invtur2024 conference

Proceedings of the INVTUR Conference 2024
8-10 May 2024, University of Aveiro, Portugal
Pages 191 – 194

Generative AI as Sentiment Analysis Tool in Hospitality

KATERINA VOLCHEK[1], STANISLAV IVANOV[1] & CELSO BRITO[1]

[1]Deggendorf Institute of Technology

Contacting author: katerina.volchek@th-deg.de

Keywords | GenAI, ChatGPT, Research Methods, Tourism, Hospitality

Objectives | Customer reviews on social media and dedicated websites are an essential source of information for hospitality companies that allows them to learn what customers think about their services and those of their competitors (Olorunsola et al., 2023; Perez-Aranda et al., 2021; Veloso & Gomez-Suarez, 2023). By analysing online reviews, hospitality managers can get insights into the opinions of customers and what they liked or disliked about the services they used. Moreover, online reviews are a helpful source of information about changing trends in customer perceptions and preferences. The evaluation of customer reviews can be implemented manually or with the help of specialised software (Tetzlaff et al., 2019). Sentiment analysis is a process of analysing language to interpret subjective evaluations, emotions and points of view (Taboada, 2016). Humans associate their opinions and emotions with specific linguistic structures in their daily lives. Sentiment analysis, therefore, addresses three tasks, including establishing whether the content represents a fact or a subjective opinion, determining its polarity (i.e. positive vs negative), and analysing the degree of polarity/ intensity of a sentiment (Cambria et al., 2017). The importance of sentiment analysis for research and practice motivated the proliferation of multiple analytical methods and tools, including the automation of the analysis with Artificial Intelligence (AI). The advantage of manual sentiment analysis is its high achievable validity and reliability. However, special linguistic skills, substantial time and cross-validation are required to prevent subjectivity bias (Sotiriadou et al., 2014). A range of tools that utilise supervised learning are trained for automatic differentiation between negative, positive and neutral (if required) emotions. The advantage of supervised learning algorithms is the achievable speed of analysis alongside the relatively high accuracy of the results. The limitation of the machine learning algorithms for sentiment analysis is the dependence of the analysis validity on the context. A new training might be required for a new content source (Taboada, 2016). Recently, ChatGPT introduced a plugin for sentiment analysis (OpenAI.com, 2023). It uses a class of machine learning, called Natural Language Processing (NLP) that is trained to understand a language autonomously. Due to its cost, speed and intuitive interface, a range of studies has already applied ChatGPT for sentiment analysis of user-generated content (e.g. Adeshola & Adepoju, 2023; Chumakov et al., 2023). The launch of ChatGPT in November 2022 opened a new opportunity for the analysis of texts that do not require significant digital skills necessary to use effectively and efficiently other software packages for textual analysis. However, its performance in comparison to other methods remains largely underexplored (Fatouros et al., 2023). This ongoing study aims to evaluate Generative AI (GenAI) as a methodological tool for sentiment analysis in the context of Tourism & Hospitality. It tests the effectiveness of ChatGPT vs NVivo auto coding vs manual analysis by comparing the validity and reliability of the results.

Methodology | The study aims to analyse the effectiveness of GenAI as a tool for a sentiment analysis. A hotel review from Booking.com were users as a content for the sentiment analysis to ensure that the sentiment is derived from subjective tourist opinions. The study selected 3 hotels in various geographical locations, but with similar type of travel (mainly, a short stay by business tourists) and with similar services robots employed by those hotels. All guest reviews related to the service robots, submitted in English, were extracted regardless of the review date of submission. The preliminary analysis included the comparison of the sentiment analysis of the 160 reviews. First, the codes that characterise hotel service robots, were manually extracted to ensure data validity. The polarity of the sentiments as well as its degree was determined with three diverse methods. First, manual coding was replicated by two trained researchers. Then NVivo 20 auto coding and ChatGPT with a prompt *"Run a sentiment analysis for each row of the table. The sentiment would range from "very positive", "moderately positive", "neutral", "moderately negative", to "very negative".",* were replicated three times. The analysis of GenAI as a methodological tool consisted of the comparison of the sentiment analysis, done by ChatGPT 3.5 vs NVivo vs manual coding. The number of errors in determining the polarity of the sentiment and its degree, were calculated for both types of the automated analysis and compared with the results of the manual coding, which have been used as a baseline.

Main results and contributions | The preliminary findings demonstrate that ChatGPT performs better than NVivo auto coding but is worse than the manual analysis. In comparison to manual coding, ChatGPT provided identical results in determining the polarity of a sentiment (i.e. negative vs neutral vs positive). At the same time, 12% of the codes, generated by NVivo, contained a wrongly identified sentiment. Regarding the degree of polarity (i.e. very negative vs moderately negative), ChatGPT generated a total of 24% of erroneous codes. NVivo-generated results were comparable with 28% of wrongly attributed degree of polarity in comparison to the manual coding.

Limitations | The key limitations of the study are related to the validity of each method and their comparability. Thus, the manual sentiment analysis required cross-validation, the supervised learning models do range in terms of their relevance for the context, and GenAI represents a "black box" with the model being unavailable for researchers' evaluation. Therefore, more studies in various contexts and with the application of different sentiment analysis methods are required before claiming GenAI as a reliable tool for the analysis.

Conclusions | The preliminary findings indicate that ChatGPT 3.5 performs substantially better than NVivo auto coding. However, manual evaluation of the sentiment still provides more valid results. While research automation can substantially decrease the time and the efforts of the researchers, a broader scale of research is required to understand how to ensure the validity and reliability of the GenAI tools.

References

Adeshola, I., & Adepoju, A. P. (2023). The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments*, 1-14. https://doi.org/10.1080/10494820.2023.2253858

Cambria, E., Das, D., Bandyopadhyay, S., & Feraco, A. (2017). Affective computing and sentiment analysis. *A practical guide to sentiment analysis*. In *Socio-affective computing* (pp. 1–10). https://doi.org/10.1007/978-3-319-55394-8_1

Chumakov, S., Kovantsev, A., & Surikov, A. (2023). Generative approach to Aspect Based Sentiment Analysis with GPT Language Models. *Procedia Computer Science*, *229*, 284-293. https://doi.org/10.1016/j.procs.2023.12.030

Fatouros, G., Soldatos, J., Kouroumali, K., Makridis, G., & Kyriazis, D. (2023). Transforming sentiment analysis in the financial domain with ChatGPT. *Machine Learning with Applications*, *14*, 100508. https://doi.org/10.1016/j.mlwa.2023.100508

Olorunsola, V. O., Saydam, M. B., Lasisi, T. T., & Eluwole, K. K. (2023). Customer experience management in capsule hotels: a content analysis of guest online review. *Journal of Hospitality and Tourism Insights*, *6*(5), 2462–2483. https://doi.org/10.1108/jhti-03-2022-0113

OpenAI.com. (2023). *Sentiment Analysis - New plugin for ChatGPT*. https://community.openai.com/t/sentiment-analysis-new-plugin-for-chatgpt/322335

Perez-Aranda, J., Guerreiro, M., & Mendes, J. (2021). Drivers of eWOM intensity: differences between hoteliers' perception and real reviews. *European Journal of Tourism Research*, *27*, 2708. https://doi.org/10.54055/ejtr.v27i.2138

Sotiriadou, P., Brouwers, J., & Le, T. (2014). Choosing a qualitative data analysis tool: A comparison of NVivo and Leximancer. *Annals of Leisure Research*, *17*(2), 218-234. https://doi.org/10.1080/11745398.2014.902292

Taboada, M. (2016). Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics*, *2*(1), 325-347. https://doi.org/10.1146/annurev-linguistics-011415-040518

Tetzlaff, L., Rulle, K., Szepannek, G., & Gronau, W. (2019). A customer feedback sentiment dictionary: Towards automatic assessment of online reviews. *European Journal of Tourism Research*, *23*, 28-39. https://doi.org/10.54055/ejtr.v23i.387

Veloso, M., & Gomez-Suarez, M. (2023). Customer experience in the hotel industry: a systematic literature review and research agenda. *International Journal of Contemporary Hospitality Management*, *35*(8), 3006–3028. https://doi.org/10.1108/IJCHM-04-2022-0517