

# Leveraging Knowledge Graphs with Large Language Models for Classification Tasks in the Tourism Domain

Andrea Cadeddu<sup>1</sup>, Alessandro Chessa<sup>1</sup>, Vincenzo De Leo<sup>1,2</sup>, Gianni Fenu<sup>2</sup>, Enrico Motta<sup>3</sup>, Francesco Osborne<sup>3,4</sup>, Diego Reforgiato Recupero<sup>2</sup>, Angelo Salatino<sup>3</sup> and Luca Secchi<sup>1,2,\*</sup>

<sup>1</sup>Linkalab s.r.l., Cagliari, Italy

<sup>2</sup>Department of Mathematics and Computer Science, University of Cagliari, Cagliari, Italy

<sup>3</sup>Knowledge Media Institute, The Open University, Milton Keynes, United Kingdom

<sup>4</sup>Department of Business and Law, University of Milano Bicocca, Milan, Italy

## Abstract

Online platforms, serving as the primary conduit for travelers to seek, compare, and secure travel accommodations, require a profound understanding of user dynamics to craft competitive and enticing offerings. Concurrently, recent advancements in Natural Language Processing, particularly large language models, have made substantial strides in capturing the complexity of human language. Simultaneously, knowledge graphs have become a formidable instrument for structuring and categorizing information. This paper introduces a cutting-edge deep learning methodology integrating large language models with domain-specific knowledge graphs to classify tourism offers. It aims at aiding hospitality operators in understanding their accommodation offerings' market positioning, taking into account the visit propensity and user review ratings, with the goal of optimizing the offers themselves and enhancing their appeal. Comparative analysis against alternative methods on two datasets of London accommodation offers attests to our approach's effectiveness, demonstrating superior results.

## Keywords

Knowledge Graphs, Natural Language Processing, BERT, Classification Tasks, Feature Engineering, Tourism, Hospitality

## 1. Introduction

In the age of digital transition, online platforms are a vital tool for travelers to explore and reserve travel accommodations. Yet, with the surge of information, users can find it challenging to choose the optimal option. Moreover, individual traveler preferences, such as location, amenities, and price, further complicate this task. Hence, understanding these dynamics is crucial for online platforms and revenue managers in the hospitality sector to better position their accommodations in an ever-competitive market [1]. Advancements in Natural Language Processing (NLP), notably large language models based on transformers, have greatly improved

---


*Workshop at ISWC 2023 on Deep Learning for Knowledge Graphs, 6-7 November, 2023, Athens*

\*Corresponding author.

✉ andrea.cadeddu@linkalab.it (A. Cadeddu); alessandro.chessa@linkalab.it (A. Chessa); vincenzo.deleo@linkalab.it (V. D. Leo); fenu@unica.it (G. Fenu); enrico.motta@open.ac.uk (E. Motta); francesco.osborne@open.ac.uk (F. Osborne); diego.reforgiato@unica.it (D. R. Recupero); angelo.salatino@open.ac.uk (A. Salatino); luca.secchi@linkalab.it (L. Secchi)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

automatic human language comprehension [2]. Similarly, knowledge graphs (KG) have gained recognition as valuable tools for structured information organization [3]. They provide a semantic representation of all significant entities within a domain, serving as a powerful resource for conveying valuable information to intelligent services.

However, the integration of these technologies presents challenges, primarily in blending unstructured and structured data and correctly encoding knowledge graph information [4].

This paper introduces a novel deep-learning methodology integrating large language models and domain knowledge graphs for tourism offers classification. Specifically, it augments a transformer model with a knowledge graph generated from Airbnb data to accurately categorize accommodation descriptions. We evaluated our approach against a BERT (Bidirectional Encoder Representations from Transformers) classifier and a baseline logistic regression classifier on a dataset of over 15,000 accommodation offers, yielding excellent results.

## 2. Related Work

Integrating structured knowledge into deep learning architectures has been the focal point of numerous studies. Typically the source of knowledge is encoded as knowledge graphs, as they facilitate reasoning and can be refined and enhanced by employing diverse techniques for knowledge graph completion. Furthermore, KG about tourism can be generated from a broad spectrum of data sources utilizing information extraction pipelines [5, 6, 7, 8].

In the realm of NLP, a significant body of research has been dedicated to integrate specialized knowledge with Language Models [4], such as BERT [2]. For example, Liu et al. [9] proposed to extend BERT by injecting knowledge graphs triples directly into the input text, whereas Ostendorff et al. [10] combines knowledge graphs embeddings and metadata to complement the information presented to BERT through text.

Previous studies in peer-to-peer accommodation business (like Airbnb) were focused on the pricing issues [11] or in detecting the booking likelihood [12]. However, these methods do not consider the textual description of each accommodation and focus on numeric features, missing an important factor in the tourist's choices. Consequently, we introduced a novel approach that can also process textual descriptions leveraging the combination of knowledge graphs and Language Models within the context of tourism, with the aim of addressing practical use cases, such as accommodation classification.

## 3. Background

Our solution was designed to support the optimization of an accommodation offer by solving two classification tasks derived from collaborative discussions with stakeholders and revenue managers<sup>1</sup>. The first task is *visit propensity* classification which aims to predict if the accommodation would be visited or not. This is also called the booking likelihood in other studies [12]. Airbnb allows a user to write a review only after the visit. Therefore, if the accommodation

---

<sup>1</sup>Within the tourism industry, an individual responsible for maximizing the performance of an accommodation is commonly known as a "Revenue Manager" or a "Revenue Optimization Manager."

received at least one review (visit) in the previous year, the label is set to *high propensity*; otherwise to *low propensity*. The second task is *review rating* classification, which aims to predict if an accommodation would have a high review rating ( $> 4.5$  over a 1-5 Likert scale). They are encoded as binary classification tasks, serving as practical checklists for revenue managers<sup>2</sup>.

The resulting characterization of an accommodation equips users with the capability to scrutinize the quality of the offer and investigate an array of strategies for its enhancement. Furthermore, it enables assessing how potential changes would impact the predicted dimensions.

To support the methodology described in this paper, we adopted the following resources:

- Tourism Analytic Ontology (TAO) [5], an ontology designed to describe the complex dynamics of the tourism domain and support intelligent services in this space, that we used to model accommodation’s data.
- London Tourism Knowledge Graph (TKG), a knowledge graph based on TAO and built by following the methodology introduced in [5] and applied in [6]. It covers Airbnb’s London accommodations and is based on open data from the Inside Airbnb project<sup>3</sup>. It was used as a source for factual knowledge.
- DBpedia Spotlight [13], a cross-domain entity linking tool built upon DBpedia public knowledge graph, we used it for extracting supplementary information from the descriptions of Airbnb accommodations and extend the Knowledge Graph.
- BERT, a well-known language model renowned for its ability to capture extensive contextual representations of both words and sentences [2]. We leveraged BERT as the core element for our text classification system.

Since TKG is stored in a triple-store database, we leverage the SPARQL language to extract all relevant data. Subsequently, a data pipeline generates three distinct datasets that are utilized for feature engineering. *Dataset 1* associates each accommodation with its description text and various properties represented as numbers, dates, or Boolean flags. *Dataset 2* associates each accommodation with its included amenities<sup>4</sup> expressed using TAO classes. *Dataset 3* associates each accommodation with related DBpedia entities, which are extracted from accommodation descriptions using DBpedia Spotlight.

## 4. Methodology

We introduce a methodology that combines a Transformer model for text processing and a Multi-Layer Perceptron (MLP), used for incorporating additional features. This solution allows knowledge infusion from TKG and uses *Datasets 1, 2 and 3* to produce the following features: (i) *textual features*, serving as a natural input for BERT, are derived from accommodation descriptions in *Dataset 1* using a BERT tokenizer<sup>5</sup>, (ii) *numerical features* are produced from numerical

---

<sup>2</sup>The use of just two states high/low is a design choice that gives managers good confidence in pursuing simple and concrete objectives thanks to high accuracy levels in the predictions.

<sup>3</sup>See <http://insideairbnb.com/about>.

<sup>4</sup>Amenities refer to additional services, features, or facilities provided to guests during their stay.

<sup>5</sup>The tokenizer is responsible for dividing input text into individual tokens and applying additional tokenization techniques, such as splitting words into subwords or adding special tokens for tasks like sentence classification or question answering.

data in *Dataset 1*, like the number of rooms and beds, and represented as a  $n$  dimensional vector<sup>6</sup> whose elements are normalized to have all values within the range  $[0, 1]$ , (iii) *categorical features* i.e., the list of amenities for each accommodation, from *Dataset 2*, are converted into numeric vectors (of size 145) through the process of one-hot encoding, and (iv) *linked entities*, which are DBpedia entities extracted from the descriptions and stored in *Dataset 3*, are also transformed into numeric vectors (of size 625) through a one-hot encoding process after those associated with less than 100 accommodations are filtered out.

The training phase employs an end-to-end optimization for each classification task. It involves fine-tuning the BERT transformer on the description set, and training the MLP from scratch on the BERT output combined with all other additional features. More in detail, the tokenized text is processed by BERT using the English uncased model<sup>7</sup>. The [CLS] hidden vector state of size 768 is used as the BERT output. The  $\tanh$  output from the pooling layer tied to BERT is scaled between 0 to 1 to match other non-textual feature vectors. Numeric features are normalized real number vectors, and categorical and linked entities' features are one-hot encoded. The four vectors are concatenated and fed into the MLP, composed of two layers with 1024 units each and a ReLU activation function. All MLP layers undergo dropout (with default probability  $p=0.1$ ) during training to counter overfitting. The output layer of the MLP is a Sigmoid layer, providing the classification probability output.

## 5. Evaluation

To evaluate the proposed classification methodology, we created a balanced dataset for each of the two classification tasks, leveraging the full datasets defined in Section 3. We split the balanced data set of each task into three parts: train, validation, and test set. Subsequently, starting from the complete training set, we produced four distinct training sets for each task, progressively increasing in size to contain 3000, 6000, 9000, and 12000 accommodations. The objective of this process is to evaluate the influence of varying data quantities on the performance of the classification tasks. We used the same validation and test set for each task with a fixed size of 1800 elements in order to obtain comparable results. We assessed the performance using precision, recall, and F1 score.

To ensure robustness, we performed the training and hyperparameter tuning process five times, resulting in five model versions for each experiment. Every model variant was assessed using the test set, and the average metric values were subsequently computed. This approach was employed to account for the potential accuracy fluctuations observed in previous studies when fine-tuning BERT-based models multiple times on the same dataset with varying random seeds [2, 14].

We evaluated three approaches:

1. *Logistic Regression*, a Logistic Regression Classifier used as a baseline<sup>8</sup>.

---

<sup>6</sup>With  $n = 15$  for review rating classification task and  $n = 12$  for visit propensity classification task.

<sup>7</sup>See Hugging Face repository <https://huggingface.co/bert-base-uncased>

<sup>8</sup>It is a network composed of one hidden layer, with a single unit, ReLU activation, and a Sigmoid layer for binary classification probabilities. Unable to process text, it is fed only numerical, categorical, and linked entity features.

**Table 1**

Visit propensity classification task: average values for F1-score, precision and recall results.

max_train_size experiment	f1-score %				precision %				recall %			
	3000	6000	9000	12000	3000	6000	9000	12000	3000	6000	9000	12000
LOGISTIC REGRESSION	72.94	76.30	78.05	78.68	73.09	76.74	78.06	79.08	72.98	76.40	78.05	78.76
BERT	64.44	67.29	66.97	67.92	65.57	66.82	66.96	67.57	64.76	68.14	67.61	68.53
BERT complemented with KG	<b>85.16</b>	<b>85.67</b>	<b>85.63</b>	<b>86.02</b>	<b>85.45</b>	<b>85.86</b>	<b>86.10</b>	<b>86.32</b>	<b>85.21</b>	<b>85.70</b>	<b>85.70</b>	<b>86.06</b>

**Table 2**

User review score classification task: average values for F1-score, precision and recall results.

max_train_size experiment	f1-score %				precision %				recall %			
	3000	6000	9000	12000	3000	6000	9000	12000	3000	6000	9000	12000
LOGISTIC REGRESSION	56.64	61.99	64.75	65.62	56.67	62.00	64.76	65.72	56.65	62.00	64.75	65.64
BERT	61.97	63.90	64.16	64.52	63.07	63.90	64.44	64.81	61.64	64.61	64.09	64.70
BERT complemented with KG	<b>69.76</b>	<b>69.43</b>	<b>69.38</b>	<b>69.78</b>	<b>70.47</b>	<b>70.67</b>	<b>70.98</b>	<b>71.51</b>	<b>69.90</b>	<b>69.70</b>	<b>69.74</b>	<b>70.16</b>

2. *BERT*, a BERT-based uncased model, trained on textual features<sup>9</sup>.
3. *BERT complemented with KG*, the methodology introduced in Section 4.

Table 1 presents the findings for the *visit propensity* classification task. The BERT model complemented with the KG outperforms all other approaches, showcasing its superior performance due to its capability to incorporate diverse features.

The results for *review rating* classification task are reported in Table 2. The task of *review rating* appears to be more demanding than the *visit propensity* classification task, as indicated by the relatively lower F1-score. This can be attributed to the nature of user reviews, which are provided post-visit and influenced by the actual experiences encountered by the users. Predicting these experiences solely based on the available accommodation information poses a significant challenge for the model.

As before, the experimental outcomes show that our approach performs better than the vanilla BERT model. This is because the proposed approach can take advantage of all available features (textual, numerical, categorical, and linked entities).

## 6. Conclusions

This paper introduces a novel methodology that integrates language models and knowledge graphs to enhance two classification tasks about accommodation offers in the tourism domain. To improve classification accuracy, we combined BERT with a knowledge graph to provide numeric data, categorical information, and linked entities. Our approach outperformed BERT with a mean increase of 12.5% points in F1 score.

Future research directions will focus on enhancing the methodology for multi-class classification and regression and explore how effectively a classifier trained for a specific tourist location, such as London, could be transferred and applied to a different destination.

<sup>9</sup>The pooled output from BERT [CLS] token is fed to a final inner layer with one unit followed by a Sigmoid layer.

## References

- [1] L. González-Serrano, P. Talón-Ballester, Revenue Management and E-Tourism: The Past, Present and Future, Springer International Publishing, Cham, 2020, pp. 1–28. URL: [https://doi.org/10.1007/978-3-030-05324-6\\_76-1](https://doi.org/10.1007/978-3-030-05324-6_76-1). doi:10.1007/978-3-030-05324-6\_76-1.
- [2] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, A. I. Language, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *Naacl-Hlt 2019 (2018)* 4171–4186.
- [3] C. Peng, F. Xia, M. Naseriparsa, F. Osborne, Knowledge graphs: opportunities and challenges, *Artificial Intelligence Review (2023)* 1–32.
- [4] J. Yang, G. Xiao, Y. Shen, W. Jiang, X. Hu, Y. Zhang, J. Peng, A survey of knowledge enhanced pre-trained models, *CoRR abs/2110.00269 (2021)*. URL: <https://arxiv.org/abs/2110.00269>. arXiv:2110.00269.
- [5] A. Chessa, G. Fenu, E. Motta, F. Osborne, D. Reforgiato Recupero, A. Salatino, L. Secchi, Data-driven methodology for knowledge graph generation within the tourism domain, *IEEE Access 11 (2023)* 67567–67599. doi:10.1109/ACCESS.2023.3292153.
- [6] A. Chessa, G. Fenu, E. Motta, F. Osborne, D. R. Recupero, A. Salatino, L. Secchi, Enriching Data Lakes with Knowledge Graphs, *CEUR Workshop Proceedings 3184 (2022)* 123–131.
- [7] R. Troncy, G. Rizzo, A. Jameson, O. Corcho, J. Plu, E. Palumbo, J. C. Ballesteros Hermida, A. Spirescu, K. D. Kuhn, C. Barbu, M. Rossi, I. Celino, R. Agarwal, C. Scanu, M. Valla, T. Haaker, 3cixty: Building comprehensive knowledge bases for city exploration, *Journal of Web Semantics 46-47 (2017)* 2–13. URL: <http://dx.doi.org/10.1016/j.websem.2017.07.002>. doi:10.1016/j.websem.2017.07.002.
- [8] D. Gazzè, A. L. Duca, A. Marchetti, M. Tesconi, An overview of the tourpedia linked dataset with a focus on relations discovery among places, *ACM International Conference Proceeding Series 16-17-Sept (2015)* 157–160. doi:10.1145/2814864.2814876.
- [9] W. Liu, P. Zhou, Z. Zhao, Z. Wang, Q. Ju, H. Deng, P. Wang, K-BERT: enabling language representation with knowledge graph, *CoRR abs/1909.07606 (2019)*. URL: <http://arxiv.org/abs/1909.07606>. arXiv:1909.07606.
- [10] M. Ostendorff, P. Bourgonje, M. Berger, J. M. Schneider, G. Rehm, B. Gipp, Enriching BERT with knowledge graph embeddings for document classification, *CoRR abs/1909.08402 (2019)*. URL: <http://arxiv.org/abs/1909.08402>. arXiv:1909.08402.
- [11] L. R. Tang, J. Kim, X. Wang, Estimating spatial effects on peer-to-peer accommodation prices: Towards an innovative hedonic model approach, *International Journal of Hospitality Management 81 (2019)* 43–53. URL: <https://www.sciencedirect.com/science/article/pii/S0278431918306984>. doi:<https://doi.org/10.1016/j.ijhm.2019.03.012>.
- [12] M. A. Afrianto, M. Wasesa, Booking prediction models for peer-to-peer accommodation listings using logistics regression, decision tree, k-nearest neighbor, and random forest classifiers, *JISEBI 6 (2020)* 123–32.
- [13] P. N. Mendes, M. Jakob, A. García-Silva, C. Bizer, DBpedia spotlight: Shedding Light on the Web of Documents, in: *Proc. of the 7th Int. Conf. on Semantic Systems - I-Semantics '11*, ACM Press, New York, New York, USA, 2011, pp. 1–8. doi:10.1145/2063518.2063519.
- [14] J. Dodge, G. Ilharco, R. Schwartz, A. Farhadi, H. Hajishirzi, N. Smith, Fine-Tuning Pre-trained Language Models: Weight Initializations, Data Orders, and Early Stopping, *CoRR abs/2002.0 (2020)*. arXiv:arXiv:2002.06305v1.