Toward a framework for building a decision support system for customer satisfaction assessment in tourism sector

Adnane Souha*, Lamya Benaddi†, Charaf Ouaddi‡, El mahi Bouziane§ and Abdeslam Jakimi¶
GL-ISI Team, Dep. of Informatics, FST Errachida, UMI-Meknes
Errachidia, Morocco
Email: *ad.souha@edu.umi.ac.ma, †l.benaddi@edu.umi.ac.ma, ‡c.ouaddi@edu.umi.ac.ma
§bouzianeelmahi@gmail.com, ¶ajakimi@yahoo.fr

***Abstract*—This paper presents an artificial intelligence solution aimed at providing hotel managers with valuable insights based on tourist feedback. Our approach utilizes two key natural language processing tasks: text classification and sentiment analysis. Through text classification, we can identify essential factors such as comfort, quality/price ratio, and cleanliness. Meanwhile, sentiment analysis enables us to extract the underlying sentiment expressed in customer comments. We proposed a model to solve the two tasks, which we validated for sentiment analysis using the Twitter Sentiment Analysis dataset. Ultimately, our solution provides insights into the factors requiring improvement, enabling hotel managers to make informed decisions that enhance the overall guest experience.**

***Index Terms*—hotel managers, AI, Information extraction, text classification, sentiment analysis**

I. INTRODUCTION

In today's hotel industry, understanding and acting on customer feedback is crucial to maintaining high standards of service and visitor satisfaction. This paper presents an innovative Artificial Intelligence (AI) solution designed for hotel managers aimed at extracting valuable information from tourist comments. Using the power of natural language processing (NLP) tasks, particularly text classification and sentiment analysis, our approach focuses on the nuanced aspects of customer feedback to reveal actionable insights.

The main focus of our approach revolves around two key NLP tasks. Firstly, through text classification, we identify and categorize the critical factors that significantly impact the customer experience, including comfort, price-quality ratio, and cleanliness. Secondly, through sentiment analysis, we explore the feelings expressed in these textual comments, discovering customers' underlying sentiments and perceptions of different aspects of their stay.

The use of a combination of text classification and sentiment analysis gives hotel managers a holistic understanding of which areas are excelling and which need improvement. By transforming complex text data into actionable insights, our AI solution enables decision-makers to implement targeted strategies that improve the overall guest experience. With informed decision-making, hotels can proactively address concerns, optimize service delivery, and cultivate lasting visitor loyalty. This paper presents state-of-the-art text classification and sentiment analysis and describes the methodology for implementing the system.

We have organized the sections of our paper according to the following structure: section 2 presents some related work in the field of text classification and sentiment analysis. Section 3 deals with the methodology adopted, the architecture of the model, and our first results for the sentiment analysis task. Section 4 concludes the paper and presents our future work.

II. RELATED WORK

This section reviews related work on text classification and sentiment analysis.

A. Text classification

Text classification is a fundamental task in Natural Language Processing (NLP) that plays a vital role in organizing text into predetermined categories or classes. Its applications are diverse and encompass a range of fields, such as sentiment analysis, document categorization, spam detection, user intent prediction, and topic classification. Over time, the task of text classification has undergone significant transformations, as traditional methods relied on creating manual features. Question Classification was previously accomplished through rule-based approaches such as keyword detection, which required time-consuming feature engineering [1].

The domain of text classification has witnessed a significant transformation with the advent of machine learning. The ability of neural networks to perform advanced language modeling has revolutionized this area, enabling them to automatically identify relevant textual information without requiring manual feature specification. This capability empowers them to capture intricate patterns and relationships between words, thereby substantially enhancing the accuracy of text classification systems.

Recurrent Neural Networks (RNNs) are a class of neural networks that have been widely used in text classification due to their ability to capture long-term dependencies in text data. Among the RNNs, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are the most frequently employed variants [2]. The LSTM and GRU architectures have proven to be effective in capturing and retaining critical contextual information of text data.

In addition, some models combine the strengths of LSTM and Convolutional Neural Network (CNN) architectures, as demonstrated in [3]. Such hybrid models have achieved superior performance in text classification tasks. In recent times, the domain of NLP has undergone a revolution due to transformers. Models like GPT (Generative pre-trained transformers) [4] and BERT (Bidirectional Encoder Representations from Transformers) [5], along with their variations, are pre-trained on extensive text collections and have demonstrated impressive performance across different natural language processing tasks, including text classification. These models have been utilized with large textual datasets, allowing them to comprehend natural language's semantics, syntax, and context. They possess the adaptability to tailor to specific text classification objectives even with limited datasets, enhancing their flexibility and accuracy.

Different sets of training and test data are used for the text classification task:

- AG News [6].
- DBpedia [7].
- TREC [8]: (TREC-6) and (TREC-50).

B. Sentiment analysis

Sentiment analysis (SA) is a technique used in Natural Language Processing (NLP) to analyze and extract information about individual opinions and emotions. It also determines the polarity of such information, whether it is positive, negative, or neutral [9]. SA is widely used in various domains such as healthcare, policy formulation, and customer feedback assessment. There are different approaches to performing sentiment analysis, including traditional, deep, and transfer learning methodologies [10].

Traditional sentiment analysis methods include rule-based approaches using emotional dictionaries [11], [12] and approaches based on supervised machine-learning techniques [13], [14]. Classifiers such as Naive Bayes [13] or SVM [14] are the most commonly used.

Deep learning has made significant strides in the field of NLP, particularly in sentiment analysis tasks, showcasing enhanced performance. Neural network architectures dedicated to sentiment analysis can be categorized into CNN-based, RNN-based, and hybrid neural network models [10]. While CNNs are traditionally linked with image processing, their effectiveness in NLP has been demonstrated, yielding impressive outcomes in text analysis [15], [16]. RNNs are commonly utilized in NLP due to their ability to handle sequential information and manage positional relationships and dependencies within text. For instance, Day and Lin [17] achieved superior results in analyzing Chinese customer reviews on Google Play using an LSTM model compared to Naive Bayes and SVMs. Hybrid models amalgamate a neural network (CNN or RNN) with other mechanisms like the attention mechanism [18] to enhance performance further.

Transfer learning is a technique that involves using pre-existing knowledge to improve the performance of a model on a related task. Instead of starting the training process from scratch for each task, transfer learning involves transferring knowledge gained from a source task to a target task. This approach utilizes insights that have been previously acquired, often from extensive tasks, to enhance the model's effectiveness and adaptability on subsequent specific tasks.

In contemporary times, pre-trained language models such as BERT, Roberta, Distilbert [19], XLNet [20], and ULMFiT [21] have exhibited significant advancements in the field of NLP. These models have the capacity to effectively capture implicit semantic relationships between words in a given text, thereby enhancing the performance of several NLP tasks. Consequently, these models have been widely used in various NLP applications and have contributed to the progress of the field.

Different datasets are used for the sentiment analysis task:

- IMDb (Internet Movie Database) [22].
- SST (Stanford Sentiment Treebank) [23].
- Yelp [24]

III. TOWARDS A DECISION-SUPPORT SYSTEM FOR HOTEL MANAGERS

In this section, we will outline our methodology and describe the architecture of the model used to solve the task. Finally, we will present our first results for the sentiment analysis task.

A. Methodology

Our approach is founded on the integration of two NLP tasks to extract meaningful information from the visitors' comments. The methodology comprises several key stages, which are as follows:

1) **Factor recognition:** Using the model described in subsection B, we automatically identify and classify comments based on relevant factors such as comfort, price-quality ratio, and cleanliness.
2) **Sentiment Analysis:** The same model employed for factor recognition will be utilized for sentiment analysis. This model has been specifically designed to categorize sentiment expressed within comments into one of three distinct classifications: positive, negative, and neutral.
3) **Correlation and Synthesis of Results:** By cross-referencing the results of factor recognition with those of sentiment analysis, we obtain a complete overview of the strengths and weaknesses perceived by visitors.

Fig. 1 shows an overview of our proposed system.

B. Description of the model

After conducting a literature review on text classification tasks, we have found that the BERT model and its variants

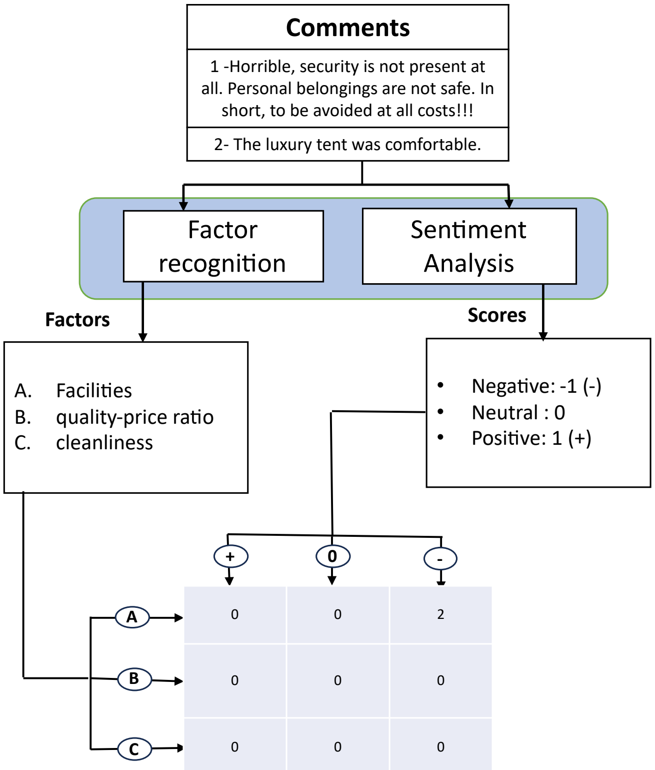

Fig. 1. Overview of our system

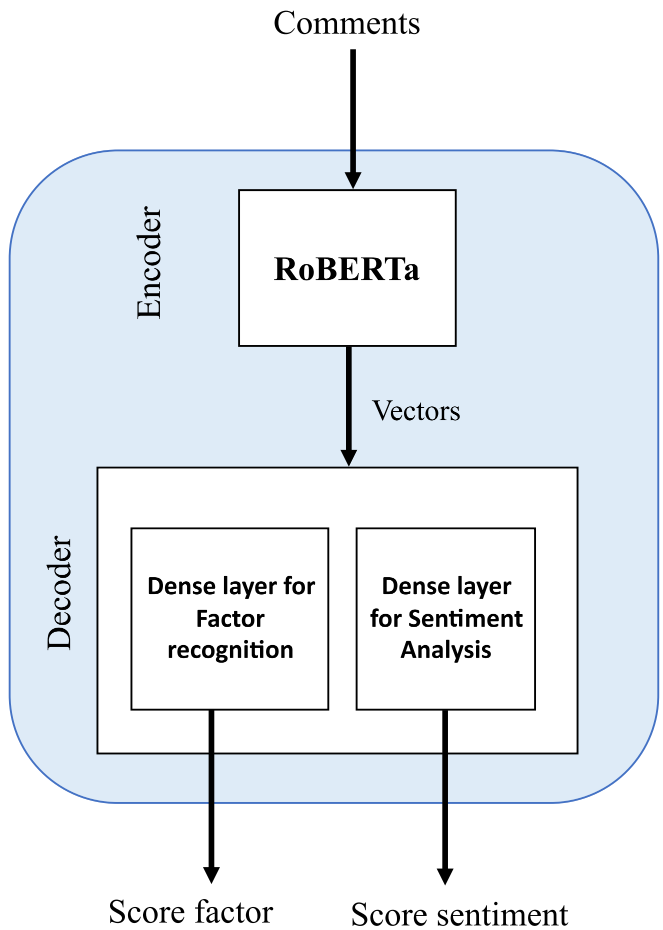

Fig. 2. Overview of the model

(such as ALBERT and RoBERTa) have shown good performance. Our model is based on the concept of fine-tuning a pre-trained model, specifically RoBERTa, known for its advanced contextual understanding capabilities in various NLP tasks such as intent classification [25]. This architecture is deployed to perform two essential tasks: factor recognition and sentiment analysis. Our proposed model, illustrated in Fig. 2 provides an overview of the model we plan to use to solve the task. The proposed model consists of three primary layers. Firstly, the input text is encoded using the RoBERTa model, with the aim of capturing and contextualizing the semantic relationships present in the data. Secondly, a dense layer is employed which uses the Rectified Linear Unit (ReLU) activation function for factor recognition. This layer serves to transform the encoded representations into probability scores that correspond to specific factor classes. The final layer is dedicated to performing sentiment analysis, and is similar to the second layer, but is designed specifically for this task.

IV. First results

For the sentiment analysis task, we used a part of the Twitter Sentiment Analysis Dataset [26]. This is an entity-level sentiment analysis dataset on Twitter. Given a message and an entity, the task is to assess the sentiment of the message towards the entity. We have pre-processed part of it to adapt it to our task. We have kept only 3 classes: positive, negative, and neutral, and we have eliminated the examples corresponding to the 'irrelevant' class. The following are the characteristics of the dataset obtained after pre-processing and splitting:

- Size of train set: 32145 (91.85%)
- Size of validation set: 1692 (4.83%)
- Size of test set: 1047 (3.32%)
- Vocabulary Size: 33953
- Number of Tags: 3

These are the hyperparameters that were used to train the model:

- Learning rate: We tested with three values - 1e-5, 3e-5, and 1e-6.
- Number of epochs: 10.
- Batch size: 32.
- Hidden size: 768
- Dropout: 0.2.

Table I summarizes the results obtained for the sentiment analysis task on the validation set. The top of these results is achieved by adjusting the learning rate (LR) to 1e-5. This configuration results in an accuracy of 0.8221 and a score of 0.8209 for the F1-score metric. Fig.3 illustrates the model's training and validation accuracy graphs. Significant progress was achieved in sentiment analysis task through the fine-tuning process of the RoBERTa model, showcasing promising outcomes. Further improvements in model performance can be attained by training the model over a substantial number of epochs and employing regularization techniques.

TABLE I
THE RESULTS OBTAINED FOR THE SENTIMENT ANALYSIS TASK USING ROBERTA

Learning_rate	Accuracy	Recall	F1_score
1e-5	0.8221	0.8198	0.8209
3e-5	0.8143	0.8109	0.8120
1e-6	0.7452	0.7427	0.7424

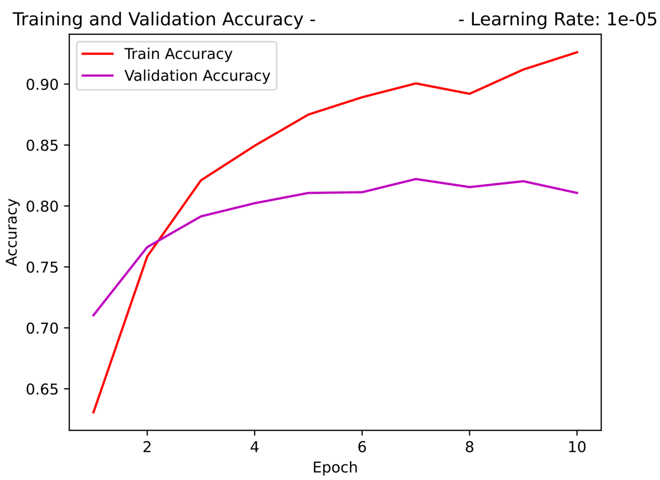

Fig. 3. Illustration of the model's training and validation accuracy graphs.

To train the model for the factor recognition task, we will need to build and annotate a dataset.

V. CONCLUSION

In this paper, we have proposed a methodology using AI for the implementation of a system that will provide hotel managers with knowledge about the degree of satisfaction of their customers. We presented an overview of solving two NLP tasks: text classification and sentiment analysis. By applying a fine-tuning methodology, we were able to evaluate our proposed model on the sentiment analysis task. The results show that RoBERTa is a promising choice for this task. This study offers many prospects for future work, including the construction of a dataset to solve the factor recognition task, the extension of our analysis to other pre-trained models such as GPT-4, Llama2, etc., and finally the full implementation of the approach to offer a tool to help hotel managers make decisions.

ACKNOWLEDGMENT

This work was supported by the Ministry of Higher Education, Scientific Research and Innovation, the Digital Development Agency (DDA) and the CNRST of Morocco (Alkhawarizmi/2020/32).

REFERENCES

[1] H. T. Madabushi and M. Lee, "High accuracy rule-based question classification using question syntax and semantics," in *Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical Papers*, 2016, pp. 1220–1230.

[2] M. Zulqarnain, R. Ghazali, M. G. Ghouse, and M. F. Mushtaq, "Efficient processing of gru based on word embedding for text classification," *JOIV: International Journal on Informatics Visualization*, vol. 3, no. 4, pp. 377–383, 2019.

[3] R. Johnson and T. Zhang, "Semi-supervised convolutional neural networks for text categorization via region embedding," *Advances in neural information processing systems*, vol. 28, 2015.

[4] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever *et al.*, "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[6] "AG's Corpus of News Articles," Available at http://groups.di.unipi.it, last visited on 2 November, 2023.

[7] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. Van Kleef, S. Auer *et al.*, "Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia," *Semantic web*, vol. 6, no. 2, pp. 167–195, 2015.

[8] E. M. Voorhees, D. M. Tice *et al.*, "The trec-8 question answering track evaluation." in *TREC*, vol. 1999, 1999, p. 82.

[9] S. R. Goniwada, "Sentiment analysis," in *Introduction to Datafication: Implement Datafication Using AI and ML Algorithms*. Springer, 2023, pp. 165–184.

[10] R. Liu, Y. Shi, C. Ji, and M. Jia, "A survey of sentiment analysis based on transfer learning," *IEEE access*, vol. 7, pp. 85 401–85 412, 2019.

[11] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 168–177.

[12] F. Bravo-Marquez, E. Frank, and B. Pfahringer, "Building a twitter opinion lexicon from automatically-annotated tweets," *Knowledge-Based Systems*, vol. 108, pp. 65–78, 2016.

[13] S. Yoshida, J. Kitazono, S. Ozawa, T. Sugawara, T. Haga, and S. Nakamura, "Sentiment analysis for various sns media using naïve bayes classifier and its application to flaming detection," in *2014 IEEE Symposium on Computational Intelligence in Big Data (CIBD)*. IEEE, 2014, pp. 1–6.

[14] L. Ting-Ting and J. Dong-hong, "Sentiment analysis of micro-blog based on svm and crf using various combinations of features." *Application Research of Computers/Jisuanji Yingyong Yanjiu*, vol. 32, no. 4, 2015.

[15] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[16] A. Conneau, H. Schwenk, L. Barrault, and Y. Lecun, "Very deep convolutional networks for text classification," *arXiv preprint arXiv:1606.01781*, 2016.

[17] M.-Y. Day and Y.-D. Lin, "Deep learning for sentiment analysis on google play consumer review," in *2017 IEEE international conference on information reuse and integration (IRI)*. IEEE, 2017, pp. 382–388.

[18] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, 2016, pp. 1480–1489.

[19] A. S. Talaat, "Sentiment analysis classification system using hybrid bert models," *Journal of Big Data*, vol. 10, no. 1, pp. 1–18, 2023.

[20] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "Xlnet: Generalized autoregressive pretraining for language understanding," *Advances in neural information processing systems*, vol. 32, 2019.

[21] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," *arXiv preprint arXiv:1801.06146*, 2018.

[22] A. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*, 2011, pp. 142–150.

[23] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a

sentiment treebank," in *Proceedings of the 2013 conference on empirical methods in natural language processing*, 2013, pp. 1631–1642.

[24] Y. D. Challenge, "Yelp open dataset," https://www.yelp.com/dataset, accessed 11 November, 2023.

[25] A. Souha, C. Ouaddi, L. Benaddi, and A. Jakimi, "Pre-trained models for intent classification in chatbot: Comparative study and critical analysis," in *2023 6th International Conference on Advanced Communication Technologies and Networking (CommNet)*. IEEE, 2023, pp. 1–6.

[26] "Twitter sentiment analysis," https://www.kaggle.com/datasets/jp797498e/twitter-entity-sentiment-analysis, accessed 15/04/2024.