

ANÁLISE E PREVISÃO DE SENTIMENTOS EM AVALIAÇÕES TURÍSTICAS: CLASSIFICAÇÃO BASEADA EM REDE NEURAL ARTIFICIAL SOBRE COPACABANA

Edmar Mateus de Freitas Bortoletto¹

Robson Henrique Zerbini Barroso²

Paulo Roberto da Silva Ruiz³

Resumo

Neste estudo é explorada a utilização de redes neurais artificiais do tipo *Long Short-Term Memory* (LSTM) na análise de sentimentos referente a comentários da plataforma *TripAdvisor* sobre a Praia de Copacabana, no Rio de Janeiro. As LSTMs, pertencentes à categoria de redes neurais artificiais recorrentes, são reconhecidas por sua capacidade de processar sequências de texto, tornando-as ferramentas ideais para classificar opiniões expressas em avaliações. Com base nos dados foram identificadas tendências nas avaliações, categorizando-as como positivas, negativas ou neutras. Além disso, foi possível discernir quais aspectos específicos da Praia de Copacabana eram frequentemente elogiados, criticados ou simplesmente mencionados sem uma conotação emocional clara. Por meio da base de dados foi desenvolvido um modelo de classificação de comentários relativo à experiência do usuário. Os resultados evidenciam a capacidade das LSTMs em captar e categorizar nuances dos sentimentos expressos, oferecendo percepções para os interessados em aprimorar sua experiência turística. Por fim, os resultados deste trabalho também servem de referência para destinos turísticos que buscam compreender e atender melhor às expectativas de seus visitantes.

Palavras-chave: Análise de Sentimentos. Aprendizado de Máquina. Mineração de Dados. Processamento de Linguagem Natural.

Abstract

In this study, the use of artificial neural networks of the Long Short-Term Memory (LSTM) type is explored in the analysis of sentiments regarding comments on the TripAdvisor platform about Copacabana Beach, in Rio de Janeiro. LSTMs, belonging to the category of recurrent artificial neural networks, are recognized for their ability to process text sequences, making them ideal tools for classifying opinions expressed in reviews. Based on the data, trends in the evaluations were identified, categorizing them as positive, negative or neutral. Furthermore, it was possible to discern which specific aspects of Copacabana Beach were frequently praised, criticized or simply mentioned without a clear emotional connotation. Using the database, a comment classification model was developed regarding user experience.

¹ Graduando em Tecnologia em Ciência de Dados pela Fatec Adamantina/SP. Endereço eletrônico: edmar.bortoletto@fatec.sp.gov.br.

² Graduando em Tecnologia em Ciência de Dados pela Fatec Adamantina/SP. Endereço eletrônico: robson.barroso@fatec.sp.gov.br.

³ Docente do curso de Ciência de Dados na Fatec Adamantina/SP. Endereço eletrônico: paulo.ruiz2@fatec.sp.gov.br.

The results highlight the ability of LSTMs to capture and categorize nuances of expressed feelings, offering insights for those interested in improving their tourist experience. Finally, the results of this work also serve as a reference for tourist destinations that seek to better understand and meet the expectations of their visitors.

Keywords: *Data Mining. Machine Learning. Natural Language Processing. Sentiment Analysis*

1 Introdução

O crescente interesse no turismo e na hospitalidade no Brasil destaca a necessidade de compreender melhor a experiência e satisfação dos visitantes em destinos turísticos populares. Plataformas, como a do *TripAdvisor*, tornaram-se lugares onde turistas de todo o mundo compartilham suas experiências, fornecendo informações valiosas tanto para futuros visitantes quanto para gestores de destinos e negócios turísticos (TRIPADVISOR, 2023). Entretanto, com milhares de comentários sendo postados diariamente, analisar manualmente cada avaliação torna-se impraticável. Surge, então, a demanda por métodos automatizados e precisos para extrair informações desses dados.

Este artigo adota um ponto turístico emblemático do Brasil: a renomada Praia de Copacabana, no Rio de Janeiro, conhecida mundialmente por sua beleza e vida cultural vibrante. Dadas as características desse destino, surge uma questão: Como os visitantes percebem e avaliam esse local? E mais profundamente, como podemos, de forma automática, categorizar e prever esses sentimentos com base em avaliações escritas?

Portanto, o objetivo deste trabalho é empregar a arquitetura de redes neurais artificiais (RNA) *Long Short-Term Memory (LSTM)* para classificar o sentimento das avaliações postadas no *TripAdvisor* sobre esse destino turístico e, com base nesses sentimentos, desenvolver um modelo de classificação para realizar uma previsão de sentimentos positivos, negativos ou neutros. A escolha da *LSTM* (GOODFELLOW et al., 2016) é justificada por sua capacidade comprovada em lidar com sequências de texto e capturar nuances e dependências de longo prazo dentro dos dados, algo essencial quando se trata de avaliações escritas, onde o contexto e a sequência das palavras são utilizados para entender o sentimento em geral.

2 Referencial teórico e trabalho correlatos

A trajetória das RNAs no campo da aprendizagem de máquina tem sido marcada por inovações e desafios. Uma das primeiras tentativas de modelar matematicamente o neurônio biológico surgiu com McCulloch e Pitts (1943), os quais foram pioneiros em propor uma concepção binária de neurônios artificiais. Eles estabeleceram um marco crucial, lançando as bases para o desenvolvimento de redes neurais ao delinear como um neurônio pode receber entradas, processá-las e, em seguida, emitir uma saída.

Depois de pouco mais de uma década, Rosenblatt (1958) introduziu o *Perceptron*, um algoritmo inovador que refletiu uma das primeiras realizações práticas da visão de um neurônio artificial. A promessa do *Perceptron* residia em sua capacidade de aprender e adaptar-se autonomamente. No entanto, seu brilho foi ofuscado pelas críticas construtivas de Minsky e Papert (1969), que destacaram limitações significativas no modelo, notadamente sua incapacidade de processar funções não linearmente separáveis.

A renovação do interesse pelas RNAs foi catalisada pela introdução do método de retropropagação (*backpropagation*). Rumelhart et al. (1986) apresentaram esse algoritmo que, ao permitir a propagação de erros de trás para frente através da rede, tornou viável o treinamento eficiente de RNAs multicamadas. Esse avanço foi decisivo, proporcionando meios para superar os desafios associados aos modelos anteriores.

Contudo, foi na virada do milênio que se presenciou um avanço estrutural no campo. Hinton et al. (2006) introduziram técnicas revolucionárias para treinar redes neurais profundas, desafiando e superando as barreiras tradicionais associadas à quantidade de camadas da rede. Sua contribuição redefiniu as possibilidades no domínio das redes neurais, estabelecendo um novo padrão para o desenvolvimento e aplicação de modelos de aprendizado profundo.

Muitas vezes, as informações relevantes em uma sequência não são apenas o que é observado agora, mas sim o que foi observado em momentos anteriores. No campo do Processamento de Linguagem Natural (PLN), é amplamente reconhecido que o contexto anterior em uma sequência desempenha um papel fundamental na

compreensão do significado atual (ADI et al., 2018). É aqui que entram as redes neurais *LSTM*.

As *LSTM* foram introduzidas por Hochreiter e Schmidhuber (1997) e representam uma evolução das redes neurais recorrentes. Ao contrário das RNA tradicionais, que muitas vezes enfrentam dificuldades ao tentar aprender dependências de longo prazo devido a problemas como o desaparecimento do gradiente, a *LSTM* foi projetada especificamente para lidar com essa deficiência (GOODFELLOW et al., 2016).

O núcleo da *LSTM* é a célula de memória, que é capaz de armazenar, ler e escrever informações ao longo do tempo. Além disso, elas possuem portas de entrada, esquecimento e saída, que controlam como as informações fluem para dentro e fora da célula de memória. Essa habilidade de lembrar e esquecer seletivamente torna as *LSTM* particularmente adequadas para tarefas sequenciais, como a análise de sentimentos (DENG; LIU, 2018).

Na análise de sentimentos, as sequências de palavras nas sentenças carregam sentimentos e emoções que se desenvolvem ao longo do texto. As *LSTM*, com sua capacidade de lembrar as nuances e subtópicos ao longo de uma sequência, são extremamente úteis. Elas podem capturar sutilezas em textos longos, reconhecendo, por exemplo, uma reviravolta no sentimento de um comentário que começa de forma positiva, mas termina de forma negativa (GOODFELLOW et al., 2016).

Em resumo, as redes *LSTM* têm se mostrado uma ferramenta poderosa na arena da análise de sentimentos. Elas não apenas entendem a sequência de palavras, mas também a emoção e o contexto que as palavras carregam ao longo do tempo. Seu *design* inovador, que desafia os problemas que as RNA tradicionais enfrentam, as posiciona como um pilar fundamental no campo do PLN (DENG; LIU, 2018).

Trabalhos recentes comprovam a eficiência das RNAs na análise de sentimentos. Por exemplo, o estudo conduzido por Araujo e Fernandes (2021) sobre a análise de *tweets* ligados ao mercado de ações brasileiro conseguiu uma precisão de 74%. Esse é apenas um exemplo de como a pesquisa em análise de sentimentos tem se diversificado e aprofundado em múltiplos contextos. Outro trabalho, realizado por Crescencio et al. (2020), focou nos comentários sobre a *Oktober Fest* em

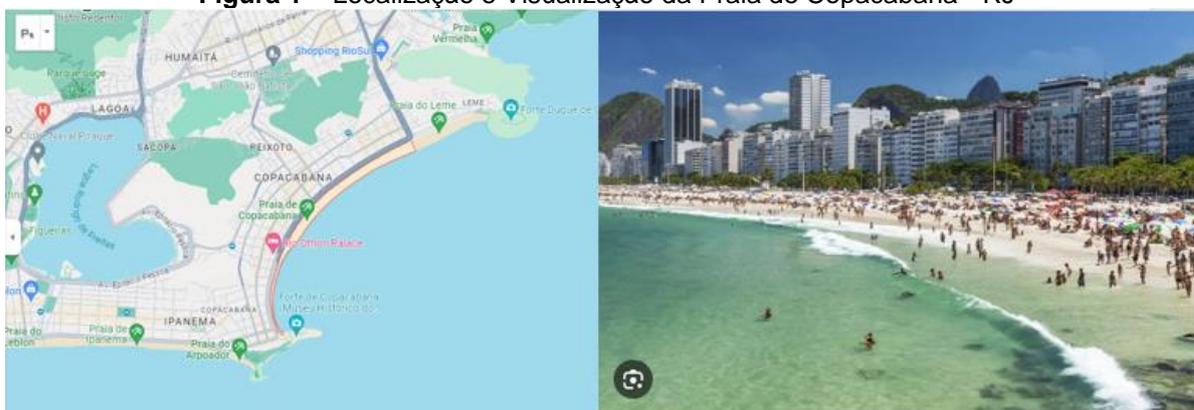
Blumenau, alcançando acurácia de 90%. Por outro lado, o estudo de Vieira do Carmo Neto (2022) centrado em *reviews* de produtos em e-commerce brasileiros obteve uma acurácia de 86%. Já Silva Albuquerque (2022) examinou sentimentos em três bases de dados distintas: *B2W*, *Buscape* e *Olist*, com resultados superiores a 90%.

3 Materiais e métodos

3.1 Materiais

Neste trabalho, é utilizada uma base de dados obtida na plataforma *Kaggle* (KAGGLE, 2023) relacionada a diversos pontos turísticos. As avaliações escolhidas para análise foram aquelas relacionadas à Praia de Copacabana, situada no Rio de Janeiro. O arquivo é composto pelos comentários, notas e datas de avaliação e totalizam 7.664 avaliações. A Praia de Copacabana, que está localizada na zona sul do Rio de Janeiro (Figura 1), foi apelidada pela população de Princesinha do Mar, em referência a uma canção de João de barro, O Braguinha e Alberto Ribeiro. Essa é considerada uma das praias mais famosas do mundo (RIOTUR, 2023).

Figura 1 – Localização e Visualização da Praia de Copacabana - RJ



Fonte: Google Maps (2023).

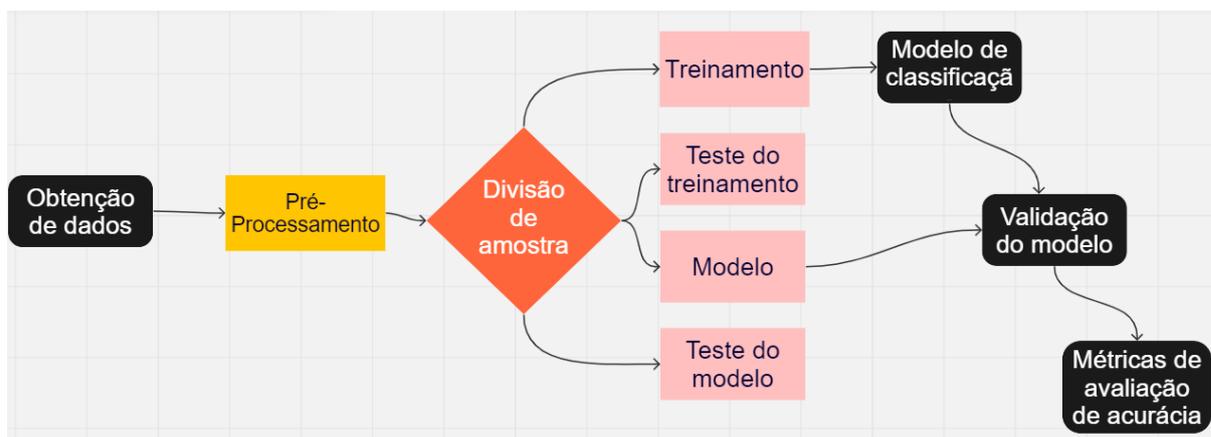
O processamento e análise dos dados foi realizado no *Google Colaboratory* (COLAB), uma plataforma baseada em nuvem que oferece um ambiente de execução interativo para a linguagem de programação Python. O COLAB tem se mostrado uma ferramenta valiosa na área de ciência de dados e aprendizado de máquina, permitindo que pesquisadores e profissionais acessem recursos

computacionais robustos sem a necessidade de configurações locais complexas (GOOGLE, 2023).

3.2 Métodos

Neste trabalho, foi adotada uma abordagem estruturada para organizar e executar a metodologia, sendo que a sequência dos métodos são fundamentais para a qualidade e eficácia do projeto. Para atingir esse objetivo, foi empregado um fluxograma que delinea claramente a ordem e a relação entre cada etapa do processo. Esse esquema, serve como um guia visual para ajudar a manter um fluxo lógico e consistente em todo o trabalho, descrevendo como cada parte do projeto se conecta e como os resultados de uma etapa influenciam as decisões e ações subsequentes (Figura 2).

Figura 2: Fluxograma metodológico



Fonte: elaborado pelos autores.

3.2.2 Obtenção dos Dados

O ponto de partida para a aquisição dos dados foi a plataforma *Kaggle*, reconhecida por ser uma rica fonte de dados. Nesse ambiente foi identificado o conjunto de dados intitulado Dados atrações turísticas BR, disponibilizado por Filho (2023). Esse conjunto, em particular, contém avaliações do *TripAdvisor* para diversos pontos turísticos, incluindo aquelas relacionadas à Praia de Copacabana, que formam o núcleo da análise deste trabalho.

3.2.3 Pré-Processamento

Nessa fase, a base de dados foi importada para o COLAB e utilizadas as bibliotecas essenciais para a leitura dos dados. Foi desenvolvida uma função para categorizar os sentimentos com base no atributo notas, que na base de dados oscila entre 0 e 5. O critério adotado foi o seguinte: notas abaixo de 2 foram classificadas como negativas, notas exatamente 3 como neutras, e notas 4 e 5 foram categorizadas como positivas. A partir deste procedimento, introduziu-se no *dataframe* uma nova coluna denominada sentimento, servindo para registrar essa classificação (Figura 3).

Figura 3 – Estrutura da base de dados.

	comentario	nota	data	sentimento
0	A praia é linda, tem algumas águas vivas na pr...	5.0	mar de 2023	Positivo
1	A natureza juntas no quadro a praia ao longe a...	5.0	abr de 2022	Positivo
2	Além de linda, nos passou a sensação de segura...	5.0	mar de 2022	Positivo
3	Fui na praia no carnaval 2023 e furtaram uma f...	2.0	fev de 2023	Negativo
4	Boa noite tive o privilégio de conhecer Copaca...	5.0	fev de 2023	Positivo
5	Fui na sexta-feira 10 de fevereiro de 2023. Fu...	3.0	fev de 2023	Neutro
6	Belíssima sem dúvida! Só que desta vez ficamos...	4.0	jan de 2023	Positivo
7	Minha viagem foi maravilhosa , acima da minhas...	5.0	jan de 2023	Positivo
8	Uma beleza sem igual, praia gostosa, tem ótima...	5.0	dez de 2022	Positivo
9	Fui a praia de Copacabana com a expectativa às...	3.0	jan de 2023	Neutro

Fonte: elaborado pelos autores.

Para ser possível a leitura de texto pelo computador, em PLN é necessário promover a conversão de dados textuais em formatos numéricos e tabulares. Para tal, inicialmente foi empregada a técnica de *tokenização*, a qual separa cada palavra do texto por meio da função *tokenizer()*. Essa função define um limite de 100 palavras-chave usando *num_words=100*. Isso especifica o número máximo de palavras únicas a serem consideradas. Ao aplicar a função *fit_on_texts()*, um vocabulário é construído, tabulando as palavras mais frequentes em cada comentário. A seguir, é realizada a conversão do texto em uma representação

numérica por meio da função `token.texts_to_sequences()`, transformando as palavras em seus respectivos números inteiros baseados no vocabulário criado anteriormente. Para garantir a uniformidade necessária para o processamento, as sequências são padronizadas usando `pad_sequences`, com `padding=post` e `maxlen=100`, onde elas são ajustadas para ter um comprimento uniforme de 100 palavras, adicionando zeros à direita quando necessário (Figura 4).

Figura 4 – Função de *tokenização* e conversão de dados.

```
token = Tokenizer(num_words=100)
token.fit_on_texts(data['comentario'].values)
X = token.texts_to_sequences(data['comentario'].values)
X = pad_sequences(X, padding="post", maxlen=100)
print(X)
```

```
[[ 3  5  4 ...  0  0  0]
 [ 3 17  3 ...  0  0  0]
 [ 1 33 72 ...  0  0  0]
 ...
 [12 44 65 ...  0  0  0]
 [ 3  5 21 ...  0  0  0]
 [ 4  6  9 ...  0  0  0]]
```

Fonte: elaborado pelos autores.

3.2.4 Divisão de amostras: Treinamento e Teste

Antes da criação do modelo de classificação, é necessário dividir os dados disponíveis em conjuntos específicos para treinamento e teste. Esta divisão assegura uma avaliação justa do desempenho do modelo em dados que não foram utilizados no treinamento. Por meio da função `train_test_split()`, o conjunto de dados é separado, reservando-se 30% para teste e 70% para o treinamento do modelo.

3.2.5 Modelo de classificação

A Figura 5 apresenta a criação das camadas da RNA para a realização do treinamento. A primeira delas é a *embedding*, a qual corresponde a entrada dos dados. A segunda camada corresponde à uma *spatialDropout*, que possui a função de diminuir a dimensionalidade dos dados. A seguir é criada a camada *LSTM* com 196 neurônios, além de parâmetros como *dropout* e as funções de ativação. Por fim,

é criada a última camada, do tipo densa possuindo 3 neurônios, os quais representam as classes negativo, neutro e positivo.

Figura 5 – Criação das camadas da RNA.

```
modelo = Sequential()  
modelo.add(Embedding(input_dim=len(token.word_index),output_dim=128,input_length=X.shape[1]))  
modelo.add(SpatialDropout1D(0.2))  
  
modelo.add(LSTM(units=196, dropout=0.2, recurrent_dropout=0, activation='tanh',  
                recurrent_activation='sigmoid', unroll=False,use_bias=True))  
modelo.add(Dense(units=3,activation="softmax"))
```

Fonte: elaborado pelos autores.

Durante o processo de treinamento, a rede é ensinada, usando um conjunto de dados, ajustando seus pesos para minimizar a função de perda escolhida (HAGAN et al., 1996). Dessa forma, o modelo é ajustado com base nos dados de treinamento. No caso presente, foram definidas 10 épocas de treinamento com um tamanho de lote de 30. Além disso, os dados de teste são usados como conjunto de validação, permitindo monitorar o desempenho do modelo em dados não treinados durante cada época.

3.2.6 Validação de Modelo

Na avaliação de modelos de aprendizado de máquina, especialmente aqueles destinados às tarefas de classificação, recorre-se a métricas específicas que oferecem uma visão clara do desempenho do modelo em diferentes aspectos. A acurácia, por exemplo, fornece uma medida geral da capacidade do modelo de fazer previsões corretas, mas pode não ser suficientemente detalhada em situações em que as classes são desbalanceadas (RASCHKA; MIRJALILI, 2019).

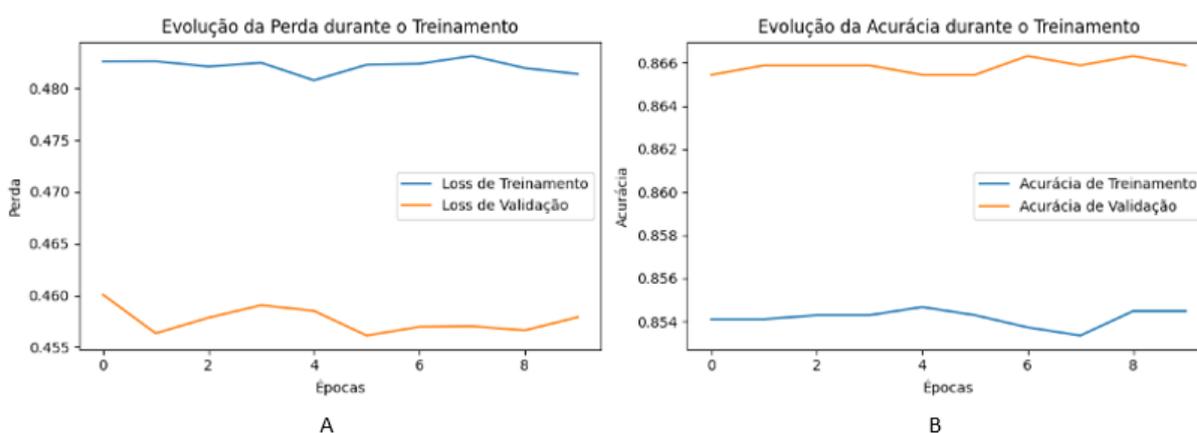
Por outro lado, a precisão e o *recall*, são métricas que dão uma perspectiva mais granular sobre a eficácia do modelo. A precisão é uma métrica crucial quando as consequências de Falsos Positivos são significativas, enquanto o *recall* é mais relevante quando é fundamental capturar todos os casos positivos possíveis (PROVOST; FAWCETT, 2013). Além dessas, o *F1-Score* é outra métrica que busca balancear precisão e *recall*, sendo particularmente útil quando se quer uma métrica composta que leve em consideração ambos os aspectos.

Adicionalmente, para obter uma visualização abrangente do desempenho do modelo em todas as classes, especialmente em situações de classificação multiclasse, a matriz de confusão se torna uma ferramenta indispensável (ADI et al., 2017). Ela permite uma análise detalhada dos acertos e erros do modelo, categorizados por classes reais e previstas.

4 Resultados e discussões

A Figura 6 apresenta o comportamento do treinamento do modelo, a partir da evolução da perda e acurácia. A Figura 6A revela que a perda do treinamento foi superior a 0.480 em todas as épocas, já aquela verificada pela validação manteve-se abaixo de 0.460. No caso da acurácia, pela Figura 6B verifica-se que a validação apresentou índice entre 0.864 e 0.866, superando a acurácia do treinamento. Estas análises são importantes para compreender como ocorreu a evolução do treinamento ao longo das épocas no contexto da RNA.

Figura 6 – Acurácia do treinamento. (A) Perda durante o treinamento. (B) Acurácia durante o treinamento.

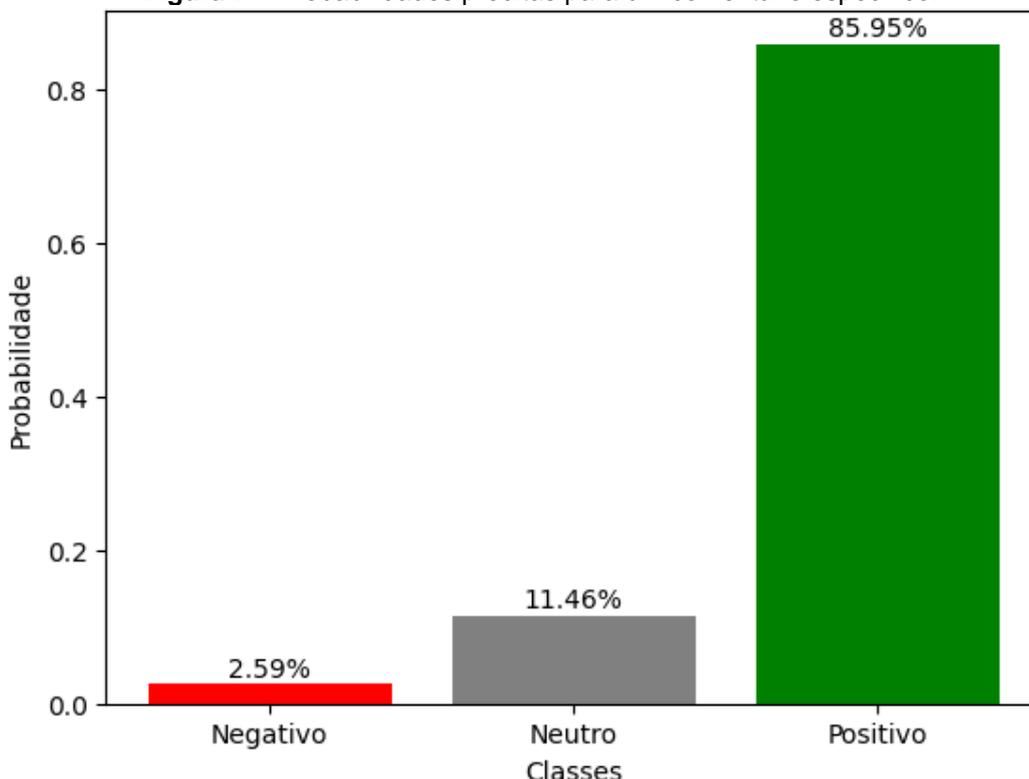


Fonte: elaborado pelos autores.

As saídas do modelo são em forma de vetores, os quais representam as probabilidades previstas para as três classes distintas deste estudo. Por exemplo, considerando o vetor [0.02589225 0.11461664 0.85949105], o modelo estima que a probabilidade do comentário em questão pertencer à classe negativo é de 2,59%, à classe neutro é de 11,46% e à classe positivo é de 85,95%, como apresentado na Figura 7. Com base nessa distribuição de probabilidade, a classificação do modelo é positiva para este comentário específico, visto que ela apresenta a maior

probabilidade entre as três. Foi dessa maneira que os comentários foram classificados, ou seja, considerando a maior probabilidade em relação as classes.

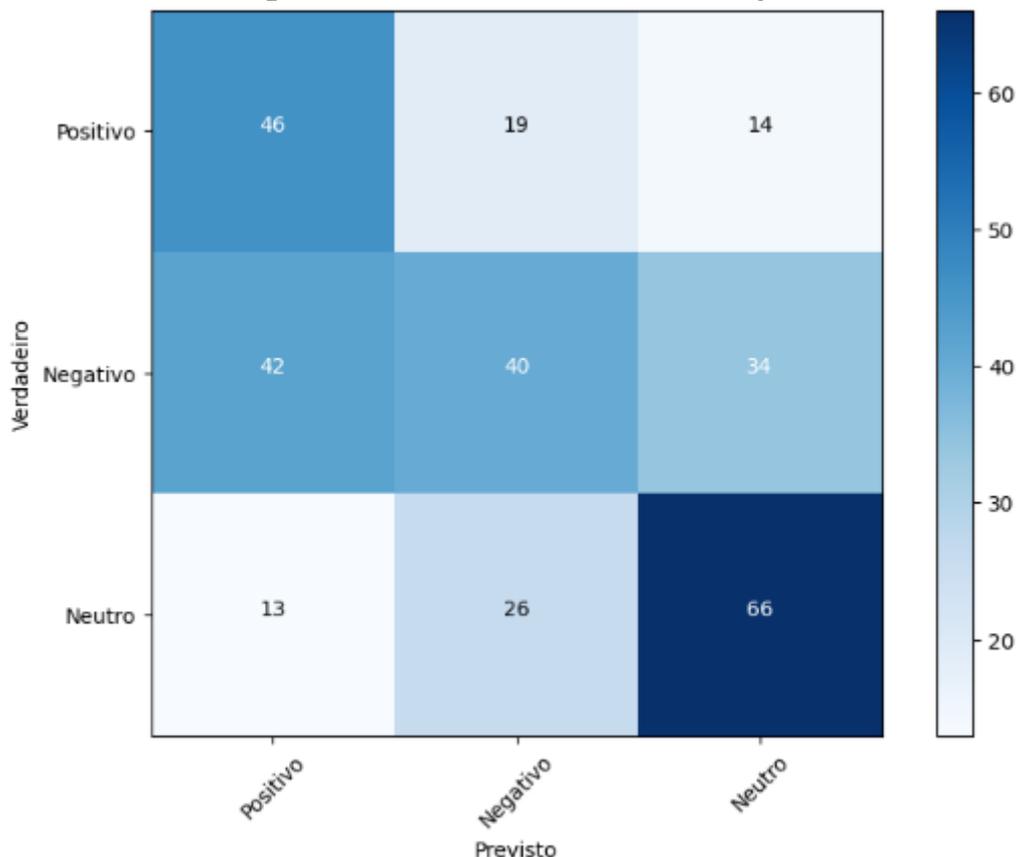
Figura 7 – Probabilidades previstas para um comentário específico.



Fonte: elaborado pelos autores.

A Figura 8 apresenta a matriz de confusão da classificação. As linhas da matriz representam as classes verdadeiras, enquanto as colunas representam as classes previstas pelo modelo. Neste caso, a matriz está disposta em uma ordem 3x3, indicando que existem três classes possíveis: Positivo, Negativo e Neutro. Os valores da diagonal principal da matriz (46, 40 e 66) indicam as previsões corretas para cada uma das classes, enquanto os outros valores representam os diferentes tipos de erros cometidos pelo modelo.

Figura 8 – Matriz de confusão da classificação



Fonte: elaborado pelos autores.

Os resultados deste trabalho ressaltam o desafio inerente à obtenção de alta acurácia em modelos de análise de sentimentos, evidenciando a influência de diversos fatores, como a natureza do texto analisado, a etapa de pré-processamento e as particularidades do modelo implementado.

A acurácia de 0.50, refletindo um desempenho na ordem de 50%, pode suscitar dúvidas sobre a eficácia do modelo, especialmente considerando a natureza particular dos dados. Em termos gerais, uma acurácia de 50% pode ser interpretada como um desempenho equivalente a uma escolha aleatória, indicando que o modelo não supera um nível de predição casual.

No que diz respeito à pontuação de 1.54 na questão do modelo, é crucial destacar que esse valor pode indicar que o modelo atual não está devidamente ajustado aos dados. Uma pontuação mais baixa sugere uma falta de correspondência entre as previsões do modelo e os valores reais, indicando a necessidade de uma otimização do modelo ou mesmo uma reconsideração da escolha da arquitetura, como o uso da LSTM.

Para alcançar uma precisão acima de 90%, será essencial explorar a fundo as limitações identificadas no estudo atual. Uma análise mais detalhada da qualidade dos dados, inclusão de características relevantes não consideradas anteriormente, e ajustes significativos na arquitetura do modelo são passos críticos para melhorar o desempenho preditivo. Adicionalmente, a implementação de técnicas avançadas, como o ajuste fino de hiper parâmetros e a utilização de *embeddings* pré-treinados, pode ser considerada para aprimorar ainda mais os resultados.

Considerações finais

Ao longo da pesquisa, foi realizada uma análise dos comentários dos visitantes da Praia de Copacabana, no Rio de Janeiro, utilizando a avançada técnica de RNA do tipo LSTM para classificação de dados no contexto da análise de sentimentos. As informações obtidas não se limitaram apenas às experiências individuais dos turistas, servindo como um indicativo abrangente da qualidade da experiência turística em Copacabana, considerando variáveis como infraestrutura, ambiente e atendimento.

A relevância desse trabalho destaca-se em suas múltiplas aplicações. Primeiramente, ressalta-se o potencial da análise de sentimentos na gestão e planejamento de destinos turísticos. No entanto, é crucial reconhecer que, apesar da metodologia avançada e da abordagem perspicaz, os resultados obtidos são inferiores àqueles apresentados por outros trabalhos. Essa diferença pode ser atribuída à base de dados utilizada, bem como às técnicas de pré-processamento dos dados.

Cabe destacar considerações sobre se os resultados foram impactados pela fonte de dados ou pelo método aplicado no processamento e análise. A qualidade e representatividade da base de dados do estudo, obtida possivelmente de fontes como o *Kaggle*, influenciaram significativamente os resultados. Além disso, uma análise crítica do método de processamento e análise adotado é imperativa para compreender plenamente as razões por trás do desempenho abaixo do esperado.

Como sugestão de trabalhos futuros, podem ser citados:

- Aprimorar a técnica de pré-processamento dos dados, com critérios mais assertivos de classificação inicial dos comentários.
- Comparação de diferentes técnicas de pré-processamento como *One Hot Encoding* e *Embeddings*.
- Utilização de mais camadas ocultas objetivando a melhora da acurácia.
- Aumentar a base de dados com comentários relativos a outros pontos turísticos brasileiros.
- Realizar análises mais aprofundadas das percepções dos visitantes para possibilitar tomadas de decisões mais assertivas por parte dos administradores dos pontos turísticos.

Essas reflexões e sugestões enriquecem não apenas a compreensão dos resultados, mas também fornecem informações para pesquisas futuras. Identificar as limitações associadas à fonte de dados e à metodologia é crucial para aprimorar futuros estudos e contribuir para a evolução da análise de sentimentos em contextos turísticos.

Referências

ADI, Y.; KERMANI, E.; BELINKOV, Y.; LAVI, O.; GOLDBERG, Y. **Analysis of sentence embedding models using prediction tasks in natural language processing**. IBM Journal of Research and Development, [s.i.], v. 61, n. 4/5, p. 3:1-3:9, jul. -set. 2017.

ARAÚJO, V. L.; FERNANDES, B. J. Análise de Sentimentos de Textos do Twitter sobre Mercado de Ações Brasileiro. **Revista de Engenharia e Pesquisa Aplicada**, Recife, v. 6, n. 5, p. 18-26, Novembro, 2021.

CRESCENCIO, M.; GONÇALVES, A. L.; TODESCO, J. L. Um processo de classificação de texto: análise de sentimento das opiniões no tripadvisor sobre a atração Oktoberfest Blumenau. **Anais do Congresso Internacional de Conhecimento e Inovação – ciki**, [S. l.], v. 1, n. 1, 2020.

DENG, L.; LIU, Y. A Joint Introduction to Natural Language Processing and to Deep Learning. In: DENG, L.; LIU, Y. (Ed.). **Deep Learning in Natural Language Processing**. Singapore: Springer, 2018.

FILHO, J. **Dados atrações turísticas BR**. Kaggle, 2023. Disponível em: <<https://www.kaggle.com/datasets/jeffersonsilho/dadosatracoesturisticasbr>>. Acesso em: 12 set. 2023.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. **Angewandte Chemie International Edition**, 6(11), 951–952. Massachusetts Institute of Technology. 2016.

GOOGLE. **Google Colaboratory**. Disponível em: <<https://colab.research.google.com/>>. Acesso em: 28 ago. 2023.

GOOGLE MAPS. **Praia de Copacabana, Rio de Janeiro, RJ**. Disponível em: <<https://maps.app.goo.gl/LE9kWNakpASCub1z7>>. Acesso em: 15 set. 2023.

HAGAN, M. T.; DEMUTH, H. B.; BEALE, M. **Neural Network Design**. 1. Ed. Boston: PWS Publishing Co., 1996. ISBN 0534943322.

HINTON, G. E.; OSINDERO, S.; TEH, Y. A Fast-Learning Algorithm for Deep Belief Nets. **Neural Computation**, Toronto, v. 18, n. 7, p. 1527-1554, 2006.

HOCHREITER, S.; SCHMIDHUBER, J. **Long Short-Term Memory**. **Neural Computation**, [s.i.], v. 9, n. 8, p. 1735-1780, 15 nov. 1997.

KAGGLE. **Kaggle: Your machine learning and data science community**, sd. Disponível em: <<https://www.kaggle.com>>. Acesso em: 01 set. 2023.

MCCULLOCH, W., PITTS, W. A Logical Calculus of the Ideas Immanent in Nervous Activity. **The Bulletin of Mathematical Biophysics**, vol. 5, no. 4, 1943.

MINSKY, M.; PAPERT, S. A. **Perceptrons: An introduction to computational geometry**. [S.l.]: MIT press, 2017.

PROVOST, F; FAWCETT, T. **Data Science for Business: What you need to know about data mining and data-analytic thinking**. 1. Ed. Sebastopol, CA: O'Reilly Media, Inc., 2013.

RIOTUR. **Copacabana e Leme**. Rio de Janeiro: RIOTUR, [s.d.]. Disponível em: <https://riotur.rio/que_fazer/copacabana-e-leme/>. Acesso em: 10 set. 2023.

ROSENBLATT, F. **The perceptron: a probabilistic model for information storage and organization in the brain**. Psychological review, American Psychological Association, v. 65, n. 6, p. 386, 1958.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. **Learning representations by back-propagating errors**. Nature, Nature Publishing Group, v. 323, n. 6088, 1986.

RASCHKA, Sebastian; MIRJALILI, Vahid. **Python Machine Learning**. 2nd ed. Packt Publishing Ltd, 2019. 772 p. ISBN 9781789958294.

SILVA ALBUQUERQUE, C.L. **Análise de sentimento sobre comentários em sites de e-commerce no idioma português/BR**. Trabalho de Conclusão de Curso (Sistemas de Informação) - Centro Universitário Christus, Fortaleza, 2022.

TRIPADVISOR. **Sobre o Tripadvisor.** Disponível em:
<<https://tripadvisor.mediaroom.com/br-about-us>>. Acesso em: 09 set 2023.

VIEIRA DO CARMO NETO, F. **Análise de sentimentos de reviews de produtos em e-commerces brasileiros.** Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) - Universidade Federal de Alagoas, Campus Arapiraca, Arapiraca, 2022.