metzler, D., & Strohman, T. (2010). *Search engines: Information retrieval in practice* (Vol. 283). AddisonWesley.
- Gândara, J. M. G., Brea, J. A. F., & Manosso, F. C. (2013). Calidad de da experiencia en los Hoteles Termales de Galicia, España Un análisis a través de la reputación online. *Estudios y Perspectivas en Turismo*, 22(3), 492–525.
- Gosling, M. de S., Silva, J. A., Mendes, J., Coelho, M. de F., & Brener, I. (2016). Experiência turística em museus: percepções de gestores e visitantes. *Tourism & Management Studies*, 12(2), 107–116. <https://doi.org/10.18089/tms.2016.12212>
- Guissoni, R., Alencar, D. G., & Gândara, J. M. (2019). Experience tourism in ParanáBrazil: An analysis of tourists looking for business, urban and sports experiences. *Journal of Tourism and Development*, 32.
- Hanefors, M., Mossberg, L. (2003) Searching for the extraordinary meal experience. *Journal of Business and Management*, 9(3), 249–270.
- Heimerl, F., & Gleicher, M. (2018). Interactive analysis of word vector embeddings. *Computer Graphics Forum*, 37(3), 253–265. <https://doi.org/10.1111/cgf.13417>
- Hochreiter, S., & Schmidhuber, J. (1997). Long shortterm memory. *Neural Computation*, 9(8), 1735–1780.
- Hu, T., & Chen, H. (2024). Destination experiencescape for coastal tourism: A social network analysis exploration. *Journal of Outdoor Recreation and Tourism*, 46, Article 100747.

- Jang, B. et al. (2020). BiLSTM model to increase accuracy in text classification: Combining Word2vec, CNN and attention mechanism. *Applied Sciences*.
- Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. (Relatório técnico). Cornell University.
- Kennedy, O. et al. (2015). Ngram based text categorization method for improved data mining. *Journal of Information Engineering and Applications*, 5(8), 35–43.
- Kim, H., So, K. K. F., Shin, S., & Li, J. (2024). Artificial Intelligence in hospitality and tourism: Insights from industry practices, research literature, and expert opinions. *Journal of Hospitality & Tourism Research*, 1–20. <https://doi.org/10.1177/10963480241229235>
- Köksal, O. (2022). A comparative text classification study with deep learning-based algorithms [Paper presentation]. *ASELSAN 9th International Conference on Electrical and Electronics Engineering*.
- Law, R., Buhalis, D., & Cobanoglu, C. (2014). Progress on information and communication technologies in hospitality and tourism. *International Journal of Contemporary Hospitality Management*, 26, 797–824.
- Li, S., Xu, D., Liu, Y., Wang, R., & Zhang, J. (2022). Identification method of influencing factors of hospital catering service satisfaction based on decision tree algorithm. *Applied Bionics and Biomechanics*, 2022, Article 6293904. <https://doi.org/10.1155/2022/6293904>
- Loureiro, S. M. C. (2014). The role of the rural tourism experience economy in place attachment and behavioral intentions. *International Journal of Hospitality Management*, 40, 1–12. <https://doi.org/10.1016/j.ijhm.2014.02.010>
- Lozano, S., & Gutiérrez, E. (2018). A complex network analysis of global tourism flows. *International Journal of Tourism Research*, 20, 588–604. <https://doi.org/10.1002/jtr.2208>
- Mateiro, B. M. de J. (2018). O contributo da experiência turística para o desenvolvimento do turismo nos destinos rurais: uma revisão da literatura. *PASOS. Revista de Turismo y Patrimonio Cultural*, 16(4), 939–956. <https://doi.org/10.25145/j.pasos.2018.16.066>
- Mondo, T. S., Talini, M. C., & Fiates, G. G. S. (2016). A qualidade de serviços em atrativos turísticos de Florianópolis à luz da teoria do turismo de experiência. *Revista de Turismo Contemporâneo*, 4(2). <https://doi.org/10.21680/2357-8211.2016v-4n2ID8302>
- Mullen, L. A. et al. (2018). Fast, Consistent Tokenization of Natural Language Text. *The Journal of Open Source Software*, 3(23), 655, <https://doi.org/10.21105/joss.00655>.
- Munzert, S.; Rubba, C.; Meibner, P.; Nyhuis, D. (2014). *Automated data collection with R: A practical guide to web scraping and text mining*. John Wiley and Sons.
- Ogada, K., Mwangi, W., & Cheruiyot, W. (2015). Ngram based text categorization method for improved data mining. *Journal of Information Engineering and Applications*, 5(8), 35–43.
- Pezzi, E., & Vianna, S. L. G. (2015). A Experiência Turística e o Turismo de Experiência: um estudo sobre as dimensões da experiência memorável. *Revista Turismo em Análise*, 26(1), 165–187. <https://doi.org/10.11606/issn.1984-4867.v26i1p165-187>
- Pine, B. J., & Gilmore, J. H. (1998). Welcome to the experience economy. *Harvard Business Review*, 76(4), 97–105.
- Pine, B. J., & Gilmore, J. H. (1999). *The experience economy: Work is a theatre & every business a stage*. Harvard Business Press.

- Rahman, M. H., et al. (2018). An investigation and evaluation of Ngram, TFIDF and ensemble methods in sentiment classification [Conference paper]. *Second International Conference (ICACDS)*, Dehradun, India.
- Santos, M. L., Alencar, D. G., Andrade e Souza, A., & Gândara, J. M. G. (2018). Tendências do consumidor em turismo: a oferta no Paraná (Brasil) para demandantes de experiências. *Investigaciones Turísticas*, 16, 143. <https://doi.org/10.14198/INTURI2018.16.08>
- Silva, E. M. C., & Trentin, F. (2018). Turismo de Experiência: L'Arte Ceccato Vila Flores. *Caderno Virtual de Turismo*, 18(3). <https://doi.org/10.18472/cvt.18n3.2018.1351>
- Singh, G., Kumar, B., Gaur, L., & Tyagi, A. (2019). Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification. *ICACTM - International Conference on Automation, Computational and Technology Management*, 593–596. <https://doi.org/10.1109/ICACTM.2019.8776800>
- Tocantins, G. U. C. (2022) *Parque Estadual do Jalapão: Tocantins*. <http://gesto.to.gov.br/uc/45/>
- Torre, O. De La (1992). *El turismo fenomeno social*. Fondo de Cultura Económica.
- TripAdvisor (2022). *TripAdvisor: mais de um bilhão de avaliações de hotéis, atrações, restaurantes e muito mais*. <https://www.tripadvisor.com.br/>
- Venzatesh, Ranjitha, K. V. (2018). Classification and Optimization Scheme for Text Data using Machine Learning Naïve Bayes Classifier. 2018 IEEE World Symposium on Communication Engineering (WSCE), Singapore, 2018, pp. 33-36, <https://doi.org/10.1109/WSCE.2018.8690536>
- Violos, J., Tserpes, K., Varlamis, I., & Varvarigou, T. (2018). Text classification using the ngram graph representation model over high frequency data streams. *Frontiers in Applied Mathematics and Statistics*, 4, Article 41. <https://doi.org/10.3389/fams.2018.00041>
- Yadav, S., & Shukla, S. (2016). Analysis of kfold crossvalidation over holdout validation on colossal datasets for quality classification. In *IEEE 6th International Conference on Advanced Computing*.
- Yang, L., & Shami, A. (2020). *On hyperparameter optimization of machine learning algorithms: Theory and practice*. Department of Electrical and Computer Engineering, University of Western Ontario.
- Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*.
- Ying, Y. (2021). Effectiveness of the news text classification test using the Naïve Bayes classification text mining method. *Journal of Physics: Conference Series*.

Recebido em: 19 mar. 2024

Aprovado em: 26 abr. 2025