

Predicting sentiment and rating of tourist reviews using machine learning

Analyzing
tourist reviews
using machine
learning

Karlo Puh and Marina Bagić Babac

*Faculty of Electrical Engineering and Computing, University of Zagreb,
Zagreb, Croatia*

Received 21 February 2022
Revised 16 April 2022
12 June 2022
13 June 2022
16 June 2022
Accepted 16 June 2022

Abstract

Purpose – As the tourism industry becomes more vital for the success of many economies around the world, the importance of technology in tourism grows daily. Alongside increasing tourism importance and popularity, the amount of significant data grows, too. On daily basis, millions of people write their opinions, suggestions and views about accommodation, services, and much more on various websites. Well-processed and filtered data can provide a lot of useful information that can be used for making tourists' experiences much better and help us decide when selecting a hotel or a restaurant. Thus, the purpose of this study is to explore machine and deep learning models for predicting sentiment and rating from tourist reviews.

Design/methodology/approach – This paper used machine learning models such as Naïve Bayes, support vector machines (SVM), convolutional neural network (CNN), long short-term memory (LSTM) and bidirectional long short-term memory (BiLSTM) for extracting sentiment and ratings from tourist reviews. These models were trained to classify reviews into positive, negative, or neutral sentiment, and into one to five grades or stars. Data used for training the models were gathered from TripAdvisor, the world's largest travel platform. The models based on multinomial Naïve Bayes (MNB) and SVM were trained using the term frequency-inverse document frequency (TF-IDF) for word representations while deep learning models were trained using global vectors (GloVe) for word representation. The results from testing these models are presented, compared and discussed.

Findings – The performance of machine and learning models achieved high accuracy in predicting positive, negative, or neutral sentiments and ratings from tourist reviews. The optimal model architecture for both classification tasks was a deep learning model based on BiLSTM. The study's results confirmed that deep learning models are more efficient and accurate than machine learning algorithms.

Practical implications – The proposed models allow for forecasting the number of tourist arrivals and expenditure, gaining insights into the tourists' profiles, improving overall customer experience, and upgrading marketing strategies. Different service sectors can use the implemented models to get insights into customer satisfaction with the products and services as well as to predict the opinions given a particular context.

Originality/value – This study developed and compared different machine learning models for classifying customer reviews as positive, negative, or neutral, as well as predicting ratings with one to five stars based on a TripAdvisor hotel reviews dataset that contains 20,491 unique hotel reviews.

Keywords Sentiment analysis, Machine learning, Deep learning, Customer reviews, Tourism

Paper type Research paper

Introduction

Customer experience and opinion are crucial for the enhancement of the tourism industry. Therefore, this industry has already largely adapted to information and communication technologies and the advent of big data (Madyatmadja *et al.*, 2021). Currently, many tourist services are available online such as booking websites (Manosso and Domareski Ruiz, 2021).

© Karlo Puh and Marina Bagić Babac. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>



Journal of Hospitality and Tourism
Insights
Emerald Publishing Limited
2514-9792
DOI 10.1108/JHTI-02-2022-0078

Since tourists use a lot of websites and social media to leave their personal opinions or comments on a specific place or service, customer reviews have become a significant factor when deciding which possible hotels or restaurants to visit (Neidhardt *et al.*, 2017). For example, a number of all reviews on TripAdvisor overpassed a total of 884 million in 2020 (Statista, 2020). Information obtained from these reviews is important to other tourists but also to service providers who can then note key aspects that make their hotel/restaurant good or bad (Sumarsono *et al.*, 2018).

In parallel with the huge increase in the number of online user reviews, there is also a growing need for automated processing of these huge amounts of data because it is impossible for humans to read and analyze all these reviews on their own (Gour *et al.*, 2021). Sentiment analysis is a technique used by natural language processing to identify and extract information in data (Collobert and Weston, 2008). In most cases, it means to determine whether a review expresses *positive*, or *negative* sentiment (Barbierato *et al.*, 2021). Although there is much research in sentiment analysis of tourist reviews over the past decade, most of the research is limited to *positive/negative* classification. Fewer studies include *neutral* review sentiment in addition to *positive/negative* (Wadhe and Suratkar, 2020), which is a more demanding task and is included in this study. Adding *neutral* in sentiment classification is important because it gives us additional useful information. A neutral comment is usually an indicator of concern since the customer can easily turn positive or negative. Thus, taking into consideration neutral comments can help one to increase the number of satisfied customers since it is easier to turn a neutral experience into a positive than a negative one. Moreover, even fewer studies include rating classification and prediction based on tourist reviews (Harrag *et al.*, 2019), which are also analyzed in this study. Performing rating prediction is useful when one has a lot of customers' comments and wants to process them fast. That way we can easily visualize data (customer satisfaction) and quickly see if drastic changes need to be made. Furthermore, it enables us to sort comments based on their importance. It makes sense to first act on comments rated with the lowest score and make our way up. Predicting specific ratings rather than sentiment is useful when we want to get more detailed information on some factors, which make a customer's rating great or poor. Then, we can find common features that make customers' experience poor and improve them in the future but also see what is being done well. In addition, the studies that test the performance of sentiment analysis are rare in the tourism and hospitality domain (Mehraliyev *et al.*, 2022), thus our study also contributes to filling this gap.

In this paper, we have analyzed sentiment and ratings of a specific place or service expressed in customer reviews on TripAdvisor to predict tourist satisfaction. We have conducted sentiment and rating classification using different methods ranging from machine learning algorithms like Naïve Bayes and support vector machines (SVMs) to deep learning methods. Experimental results have shown that deep learning methods based on bidirectional long short-term memory (BiLSTM) outperformed other implemented methods. Based on results from his study, tourist service providers can easily and quickly process a lot of data and get very accurate customer feedback, since user-generated content is regarded as the most influential content in the tourism industry. There are other noteworthy benefits of sentiment analysis like shaping company marketing strategies, classification of textual data and providing overall better service.

Literature review

Sentiment analysis has been performed with a variety of techniques over the last decade including lexicon-based and machine-learning and deep-learning-based techniques (Jurafsky and Martin, 2000).

Lexicon-based sentiment analysis

For lexicon-based sentiment analysis, a sentiment relates to its semantic value and the intensity of each word in the sentence, which requires a pre-defined lexicon to classify positive and negative words (Bagić Babac and Podobnik, 2016). Generally, a text item is treated as a bag of words, and after scoring each word, the sentiment is obtained by a certain pooling operation such as taking an average of individual word scores.

Today many of these lexicon-based approaches are automated, such as using the TextBlob (Loria, 2018), a Python library for natural language processing (NLP). Larasati *et al.* (2020) used TextBlob to obtain sentiment analysis scores from eight tourist websites, which confirmed most of the visitors' sentiments were positive. In addition, a lexicon-based approach has been used to evaluate consumers' sentiment toward several well-known technological brands (Mostafa, 2013), and sentiment analysis confirmed a generally positive consumer sentiment. Tan and Wu (2011) utilized a lexical database for extracting hotel reviews from Ctrip based on the random walk algorithm for the automated generation of a specific-domain sentiment lexicon. Serna *et al.* (2016) made use of the WordNet lexical database to obtain emotions from Twitter mentioning two holiday periods. In addition, Kang *et al.* (2012) proposed a replacement senti-lexicon for the sentiment analysis of building reviews based on an improved Naïve Bayes algorithm.

It should be noted that most of the lexicon-based approaches are built upon, so-called, general-purpose lexicons (Avdić and Bagić Babac, 2021). Bagherzadeh *et al.* (2021) developed two specific lexicons, namely weighted and manually selected lexicons, which were tested and validated by applying classification accuracy metrics to the TripAdvisor data. Their approach outperformed a SentiWords lexicon-based method and a Naïve Bayes machine-learning algorithm in classifying sentiment.

Machine learning approach to sentiment analysis

In the supervised machine learning approach to sentiment analysis in tourism, a variety of classifiers were used (Waghmare and Bhala, 2020). One of the techniques, called *Naïve Bayes*, was used in research on sentiment analysis on hotel reviews using a *Multinomial Naïve Bayes* (NBM) classifier (Farisi *et al.*, 2018). In that study, the authors provided a solution for classifying customer reviews as positive or negative using a NBM classifier using a bag of words to extract features after data preprocessing, which resulted in an average F1 score of more than 91% in experimental results. Likewise, Afzaal *et al.* (2019) have shown high accuracy of NBM, that is NBM correctly classified 88.08% of the aspects of the restaurants' reviews dataset and achieved 90.53% accuracy in the hotels' reviews dataset.

Another well-known technique called *Support Vector Machine* (SVM) was used in research on sentiment analysis model for hotel reviews based on supervised learning (Shi and Li, 2011). That paper discusses sentiment analysis using SVM with the *Term Frequency-Inverse Document Frequency* (TF-IDF) and a bag of words. After conducting the experiments, the results showed that TF-IDF was more effective. TF-IDF resulted in 87.2% and a bag of words 86.4% on the F1 score. In addition, Prameswari *et al.* (2017) used a similar approach combining TF-IDF with SVM for sentiment analysis of hotel reviews and achieved an accuracy of 78%.

Deep learning approach to sentiment analysis

Although previous techniques have given satisfying results, in recent years deep learning is getting more and more used for sentiment analysis and natural language processing tasks (Faralli *et al.*, 2021). A paper on *bidirectional recursive neural networks* (RNNs) for token-level labeling with structure (Irsoy and Cardie, 2013) proposed an extension to RNN to carry out labeling tasks at the token level that improves sentiment analysis accuracy. Ramadhani *et al.* (2021) used long short-term memory (LSTM) architecture to classify tourist reviews and

achieved the best accuracy result of 84%. Furthermore, [Baziotis et al. \(2017\)](#) presented LSTM based model augmented with an attention mechanism. Using that model, they ranked very high at SemiEval-2017 Task 4 “Sentiment Analysis in Twitter”. [Xu et al. \(2019\)](#) proposed a method based on BiLSTM and compared it with other sentiment analysis methods like convolution neural network (CNN), RNN, LSTM and Naïve Bayes. The conclusion of the experiments was that the proposed BiLSTM gave better results on F1 score, recall, and higher accuracy.

In addition to memory-based neural networks, CNNs have also shown satisfactory results in sentiment analysis. Based on a dataset of travel destination reviews, [Huang \(2021\)](#) implemented a sentiment classification model based on a CNN and compared it with several other machine learning models, and the CNN model had the highest accuracy of sentiment classification, reaching 91.6%,

Rating prediction from tourist reviews

While there are many studies of sentiment analysis and rating prediction in the various domains of interest ([Harrag et al., 2019](#)), fewer studies provide a framework for analyzing and predicting ratings from tourist reviews based on machine and deep learning.

[Leal Gonzalez-Velez Malheiro et al. \(2017\)](#) used multiple linear regression to calculate an overall rating to estimate the remaining ratings (feature variables) for HotelExpedia and TripAdvisor datasets. While evaluating the performance of five machine learning algorithms, namely Decision Trees, SVMs, Neural Networks, Random Forest and Naïve Bayes algorithms for predicting Google user review rating on the travel experience, [Hossain and Das \(2020\)](#) showed that SVMs provided better results than other algorithms. In addition, [Leal et al. \(2018\)](#) suggested that the rating prediction can be further advanced by using online processing and post-filtering to improve accuracy in online recommendations.

Overall, it can be concluded that one of the fruitful ways to conduct sentiment and rating analysis and prediction is using natural language processing and machine learning ([Abadi et al., 2016](#)). The main goal of this paper is to develop, implement and test machine learning models using data from TripAdvisor, which are capable of classifying customer reviews as positive, negative, or neutral, and predicting customer review ratings with one to five stars.

Research methodology

Data preprocessing

Preprocessing is one of the most important steps when performing any NLP task ([Bagić Babac and Podobnik, 2016](#)). Basically, preprocessing means bringing the text into a clean form and making it ready to be fed into the model. When it comes to data preprocessing, there are many useful techniques. Specifically in this paper, tokenization is the first step in preprocessing. Tokenization means splitting a sentence into a list of words. After tokenization, removing stop words comes as the next step. Stop words are words that are commonly used in any language. If we take for example English, stop words are words such as “is”, “the”, “and”, “a”, etc. Those words are considered unimportant in natural language processing, so they are being removed. Next comes the process of transforming a word into its root or lemma called lemmatization. An example of that would be “swimming” to “swim”, “was” to “be” and “mice” to “mouse”. Considering that machines treat the lower and upper case differently, all the words will be lower-cased for better interpretation. Finally, all punctuation is being removed which contributes to noise reduction and getting rid of useless data.

To perform preprocessing tasks, *spaCy* was used, an open-source library for advanced natural language processing in Python. It is multilingual, but for this project, only English

was needed. After loading data for the English language, *spaCy* enables us to perform tokenization, lemmatization and stop word removal quite easily. Examples of using *spaCy* and the explained techniques are shown in Table 1.

Word representation

Since computers do not understand words or their context, it is needed to convert text into the appropriate, machine-interpretable form. Word embeddings are mathematical representations of words that give similar representations to words that have a similar meaning (Mikolov *et al.*, 2013). In other words, those representations model the semantic meaning of words. Specifically, those representations are vectors that are positioned in space in such a way that vectors closer to each other have more similar semantic meanings.

The word representation used in this research is called *global vectors* (GloVe) for word representation as introduced by Pennington *et al.* (2014). Since then, it gained popularity due to its good performance and simplicity. The GloVe is a log-bilinear model with a weighted least-squares objective trained on a global word-word co-occurrence matrix. That matrix shows words' co-occurrence frequency with one another in a given corpus. The main idea behind GloVe is that ratios of word-word co-occurrence probabilities encode meaning, as shown with an example in Table 2.

If we investigate the example shown in Table 2, we can see some actual probabilities from a six billion word, i.e. token corpus. The table shows how the word *ice* co-occurs more frequently with *solid*, but steam co-occurs more with *gas*. Furthermore, if we look at the word *water*, we can see that both *ice* and steam co-occur with it frequently because it is their shared property.

Another way used for representing words by vectors is *Term Frequency Inverse Document Frequency* (TF-IDF). It is commonly used in NLP tasks because it takes into consideration the relevance of a word in a document and scales it across all documents in a specific corpus. TF-IDF is calculated by multiplying two metrics, namely term frequency, and inverse document frequency (IDF). Term frequency (TF) is the number of times a specific word (term t) appears in a document (d) divided by the total number of words in a document as shown in Eq. (1) (Jurafsky and Martin, 2000).

Reviewed text	Preprocessed text
Rude people. Do not stay, despite the fact cool hotel, the place sucks, rudest people, are disappointed	"rude", "people", "stay", "despite", "fact", "cool", "hotel", "place", "suck", "rude", "people", "disappointed"
Great location, jr. suite is great, clean comfortable, close pike. Market in walking distance, breakfast nice and fresh	"great", "location", "jr", "suite", "great", "clean", "comfortable", "close", "pike", "market", "walking", "distance", "breakfast", "nice", "fresh"
Enjoyed the hotel. Location and service costs are excellent, good room. Recommend	"enjoy", "hotel", "location", "service", "cost", "excellent", "good", "room", "recommend"

Table 1.
Examples of data
preprocessing

Probability and ratio	$k = \text{solid}$	$k = \text{gas}$	$k = \text{water}$	$k = \text{fashion}$
$P(k \text{ice})$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}	1.7×10^{-5}
$P(k \text{steam})$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}	1.8×10^{-5}
$P(k \text{ice})/P(k \text{steam})$	8.9	8.5×10^{-2}	1.36	0.96

Table 2.
Example of co-
occurrence probability
ratios (Pennington
et al., 2014)

$$tf(t, d) = \frac{\text{count of } t \text{ in } d}{\text{number of words in } d} \quad (1)$$

Inverse document frequency (IDF) measures how important a word is in the whole corpus. For frequent words, IDF will be low. This value is calculated by dividing the total number of documents by the number of documents that contain a specific word (document frequency). For mathematical reasons (to avoid division by zero and value explosion) final formula looks as follows:

$$idf(t) = \log\left(\frac{N}{df + 1}\right) \quad (2)$$

Finally, to calculate TF-IDF for the specific term we multiply those two values.

$$tf - idf = tf(t, d) \cdot \log\left(\frac{N}{df + 1}\right) \quad (3)$$

The main difference between these two described vectorization methods is that TF-IDF is easier to use, but GloVe carries semantic meaning and can understand the context better.

Sentiment analysis using machine learning

For the purposes of this study, Naïve Bayes and SVMs were chosen as frequently used machine learning algorithms in data science (Poch Alonso and Bačić Babac, 2022).

Naïve Bayes is one of the most commonly used methods in natural language processing tasks. It is based on the Bayes theorem which calculates the probability of a specific event based on prior knowledge using the next equation:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (4)$$

where $P(c|x)$ is a posterior probability of a class, $P(c)$ is the prior probability of a class, $P(x)$ is the prior probability of the predictor, and $P(x|c)$ is the conditional probability that the predictor is a given class.

SVM is a machine learning algorithm that uses a hyperplane to separate different classes of data. A hyperplane is a subspace that is always one dimension less than its parent dimension. For example, if we were in a two-dimensional space then a hyperplane would be a line. The main goal of this algorithm is to find the hyperplane that has the largest distance (margin) between the hyperplane and the nearest data called support vectors. New data is being classified based on which side of the hyperplane they are located. Furthermore, the larger the margin is, the more confidence we have in determining data class.

Sentiment analysis using deep learning

When it comes to deep learning in natural language processing, the first thing that we think of is a recurrent neural network or RNN. The idea behind RNN is to be able to process arbitrary length data while keeping track of its order. Since that approach has some big flaws, for example not being able to capture long-distance semantic connections or vanishing gradient problems, another type of neural network was used. LSTM is a type of recurrent neural network that overcomes previously mentioned problems. To do that, LSTMs use four, instead of one neural network layer (Sherstinsky, 2020).

The reason why LSTMs work so well is their ability to add or remove information to the cell state. Structures called gates enable them that kind of behavior. Gates are different neural networks that consist of a sigmoid layer and a pointwise multiplication

operation. The core idea behind that is to forget or update data because the sigmoid layer squishes values between 0 and 1. That way the network can learn what data is relevant or irrelevant and decide to keep it or forget it. The first gate is called the forget gate and it decides which information to keep or discard. The step is demonstrated in Eq. (5), where h_{t-1} and x_t are the inputs to LSTM, W_f is the weight, and b_f is the bias (Hochreiter and Schmidhuber, 1997).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (5)$$

Next, we want to update the cell state. The second gate, called the input gate, also using the sigmoid layer decides which values to update. Afterward, we combine the result of the input gate with the tanh layer to create the update on the cell state (Hochreiter, 1998).

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (6)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (7)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (8)$$

Specifically, to update the cell state, we multiply the old cell state by the forget gate, then add it with the input gate multiplied with \tilde{C}_t . Described process is shown in Eq. (8). Finally, we have the output gate. Its job is to calculate the next hidden state. As Eq. (9) shows, we first pass the current and the previously hidden state through the sigmoid. Then, to get the output, we put the cell state through tanh and multiply it by the previously calculated sigmoid output. As a result of everything mentioned, we get the new hidden state shown in Eq. (10). In the end, the new hidden state and the cell state are carried over to the next cell (Hochreiter and Schmidhuber, 1997).

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (9)$$

$$h_t = o_t * \tanh(C_t) \quad (10)$$

Described LSTM model achieves much better results than traditional RNN (Sherstinsky, 2020) but there is still a place for an upgrade. We have seen that LSTM uses information from the past, meaning that the current state depends on the information before that moment. In order to have more contextual information in every moment, i.e. increase the amount of networks information, we use BiLSTM. BiLSTM consists of two LSTMs, each one of them going in a different direction. The first one goes forward (from past to the future) and the second one goes backward (from future to past). That kind of architecture enables us to understand the context much better.

Besides RNNs, CNNs have been commonly used for text classification and sentiment analysis tasks, although they are more known for working with images. The difference here is that one-dimensional (1D) convolution is being used instead of two-dimensional (2D) like with images as inputs. One of the biggest CNN's advantages is that they are translation invariant. It basically means that when some pattern is learned, CNN can recognize it later at any other different position. Just as 2D convolution, 1D convolution includes many kernels with weights that are learned through the training process. Those kernels are designed to generate an output by looking at the word and its surroundings. That way, since similar words have similar vector representations, convolution will produce a similar value. In practice, those convolutional layers are combined with pooling layers that discard less relevant information (Kuhn and Johnson, 2013).

Model architectures for machine learning

For conducting sentiment analysis, a few different methods and architectures were proposed. First, we implemented two machine learning algorithms, namely MNB and SVM. In these machine learning approaches, we used TF-IDF for word representations.

After that, we implemented deep learning models using the GloVe for word representations. Our first deep learning model is based on a 1D CNN, i.e. it consists of three 1D convolutional layers combined with dropout and max-pooling layers with three linear layers followed by *softmax* in the end. Described architecture is shown in [Figure 1](#).

Furthermore, we implemented a model architecture that consists of two stacked BiLSTMs followed by three linear layers with a *softmax* function at the end. The model's architecture is shown in [Figure 2](#). For this model, word representations are provided by GloVe, thus the word embeddings are used as the inputs of BiLSTM. After passing word embeddings through two BiLSTM layers and text feature extraction, vectors are used as inputs into three linear neural network layers with *ReLU* activation functions to perform text classification. Lastly, the output is passed through the *softmax* function to convert the numerical output into the range $[0, 1]$ representing the probabilities of each class.

In addition, another model has the same architecture as the one shown in [Figure 2](#), but we used normal LSTMs instead of BiLSTMs.

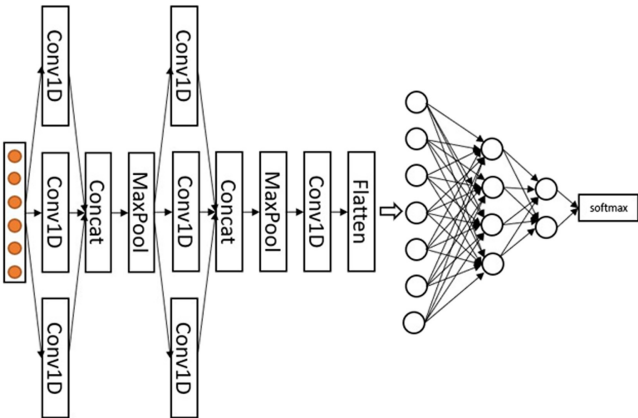


Figure 1.
Proposed CNN model
architecture

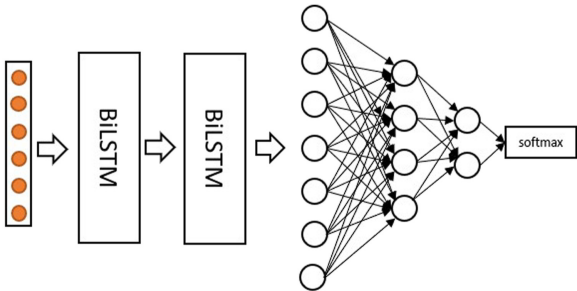


Figure 2.
BiLSTM model
architecture

Experimental results

For the purpose of training the models to achieve good performance in practice, the dataset has to be convincing (Cvitanović and Babić Babac, 2022). Having that in mind, data was extracted from TripAdvisor, the world's largest travel platform that today has over 860 million reviews and opinions (Alam *et al.*, 2016a, b). This study utilized a dataset called TripAdvisor Hotel Reviews that contains 20,491 unique hotel reviews graded from one to five stars by guests (Alam *et al.*, 2016a, b). For training purposes, the dataset was split into three parts, that is training, evaluation and testing subsets in the ratio of 70% for the training, 10% for evaluation and 20% for the testing subset (Kuhn and Johnson, 2013).

Since the used dataset consists of reviews and their scores which are grades from one to five, the machine learning models are first trained to predict the exact grade based on the review text. MNB algorithm resulted in 46% accuracy on the test data while SVM managed to outperform Naïve Bayes and achieve the accuracy of 55%.

After machine learning algorithms, deep learning models were trained and tested. Given that grid search is quite exhaustive and time-consuming, a random search was used in the process of setting hyperparameters for training (Vrigazova, 2021). Furthermore, an early stopping mechanism was used and the model with the smallest loss in the evaluation data was saved (Marrese-Taylor *et al.*, 2014). Also, to prevent the model from overfitting, we used the dropout mechanism. In addition, a technique to prevent exploding gradients problem called gradient clipping was used too. Through all training processes, the batch size of 16 examples was constant. Considering the problem of predicting the score using text review can be treated as a classification task, a loss function called categorical cross-entropy was implemented. Categorical cross-entropy is one of the most popular loss functions when it comes to multi-class classification (Neidhardt *et al.*, 2017). It is shown in Eq. (11), where \hat{y}_i is the i -th value in the model prediction and y_i is the true label value.

$$Loss = - \sum_{i=1}^5 y_i \cdot \log(\hat{y}_i) \quad (11)$$

Finally, an optimization algorithm for stochastic gradient descent called Adam was chosen for model training.

The highest accuracy that the 1D CNN managed to achieve after conducting a random search for setting hyperparameters was 62%. Furthermore, the stacked LSTM model performed expectedly better than CNN based model. The best model with LSTM architecture managed to achieve 66% accuracy. Finally, a stacked BiLSTM model outperformed other models by achieving 72% accuracy. Table 3 shows experimental results of the BiLSTM based model on test data for specific hyperparameters combinations. Figure 3 shows losses in evaluation and training data during the training of the best-performing model.

Finally, we can compare the experimental results of all the above classification methods. An overview of those results is shown in Table 4. The "Rating task" column summarizes the previously explained results, i.e. classifying reviews into five classes (or grades) from one to

Table 3.				Experimental results of the BiLSTM model for different hyperparameters: learning rate, dropout, and gradient clipping norm
Learning rate	Dropout	Clip norm	Accuracy	
0.001	0.33	0.33	0.69	
0.0001	0.33	0.33	0.49	
0.001	0.5	0.5	0.72	
0.001	0.65	0.65	0.71	

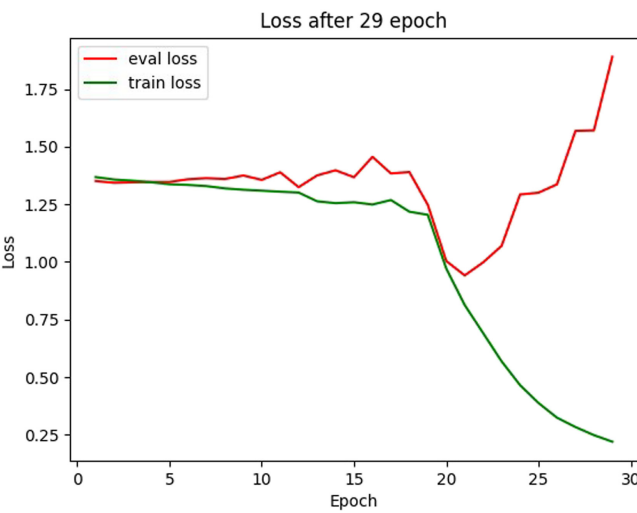


Figure 3. Visualization of loss during training on training and evaluation data for the best performing BiLSTM model

	Model	Rating task	Sentiment task
Table 4. The highest accuracies achieved by each model for the rating and sentiment tasks	Naïve Bayes	0.46	0.73
	SVM	0.55	0.80
	1D CNN	0.62	0.85
	LSTM	0.66	0.87
	BiLSTM	0.72	0.89

	Learning rate	Dropout	Clip norm	Accuracy
Table 5. Experimental results for different hyperparameters: learning rate, dropout, and gradient clipping norm	0.001	0.33	0.33	0.85
	0.001	0.45	0.45	0.89
	0.001	0.65	0.75	0.84
	0.001	0.55	0.55	0.86

five, and the “Sentiment task” column shows the results from classifying reviews into three classes representing positive, negative and neutral customer experience. During the process of calculating the scores, a review is considered positive if it has a score greater than 3, neutral if the score is 3, and negative if the score is less than 3.

Table 5 shows the results for different hyperparameters combinations of the BiLSTM model proposed for sentiment classification, while Figure 4 shows how evaluation and training loss behave during the training process of the model with the highest accuracy.

From the results presented in this section, it can be concluded that deep learning models delivered better overall performance than the existing classical machine learning approaches. It has been shown that by leveraging the BiLSTM-based model architecture with touristic opinion data, higher accuracy in predictions may be obtained. This model’s high accuracy and efficiency can help further improve the hotel or tourism industry in better understanding

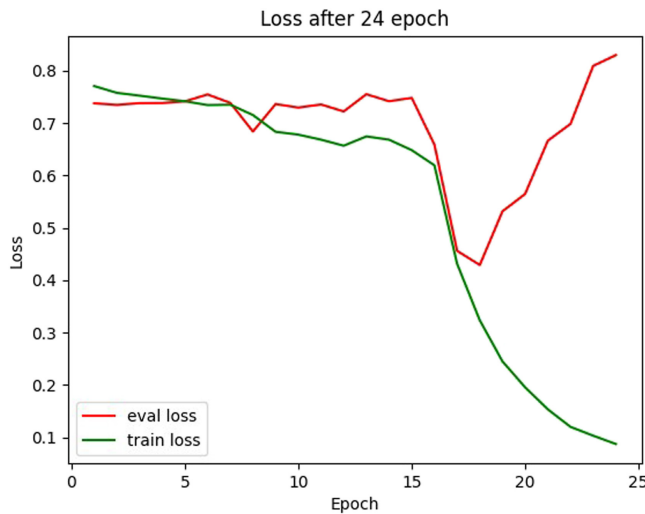


Figure 4.
Visualization of loss
during the training
process on training and
evaluation data for the
best performing model

the requirements and expectations of tourists, which benefits both customers and touristic organizations and businesses.

Although deep learning models outperform other machine learning models in these multi-class predicting tasks, it can be also noticed that the results from the sentiment task also seem satisfactory in certain settings, e.g. 80% for SVM given the fact that simpler pre-processing and less memory consumption were used. Thus, during the decision-making process in a particular setting or environment, one can balance between achieving higher efficiency and accuracy versus utilizing less computational resources. However, for a more complex task such as rating prediction, deep learning models provide significantly better accuracy compared to some other models that do not even provide adequate accuracy (e.g. Naïve Bayes with below 50% accuracy).

In the comparison of our results to the results of others (Gitto and Mancuso, 2017; Mehta *et al.*, 2021; Wang *et al.*, 2022), it can be noted that there are differences in the size, quality, and purpose of a particular dataset and different uses and implementations of sentiment analysis. In addition, there are various approaches to calculate whether the sentiment is positive, negative, or neutral, e.g. a study that explored the cruise experiences (Wang *et al.*, 2022) considered a comment as negative “if a comment’s positive score was less than or equal to two times the absolute value of its negative score”. Moreover, a recent survey on the sentiment analysis in hospitality and tourism (Mehraliyev *et al.*, 2022) has reported that the studies that test the performance of sentiment analysis are rare, thus our results contribute to filling this gap.

Discussion

Conclusions

Given the vast amount of data on people’s individual opinions, there is a need to develop and improve existing sentiment analysis tools. These tools not only serve the individuals as a recommender on how to optimize their choices of services to use, but also to decision-makers in improving the quality of their services. The long-term implications of the knowledge gained by these sentiment tools may influence tourism development and the engagement of tourist stakeholders. Our contribution in the form of proposed models can indicate a plausible

further direction for developing more robust and accurate models for sentiment and rating classification.

More specifically, this study provides an insight into how to apply machine and deep learning models for sentiment analysis on tourist reviews. It showed that the BiLSTM model outperformed in both sentiment and rating classification tasks. Specifically, in our BiLSTM model, data were first passed through two stacked BiLSTMs whose job was to gather contextual information followed by three linear layers that perform classification. Models were trained to classify reviews first into five and later into three classes with GloVe used for word representation. As result, the best performing model for five classes achieved 72% accuracy while the best model for three classes surpassed accuracy by 89%. For methodological comparison, other models based on machine learning called Naïve Bayes and SVMs were implemented as well as other deep learning models like 1D convolution and LSTM. Experimental results have shown that the deep learning model based on BiLSTM achieved the best results in both tasks.

Our results confirmed that deep neural network algorithms are more accurate than machine learning algorithms (Waghmare and Bhala, 2020). Deep neural networks in general need less time as less human intervention is needed, and they perform automated feature extraction. However, to produce appropriate accuracy and performance, they require larger amount of data than machine learning algorithms and the training costs are also high.

Theoretical implications

In recent research based on customer comments in the hospitality and tourism field, the three themes were identified as most relevant, those are behavior, social media, and marketing related to user-generated data (Mukhopadhyay *et al.*, 2022). In addition to the use of sentiment analysis to gain insights from these user-generated data, various text mining techniques are used depending on a particular research goal, e.g. a recent study that analyzed online reviews from TripAdvisor has applied sentiment analysis, clustering, topic modeling, and machine learning algorithms for real-time classification (Gour *et al.*, 2021). Furthermore, sentiment variables were investigated “not only as independent but also as dependent variables” (Mehraliyev *et al.*, 2022). For instance, Kim and Han (2022) used regression analysis to understand the impacts of the length of stay at hotels on online reviews.

A systematic review of sentiment analysis literature in hospitality and tourism from methodological and thematic perspectives confirmed that “testing the performance of sentiment analysis was uncommon” (Mehraliyev *et al.*, 2022), and our study contributes to filling this gap by providing performance results of sentiment analysis based on different machine and deep learning models. While most studies use sentiment analysis as a tool to find insights into customer opinions, our study provides a methodological framework to create and customize sentiment analysis models based on machine and deep learning approaches.

Sentiment analysis theoretics might find fruitful insights from methodological aspects of this study, for instance, when investigating an appropriate model architecture for the particular purpose and domain as well as fine-tuning the parameters of the machine and deep learning models. This study provides detailed methodological insight into several different models, their architectures and complete training and testing processes. Furthermore, different word representation models like TF-IDF and Glove are implemented and compared. It can also be noted that our approach goes beyond hospitality and tourism domain.

Another direction for exploring optimal models for rating prediction is the use of other features as input to the model. Additional features may also be learned from the text, e.g. by content analysis (Wang *et al.*, 2022), or topic analysis (Gour *et al.*, 2021).

Furthermore, the touristic insights made from these models may provide a basis for the understanding of tourist behavior patterns and setting up a theoretical framework for shaping public opinion, i.e. distilling variables that contribute to making opinions. Such a framework can contribute to establishing short-, mid-, or long-term marketing and other relevant strategies for companies and organizations given a particular touristic context.

Practical implications

State-of-the-art natural language processing technologies enable high-quality, fast and efficient analysis of the text that enables a holistic and deep understanding of user experiences on various topics. As tourists often use the online reviews of others as a primary source of information (Kim and Han, 2022), the insights from the analyses of online reviews are a valuable resource of knowledge for many stakeholders in tourism. Sentiment analysis techniques in the tourism industry can be utilized for forecasting tourist expenditure and the number of arrivals and gaining insights into the tourists' profiles. Besides that, they can process a large amount of data and give priceless feedback since it is produced based on user-generated content.

There are other meaningful benefits of sentiment analysis in tourism like using that information for improving overall customer experience and upgrading marketing strategies. Thus, obtaining highly accurate results from sentiment analysis can be used as a resourceful and purposeful crowd intelligence summary that can help decision-makers and different stakeholders in tourism in strategy planning, decision making, and marketing activities. Different service sectors can use the implemented models in different domains of interest as a means of satisfaction recognition of their own products and services as well as a tool to predict the opinions on upcoming challenges.

Limitations and future research

There are several limitations regarding the sentiment analysis research of user-generated data. First, in our study, we assumed the appropriate credibility of all explored reviews in our dataset. However, there is a growing field of research that investigates possible fake information, thus warning both the users and service providers about the continuous need for updating and analyzing the factors that can influence the perceived credibility and quality of online information (Reyes-Menendez *et al.*, 2019). It has been shown that the deep learning approach such as presented in this study achieves promising results even in fake information detection (Cvitanović and Babić Babac, 2022). In addition, aspect-based sentiment classification methods have shown promising results in suppressing the noise within this kind of data (Afzaal *et al.*, 2019).

Another limitation is the use of a single data resource, so future avenues of this research can be in the direction of data enlargement as well as the increase of data sources. Data from other social media have already proven to be beneficial for gaining valuable insights from tourists (Serna *et al.*, 2016). In addition, exploration of more recent and more rich datasets should enable a comparison with the current data and results. Different data types coming from various domains such as transport, environment, weather, etc. may be combined with sentiment scores to investigate the unexplored patterns. Furthermore, review classification using natural language processing has future scope for handling multilingual review classification.

Regarding future work, possible improvements can be realized using transformers, a new architecture based on the attention mechanism and transfer learning (Zhang *et al.*, 2018). Transfer learning enables more efficient leveraging of computational resources and overcomes the limitations such as the need for large, labeled data for training accurate models like recurrent or CNNs. Training these models is less time-consuming and expensive, and the trained model could be easily adapted to a new task, e.g. with a different set of labels (Vaswani *et al.*, 2017).

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. and Zheng, X. (2016), "Tensorflow: large-scale machine learning on heterogeneous distributed systems", *OSDI'16: Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation*, November 2016, pp. 265-283.
- Afzaal, M., Usman, M. and Fong, A. (2019), "Tourism mobile app with aspect-based sentiment classification framework for tourist reviews", *IEEE Transactions on Consumer Electronics*, Vol. 65 No. 2, pp. 233-242.
- Alam, M.H., Ryu, W.-J. and Lee, S. (2016a), "Joint multi-grain topic sentiment: modeling semantic aspects for online reviews", *Information Sciences*, Vol. 339, pp. 206-223.
- Alam, M.H., Ryu, W.-J. and Lee, S. (2016b), "Joint multi-grain topic sentiment: modeling semantic aspects for online reviews", *TripAdvisor Hotel Review Dataset*, available at: <https://zenodo.org/record/1219899#.YeNupP7MKUk> (accessed 8 January 2022).
- Avdzić, D. and Bagić Babac, M. (2021), "Application of affective lexicons in sports text mining: a case study of FIFA world cup 2018", *South Eastern European Journal of Communication*, Vol. 3 No. 2, pp. 23-33.
- Bagherzadeh, S., Shokouhyar, S., Jahani, H. and Sigala, M. (2021), "A generalizable sentiment analysis method for creating a hotel dictionary: using big data on TripAdvisor hotel reviews", *Journal of Hospitality and Tourism Technology*, Vol. 12 No. 2, pp. 210-238.
- Bagić Babac, M. and Podobnik, V. (2016), "A sentiment analysis of who participates, how and why, at social media sports websites: how differently men and women write about football", *Online Information Review*, Vol. 40 No. 6, pp. 814-833.
- Barbierato, E., Bernetti, I. and Capecci, I. (2021), "Analyzing TripAdvisor reviews of wine tours: an approach based on text mining and sentiment analysis", *International Journal of Wine Business Research*, Vol. 34 No. 2, pp. 212-236, doi: [10.1108/IJWBR-04-2021-0025](https://doi.org/10.1108/IJWBR-04-2021-0025).
- Baziotis, C., Pelekis, N. and Doukeridis, C. (2017), "DataStories at SemEval-2017 task 4: deep LSTM with attention for message-level and topic-based sentiment analysis", *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, Association for Computational Linguistics, Vancouver, pp. 747-754.
- Collobert, R. and Weston, J. (2008), "A unified architecture for natural language processing: deep neural networks with multitask learning", *Proceedings of the 25th International Conference on Machine Learning, ICML '08*, ACM, New York, NY, USA, pp. 160-167.
- Cvitanović, I. and Bagić Babac, M. (2022), *Deep Learning with Self-Attention Mechanism for Fake News Detection, Combating Fake News with Computational Intelligence Techniques*, in Lahby, M., Pathan, A.S.K., Maleh, Y. and Yafooz, W.M.S. (Eds), Springer, Switzerland, pp. 205-229.
- Faralli, S., Rittinghaus, S., Samsami, R., Distante, D. and Rocha, E. (2021), "Emotional intensity-based success prediction model for crowdfunded campaigns", *Information Processing and Management*, Vol. 58, No. 102394.
- Farisi, A.A., Sibaroni, Y. and Al Faraby, S. (2018), "Sentiment analysis on hotel reviews using multinomial naïve Bayes classifier", *Journal of Physics: Conference Series, The 2nd International Conference on Data and Information Science*, Bandung, Indonesia, 15-16 November 2018, Vol. 1192.
- Gitto, S. and Mancuso, P. (2017), "Improving airport services using sentiment analysis of the websites", *Tourism Management Perspectives*, Vol. 22, pp. 132-136.
- Gour, A., Aggarwal, S. and Erdem, M. (2021), "Reading between the lines: analyzing online reviews by using a multi-method Web-analytics approach", *International Journal of Contemporary Hospitality Management*, Vol. 33 No. 2, pp. 490-512.

-
- Harrag, F., Alsalman, A. and Alqahtani, A. (2019), "Prediction of reviews rating: a survey of methods, techniques and hybrid architectures", *Journal of Digital Information Management*, Vol. 17 No. 3, pp. 164-178.
- Hochreiter, S. (1998), "The vanishing gradient problem during learning recurrent neural nets and problem solutions", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 6 No. 2, pp. 107-116.
- Hochreiter, S. and Schmidhuber, J. (1997), "Long short-term memory", *Neural Computation*, Vol. 9 No. 8, pp. 1735-1780.
- Hossain, E. and Das, S. (2020), "A machine learning-based approach to predict travel experience based on tourist's rating reviews", *Port City International University Journal*, Vol. 7 Nos 1-2, pp. 9-16.
- Huang, T. (2021), "Research on sentiment classification of tourist destinations based on convolutional neural network", *2021 IEEE 3rd Eurasia Conference on IOT, Communication and Engineering (ECICE)*, pp. 358-361.
- Irsoy, O. and Cardie, C. (2013), "Bidirectional recursive neural networks for token-level labeling with structure", *NIPS Deep Learning Workshop*, 2013.
- Jurafsky, D. and Martin, J.H. (2000), *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Prentice-Hall, Upper Saddle River, NJ.
- Kang, H., Yoo, S.J. and Han, D. (2012), "Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews", *Expert Systems with Applications*, Vol. 39 No. 5, pp. 6000-6010.
- Kim, J.M. and Han, J. (2022), "Impact of the length of stay at hotels on online reviews", *International Journal of Contemporary Hospitality Management*, Vol. 34 No. 4, pp. 1249-1269, doi: [10.1108/IJCHM-05-2021-0659](https://doi.org/10.1108/IJCHM-05-2021-0659).
- Kuhn, M. and Johnson, K. (2013), *Applied Predictive Modeling*, Springer, New York, Vol. 26, p. 13.
- Larasati, A., Sayono, J., Purnomo, A., Mohamad, E., Farhan, M. and Rahmawati, P. (2020), "Applying web mining and sentiment analysis to assess tourists review on Batu City tourist destination", *2020 4th International Conference on Vocational Education and Training (ICOVET)*, pp. 63-68.
- Leal, F., Malheiro, B. and Burguillo, J.C. (2018), "Analysis and prediction of hotel ratings from crowdsourced data", *WIREs Data Mining and Knowledge Discovery*, Vol. 9 No. 2, e1296.
- Leal, F., Gonzalez-Velez, H., Malheiro, B. and Burguillo, J.C. (2017), "Profiling and rating prediction from multi-criteria crowd-sourced hotel ratings", in Zoltay Paprika, Z. *et al.*, (Eds), *ECMS 2017 Proceedings*, European Council for Modeling and Simulation.
- Loria, S. (2018), "TextBlob documentation", Release 0.15, Vol. 2.
- Madyatmadja, E.D., Pristinella, D., Rahardja, N. and Ginting, R.B. (2021), "Smart tourism services: a systematic literature review", *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)*, pp. 329-333.
- Manosso, F.C. and Domareski Ruiz, T.C. (2021), "Using sentiment analysis in tourism research: a systematic, bibliometric, and integrative review", *Journal of Tourism, Heritage and Services Marketing*, Vol. 7 No. 2, pp. 17-27, doi: [10.5281/zenodo.5548426](https://doi.org/10.5281/zenodo.5548426).
- Marrese-Taylor, E., Velásquez, J.D. and Bravo-Marquez, F. (2014), "A novel deterministic approach for aspect-based opinion mining in tourism products reviews", *Expert Systems with Applications*, Vol. 41 No. 17, pp. 7764-7775.
- Mehrliyev, F., Chan, I.C.C. and Kirilenko, A.P. (2022), "Sentiment analysis in hospitality and tourism: a thematic and methodological review", *International Journal of Contemporary Hospitality Management*, Vol. 34 No. 1, pp. 46-77, doi: [10.1108/IJCHM-02-2021-0132](https://doi.org/10.1108/IJCHM-02-2021-0132).
- Mehta, M.P., Kumar, G. and Ramkumar, M. (2021), "Customer expectations in the hotel industry during the COVID-19 pandemic: a global perspective using sentiment analysis", *Tourism Recreation Research*, pp. 1-18, doi: [10.1080/02508281.2021.1894692](https://doi.org/10.1080/02508281.2021.1894692).

- Mikolov, T., Yih, W. and Zweig, G. (2013), "Linguistic regularities in continuous space word representations", *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, Atlanta, Georgia, pp. 746-751.
- Mostafa, M.M. (2013), "More than words: social networks' text mining for consumer brand sentiments", *Expert Systems with Applications*, Vol. 40, pp. 4241-4251.
- Mukhopadhyay, S., Pandey, R. and Rishi, B. (2022), "Electronic word of mouth (eWOM) research – a comparative bibliometric analysis and future research insight", *Journal of Hospitality and Tourism Insights*, Vol. ahead-of-print No. ahead-of-print, doi: [10.1108/JHTI-07-2021-0174](https://doi.org/10.1108/JHTI-07-2021-0174).
- Neidhardt, J., Rümmele, N. and Werthner, H. (2017), "Predicting happiness: user interactions and sentiment analysis in an online travel forum", *Information Technology Tourism*, Vol. 17 No. 16, pp. 101-119.
- Pennington, J., Socher, R. and Manning, C.D. (2014), "GloVe: global vectors for word representation", *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014. doi: [10.3115/v1/D14-1162](https://doi.org/10.3115/v1/D14-1162).
- Poch Alonso, R. and Bagić Babac, M. (2022), "Machine learning approach to predicting a basketball game outcome", *International Journal of Data Science*, Vol. 7 No. 1, available at: <https://www.inderscience.com/info/ingeneral/forthcoming.php?jcode=ijsds>.
- Prameswari, P., Zulkarnain, Surjandari, I. and Laoh, E. (2017), "Mining online reviews in Indonesia's priority tourist destinations using sentiment analysis and text summarization approach", *2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)*, pp. 121-126.
- Ramadhani, A., Sutoyo, E. and Widartha, V.P. (2021), "LSTM-based deep learning architecture of tourist review in Tripadvisor", *2021 Sixth International Conference on Informatics and Computing (ICIC)*, pp. 1-6.
- Reyes-Menendez, A., Saura, J.R. and Martinez-Navalon, J.G. (2019), "The impact of e-WOM on hotels management reputation: exploring TripAdvisor review credibility with the ELM model", *IEEE Access*, Vol. 7, pp. 68868-68877.
- Serna, A., Gerrikagoitia, J.K. and Bernabé, U. (2016), "Discovery and classification of the underlying emotions in the user-generated content (UGC)", in Inversini, A. and Schegg, R. (Eds), *Information and Communication Technologies in Tourism 2016*, Springer, Cham.
- Sherstinsky, A. (2020), "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network", *Physica D: Nonlinear Phenomena*, Vol. 404, 132306.
- Shi, H. and Li, X. (2011), "A sentiment analysis model for hotel reviews based on supervised learning", *2011 International Conference on Machine Learning and Cybernetics*, pp. 950-954.
- Statista, "Number of user reviews and opinions on Tripadvisor worldwide 2014-2020", available at: <https://www.statista.com/statistics/684862/tripadvisor-number-of-reviews/> (accessed 21 January 2022).
- Sumarsono, D., Sudardi, B., Wanto and Abdullah, W. (2018), "The influence of TripAdvisor application usage towards hotel occupancy rate in Solo", *Journal of Physics: Conference Series, 1st International Conference on Advance and Scientific Innovation*, Vol. 1175, Medan, Indonesia, 23-24 April 2018.
- Tan, S. and Wu, Q. (2011), "A random walk algorithm for automatic construction of domain-oriented sentiment lexicon", *Expert Systems with Applications*, Vol. 38 No. 10, pp. 12094-12100.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. (2017), "Attention is all you need", *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*, Curran Associates, Red Hook, NY, USA, pp. 6000-6010.
- Vrigazova, B. (2021), "The proportion for splitting data into training and test set for the bootstrap in classification problems", *Business Systems Research*, Vol. 12 No. 1, pp. 228-242.

-
- Wadhe, A.A. and Suratkar, S.S. (2020), "Tourist place reviews sentiment classification using machine learning techniques", *2020 International Conference on Industry 4.0 Technology (I4Tech)*, pp. 1-6.
- Waghmare, K.A. and Bhala, S., K. (2020), "Survey paper on sentiment analysis for tourist reviews", *2020 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1-4.
- Wang, S., Chu, T., Li, H. and Sun, Q. (2022), "Cruise vacation experiences for Chinese families with young children", *Tourism Review*, Vol. 77 No. 3, pp. 815-840, doi: [10.1108/TR-08-2021-0394](https://doi.org/10.1108/TR-08-2021-0394).
- Xu, G., Meng, Y., Qiu, X., Yu, Z. and Wu, X. (2019), "Sentiment analysis of comment texts based on BiLSTM", *IEEE Access*, Vol. 7, pp. 51522-51532.
- Zhang, J., Luan, H., Sun, M., Zhai, F., Xu, J., Zhang, M. and Liu, Y. (2018), "Improving the transformer translation model with document-level context", *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Brussels, pp. 533-542.

Further reading

- Kim, Y. (2014), "Convolutional neural networks for sentence classification", *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics, pp. 1746-1751.
- Malthouse, E.C., Haenlein, M., Skiera, B., Wege, E. and Zhang, M. (2013), "Managing customer relationships in the social media era: introducing the social CRM house", *Journal of Interactive Marketing*, Vol. 27, pp. 270-280.
- Moliner-Velázquez, B., Fuentes-Blasco, M. and Gil-Saura, I. (2022), "Antecedents of online word-of-mouth reviews on hotels", *Journal of Hospitality and Tourism Insights*, Vol. 5 No. 2, pp. 377-393.
- Tsai, C.F., Chen, K., Hu, Y.H. and Chen, W.K. (2020), "Improving text summarization of online hotel reviews with review helpfulness and sentiment", *Tourism Management*, Vol. 80, doi: [10.1016/j.tourman.2020.104122](https://doi.org/10.1016/j.tourman.2020.104122).
- Zhang, Z., Ye, Q. and Law, R.Y.L. (2010), "The impact of e-word-of-mouth on the online popularity of restaurants: a comparison of consumer reviews and editor reviews", *International Journal of Hospitality Management*, Vol. 29, pp. 694-700.

Corresponding author

Marina Bagić Babac can be contacted at: marina.bagic@fer.hr