JOURNAL OF SCIENCE
OF LAC HONG UNIVERSITY
JSLHU

ISSN: 2525 - 2186
Tạp chí Khoa học Lạc Hồng, **2025**, 20, 019-024

TỐI ƯU HÓA CÁC LỚP ẨN TRONG MẠNG LSTM ĐỂ TẠO TÓM TẮT CHI TIẾT ĐÁNH GIÁ KHÁCH SẠN

Lê Quốc Bảo[1*], Lê Phương Long[2], Nguyễn Thanh Sơn[2], Đặng Đăng Khoa[2]
[1]*Khoa học máy tính và Công nghệ thông tin, Trường Đại học Quang Trung, Bình Định, Việt Nam*
[2]*Trường Đại học Lạc Hồng, Số 10 Huỳnh Văn Nghệ, phường Bửu Long, Biên Hòa, Đồng Nai, Việt Nam*
* Tác giả liên hệ: *lqbao@qtu.edu.vn*

THÔNG TIN BÀI BÁO

Ngày nhận:	11/2/2025
Ngày hoàn thiện:	7/3/2025
Ngày chấp nhận:	13/3/2025
Ngày đăng:	15/3/2025

TỪ KHÓA

Phân tích tình cảm;
Mạng lưới nơ-ron sâu;
LSTM;
Đánh giá khách sạn.

TÓM TẮT

Các nền tảng đặt phòng khách sạn trực tuyến thường thiếu khả năng cung cấp các bản tóm tắt chi tiết về đánh giá của người dùng trên các lĩnh vực dịch vụ chính như thực phẩm, phòng ở, chất lượng dịch vụ và vị trí. Nghiên cứu này giới thiệu một giải pháp đột phá sử dụng mạng Long Short-Term Memory (LSTM) với cấu hình lớp ẩn được tối ưu hóa. Với F1- score đạt 75,29%-tăng 10,18% so với mô hình LSTM tiêu chuẩn. Mô hình đã chứng minh được hiệu quả trong việc tạo ra các bản tóm tắt đánh giá theo từng khía cạnh cụ thể. Tiến bộ này mang lại cho người dùng trải nghiệm đặt phòng khách sạn thông minh và hiệu quả hơn, đồng thời giải quyết các hạn chế trong hệ thống đánh giá hiện tại.

OPTIMIZING HIDDEN LAYERS IN LSTM NETWORKS TO GENERATE DETAILED SUMMARIES OF HOTEL REVIEWS

Le Quoc Bao[1*], Le Phuong Long[2], Nguyen Thanh Son[2], Dang Dang Khoa[2]
[1]*Faculty of Computer Science and Information Technology, Quang Trung University, Binh Dinh, Vietnam*
[2]*Lac Hong University, No 10 Huynh Van Nghe Str., Buu Long Ward, Bien Hoa, Dong Nai, Viet Nam*
*Corresponding Author: *lqbao@qtu.edu.vn*

ARTICLE INFO

Received:	Feb 11st, 2025
Revised:	Mar 7th, 2025
Accepted:	Mar 13rd, 2025
Published:	Mar 15th, 2025

KEYWORDS

Sentiment analysis;
Deep neural network;
LSTM;
Hotel review.

ABSTRACT

Online hotel booking platforms often lack the ability to provide detailed summaries of user reviews across key service areas, such as food, accommodation, service quality, and location. This study introduces a breakthrough solution using Long Short-Term Memory (LSTM) networks with optimized hidden layer configurations. With an F1-score of 75.28%—an increase of 10.18% compared to the standard LSTM model—the model has proven effective in generating review summaries for specific aspects. This advancement offers users a smarter and more efficient hotel booking experience while addressing the limitations of current review systems.

Doi: https://doi.org/10.61591/jslhu.20.623
Available online at: https://js.lhu.edu.vn/index.php/lachong

1. INTRODUCTION

Hotel quality is a critical factor for users when making booking decisions. Advances in information technology have simplified the evaluation process, with most online booking platforms providing user reviews [1]. However, these reviews often remain static and lack detailed insights into specific aspects of hotel quality [2].

Sentiment analysis has emerged as a powerful tool to transform review data into valuable insights by categorizing sentiments as positive, negative, or neutral [3]. While document and sentence-level sentiment analysis provide an overall sentiment classification, they fail to identify specific features contributing to the feedback [4]. Aspect-Based Sentiment Analysis (ABSA) addresses these limitations by linking sentiments to distinct aspects, such as service, cleanliness, or location, offering a more granular understanding of reviews [5].

Moreover, the ABSA is a specialized sentiment analysis method that evaluates specific entities and their corresponding aspects [6]. It consists of three key tasks: aspect term extraction, category detection, and sentiment polarity classification [7]. Over time, extensive research has explored ABSA's applications in domain-specific and open-domain reviews, including in foreign language contexts, including clothing retail [8], restaurants [9], and e-commerce [10], showcasing its versatility in real-world applications.

In addition, research in [11] incorporating classical machine learning techniques with feature extraction for ABSA introduced a Support Vector Machine (SVM) model, which yielded an F1- score of 0.92. Further exploration of ABSA focused on three tasks: classifying aspect categories, extracting opinion targets, and determining sentiment polarity [12]. The study applied a Convolutional Neural Network (CNN) for aspect category and sentiment polarity classification, while Conditional Random Fields (CRF) and Bidirectional Long Short-Term Memory (Bi-LSTM) networks were employed to extract opinion targets. This powerful model achieved F1-scores of 0.87 for classifying aspect categories, 0.78 for extracting opinion targets, and 0.764 for identifying sentiment polarity.

In particular, the ABSA has been applied across various domains to address specific analytical challenges. Al-Smadi et al. [13] focused on examining reviews in Arabic hotels using ABSA methods with SVM excelled in binary classification tasks, whereas RNN was more effective in managing sequential data during training and evaluation. The author Akhtar et al. [14] combined aspect identification with topic modeling in ABSA for TripAdvisor hotel reviews, extracting meaningful insights from user-generated content. Alqaryouti et al. [15] explored government application reviews using a hybrid lexicon and rule-based ABSA method, effectively addressing implicit aspects, negation, and sentiment ambiguity. Liu et al. [16] introduced the Recurrent Memory Neural Network (ReMemNN), an enhanced memory neural network incorporating embedding adjustment and attention mechanisms, achieving superior results across datasets. Similarly, Liu et al. [17] proposed the Gated Alternate Neural Network (GANN), designed to overcome limitations of traditional neural network architectures and memory-based models by leveraging convolutional and max-pooling layers to handle local features and resolve long-term dependency issues, achieving performance.

2. METHODOLOGY

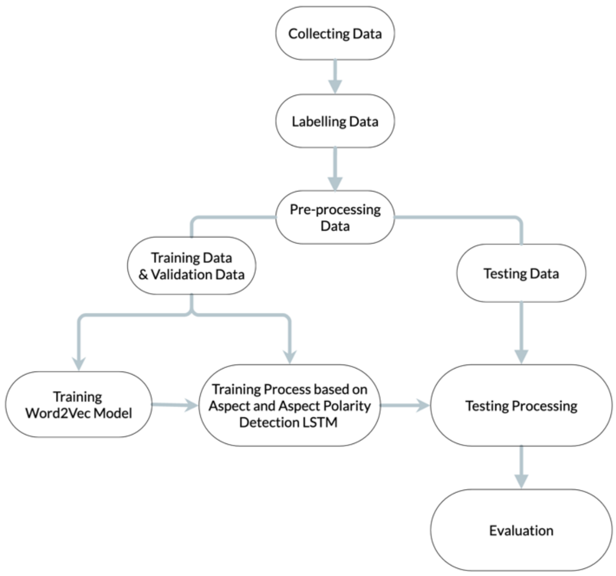

Figure 1. *Overview of the Research Methodology*

Fig. 1 depicts the investigation stages of the research. Data gathering starts with collecting data from the website Traveloka. The collected data is pre-processed to prepare the "word2vec model" utilized for feature extraction. The extracted features are integrated into the model during training, with the data divided into training and validation sets. Models are then trained using different parameter combinations, and their results are assessed during testing, which determines how well each model constructed performs. The training data was sourced from the Traveloka website. Each model developed is evaluated for its effectiveness during the assessment phase. To analyze the outcomes, a confusion matrix is used, with the micro-average F1-score being the primary metric for measurement.

2.1 Data Collection, Labeling, and Splitting

The data was collected using Selenium to crawl reviews from the Traveloka website. A total of 2,700 reviews were gathered from ten randomly selected hotels. Additionally, 2,500 reviews from a previous study 8[1] were included, resulting in a dataset of 5,200 hotel reviews for this research. The dataset was divided into two categories: a training-validation set and a testing set.

Table 1. *The distribution of data among various aspects*

Aspect	***Count***
Room	3,185

Foods	1,398
Services	3,169
Location	1,887
Miscellaneous	1,153

The training-validation dataset comprised 5,000 reviews, including 10,283 identified aspects. The reviews were labeled around specific aspects and their associated sentiment classifications. The aspect labels included "món ăn" (food), "phòng" (room), "dịch vụ" (service), "vị trí" (location), and "khác" (miscellaneous), while sentiment polarities were categorized as positive, neutral, or negative. At least one aspect was labeled in each review.

A total of 200 reviews made up the testing dataset, annotated following the same criteria as the training validation set. Table 1 illustrates the distribution of data across various aspects. The labels were encoded using the one-hot encoding technique, with each label represented by 15 binary elements. These elements corresponded to combinations of five distinct aspects, each classified into three sentiment categories.

2.2 Data cleansing and preparation

Pre-processing refers to the preparation and cleaning of data to facilitate class [19]. This process includes several steps such as case folding, which standardizes text by converting all characters to a consistent case, and stopword removal, which eliminates insignificant words like conjunctions to enhance machine learning efficiency [20].

Stemming reduces words to their root forms, ensuring uniformity, while tokenization breaks sentences into smaller units called tokens Padding adjusts the lengths of tokenized sequences by appending dummy tokens, and vectorization converts textual data into a numerical format using a dictionary built from the dataset's vocabulary. Together, these steps create a structured and optimized dataset for machine learning applications.

2.3 Word2Vec Framework

This research made use of Word2Vec, a feature extraction framework developed by [21], known for its ability to process large volumes of data efficiently and within a short time frame. The parameters for Word2Vec in this research were determined using the skip-gram model, alongside hierarchical SoftMax as the evaluation method.

The model employs two primary techniques: Continuous Bag of Words (CBOW), which predicts a word based on its surrounding context, and Skip-Gram, which uses a target word to predict its contextual words. A vector size of 300 was utilized, reflecting the expectation that the output vector's dimensionality would grow in proportion to the importance of the dataset used.

2.4 LSTM-Driven Training and Evaluation Pipeline for ABSA

At this stage, an LSTM-based model was designed to identify both aspects and their corresponding polarities in reviews. The study utilized ABSA to perform these tasks, allowing for the detection of sentiment polarity for aspects within a review, as well as identifying the various aspects mentioned. The ABSA implementation follows standard LSTM-based architectural frameworks, as shown in Fig. 2.

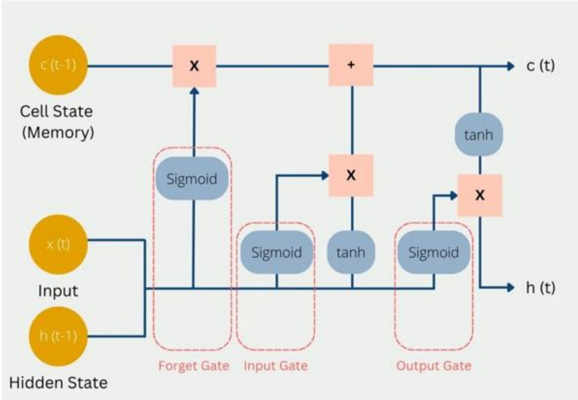

Figure 2. *Design of a Standard LSTM [22]*

LSTM, initially introduced by [22], resolves the long-term dependency problems commonly found in RNNs. It achieves this by utilizing gates and a memory cell. The primary components of LSTM include the forget gate, input gate, and candidate gate, each governed by specific equations. For instance, the input gate and memory state are defined by (1) and (2), respectively

$$\tilde{C}_t = \tanh\left(W_c x(t) + U_c h(t-1) + b_c\right) \qquad (1)$$

$$i_t = \sigma(W_i x(t) + U_i h(t-1) + b_i \qquad (2)$$

The forget gate and cell state are detailed, along with a description of the candidate gate

$$f_t = \sigma\big(W_f x(t) + U_f h(t-1) + b_f \qquad (3)$$

$$C_t = i_t * \tilde{C}_t + f_t * C_{t-1} \qquad (4)$$

$$o_t = \sigma(W_o x(t) + U_o h(t-1) + b_o \qquad (5)$$

Each LSTM cell produces an output consisting of the hidden state paired with the candidate gate. The hidden state output is computed as

$$h_t = o_t * \tanh\left(C_t\right) \qquad (6)$$

As illustrated in Fig. 3, the study adopted an LSTM model with fully connected layers. Pre-processed data is fed into an embedding layer, constructed using word pairs and word2vec vectors. The output of the LSTM layer is a matrix matching the embedding size, which is flattened into one dimension before passing through two fully connected layers. The standard LSTM architecture from [23] serves as the default model.

The final layer of the default model was modified to match the size of the data labels. Throughout the training phase, various parameter combinations were tested to identify the most efficient model setup.

The testing phase aims to identify the optimal parameter combination for the proposed model. Testing will be performed using a dataset other than the one utilized in the training phase.

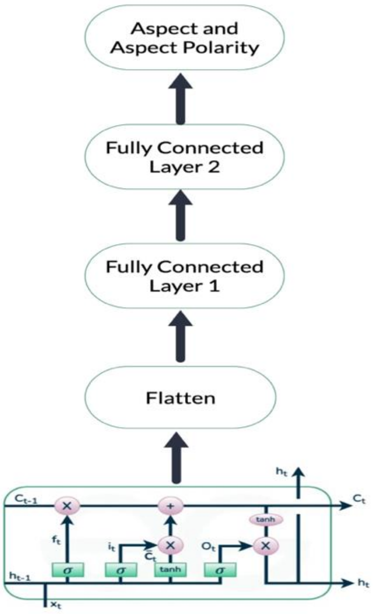

Figure 3. *Proposed Architecture*

Four scenarios were evaluated, emphasizing the selection of the optimal size and activation function for the first fully connected layer, followed by identifying the ideal size and activation function for the second fully connected layer. Table 2 provides a summary of the various parameter combinations tested.

Table 2. *Parameter Settings for the Proposed Design*

Parameters	***Values***
Layer 1	700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200
Layer 2	50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700
Functions	ReLU; tanh; and sigmoid

2.5 Performance Evaluation

This problem is well-suited for evaluation based on classes or labels, making the micro-average F1- score an appropriate choice due to its sensitivity to the most frequently occurring classes or labels [24]. A critical element in calculating the micro-average F1- score is the confusion matrix. Ideally, for any classifier, the values along the main diagonal of the confusion matrix should be significantly larger than the off-diagonal entries, resulting in an F1-measure value close to 1. The formula used to compute the F1-measure is as follows:

$$Precision_{mic} = \frac{\sum_{i=1}^{N} TP_1}{\sum_{i=1}^{N} (TP_1 + FP_1)} \quad (7)$$

$$Recall_{mic} = \frac{\sum_{i=1}^{N} TP_1}{\sum_{i=1}^{N} (TP_1 + FP_1)} \quad (8)$$

$$F1_{mic\ avg} = \frac{2 \times Precision_{mic} \times Recall_{mic}}{Precision_{mic} + Precision_{mic}}, \quad (9)$$

where TP represents true positives, FP denotes false positives, and FN indicates false negatives. For each sentiment, true positives (TP) correspond to the entries along the main diagonal of the confusion matrix, while false positives (FP) are located in the row below the diagonal, and false negatives (FN) appear in the corresponding horizontal column. According to (7) and (8), when the denominators for precision and recall are the same, the micro-averaged precision and recall values will also be identical. Consequently, the micro-averaged F1-measure, $F1_{mic\ avg.}$, from Equation (11) will be equal to both precision and recall.

The formula for F1-measure computation is as follows:

$$Precision_{mic\ avg.} = Recall_{mic\ avg.} \quad (10)$$

$$F1_{mic\ avg.} = Precision_{mic\ avg.} \quad (11)$$

3. RESULTS AND DISCUSSION

In the first scenario, the size of Fully Connected Layer 1 demonstrated an upward trend in performance as it increased from 700 to 1,200 neurons. The highest performance, reflected by a micro-average F1-measure of 0.7488, was achieved at 1,200 neurons. This improvement can be attributed to larger layers minimizing the risk of incorrect information transfer, as illustrated in Fig. 4.

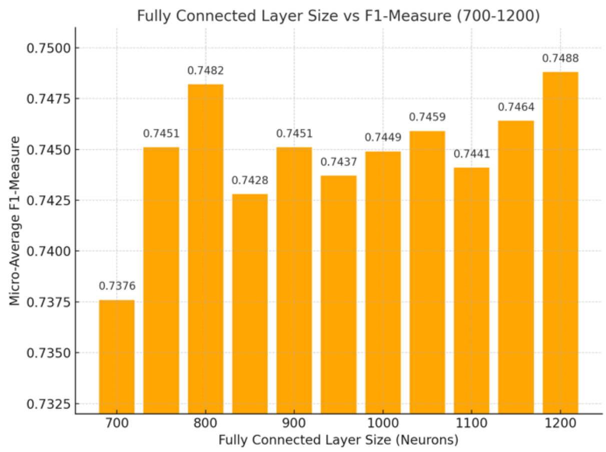

Figure 4. *Effect of Fully Connected Layer 1 Size*

In the second scenario, the performance of Fully Connected Layer 2 improved as its size increased up to 400 neurons, after which it stabilized. The best performance was observed with 600 neurons, achieving a micro-average F1-score of 0.7590, slightly higher than the 0.7564 recorded with 400 neurons. Smaller layers resulted in greater information loss, as illustrated in Fig. 5.

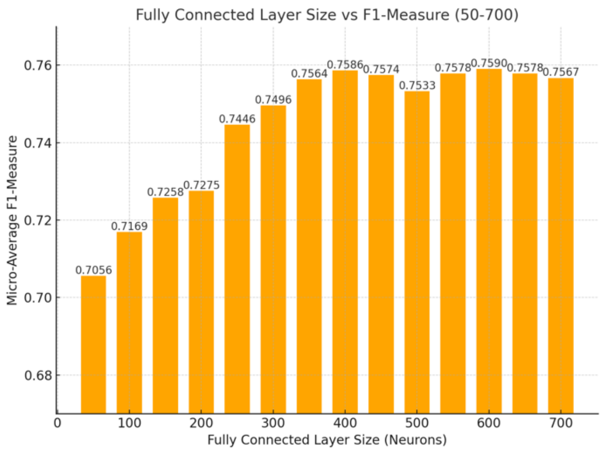

Figure 5. *Effect of Fully Connected Layer 2 Size*

In Fully Connected Layer 1, the tanh activation function demonstrated the best performance, achieving a micro-average F1-measure of 0.7462, while ReLU recorded the lowest at 0.7419. The tanh function effectively prevented premature convergence, enhancing the recognition of data patterns, as shown in Fig. 6.

For Fully Connected Layer 2, ReLU delivered the highest performance with a micro-average F1-score of 0.768 by constraining neuron values within the desired range, whereas sigmoid proved the least effective with a score of 0.72, as depicted in Fig. 7.

An analysis of layer size for Fully Connected Layer 2 revealed a performance improvement trend up to 400 neurons, beyond which it stabilized. The optimal performance was observed at 600 neurons, achieving a micro-average F1-score of 0.7590, slightly surpassing the 0.7564 performance at 400 neurons. Smaller layer sizes were linked to an increased risk of information loss between layers.

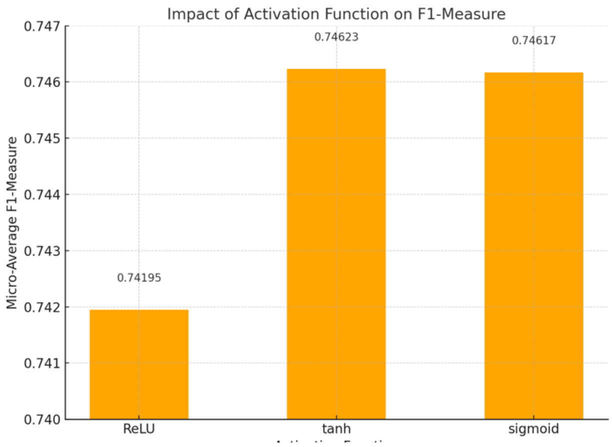

Figure 6. *Effect of Activation Functions in Fully Connected Layer 1*

The optimal model was designed using the best parameters identified from the four scenarios: 1,200 neurons with the tanh activation function for Fully Connected Layer 1, and 600 neurons with the ReLU activation function for Fully Connected Layer 2. When evaluated on the same dataset, the model achieved a micro-average F1-measure of 0.7528.

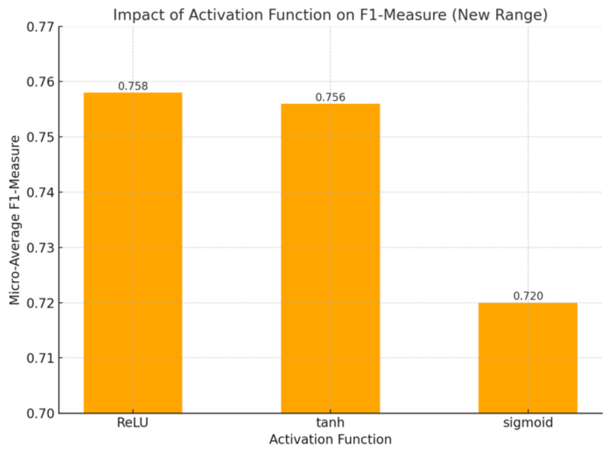

Figure 7. *Effect of Activation Functions in Fully Connected Layer 2*

In a comparative analysis, the proposed model was tested alongside the baseline LSTM architecture using the same dataset, consisting of 5,000 training and 200 testing samples. As detailed in Table 4, the proposed model demonstrated superior performance compared to the baseline. The improved performance is credited to the inclusion of two fully connected layers preceding the output layer, both utilizing the sigmoid activation function

Table 4. *Comparison of F1-Score Model*

Model	***Micro-average F1-score***
Standard Model	0.651
Proposed Model	0.757

4. CONCLUSION

A hotel quality recognition system can be developed by collecting and preprocessing review data, implementing an advanced recognition model, and extracting key aspects and sentiment information from the reviews. This processed data offers valuable and easily understandable insights for customers. The model's optimal architecture integrates an LSTM with two fully connected layers: the first layer contains 1,200 neurons with a tanh activation function, while the second layer has 600 neurons with a ReLU activation function. The proposed model achieved a micro-average F1-score of 0.7528, outperforming the baseline model by 10.16%, demonstrating its superior effectiveness.

5. REFERENCES

[1] Li H., Wang J., Zhang X. The role of online reviews in hotel booking decisions: Insights and implications. *Journal of Information Technology in Tourism* **2021**, 23(2), pp. 145-160.

[2] Phillips P., Barnes S., Zigan K., Schegg R. Understanding the impact of online reviews on hotel performance. *Journal of Travel Research* **2016**, 56(2), pp. 235–249.

[3] Birjali M., Kasri M., Beni-Hssane A. A comprehensive survey on sentiment analysis: Approaches, challenges, and trends. *Knowledge-Based Systems* **2021**, 226, 107134.

[4] Wankhade M., Rao A. C. S., Kulkarni C. A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review* **2022**, 55(7), pp. 5731–5780.

[5] Özen İ. A, Katlav E. Ö. Aspect-based sentiment analysis on online customer reviews: a case study of technology-supported hotels. *Journal of Hospitality and Tourism Technology* **2023**, 14(2), pp.102–120.

[6] Schouten K., Frasincar F. Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering* **2016**, 28(3), pp. 813-830.

[7] Kontonatsios G., Clive J., Harrison G., Metcalfe T., Sliwiak P., Tahir H., Ghose A. FABSA: An aspect-based sentiment analysis dataset of user reviews. *Neurocomputing* **2023**, 562, 126867.

[8] Huang A., Xu S., Zhang W. ABSA for clothing reviews: A case study using Naive Bayes and bag-of-words features. *Journal of Retail Analytics* **2020**, 12(2), pp.35-48.

[9] Pontiki M., Galanis D., Pavlopoulos J., Papageorgiou H., Androutsopoulos I., Manandhar S. SemEval-2014 task 4: Aspect-based sentiment analysis. *Proceedings of the 8th International Workshop on Semantic Evaluation* **2014** (SemEval 2014), pp. 27-35.

[10] Medhat W., Hassan A., Korashy H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal* **2014**, 5(4), pp. 1093-1113.

[11] Kiritchenko S., Zhu X., Cherry C., Mohammad S. NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. *Proceedings of the 7th International Workshop on Semantic Evaluation* **2014** (SemEval 2014), pp. 321-327.

[12] Poria S., Cambria E., Hazarika D., Kwok K. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems* **2016**, 108, pp. 42-49.

[13] Elnagar A., Khalifa Y. S., Einea, A. Hotel Arabic-Reviews Dataset construction for sentiment analysis applications. *In Studies in computational intelligence* **2017**, pp. 35–52.

[14] Akhtar A., Zubair N., Kumar A., Ahmad T. Aspect based Sentiment Oriented Summarization of Hotel Reviews. *Procedia Comput* **2017**. Sci., 115, pp. 563–571.

[15] Alqaryouti O., Siyam N., Monem A., Shaalan K. Aspect-based sentiment analysis using smart government review data. Appl. Comput. Informatics, **2020**.

[16] Liu N., Shen B. ReMemNN: A novel memory neural network for powerful interaction in aspect-based sentiment analysis. Neurocomputing, 395, pp. 66–67, **2020**.

[17] Liu. N., and Shen, B. "Aspect-based sentiment analysis with gated alternate neural network," Knowledge-Based Syst., 188, p. 105010, **2020**.

[18] R. P. Nawangsari, R. Kusumaningrum, and A. Wibowo, "Word2vec for Indonesian sentiment analysis towards hotel reviews: An evaluation study," Procedia Comput. Sci., 157, pp. 360–366, **2019**.

[19] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," Procedia Comput. Sci., vol. 17, pp. 26–32, **2013**.

[20] A. Kulkarni and S. Mundhe, "A theoretical review on text mining: Tools, techniques, applications and future challenges," Int. J. Innov. Res. Comput. Commun. Eng., vol. 4, no. 11, pp. 19225–19230, **2016**.

[21] T. Mikolov, G. Corrado, K. Chen, and J. Dean, "Efficient estimation of word representations in vector space," in ICLR, **2013**, pp. 1–12

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., 9(8), pp. 1735–1780, **1997**.

[23] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016, pp. 606–615.

[24] Y. Yang and X. Liu, "A re-examination of text categorization methods", **1999**.