Processamento de Linguagem Natural para a classificação de domínios da experiência no Parque Estadual do Jalapão utilizando técnicas de IA

Natural Language Processing to classify domains of experience in Jalapão State Park using AI techniques

Procesamiento de Lenguaje Natural para la clasificación de dominios de la experiencia en el Parque Estatal de Jalapão utilizando técnicas de IA

Como citar:

Maschio, Johann B., Marynowski, João E., Feger, José E. & Oliveira, Melissa S. (2026). Processamento de Linguagem Natural para a classificação de domínios da experiência no Parque Estadual do Jalapão utilizando técnicas de IA. Revista Gestão & Tecnologia, vol. 26, nº 2, p.96-118

Johann Belenda Maschio, Mestrando em Informática na Universidade Federal do Paraná.
johannmaschio@ufpr.br
https://orcid.org/0009-0006-6071-4250

João Eugenio Marynowski, Professor adjunto na Universidade Federal do Paraná
jeugenio@ufpr.br
https://orcid.org/0000-0002-0168-7217

Jose Elmar Feger, Professor Associado na Universidade Federal do Paraná
elmar@ufpr.br
https://orcid.org/0000-0002-1982-4179

Melissa Silva de Oliveira, Professora de Informação e Comunicação no Ins. Federal do Paraná
melissaoliveira.st@gmail.com
https://orcid.org/0009-0001-5679-0254

"Os autores declaram não haver qualquer conflito de interesse de natureza pessoal ou corporativa, em relação ao tema, processo e resultado da pesquisa".

Editor Científico: José Edson Lara
Organização Comitê Científico
Double Blind Review pelo SEER/OJ
Recebido em 05/06/2025 Aprovado em 18/06/2026

Resumo

Objetivo: Este estudo aborda a automatização da categorização dos domínios da experiência turística no Parque Estadual do Jalapão (PEJ) através de técnicas avançadas de Processamento de Linguagem Natural (PLN), campo da Inteligência Artificial (IA), nos comentários de turistas no TripAdvisor.

Metodologia: Avaliar o processo de criação de modelos de IA com a aplicação dos métodos: Random Forest, Support Vector Machine, Stochastic Gradient Descent (SGD), Multi-layer Perceptron e Redes Neurais Artificiais (RNA).

Originalidade: A classificação manual de comentários online sobre pontos turísticos é dificultada pelo alto volume de dados disponíveis na Internet, razão pela qual este estudo propõe uma abordagem inovadora ao aplicar IA para automatizar esta tarefa.

Principais resultados: O estudo gerou um modelo de classificação, baseado em IA, que classifica sequências de quatro palavras (quadrigramas) nos quatro domínios da experiência. A geração do modelo considerou diferentes proporções para treinamento e teste, que modificam a acurácia dos modelos, e obteve a melhor forma de divisão sendo 85% para treinamento e 15% para teste. Quanto aos métodos, os dois melhores desempenhos foram alcançados por RNA, com 78% de acurácia, e por SGD, com 76% de acurácia.

Contribuições: Amplia os estudos que aprofundam conhecimentos de ferramentas que melhorem a coleta e o tratamento de dados por meio da IA, instrumentalizando os envolvidos com o turismo na tomada de decisões. O modelo de classificação obtido tem potencial para automatizar a avaliação de experiências turísticas, facilitando sua aplicação no setor e a tomada de decisão de gestores.

Palavras-chaves: Atividade turística. Inteligência artificial. Aprendizado de máquina. Predição. Classificação.

Abstract

Objective: This study addresses the automation of categorizing the domains of tourist experiences in the Jalapão State Park (PEJ) using advanced Natural Language Processing (NLP) techniques, a field within Artificial Intelligence (AI), applied to tourist reviews on TripAdvisor.

Methodology: To evaluate the process of creating AI models through the application of the following methods: Random Forest, Support Vector Machine, Stochastic Gradient Descent (SGD), Multi-layer Perceptron, and Artificial Neural Networks (ANN).

Originality: The manual classification of online reviews about tourist attractions is hindered by the large volume of data available on the Internet. For this reason, this study proposes an innovative approach by applying AI to automate this task.

Main Results: The study produced a classification model, based on AI, capable of classifying sequences of four words (quadrigram) into the four experience domains. The model generation considered different training and testing ratios, which impacted model accuracy, and identified the optimal split as 85% for training and 15% for testing. Regarding the methods used, the two best performances were achieved by ANN, with 78% accuracy, and by SGD, with 76% accuracy.

Contributions: This research expands the body of studies that deepen knowledge of tools to improve data collection and processing through AI, equipping tourism stakeholders for better decision-making. The resulting classification model has the potential to automate the evaluation of tourist experiences, facilitating its application in the sector and supporting decision-making by managers.

Keywords: Tourist activity. Artificial intelligence. Machine learning. Prediction. Classification.

Resumen

Objetivo: Este estudio aborda la automatización de la categorización de los dominios de la experiencia turística en el Parque Estatal del Jalapão (PEJ) mediante técnicas avanzadas de Procesamiento de Lenguaje Natural (PLN), un campo de la Inteligencia Artificial (IA), aplicadas a los comentarios de turistas en TripAdvisor.

Metodología: Evaluar el proceso de creación de modelos de IA mediante la aplicación de los siguientes métodos: Random Forest, Support Vector Machine, Stochastic Gradient Descent (SGD), Perceptrón Multicapa y Redes Neuronales Artificiales (RNA).

Originalidad: La clasificación manual de comentarios en línea sobre atracciones turísticas se ve dificultada por el alto volumen de datos disponibles en Internet. Por esta razón, este estudio propone un enfoque innovador al aplicar IA para automatizar esta tarea.

Principales Resultados: El estudio generó un modelo de clasificación basado en IA, capaz de clasificar secuencias de cuatro palabras (cuadrigramas) en los cuatro dominios de la experiencia. La generación del modelo consideró diferentes proporciones para entrenamiento y prueba, lo que afectó la precisión de los modelos, y se identificó como mejor división el 85% para entrenamiento y el 15% para prueba. En cuanto a los métodos, los dos mejores desempeños fueron alcanzados por las RNA, con un 78% de precisión, y por SGD, con un 76% de precisión.

Contribuciones: Amplía los estudios que profundizan el conocimiento sobre herramientas que mejoran la recolección y el procesamiento de datos mediante IA, brindando soporte a los involucrados en el turismo en la toma de decisiones. El modelo de clasificación obtenido tiene el potencial de automatizar la evaluación de experiencias turísticas, facilitando su aplicación en el sector y la toma de decisiones por parte de los gestores.

Palabras claves: Actividad turística. Inteligencia artificial. Aprendizaje automático. Predicción. Clasificación.

1 Introdução

O turismo é um fenômeno social, cultural e econômico, que envolve a movimentação de pessoas para lugares diferentes dos de domicílio habitual, sendo relacionado às atividades e aos gastos realizados pelos turistas (UN Tourism, 2024). O ecoturismo, por sua vez, é uma forma sustentável de turismo que valoriza e conserva o patrimônio natural e cultural (Ministério do Turismo, 2010). No Brasil, o Parque Estadual do Jalapão (PEJ), localizado no estado do Tocantins, é um ponto turístico caracterizado pela prática do ecoturismo (Governo do Tocantins, 2024).

O crescimento constante do turismo e a grande movimentação de pessoas e recursos nesta área tem se beneficiado cada vez mais do uso de Inteligência Artificial (IA) para aprimorar a experiência dos viajantes e otimizar a gestão de destinos (Santos et al., 2024). A IA tem revolucionado a administração pública em destinos turísticos, modernizando a gestão governamental e os serviços oferecidos aos visitantes, surgindo como uma ferramenta poderosa para analisar e compreender as percepções dos turistas, auxiliando na tomada de decisões (Tuo, Ning & Zhu, 2021).

Com o objetivo de classificar a experiência durante o consumo de serviços, Pine e Gilmore (1999) propuseram os domínios da experiência, cujos preceitos foram adaptados para o turismo. Estudos anteriores analisaram manualmente comentários de viajantes na Internet para identificar as experiências relatadas por eles em diferentes percursos turísticos (Benetti et al., 2018; Alencar et al., 2020; Caracristi et al., 2020; Kaizer et al., 2021). Um desses estudos, precisamente o de Kaizer et al. (2021), identificou os domínios da experiência expressos pelos turistas do PEJ nos comentários do website TripAdvisor[1].

Embora os trabalhos citados sejam fundamentais para estabelecer classificações validadas, a análise manual de grandes volumes de dados disponíveis na Internet se torna inviável. Reconhecendo isso, pesquisadores propuseram o uso de IA para essa tarefa, tecnologia vista como necessária para tomada de decisões no século XXI ao superar barreiras humanas de processamento de dados (Marynowski et al., 2023; Teliukov et al., 2024). Neste contexto, o

[1] https://www.tripadvisor.com.br/

campo da IA denominado Processamento de Linguagem Natural (PLN) se mostra relevante, pois visa capacitar os computadores a compreenderem a linguagem humana, identificando intenções, contextos e significados de palavras (Caseli & Nunes, 2024).

A possibilidade de aplicação de ferramentas de IA para analisar grandes volumes de dados é designada como um desafio nos estudos de Ghesh et al. (2024), Kim et al. (2024) e Hu & Chen (2024). Ainda, asseveram que há lacunas a serem preenchidas para aplicação de IA nas investigações relacionadas com a gestão do turismo, especialmente, no que tange ao tratamento de grandes volumes de dados disponíveis na Internet e a exploração de diferentes métodos citada por Marynowski et al. (2023).

Este estudo avança na aplicação de PLN no turismo, propondo uma abordagem inovadora que utiliza técnicas de aprendizado de máquina para classificar automaticamente comentários de turistas do PEJ no TripAdvisor dentro dos domínios da experiência. São comparadas cinco técnicas: Random Forest (RF), Support Vector Machine (SVM), Stochastic Gradient Descent (SGD), Redes Neurais Artificiais (RNA) e Multi-layer Perceptron (MLP). Ao avaliar a utilização de diferentes taxas de divisão entre os conjuntos de treinamento e teste, bem como, a eficácia dos modelos, que foi medida por meio da acurácia e de matrizes de confusão, identificou o modelo mais eficiente na classificação e avaliação das experiências turísticas, auxiliando o setor a ajustar suas ofertas e aprimorar a experiência do turista.

2 Referencial teórico

A experiência turística é um fenômeno complexo e multifacetado, que pode ser classificado em diferentes domínios. Visando classificar a experiência de usuários, os autores Pine & Gilmore (1999) criaram um modelo chamado de estágios da estruturação de uma experiência o qual foi adaptado para estudos turísticos (Benetti et al., 2018; Alencar et al., 2020). O modelo oferece uma estrutura útil para entender as diversas formas de experiência turística, visto que é constituído por dois eixos: um vertical e outro horizontal (Figura 1). O eixo horizontal representa a participação ativa (active participation) e passiva (passive participation), já o eixo vertical representa a absorção (absorption) e a imersão (immersion) do usuário de serviços, no caso deste estudo o turista.

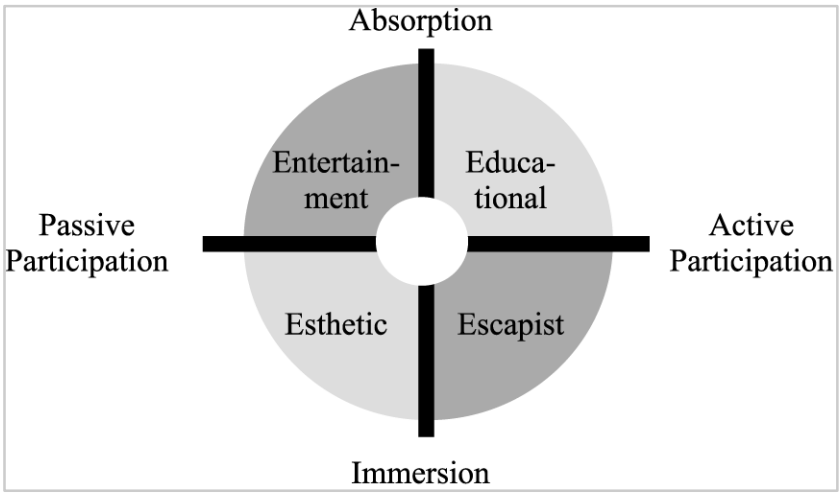


Figura 1: Pine & Gilmore (1999).

A análise das intersecções dos eixos revela os diferentes domínios da experiência. Quando a absorção se combina com a participação passiva, o resultado é o domínio do entretenimento (*entretainment*). Por outro lado, quando a absorção se junta à participação ativa, emergem experiências educativas (*educational*). Além disso, a imersão combinada com a participação passiva conduz à experiência estética (*esthetic*). Por fim, as atividades resultantes da imersão e da participação ativa levam à evasão ou escapismo (*escapist*).

A classificação dos domínios da experiência em pontos turísticos com base em comentários online é um tema recorrente em estudos acadêmicos. Benetti et al. (2018) exploraram os comentários do TripAdvisor para analisar a experiência turística em destinos culturais e naturais na Transpantaneira-MT. Alencar et al. (2020) avaliaram as características dos turistas que buscam entretenimento no Paraná, também utilizando comentários do TripAdvisor. Da mesma forma, estudos conduzidos por Caracristi et al. (2020) e Kaizer et al. (2021) utilizaram comentários do TripAdvisor para avaliar a experiência dos turistas em pontos do PEJ.

Os estudos mencionados foram conduzidos manualmente, resultando em uma abordagem lenta e custosa. Diante dessas limitações, Marynowski et al. (2023) propuseram o uso de algoritmos de IA para identificar automaticamente os domínios de experiência expressos pelos turistas. Utilizando a base de dados previamente classificada por Kaizer et al. (2021), eles

aplicaram os algoritmos Gaussian Naive Bayes, Support Vector Machine e Long Short-Term Memory para demonstrar essa estratégia.

Ao analisar comentários online, uma IA pode superar as barreiras humanas de processamento de um grande volume de dados e identificar os domínios de experiência mais valorizados pelos turistas, permitindo que as empresas do setor adaptem suas ofertas para atender às demandas específicas de cada segmento (Secondo et al., 2023; Santos et al., 2024; Teliukov et al., 2024). A personalização da experiência turística, impulsionada pela IA, pode levar a um aumento da satisfação do cliente, maior fidelização e, consequentemente, ao crescimento do setor (Kazak et al., 2020). Ao oferecer aos turistas exatamente o que eles buscam, a IA pode transformar a maneira como as pessoas viajam e experimentam o mundo.

A constante evolução da IA destaca a necessidade de replicar estudos já realizados com diferentes métodos e destinos turísticos. Com isso, poderia ser possível avaliar com maior precisão a eficácia do processo de reconhecimento dos domínios de experiência expressados pelos turistas em seus comentários online.

3 Metodologia

A análise de dados textuais para compreender a experiência do visitante tem crescente importância para o desenvolvimento de ferramentas eficientes para lidar com o grande volume de informações geradas em plataformas online. Desta forma, o propósito desta análise é avançar na pesquisa sobre a aplicação da IA no turismo, especificamente no campo de PLN, focando na classificação automatizada de comentários de turistas em quatro domínios de experiência propostos por Pine & Gilmore (1999).

Para tanto, são utilizados e comparados cinco algoritmos de aprendizado de máquina: Random Forest (RF), Support Vector Machine (SVM), Stochastic Gradient Descent (SGD), Redes Neurais Artificiais (RNA) e Multi-layer Perceptron (MLP) (Hastie et al., 2009). Esta escolha se baseia na busca por diversificar as abordagens e na eficácia demonstrada por esses modelos em tarefas de classificação de texto em diferentes contextos (Minaee et al., 2021).

A comparação desses modelos busca identificar o melhor desempenho na classificação dos domínios de experiência, auxiliando na seleção de ferramentas mais adequadas para futuras

aplicações. Ao aplicar o conhecimento sobre a aplicabilidade de diferentes modelos de PLN no turismo, esta pesquisa expande as contribuições de Marynowski et al. (2023).

A metodologia adotada neste estudo compreende cinco etapas principais: 1) composição da base de dados; 2) pré-processamento dos dados; 3) preparação do treinamento; 4) treinamento e 5) predição. Ao final, o estudo resultou em um modelo de classificação baseado em IA capaz de categorizar sequências de quatro palavras (quadrigramas) nos quatro domínios da experiência. As etapas da metodologia serão detalhadas nas subseções subsequentes.

3.1 Composição da base de dados

A base de dados utilizada é formada por 2.104 comentários de seis atrações do Parque Estadual do Jalapão (PEJ), postados no website TripAdvisor e extraídos utilizando a técnica de Web Scraping (Paula et al., 2023). A partir destes comentários, Kaizer et al. (2021) geraram e classificaram manualmente 1257 quadrigramas. O quadrigrama foi empregado pois carrega um significado considerável e pode ser identificado com frequência em um texto (Hyland, 2008). A Tabela 1 mostra alguns exemplos de quadrigramas e seus domínios de experiência.

Tabela 1
Exemplos de quadrigramas.

Quadrigrama	**Domínio**
Pedra solta portanto recomendo	Aprendizagem
Contrate guia local assim	Aprendizagem
Piquenique deliciosa falar presteza	Entretenimento
Atração não pode deixa	Entretenimento
Cristalina temperatura super agradável	Estética
Algo tão lindo assim	Estética
Recomendo entusiasmo não deixar	Evasão
Não dá vontade sair	Evasão

Fonte: Elaborada pelos autores, baseada em Kaizer et al. (2021).

3.2 Pré-processamento de dados

O banco de dados foi formado por cinco grupos, um para cada domínio da experiência e outro para os quadrigramas sem classificação. A Tabela 2 apresenta o número de

quadrigramas em cada domínio de experiência, categorizados pelos seus respectivos atrativos e domínios, incluindo também a quantidade de quadrigramas não classificados.

Tabela 2
Composição do banco de dados por atrativo e domínio

Local	Aprendizagem	Entretenimento	Estética	Evasão	Sem Classe
Cachoeira da Formiga	49	52	66	8	25
Cachoeira da Velha	49	76	39	9	1
Serra do Espírito Santo	60	53	58	12	15
Dunas	49	48	73	14	16
Mumbuca	38	129	5	0	21
Fervedouro	37	93	33	16	8
PEJ (Geral)	39	65	54	32	10

Para evitar a distinção de palavras idênticas, mas com escritas diferentes, todas as palavras foram transformadas em suas versões sem acentos e os caracteres foram convertidos para minúsculo. Essa correção foi realizada com o auxílio da função unicode [2].

2.3 Preparação para o treinamento

Há uma ampla variedade de opiniões e poucos consensos sobre a proporção ideal para dividir a base de dados entre treinamento e teste. Na tabela 3, é possível observar essa diversidade nas divisões usadas em trabalhos da área, assim como em bases de dados previamente separadas nos dois conjuntos, como o MNIST, CIFAR-100 e CIFAR-10.

Tabela 3
Trabalhos e suas divisões de treinamento e teste.

Trabalho	Treinamento & Teste
MNIST	85% / 15%

[2] https://www.python.org/

CIFAR-100	83,5% / 16,5%
CIFAR-10	83,5% / 16,5%
Comparação de técnicas de word embedding na análise de sentimentos (ZIGER, 2021)	80% / 20%
Autoregressive Convolutional Neural Networks for Asynchronous Time Series (BIŃKOWSKI et al., 2018)	80% / 20%
Um processo de classificação de texto: análise de sentimento das opiniões no TripAdvisor sobre a atração Oktoberfest Blumenau (CRESCENCIO et al., 2020)	75% / 25%
Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning (ALJUAID et al., 2022)	65% / 35%

No contexto deste trabalho, inicialmente, a base de dados foi dividida em 80% para treinamento e 20% para testes, mantendo essa proporção para cada domínio. Entretanto, na Etapa 5, foram realizados testes para avaliar a melhor divisão dos dados de treinamento e teste neste estudo. Para realizar esta divisão, utilizamos a função *train_test_split*, disponível no pacote Scikit-learn[3]. Os resultados estão apresentados na tabela 4, onde podemos observar a quantidade de quadrigramas separados para treinamento e teste em cada domínio da experiência.

Tabela 4
Banco de dados separado em treino e teste

Domínio	Treino	Teste
Aprendizagem	257	64
Entretenimento	413	103
Estética	263	65
Evasão	73	18

Como parte da preparação para o treinamento, todas as palavras contidas nos quadrigramas foram convertidas em números, gerando um tipo de índice das palavras e objetivando otimizar o processamento do modelo de IA gerado. Para a técnica RNA foi

[3] https://scikit-learn.org

empregada a função tokenizer da biblioteca Keras[4], que se integra ao TensorFlow[5]. Para os outros modelos foi empregada uma abordagem semelhante, chamada de medida TF-IDF (Term Frequency-Inverse Document Frequency), utilizando a função TfidfVectorizer da biblioteca Scikit-learn[6]. Medida tal dada pela Equação:

$$W_{x,y} = tf_{x,y} * \log(N/df_x)$$

Onde $tf_{x,y}$, df_x é o número de documentos em x e N é o número total de documentos.

O Tabela 6 mostra uma matriz de confusão genérica, representando os valores preditos e os valores de referência para cada classe. "P" denota a classe positiva e "N" a classe negativa. Os elementos da diagonal principal, VP e VN, indicam os verdadeiros positivos e verdadeiros negativos, respectivamente, representando as classificações corretas. Por outro lado, os elementos da diagonal secundária, FP e FN, representam os falsos positivos e falsos negativos, respectivamente, indicando as classificações incorretas.

Tabela 6
Matriz de Confusão

Palavra	**Valor**		
Predição		P	N
	P	VP	FP
	N	FN	VN

Tomando como base a matriz de confusão, pode-se calcular a acurácia que um modelo atingiu, encontrando seu percentual de acertos através da Equação:

$$\text{Acurácia} = (VP + VN) / (VP + VN + FP + FN)$$

[4] https://keras.io

[5] https://www.tensorflow.org

[6] http://https://scikit-learn.org/stable/

2.4 Treinamento

As técnicas de IA empregadas neste estudo foram RF, SVM, SGD, RNA e MLP. Os hiperparâmetros de cada uma das redes serão descritos a seguir. Para o treinamento dos modelos e a verificação dos melhores hiperparâmetros, será utilizado o método GridSearchCV contido na biblioteca Scikit-learn (Pedregosa et al., 2011).

O modelo RF, baseado na teoria das árvores de decisão, opera através de nós, onde cada nó pode ser um nó folha ou um nó de decisão. Para cada decisão, uma sub-árvore com uma estrutura semelhante é construída (Rezende et al., 1999).

Os hiperparâmetros treinados são, o número de árvores treinadas (*n_estimators*), o número de características consideradas quando se está procurando a melhor divisão (*max_features*), a profundidade máxima da árvore (*max_depth*) e o parâmetro para medir a qualidade da árvore (criterion). Para o *n_estimators* foram testados os seguintes valores 50, 100, 200, 500 e 1000. Quanto à *max_features*, as opções consideradas foram *sqrt* ou *log2*. Já para o *max_depth*, foram testados os valores de 4 a 10. Quanto ao critério, as alternativas foram *gini* ou *entropy*.

O modelo SVM é desenvolvido com base na teoria de aprendizado estatístico (TAE) e segue diversos princípios, como o limite de risco. O SVM busca classificadores com boa capacidade de generalização, utilizando vetores lineares e não lineares (Faceli et al., 2021). Para o SVM, o kernel utilizado foi o RBF (*Radial Bases Function*), possibilitando o treinamento de dados não-lineares.

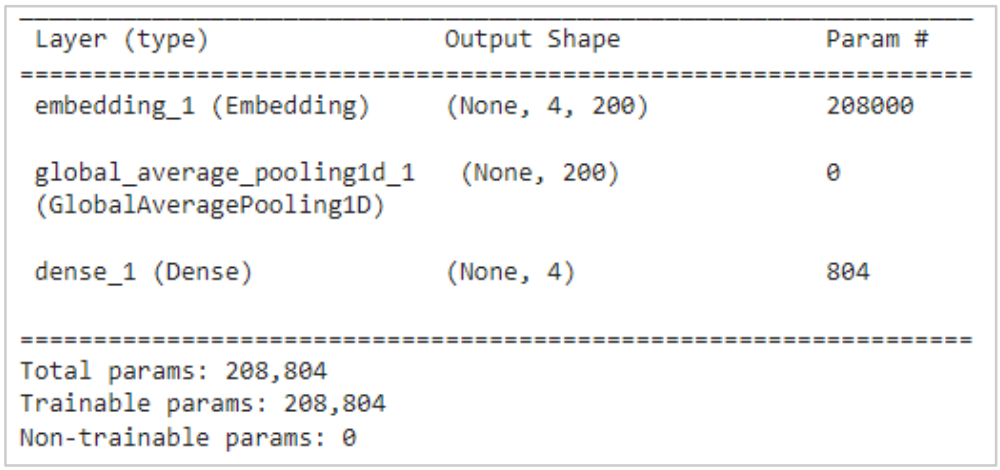
Os dois hiperparâmetros testados foram o custo C e o Gamma (coeficiente do kernel). A faixa de valores testados para C foi de 0.1, 1, 10, 100 e 1000. Quanto ao Gamma, os valores testados foram 1, 0.1, 0.01, 0.001 e 0.0001. Os algoritmos de Stochastic Gradient têm como objetivo mitigar os riscos de cada modelo, calculando os parâmetros, pesos e função de perda a cada rodada de treinamento. Dessa forma, ajustes são feitos nos pesos e uma nova rodada é iniciada (Hardt et al., 2016).

No modelo SGD, os hiperparâmetros disponíveis incluem o *loss* (função de perda), *penalty* (função de penalidade ou regularização), *alpha* (constante multiplicada pelo termo de

regularização) e o número máximo de iterações, também conhecido como épocas, *max_iter*. Os parâmetros treinados para loss foram *hinge*, *log*, *modified huber*, *perceptron* e *squared hinge*. Para o *penalty*, foram l2, l1, *elasticnet* e *none*. Para o treinamento de *alpha* foram utilizados os valores 0.0001, 0.001, 0.01, 0.1, 1 e 10, enquanto os valores de *max_iter* foram 100, 500, 1000 e 2000.

O modelo RNA, inspirado no funcionamento do sistema nervoso humano, utiliza uma densa camada de neurônios, permitindo o aprendizado e a transmissão de conhecimento para a próxima camada (Faceli et al., 2021). Na implementação com a biblioteca Keras, os hiperparâmetros utilizados foram o tamanho da camada de *Embedding*, a função de ativação, a função de otimização, o batch size e o número de épocas.

Os valores treinados para a camada de *Embedding* foram 100, 200 e 500. As funções de ativação escolhidas para o treinamento foram *softmax*, *relu* e *sigmoid*. Para o número de neurônios os valores foram 64, 128 e 256. As funções de otimização testadas foram SGD, RMSprop e Adam. Para o *batch size* foram testados os valores 8, 16, 32 e 40, enquanto os valores para o número de épocas foram 10, 30, 100 e 200. Assim, a arquitetura para o modelo final da RNA consistiu em uma camada de *Embedding*, uma camada *Global Average Pooling* e uma camada densa, conforme ilustrado na Figura 2.



Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 4, 200)	208000
global_average_pooling1d_1 (GlobalAveragePooling1D)	(None, 200)	0
dense_1 (Dense)	(None, 4)	804

Total params: 208,804
Trainable params: 208,804
Non-trainable params: 0

Figura 2: Arquitetura final para o modelo RNA.

Similarmente ao funcionamento de uma RNA, a MLP adiciona mais camadas intermediárias de neurônios, além de uma camada de saída. Todos os neurônios estão completamente conectados, onde os neurônios de uma camada estão conectados a todos os neurônios da segunda camada, e assim por diante.

No modelo MLP, os parâmetros utilizados foram *solver* (para a otimização dos pesos), *activation* (função de ativação da camada oculta), *alpha* (termo de regularização), *hidden_layer_sizes* (representando o número de neurônios na camada oculta) e *max_iter* (representando o número máximo de iterações). Os valores treinados para o solver foram *lbfgs*, *sgd* e *adam*, e para a activation foram *identity*, *logistic*, *tanh* e *relu*. Os valores de alpha foram 0.0001, 0.001, 0.01, 0.1, 1 e 10, enquanto para *hidden_layer_sizes* foram testados (100,), (100,2) e (100,3). Quanto ao *max_iter*, os valores escolhidos foram 10, 100, 200, 500, 1000 e 2000.

Objetivando a avaliação dos modelos, foi realizada a comparação de seu tempo de treinamento utilizando a divisão de 80% para treinamento e 20% para teste. O resultado desta avaliação será apresentado na Seção 4.

2.5 Predição

Após identificar o modelo com melhor acurácia, utilizando os mesmos hiperparâmetros, foi realizada uma nova bateria de testes utilizando a base de dados contendo os quadrigramas sem classificação, cujos resultados são apresentados na próxima seção. Na última etapa dos testes, realizou-se o experimento para determinar a melhor divisão da base de treinamento e teste. Nessa etapa, foram utilizados os mesmos hiperparâmetros já descritos para o modelo que apresentou melhor acurácia, juntamente com a base de dados contendo apenas os quadrigramas classificados dentro dos quatro domínios da experiência.

4 Apresentação e discussão dos resultados

A Tabela 7 apresenta o tempo de execução do treinamento de cada modelo utilizando a base de dados, com a divisão de 80% para treinamento e 20% para teste. O parâmetro n_jobs do GridSearchCV é responsável por definir o número de núcleos utilizados durante a

execução do script. Como padrão ele utiliza apenas um, se for atribuído o número -1 ele utiliza todos os núcleos do processador.

O primeiro modelo testado foi o MLP, resultando em um tempo de execução de 2 horas, 40 minutos e 7 segundos com o valor padrão de *n_jobs* e 1 hora, 59 minutos e 11 segundos com o número máximo de núcleos do processador. Todos os outros modelos foram testados com n_jobs com valores iguais a -1.

Tabela 7
Tempo de execução dos modelos

Modelo	Tempo de Execução
RF	00:00:39
SVM	00:00:03
SGD	00:00:01
MLP	01:59:11
RNA	04:37:48

Para iniciar as previsões, os melhores hiperparâmetros foram identificados utilizando o GridSearchCV. a Tabela 8 apresenta os modelos, seus hiperparâmetros e a acurácia obtida. O método que obteve a melhor acurácia foi a RNA, com 74,21% e o RF, por sua vez, obteve a menor acurácia, com 61,11%.

Tabela 8
Resultado dos experimentos GridSearchCV

Algoritmo	Hiperparâmetros	Acurácia
RF	n_estimators=50, max_features='sqrt', max_depth=10, criterion='gini'	61.11%
SVM	C=1000, gamma=0.001	69.84%
SGD	loss='hinge', alpha=0.001, penalty='elasticnet', max_iter=100	73.01%
MLP	solver='SGD', activation='tanh', alpha=0.01, hidden_layer_sizes=(100, 3), max_iter=2000	71.42%
RNA	embedding= 200, activation='softmax', batch_size=40, optimizer='adam', epochs=200	74.21%

O Tabela 9 apresenta a divisão da base em diferentes proporções para o treinamento e teste dos modelos. No primeiro teste, utilizando uma divisão de 60% para treinamento e 40% para teste, foi alcançada uma acurácia de 67,20%. Ao dividir a base em 70% para treinamento e 30% para teste, a acurácia obtida foi de 67,64%. A divisão de 80% para treinamento e 20% para teste foi empregada na primeira rodada de treinamento dos modelos, resultando em uma acurácia de 74,21%, sendo a segunda melhor divisão para a base utilizada neste trabalho. Alcançando uma acurácia de 76,19%, a divisão de 85% para treinamento e 15% para teste obteve a melhor acurácia dentre os testes realizados. Por fim, a última divisão testada foi de 90% para treinamento e 10% para teste, alcançando uma acurácia de 69,84%.

Tabela 9
Resultado das divisões da base de treinamento e teste

Treinamento	Teste	Acurácia
90%	10%	69,84%
85%	15%	76,19%
80%	20%	74,21%
70%	30%	67,64%
60%	40%	67,20%

Diante destes resultados, uma nova rodada de treinamento foi realizada, agora utilizando a base mais ajustada com a divisão de 85% para treinamento e 15% para teste.

3.1 Medida e qualidade dos modelos

A métrica de desempenho escolhida para este trabalho foi a acurácia. O melhor resultado obtido para as técnicas apresentadas foi de 77,78% para a Rede Neural Artificial, seguido pelo Stochastic Gradient Descent, Support Vector Machine, Multi-layer Perceptron e Random Forest.

Na Figura 3, os resultados são apresentados por modelo e organizados por base de dados. Para cada modelo, o primeiro valor corresponde ao treinamento realizado com a base de dados contendo os quadrigramas classificados nos quatro domínios da experiência, até daqueles não classificados, com uma divisão de 80% para treinamento e 20% para teste. O

segundo valor corresponde à base de dados apenas com os quadrigramas classificados, com uma divisão de 85% para treinamento e 15% para teste. Por fim, o terceiro valor corresponde à base de dados com os quadrigramas classificados e uma divisão de 80% para treinamento e 20% para teste.

Analisando o gráfico da Figura 3, observamos uma melhora média de 3,4% na acurácia dos primeiros quatro modelos ao utilizar a base com divisão de 85% para treinamento e 15% para teste. Além disso, notamos uma performance semelhante nos modelos RNA e SGD, com acurácias de 77,78% e 76,71%, respectivamente. Isso resulta em uma distribuição similar da matriz de confusão, que será detalhada na próxima seção.

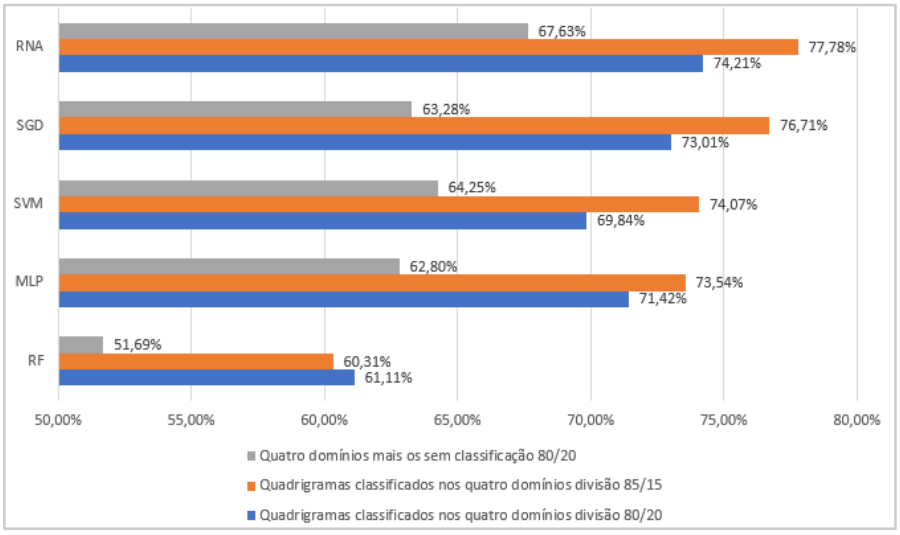


Figura 3: Gráfico comparativo das acurácias dos modelos aplicados em diferentes bases de dados

3.2 Matrizes de confusão

A matriz de confusão do modelo RNA é apresentada em todas as Tabelas ao lado das matrizes de confusão dos demais modelos. A do modelo SGD é apresentada na tabela 10. A do modelo SVM é apresentada na tabela 11. A do modelo MLP é apresentada na tabela 12, e a matriz de confusão do RF é apresentada na tabela 13.

Em todas as Tabelas, "AP" significa "Aprendizagem", "EN" significa "Entretenimento", "ES" significa "Estética" e "EV" significa "Evasão", referentes à classificação dentro dos domínios

da experiência. As linhas horizontais representam a predição e as colunas representam a referência.

O modelo RNA demonstrou melhor desempenho na predição da classe "Aprendizagem", com 28 classificações corretas, em comparação com 22 do SGD. Ambos os modelos apresentaram o mesmo número de acertos na classe "Evasão", com 6 cada. No entanto, o SGD obteve resultados ligeiramente melhores nas classes "Entretenimento" e "Estética", com 67 e 50 classificações corretas, respectivamente, em comparação com 64 e 49 do RNA. Por outro lado, o RNA apresentou um erro maior na predição das classes "Entretenimento" e "Estética", com 16 e 8 erros, respectivamente, enquanto o SGD cometeu 13 e 7 erros, respectivamente.

Tabela 10
Matriz de Confusão: RNA x SGD

	RNA					**SGD**			
Predição		AP	EN	ES	EV	AP	EN	ES	EV
	AP	28	11	4	1	22	7	3	0
	EN	13	64	3	3	15	67	4	3
	ES	0	2	49	1	4	6	50	2
	EV	0	3	1	6	0	0	0	6

Analisando as colunas de referência, o modelo SVM demonstrou capacidade igualmente eficaz na classificação da classe "Evasão", com seis acertos, em comparação com o RNA e o SGD. Nas outras classes, apresentou resultados ligeiramente inferiores, com uma diferença de quatro na classe "Aprendizado", dois na classe "Entretenimento" e um na classe "Estética".

Tabela 11
Matriz de Confusão: RNA x SVM

	RNA					**SVM**			
Predição		AP	EN	ES	EV	AP	EN	ES	EV
	AP	28	11	4	1	24	13	5	1
	EN	13	64	3	3	15	62	3	2
	ES	0	2	49	1	2	2	48	2

	EV	0	3	1	6	0	3	1	6

A matriz de confusão do MLP revela sua capacidade de previsão na classe "Entretenimento", onde obteve resultados semelhantes ao RNA, com ambos alcançando 64 acertos. Nas demais classes, o RNA mostrou-se superior, com seis acertos a mais na classe "Aprendizado", um a mais na classe "Estética" e dois a mais na classe "Evasão".

Tabela 12
Matriz de Confusão: RNA x MLP

	RNA					**MLP**			
Predição		AP	EN	ES	EV	AP	EN	ES	EV
	AP	28	11	4	1	22	12	3	0
	EN	13	64	3	3	15	64	6	5
	ES	0	2	49	1	3	3	48	2
	EV	0	3	1	6	0	1	0	4

Por fim, a matriz de confusão do RF destaca uma grande vantagem na previsão da classe "Entretenimento", com o modelo acertando 14 vezes mais que o RNA. No entanto, nas outras classes, o RF apresentou um número significativamente maior de erros, com apenas três acertos na classe "Aprendizado", 32 na classe "Estética" e um na classe "Evasão".

Tabela 13
Matriz de Confusão: RNA x RF

	RNA					**RF**			
Predição		AP	EN	ES	EV	AP	EN	ES	EV
	AP	28	11	4	1	3	1	0	0
	EN	13	64	3	3	37	78	25	10
	ES	0	2	49	1	1	1	32	0
	EV	0	3	1	6	0	0	0	1

Como registro, as execuções foram realizadas em um computador pessoal equipado com um processador Ryzen 7 3700X de 8 núcleos a 3,59 GHz, executando o sistema operacional Windows 10 Pro 21H2. As principais aplicações e bibliotecas utilizadas incluem TensorFlow versão 2.9.0, Python versão 3.9.2, Spyder versão 5.2.2 e Scikit-learn versão 1.1.2.

5 Considerações finais

O objetivo deste trabalho foi empregar técnicas de Inteligência Artificial para classificar automaticamente os domínios de experiência presentes nos comentários do website TripAdvisor para o Parque Estadual do Jalapão. Os resultados obtidos demonstram o potencial dessas técnicas na automatização da avaliação de experiências de turismo e, consequentemente, na facilitação de sua aplicação em políticas, permitindo ao setor adaptar suas ofertas para atender às demandas específicas de cada segmento, melhorando a experiência do cliente e promovendo o crescimento do setor.

Após realizar experimentos com diferentes divisões da base de dados, concluímos que a distribuição ideal para o treinamento é de 85% para treinamento e 15% para teste. Com base nessa configuração, refizemos os treinamentos e analisamos os resultados. A métrica de desempenho utilizada foi a acurácia, e constatamos que a Rede Neural Artificial obteve o melhor desempenho, alcançando 77,78%, seguida pelo Stochastic Gradient Descent, Support Vector Machine, Multi-layer Perceptron e, por último, Random Forest.

O estudo aqui apresentado contribui para os gestores de turismo no sentido de fornecer modelos confiáveis para a classificação de comentários de turistas em relação aos atrativos e destinos turísticos. Possibilita a economia de tempo para a escolha dentre os diversos modelos disponíveis no mercado, podendo se dedicar na interpretação dos resultados. Com o modelo preditivo desenvolvido, que possibilita classificar os segmentos de frases nos domínios de experiência, viabiliza a comparação com outros atrativos e com períodos anteriores. Isso auxilia no sentido de verificar o efeito de estratégias adotadas pelos gestores para melhorar o desempenho de um atrativo, por exemplo. Em próximas rodadas de coleta e processamento de dados pode avaliar se as modificações contribuíram para com o desempenho do produto.

No que tange a contribuição acadêmica, a pesquisa amplia a oferta de estudos que tratem de aprofundar conhecimentos em relação a ferramentas que melhorem a coleta e o tratamento de dados por meio da IA, a fim de instrumentalizar os envolvidos com o turismo na tomada de decisões, como indicado por Ghesh, Alexander & Davis (2024), Kim et al. (2024), Hu & Chen (2024), Marynowski et al. (2023).

Como trabalhos futuros, podem-se explorar diversas abordagens, como obter uma base de dados maior para o treinamento do modelo, especialmente com uma quantidade maior de quadrigramas na classe "Evasão". Pode-se também aplicar o modelo em uma API (Application Programming Interface) de classificação de quadrigramas nos domínios da experiência. Além disso, pode-se desenvolver um novo modelo de IA mais bem ajustado, explorando abordagens não realizadas neste trabalho, como a implementação de CNN realizadas por Silva & Serapião (2018) e testes de diversas técnicas de word embedding, conforme Ziger (2021).

Outra limitação deste estudo decorre da coleta de comentários espontâneos dos turistas, que muitas vezes não expressam sensações relacionadas com as variáveis necessárias para compreender a situação, muito menos em um formato balanceado. Isso pode restringir a possibilidade de extrapolação estatística dos resultados e demandar aprimoramentos na captura e mineração dos textos para que os resultados sejam mais próximos da realidade.
Em suma, podemos concluir que obtivemos um resultado satisfatório com os modelos RNA e SGD na classificação dos quadrigramas, porém há possibilidade de melhoria.

Referências

Alencar, D. G., dos Santos, M. L., e Souza, A. A., & Gândara, J. M. G. (6 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, 46. https://doi.org/10.14210/rtva.v21n2.p46-67

Aljuaid, H., Alturki, N., Alsubaie, N., Cavallaro, L., & Liotta, A. (2022). Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning. Computer Methods and Programs in Biomedicine, 223. https://doi.org/10.1016/j.cmpb.2022.106951

Benetti, A. C., Ozelame, Â. M. C. C., Pereira, L. A., & Tricárico, L. T. (7 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, 565–581. https://doi.org/10.25145/j.pasos.2018.16.042

Caracristi, M. de F. de A.., Feger, J. E., Silva, T. M., & Marynowski, J. E. (2021). Uma Viagem pelo Jalapão, Brasil: análise das experiências turísticas. Revista Paranaense de Desenvolvimento (RPD), 41, 89–110.

Medeiros Caseli, H., & das Graças Volpe Nunes, M. (2024). Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português (2nd ed.; H. M. Caseli & M. G. V. Nunes, Eds.). Retrieved from https://brasileiraspln.com/livro-pln/2a-edicao/

Crescencio, M., Gonçalves, A. L., & Todesco, J. L. (2020). Um processo de classificação de texto: análise de sentimento das opiniões no TripAdvisor sobre a atração Oktoberfest Blumenau. X Congreso Internacional de Conocimiento e Innovación.

Faceli, K., Lorena, AC, Gama, J., & Carvalho, ACPLF (2011). Inteligência artificial: uma abordagem de aprendizado de máquina.

Ghesh, N., Alexander, M., & Davis, A. (2024). The artificial intelligence-enabled customer experience in tourism: a systematic literature review. Tourism Review, Vol. 79. https://doi.org/10.1108/TR-04-2023-0255

Governo do Tocantins. (2024). Jalapão. https://www.to.gov.br/jalapao/w2szzqpn9qm

Hardt, M., Recht, B., & Singer, Y. (2016). Train faster, generalize better: Stability of stochastic gradient descent. 33rd International Conference on Machine Learning, ICML 2016, 3.

Hastie, T., Tibshirani, R., Friedman, J. H., & MyiLibrary. (2001). The elements of statistical learning data mining, inference, and prediction : with 200 full-color illustrations. Springer Series in Statistics.

Hu, T., & Chen, H. (6 2024). Destination experiencescape for coastal tourism: A social network analysis exploration. Journal of Outdoor Recreation and Tourism, 46, 100747. https://doi.org/10.1016/j.jort.2024.100747

Hyland, K. (1 2008). As can be seen: Lexical bundles and disciplinary variation. English for Specific Purposes, 27, 4–21. https://doi.org/10.1016/j.esp.2007.06.001

Kaizer, E. F., Caracristi, M. F. A., Feger, J. E., Marynowski, J. E., & Silva, T. M. (2021). Análise da experiência relatada pelos turistas ao visitar o Parque Estadual do Jalapão (PEJ) – TO, Brasil. Ateliê Do Turismo, 5.

Kazak, A. N., Chetyrbok, P. V., & Oleinikov, N. N. (2020). Artificial intelligence in the tourism sphere. IOP Conference Series: Earth and Environmental Science, 421. https://doi.org/10.1088/1755-1315/421/4/042020

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 and Tourism Research. https://doi.org/10.1177/10963480241229235

Marynowski, J. E., Fontana, R. M., Feger, J. E., & de Souza Lima, J. (2023). Domínios de Experiência no Turismo: uma Análise de Comentários Turísticos do Jalapão Baseada em Inteligência Artificial. https://www.anptur.org.br/anais/anais/files/20/3077.pdf

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep Learning-Based Text Classification. ACM Computing Surveys, Vol. 54. https://doi.org/10.1145/3439726

Ministério do Turismo. (2010). ECOTURISMO: Orientações Básicas. Ministério do Turismo.

Paula, J., Marynowski, J., & Feger, J. (5 2024). Coleta de Dados Turísticos com Web Scraping: Problemas, Soluções e Otimizações. iSys - Brazilian Journal of Information Systems, 17. https://doi.org/10.5753/isys.2024.3644

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12.

Pine, B. J., & Gilmore, J. H. (1999). The experience economy: work is theatre & every business a stage. Choice Reviews Online, 37. https://doi.org/doi:10.5860/choice.37-2254

Rezende, S. O., Monard, M. C., & de Carvalho, A. C. P. de L. F. (1998). Pesquisa e desenvolvimento de sistemas inteligentes para engenharia no LABIC/ICMSC/USP. Workshop de Sistemas Inteligentes Para Engenharia.

Santos, V. S., de Sousa, S. J. A., Santos, L. M. L., Filho, L. A. M. M., de Santana Porte, M., da Silva Taveira, M., & de Oliveira Alexandre, M. L. (3 2024). Inteligência Artificial nos estudos e pesquisas em Turismo no Brasil. Revista Brasileira de Pesquisa Em Turismo, 18, 2896. https://doi.org/10.7784/rbtur.v18.2896

Secundo, A., Almeida, I. C. D., & Tenório, N. (2023). Computação cognitiva nas organizações: uma investigação das ferramentas tecnológicas como apoio aos ciclos da gestão do conhecimento. Revista Tecnologia e Sociedade, 19. https://doi.org/10.3895/rts.v19n56.15380

Silva, S., & Serapiao, A. (5 2018). Ensaio em deep learning com aplicações em imagens e textos.