INTRODUCTION

Tourism is one of the contributing factors for Malaysia's economic growth before pandemic COVID-19 in 2020 and now is slowly rebounding post pandemic. Based on the Ministry of Tourism, Arts and Culture Malaysia (MOTAC), in 2018, the total number of tourist arrivals in Malaysia reached 25.83 million people and has increased to 26.1 million in 2019. As the pandemic COVID19 hits worldwide and most of the countries close their borders, Malaysia faces a major decline in the number of tourists in 2020 and 2021. However, the tourism industry started to gain tourists in 2022 as Malaysia opened their international borders, and continues to increase from 10.07 million people in 2022 to 14.47 million people visiting Malaysia in 2023.

Tourist visits are heavily influenced by global travelers' satisfaction, reviews and feedback. There is an immediate need for improvement in the tourism business through the analysis of travelers' experiences. In order to draw travelers' internal motivations, major attractions should be provided as appropriate pull factors [1]. Tourism managers should enhance their services and implement marketing strategies to better meet the factors that influence travelers' desire to visit the destinations [2].

As technology becomes more advanced over the years, with more social media platforms and review websites being provided to the public, people are using them to express their opinions and experiences about various products, services and travel destinations. This is called user-generated content (UGC) and in the context of the tourism industry, it enables travelers to share their reviews and feedback on travel-related websites [3]. Travelers usually rely on UGC shared by other travelers worldwide to take their advice or recommendations before planning their trips to certain destinations where they are not familiar with. Tourism destinations are being forced to adopt new strategies for drawing visitors in order to remain competitive due to the increased competition among tourism destinations as well as changes in the expectations and habits of tourists [4]. UGC became a data source to analyze travelers perceptions on tourist destinations. Travelers' opinions and feedback are collected through the mining of online text information and the use of sentiment analysis techniques to identify positive or negative emotions expressed in the text. This information is then used to improve the quality of the destination and ensure a balance

between social, economic, and environmental issues as well as the achievement of sustainable development [5].

Sentiment analysis provides the ability to identify the sentiments underlying reviews and classify them accordingly. [6]. Existing sentiment analysis methods in various domains use machine learning algorithms such as Naïve Bayes (NB), Support Vector Machine (SVM), maximum entropy classifier (Max. Ent.), random forest (RF) and Decision Tree (J48) [7,8]. Deep learning approaches have been more popular in recent years due to their ability to increase the accuracy of data classification, especially when thousands of examples of labeled data are present [9]. This research focuses on analyzing online travel reviews from TripAdvisor and Google Map review websites using deep learning techniques to investigate the travelers' perceptions on Langkawi, Kedah post-COVID19 and factors that influence the positive and negative feelings.

PROBLEM BACKGROUND

Travelers reviews and feedback have influence on travelers' behavior and perceptions, hence it is crucial for the tourism industry to improve services in order to attract more global tourists and meet their changing expectations. Text reviews on the internet are analyzed using sentiment analysis techniques to find positive and negative feelings that passengers have expressed. Deep learning techniques have become more and more popular because of their ability to understand the context of documents and increased accuracy in data classification. The goal of this study is to better comprehend tourists' opinions on Langkawi, Kedah, by conducting topic-based analysis on traveler reviews through the use of deep learning techniques.

RESEARCH QUESTIONS

1. What are the key topics from user reviews on the tourism industry in Langkawi, Kedah?
2. How do LSTM and BERT models perform in sentiment analysis when applied to identified topics from user reviews on Langkawi, Kedah?
3. What are the performance of different deep learning techniques for sentiment analysis?

OBJECTIVES

1. To determine the key topics from user reviews on the tourism industry in Langkawi, Kedah.
2. To develop the LSTM algorithm and fine-tune BERT for sentiment analysis on chosen topics.
3. To evaluate the performance of LSTM and BERT in sentiment analysis.

SCOPE OF THE STUDY

The goal of this research is to analyze online reviews in Langkawi, Kedah using deep learning techniques. This research's scope is as follows:

1. Online review sentiment analysis specifically in Langkawi, Kedah.
2. Web scraping from TripAdvisor and Google Map website limited to English language.
3. Employing deep learning methods for sentiment analysis to classify travelers sentiments.
4. Performance evaluation: Compare the performance of LSTM and BERT in sentiment analysis.

ASSUMPTIONS

It is assumed that there is a significant volume of user feedback about the tourism industry in Langkawi, Kedah, that can be analyzed. The reviews are assumed to reflect the general feelings of tourists visiting Langkawi, Kedah. It is anticipated that the sentiment labels used for training and evaluation (such as positive, negative, and neutral) are assigned correctly and consistently. The sentiment labels are assumed to accurately reflect the sentiment indicated in the user reviews.

SIGNIFICANCE OF THE RESEARCH

This project will provide valuable insights on the key topics and sentiment expressions from tourists traveling to Langkawi, Kedah post pandemic. This enables stakeholders, local tourism authorities, and businesses to get a better understanding of tourist experiences and preferences. By evaluating the performance of LSTM and BERT models in sentiment analysis,

this research will add to the current body of knowledge about their strengths and shortcomings, directing future research and applications in the field.

STRUCTURE OF THE THESIS

This thesis consists of four chapters explaining the working process of the project. In the first chapter, the background problem related to tourism sentiment analysis is discussed, and this project's objectives and scope of work are stated.

The second chapter focuses on the literature review done to gather beneficial information as guidance and get an overall overview in order to make this project successful. It discusses previously published papers that focus on tourism sentiment analysis, topic-based sentiment analysis and deep learning approaches for tourism sentiment classification.

The third chapter goes over the proposed research methodology for this project. It covers the proposed research framework for this research and the fourth chapter discusses the initial findings of this project.

Reference

1. Pestana, M. H., Parreira, A., & Moutinho, L. (2020). Motivations, emotions and satisfaction: The keys to a tourism destination choice. Journal of Destination Marketing & Management, 16, 100332.
2. Carvache-Franco, M., Segarra-Oña, M., & Carrascosa-López, C. (2019). Segmentation and motivations in eco-tourism: The case of a coastal national park. Ocean & Coastal Management, 178, 104812.
3. Kanwal, B., Rehman, S. U., Imran, A., Shaukat, R. S., Li, J., Alzahrani, A., ... & Alarfaj, F. K. (2023). Opinion Mining from Online Travel Reviews: An Exploratory Investigation on Pakistan Major Online Travel Services Using Natural Language Processing. IEEE Access, 11, 29934-29945.
4. Pavlić, I., Puh, B., & Mišković, L. (2020). The perception of travellers and the World Heritage Site image. Tourism: an international interdisciplinary journal, 68(2), 181-194.
5. Borrajo-Millán, F., Alonso-Almeida, M. D. M., Escat-Cortes, M., & Yi, L. (2021). Sentiment analysis to measure quality and build sustainability in tourism destinations. Sustainability, 13(11), 6015.
6. Haris, N. A. K. M., Mutalib, S., Ab Malik, A. M., Abdul-Rahman, S., & Kamarudin, S. N. K. (2023). Sentiment classification from reviews for tourism analytics. International Journal of Advances in Intelligent Informatics, 9(1), 108-120.
7. Nabiha, A., Mutalib, S., & Ab Malik, A. M. (2021, September). Sentiment analysis for informal malay text in social commerce. In 2021 2nd International Conference on Artificial Intelligence and Data Sciences (AiDAS) (pp. 1-6). IEEE.
8. Saleena, N. (2018). An ensemble classification system for twitter sentiment analysis. Procedia computer science, 132, 937-946.
9. Cao, Z., Xu, H., & Teo, B. S. X. (2023). Sentiment of chinese tourists towards malaysia cultural heritage based on online travel reviews. Sustainability, 15(4), 3478.
10. Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., ... & Danforth, C. M. (2015). Human language reveals a universal positivity bias. Proceedings of the national academy of sciences, 112(8), 2389-2394.
11. Díaz-Pacheco, Á., Guerrero-Rodríguez, R., Álvarez-Carmona, M. Á., Rodríguez-González, A. Y., & Aranda, R. (2023). A comprehensive deep learning approach for

topic discovering and sentiment analysis of textual information in tourism. Journal of King Saud University-Computer and Information Sciences, 35(9), 101746.

12. Irawan, H., Akmalia, G., & Masrury, R. A. (2019, September). Mining tourist's perception toward Indonesia tourism destination using sentiment analysis and topic modelling. In Proceedings of the 2019 4th International Conference on Cloud Computing and Internet of Things (pp. 7-12).
13. Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. Journal of the American Medical Informatics Association, 18(5), 544-551.
14. Kim, J. M., & Hyun, S. (2021). Differences in online reviews caused by distribution channels. Tourism Management, 83, 104230.
15. Arefeva, V., & Egger, R. (2022). When BERT Started Traveling: TourBERT—A Natural Language Processing Model for the Travel Industry. Digital, 2(4), 546-559.
16. Ounacer, S., Mhamdi, D., Ardchir, S., Daif, A., & Azzouazi, M. (2023). Customer sentiment analysis in hotel reviews through natural language processing techniques. International Journal of Advanced Computer Science and Applications, 14(1), 569-579.
17. Baqer, N. H., Sadiq, A. T., & Ali, Z. H. (2023). Enhancement of Sentiment Analysis in Hotel Reviews through Latent Semantic Indexing and Convolutional Neural Networks. Ingénierie des Systèmes d'Information, 28(6).
18. Ameur, A., Hamdi, S., & Ben Yahia, S. (2023). Sentiment analysis for hotel reviews: a systematic literature review. ACM Computing Surveys, 56(2), 1-38.
19. ALRASHIDI, S. M., ALANAZI, F. N., ALBALAWI, H. A., ALBALAWI, O. M., & AWADELKARIM, A. M. (2022). Machine Learning-Based Sentiment Analysis for Tweets Saudi Tourism. Journal of Theoretical and Applied Information Technology, 100(16), 5096-5109.
20. Viñán-Ludeña, M. S., & de Campos, L. M. (2022). Discovering a tourism destination with social media data: BERT-based sentiment analysis. Journal of Hospitality and Tourism Technology, 13(5), 907-921.
21. Mehra, P. (2023). Unexpected surprise: Emotion analysis and aspect based sentiment analysis (ABSA) of user generated comments to study behavioral intentions of tourists. Tourism Management Perspectives, 45, 101063.

22. Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2022). A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. IEEE Transactions on Knowledge and Data Engineering, 35(11), 11019-11038.
23. Jim, J. R., Talukder, M. A. R., Malakar, P., Kabir, M. M., Nur, K., & Mridha, M. F. (2024). Recent advancements and challenges of nlp-based sentiment analysis: A state-of-the-art review. Natural Language Processing Journal, 100059.
24. Ali, T., Omar, B., & Soulaimane, K. (2022). Analyzing tourism reviews using an LDA topic-based sentiment analysis approach. MethodsX, 9, 101894.
25. Abuzayed, A., & Al-Khalifa, H. (2021). BERT for Arabic topic modeling: An experimental study on BERTopic technique. Procedia computer science, 189, 191-194.
26. Gethsiya Raagel, K., Bagavandas, M., Sathya Narayana Sharma, K., Manikandan, P., & Muthu, C. (2023). Sentiment Analysis and Topic Modeling on Polycystic Ovary Syndrome from Online Forum Using Deep Learning Approach. Wireless Personal Communications, 133(2), 869-888.
27. Krishnan, A. (2023). Exploring the power of topic modeling techniques in analyzing customer reviews: a comparative analysis. arXiv preprint arXiv:2308.11520.
28. Gokcimen, T., & Das, B. (2024). Exploring Climate Change Discourse on Social Media and Blogs Using a Topic Modeling Analysis. Heliyon.
29. Kherwa, P., & Bansal, P. (2019). Topic modeling: a comprehensive review. EAI Endorsed transactions on scalable information systems, 7(24).
30. Du, R., Kuang, D., Drake, B., & Park, H. (2017). DC-NMF: nonnegative matrix factorization based on divide-and-conquer for fast clustering and topic modeling. Journal of Global Optimization, 68, 777-798.
31. O'callaghan, D., Greene, D., Carthy, J., & Cunningham, P. (2015). An analysis of the coherence of descriptors in topic modeling. Expert Systems with Applications, 42(13), 5645-5657.
32. Suh, S., Choo, J., Lee, J., & Reddy, C. K. (2016, December). L-ensnmf: Boosted local topic discovery via ensemble of nonnegative matrix factorization. In 2016 IEEE 16th International Conference on Data Mining (ICDM) (pp. 479-488). IEEE.

33. Syamala, M., & Nalini, N. J. (2019, July). LDA and deep learning: a combined approach for feature extraction and sentiment analysis. In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.
34. Shafqat, W., & Byun, Y. C. (2019). A recommendation mechanism for under-emphasized tourist spots using topic modeling and sentiment analysis. Sustainability, 12(1), 320.
35. Mishra, R. K., Urolagin, S., Jothi, J. A. A., Neogi, A. S., & Nawaz, N. (2021). Deep learning-based sentiment analysis and topic modeling on tourism during Covid-19 pandemic. Frontiers in Computer Science, 3, 775368.
36. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
37. Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.
38. McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426.
39. McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. J. Open Source Softw., 2(11), 205.
40. Kastrati, Z., Imran, A. S., Daudpota, S. M., Memon, M. A., & Kastrati, M. (2023). Soaring energy prices: Understanding public engagement on Twitter using sentiment analysis and topic modeling with transformers. IEEE Access, 11, 26541-26553.
41. Souza, C. N., Martínez-Arribas, J., Correia, R. A., Almeida, J. A., Ladle, R., Vaz, A. S., & Malhado, A. C. (2024). Using social media and machine learning to understand sentiments towards Brazilian National Parks. Biological Conservation, 293, 110557.
42. Bhuvaneswari, A., & Kumudha, M. (2024, April). Topic Modeling Based Clustering of Disaster Tweets Using BERTopic. In 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon) (pp. 1-6). IEEE.
43. Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.
44. Kennedy, A., & Inkpen, D. (2006). Sentiment classification of movie reviews using contextual valence shifters. Computational intelligence, 22(2), 110-125.
45. Ohana, B., & Tierney, B. (2009). Sentiment classification of reviews using SentiWordNet.

46. Alessia, D., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, tools and applications for sentiment analysis implementation. International Journal of Computer Applications, 125(3).
47. Fu, X., Yang, J., Li, J., Fang, M., & Wang, H. (2018). Lexicon-enhanced LSTM with attention for general sentiment analysis. IEEE Access, 6, 71884-71891.
48. Mercha, E. M., & Benbrahim, H. (2023). Machine learning and deep learning for sentiment analysis across languages: A survey. Neurocomputing, 531, 195-216.
49. Kumawat, S., Yadav, I., Pahal, N., & Goel, D. (2021, January). Sentiment analysis using language models: A study. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 984-988). IEEE.
50. Alharbi, B. A., Mezher, M. A., & Barakeh, A. M. (2022). Tourist reviews sentiment classification using deep learning techniques: A case study in saudi arabia. International Journal of Advanced Computer Science and Applications, 13(6).
51. Wang, Q., Sun, L., & Chen, Z. (2019, June). Sentiment analysis of reviews based on deep learning model. In 2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS) (pp. 258-261). IEEE.