Ranking Online User Reviews for Tourism Based on Usefulness

Konstantinos Tsamis
Andreas Komninos
[ktsamis,akomninos]@ceid.upatras.gr
Computer Technology Institute and Press "Diophantos"
Rio, Greece

Konstantinos Kovas
Nikolaos Zotos
[kkovas,nzotos]@knowledge.gr
Knowledge Broadband Services SA
Patras, Greece

ABSTRACT

The growth of Web 2.0 services has lead to an increase in the volume of user-generated content in the form of online user reviews. Web platforms offering users the ability to evaluate the services of hotels have increased in popularity, as a large percentage of travellers offer their feedback or read hotel reviews to assist their decision making process. Users usually do not have time to go through the sheer volume of available hotel reviews and would prefer to read the most useful ones, whereas review usefulness is subjective and depends on the reader's needs and preferences. Therefore, the need for automatically detecting hotel review helpfulness arises.

In this paper, we propose the use of features that capture both textual content and review metadata for predicting hotel review helpfulness of Greek and English reviews. A novel approach for representing text as a word embeddings-based vector is introduced and review association with certain hotel service aspects is mapped. Evaluating the performance of our approach using Machine Learning and Neural classifiers yields promising results for the review helpfulness classification task.

CCS CONCEPTS

• **Computing methodologies → Classification and regression trees**; **Supervised learning by classification**; **Information extraction**; *Neural networks.*

KEYWORDS

User-generated content, Review helpfulness classification, Machine learning, Neural networks, Natural language processing

ACM Reference Format:
Konstantinos Tsamis, Andreas Komninos, Konstantinos Kovas, and Nikolaos Zotos. 2020. Ranking Online User Reviews for Tourism Based on Usefulness. In *PCI '20: Panhellenic Conference on Informatics, November 22–25, 2020, Athens, Greece.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3437120.3437308

1 INTRODUCTION

Much of today's travel and tourism activity planning takes place online. The decision process for travellers is heavily influenced by the reviews left on various websites and apps by other travellers. Reviews can include a range of information depending on the platform they are being shared on, but users frequently are called upon to provide an overall score (e.g. 8/10), itemized scores (e.g. 4/5 stars for cleanliness), recommendations (e.g. "good for families") and free-form text of their opinion.

Reviewing is a popular activity for users of online travel planning services. It is also heavily promoted by the services themselves, in order to solicit as much information about places as possible, for the benefit of other platform users. On one such service (TripAdvisor), contained 859 million reviews and opinions in 2019 alone [3]. On one hand, the rich ecosystem of users and opinions can ultimately benefit the whole community. On the other hand, the popularity of reviewing activity is such that the sheer volume of information available to users is simply impossible for the user to digest in its entirety. Problems in online reviews relate to the number of reviews a user is prepared to read in order to form an opinion, the validity of the reviews in terms of their recency and trustworthiness of the reviewer (e.g. genuine travellers vs. fraudulent reviewers, experienced vs. inexperienced travellers), the quality of information contained therein (volume and detail of information), the user's personal preferences (e.g. being particularly concerned about specific aspects of a venue, such as cleanliness or quietness) etc.

Overall, in order to provide an effective service, a travel platform must present to their users not just all reviews, but those which are likely to be most helpful to the user. Some platforms implement this by allowing users to vote for the helpfulness of a review, but this is a manual process and could lead to helpful reviews being ignored. In any case, the notion of "helpfulness" is fluid - for example, it's not clear to a reader if past readers of the review indicated it as "helpful" while reading them before the actual trip, or after their trip was completed and a visit to the venue had actually taken place. What makes for a "helpful" review is not strictly defined and thus can be difficult to assess.

In this paper, we present research towards the automatic classification of online tourism reviews in terms of their "helpfulness". We introduce a range of metrics designed to capture aspects of trustworthiness and information quality in a review, and perform experiments in a real-world dataset derived from TripAdvisor. Our main contributions are:

- Applying both text-based and non-textual features for the review usefulness classification, on reviews written in both the Greek and English language, using a common approach for both languages.
- Using a text vector based on word-embeddings for representing review texts, instead of text-based features such as

Table 1: Factors affecting the perceived helpfulness of tourism business reviews

Review Aspect	Review Factor	References
Author	Identity disclosure	[13], [4], [12]
	Expertise	[13], [10], [12]
	Reputation	[13], [4]
	Author rating distribution	[6]
	Author gender	[10]
	Author locality	[12]
Review metadata	Review rating	[13], [10], [20], [12]
	Review age	[9]
	Review time on homepage	[9]
	Hotel star class	[9]
	Manager response	[10]
	Review helpful votes	[20]
Review text	Text readability	[13], [6], [10], [12]
	Text sentiment	[13], [6],
	Text Reading enjoyment	[13]
	Text length	[13]

unigrams or bigrams. The idea behind this, is that a document should be semantically related to the sum of its words.

2 RELATED WORK

The usefulness of online tourism reviews has been investigated in several recent publications. Examining some of the most highly cited works concerning review helpfulness determinants in tourism businesses (e.g. hotels and restaurants), we find a range of factors deemed important at a different level in each study, but overall there appears to be some commonality across most of these (Table 1). It's important to note that in each study, the results seem heavily dependent on the dataset examined by the authors. Most datasets cover only one or two cities, and hence, depending on the type of visitors this city receives, it is likely that the significance of determinants varies according to population characteristics and experience. In [19] it is also shown that review linguistic characteristics, semantic features, sentiment and usefulness can vary significantly depending on the platform where reviews are posted. Furthermore, in [8], it is shown that aspects of the service rated in a review can significantly affect the "helpfulness" rating, depending on the hotel class the review relates to. For example, text content relating to amenities is more important when a 4 or 5 star hotel is being reviewed, while lower-class hotel reviews are more helpful if they report on aspects such as convenience and value. Therefore, previous work can be considered relatively limited in generalisability. On the other hand, in [7] a meta-analysis of review helpfulness factors literature across a range of domains (e.g. tourism, e-commerce, online services) indicates that most of the determinants in Table 1 are likely to have a significant role across multiple domains. These include Review Metadata (e.g. rating, extremity, age); Reviewer-related characteristics (e.g. reputation, identity, social network); Review readability; Syntactic features; Semantic features; Lexical features (e.g. unigrams, bigrams, spelling errors).

More recently, there have been a few attempts to use machine learning (ML) approaches in order to automatically classify hotel and tourism-related business reviews in terms of their helpfulness [5]. This approach contrasts previous work, which relied mostly on the application of logistic regression models, which, however, have the advantage of being able to explain factor importance, compared to the black-box approach used in ML. Although a substantial body of work exists for other product reviews (e.g. as sold on Amazon), hotel review helpfulness has received less attention. In fact we were able to identify only a handful of papers that present related results. The earliest example is [16] where data scraped from TripAdvisor are used to classify reviews using the JRip algorithm and achieving a mean AUC score of 0.82. Features used included metrics on author reputation, text content, social activity of authors and text sentiment. The choice of these features was not grounded in other work. In [15], authors use text sentiment analysis, word emotionality, part-of-speech tagging and a range of text statistics as features using several classifiers, and find that emotionality significantly improves classification performance (>0.85 AUC, using RF-SMOTE). In [11], again a range of classifiers is employed with features extracted from a much larger dataset (1.1m TripAdvisor reviews from 5 US cities). In this study, Random Forests outperform all other classifiers (mean AUC = 0.906). Examining RF performance using features from specific categories only, they report that using just features related to the review author, the performance is very close to using all of the features together. On the other hand, work in [17] demonstrated that reasonable performance (F1 = 0.7565, using RF) can be achieved using linguistic features of the review text alone, extracted through NLP techniques. Finally, in [14], the authors combine textual and photo content of reviews to predict helpfulness. Text features are extracted with an LSTM algorithm, while photo features are extracted using CNN. The combined feature set is then used in an LSTM-based model. The results demonstrated that using text features alone, an F1-score of 0.70 is attainable, while adding the photo features improves performance to 0.78.

In our paper, we rest upon the recent observation of [8] that commentary and ratings on different aspects of service can influence the helpfulness of a review. We introduce features based on this concept, and use a range of other features found in previous literature as well, in order to predict review helpfulness using ML-classifiers. We also investigate the performance of classification for reviews in two languages, English and Greek, something which has not been attempted previously. Finally, another novel element in our work is to introduce the representation of a document as a single vector based on the use of word embeddings.

3 METHODOLOGY

3.1 Data and preprocessing

For this paper, we scraped the TripAdvisor website for hotel reviews across the whole of Greece. We collected 59,792 reviews in Greek and 65,243 reviews in English. The data collected included the review title and text, additional text (tips), review rating, travel type, additional ratings on individual hotel aspects, number of reviews posted by the author, number of helpful votes received by this review, and number of helpful votes received by the author for all their reviews.

Past literature emphasises the importance of author reputation for helpfulness labelling. We calculated the average helpful votes per review (H-Ratio) for each author. We label "helpful" those reviews that meet all of the following conditions: a) Received 4 or more helpful votes; b)the author's H-Ratio ≥ 0.5, and; c) the author's H-Ratio ≤ the review's helpful votes.

As a result, we identified 2,316 helpful reviews in Greek and 3,457 helpful reviews in English. To create balanced datasets for training and testing, we randomly selected an equal number of unhelpful reviews from each language.

Finally, we pre-processed review text, in order to remove URLs, hashtags and user mentions. Based on our previous work, we replaced emoticons that express a positive emotion with the word `"posemoji"` and negative emotion with `"negemoji"`. Consecutive punctuation marks are replaced by corresponding labels (e.g. "..." → `"multidot"`, "!!!!" →`"multiex"`). Finally, we identified negation-marking words in each language (e.g. in English, "not, don't, wasn't") and replaced them with the labels `"grnot"` and `"not"` respectively. This pre-processing is necessary for the sentiment analysis used in the features described in the next section.

3.2 Selected Features

3.2.1 Text-based features. Since text statistics such as length and word count do not appear to play major roles in helpfulness according to literature, we emphasised use of qualitative and semantic features. The following were selected:

- Word Embeddings: We use the word2vec approach to model review text into a vector space. For that purpose, we opted to use *fastText* [1] pre-trained word vectors, with dimension 300, for both languages. The vector space produced by these models, includes words in such a manner so that semantically related words are placed in close proximity. We merge the review title, main text and additional text in tips into one document for each review. All the word vectors of words occurring in the text, are added together and then normalised by dividing the resulting vector by its length. To create the vectors, we take into account the term occurrence rather than term frequency.
- Adjective ratio: Using a Part-Of-Speech tagger for each of the two languages, we identify the percentage of adjectives over the number of words in the whole text of a review.
- Readability: To assess review readability, we use the Guiraud's R metric. This is calculated as the number of different parts-of-speech in a text, divided by the square root of the total number of words in the text.
- Subjective sentences: A sentence in the review text which expresses a positive or negative opinion is termed a subjective sentence. We use a previously developed sentiment analysis algorithm for English and Greek [18] to identify the number and the percentage of subjective sentences over the total number of sentences in the text.
- Aspect-based similarity: As per [8], reviews can be more helpful if they pertain to specific service aspects of the hotel. We selected the aspects of "Price", "Service quality", "Location, Rooms", "Cleanliness", "Sleep quality" and the generic aspect "Hotel" as concepts that a review could contain. We then calculated word embedding vectors for each of these aspects (in both languages) and used the cosine similarity measure to identify how "close" each review is to each of these aspects.

3.2.2 Additional features. Although literature emphasises the importance of author-related metrics, we chose not to employ these, since we are interested in the objective usefulness of a review based on its content, rather than the identity or reputation of its author. Thus, we incorporated the following features:

- Rating: the overall rating given to the hotel by a reviewer
- Additional Ratings: the number of ratings to individual hotel aspects given by the reviewer
- Travel type: The self-reported type of trip the review relates to, encoded as *[0=None, 1=Alone, 2=Couple, 3=Family, 4=Friends, 5=Business]*

4 IMPLEMENTATION

We designed and implemented our method for use as a web API for automatically detecting the helpfulness of reviews. The system consists of three main phases; (a) text pre-processing, (b) features calculation, (c) review classification. In the text pre-processing step, text goes through all the pre-processing tasks that we described earlier. During the features calculation phase, the feature vector of the review is created by calculating all the features. Text is split into tokens and for each unique token search is performed in the appropriate pre-trained word vectors model, based on the text's language, in order to construct the text vector. Aspect-based similarities are then calculated, using cosine similarity.

Next, the sentiment polarity of each sentence is determined using the sentiment classification algorithm. Positive and negative sentences are added together to form the subjective sentences feature, while also the percentage of subjective sentences is calculated. After identifying the part-of-speech for each word of the text using a POS tagger, Guiraud's R and percentage of adjectives is computed.

Finally, during the last phase, the feature vector of the review (including non-textual features) is fed into an SVM classifier that uses pre-trained models for each language, which have been trained using the balanced Trip Advisor hotel review datasets. For the SVM classifier, we opted to use the `libsvm` implementation in the `php-ml` [2] library, and optimized its parameters using grid search.

5 RESULTS

5.1 Experiment 1 - basic ML classifiers

To evaluate the performance of our method for predicting the helpfulness of online hotel reviews, we conducted a series of experiments using several ML classifiers on the balanced Trip Advisor datasets for both languages. The Greek dataset contains 4,632 hotels reviews, whereas the English dataset consists of 6,914 hotel reviews. Both datasets contain an equal number of heplful and non-helpful reviews. We used three ML algorithms in order to assess their performance in the helpfulness classification task, as per [11]; SVM, Decision Tree and Random Forest. The RapidMiner environment was used to perform the relevant modelling and evaluation tasks.

For selecting the appropriate hyperparameters, we deployed a grid-search optimization process over 20% of the training set from each language, to avoid overfitting. The parameters determined by this process were then used to perform classification in the whole dataset, using 10-fold cross validation. In Table 2, we present the parameters for each combination of algorithm and dataset.

Table 2: ML algorithm parameters

Dataset	Algorithm	Parameters
Greek Reviews	SVM	Kernel: rbf; C : 0.251; γ : 0.167; ϵ : 0.01
	Decision Tree	Criterion: Gain Ratio; Max. Depth: 10; Confidence: 1.0$E-7$; Min. Leaf Size: 6
	Random Forest	N. Trees: 22; Criterion: Gini Index; Max. Depth: 90
English Reviews	SVM	Kernel: rbf; C : 15848.9; γ : 0.002; ϵ : 0.01
	Decision Tree	Criterion: Accuracy; Max. Depth: 10; Confidence: 1.0E-7; Min. Leaf Size: 1
	Random Forest	N. Trees: 61; Criterion: Information Gain; Max. Depth: 29

After pre-processing the text and computing the features using the implementation presented earlier, we feed the ML-Classifiers using all the available textual and non-textual features, to predict review helpfulness of hotel reviews for both languages. The evaluation metrics we used to assess classification performance are the Accuracy, Precision, Recall and F-score of the classification. In Table 3, we present the average evaluation metrics by ML algorithm for the classification of hotel reviews, but for the rest of this section we focus on *F-score*, to compare directly with [11], and *Accuracy*, since our dataset is balanced. As we can see, for Greek reviews, SVM performs better than DT and RF achieving an average Accuracy and F-score of 79,77% and 79,84%, respectively. DT is second best coming close to the performance of SVM, while SVM and DT outperform RF by a healthy margin. For English reviews, results show that, as before, SVM achieves the best performance with an average Accuracy of 80,46% and F-Score of 80.48%, with RF coming second and DT being outperformed by both SVM and RF.

Table 3: Evaluation Metrics By ML Algorithm

Greek Reviews				
Algorithm	**Accuracy**	**Precision**	**Recall**	**F-score**
SVM	79.77% (σ = 1.78%)	79.90% (σ = 1.81%)	79.77% (σ = 1.78%)	79.83%
DT	78.65% (σ = 1.58%)	78.85% (σ = 1.57%)	78.65% (σ = 1.57%)	78.75%
RF	73.53% (σ = 2.09%)	73.68% (σ = 2.07%)	73.53% (σ = 2.09%)	73.60%
English Reviews				
Algorithm	**Accuracy**	**Precision**	**Recall**	**F-score**
SVM	80.46% (σ = 1.65%)	80.50% (σ = 1.67%)	80.46% (σ = 1.65%)	80.48%
DT	75.07% (σ = 1.71%)	75.14% (σ = 1.72%)	75.07% (σ = 1.71%)	75.10%
RF	77.23% (σ = 1.56%)	77.44% (σ = 1.57%)	77.24% (σ = 1.56%)	77.34%

Next, comparing the classification performance of each algorithm per language, results show that SVM's performance is similar for both languages, with slightly better performance for English reviews. DT performs better for Greek reviews by a margin of approximately 3.6% in F-score compared to English reviews. On the other hand, RF achieves better results for the English language with the difference in F-score being approximately 4.7%

The next step of the evaluation process, was to perform classification using different sets of features in order to assess the effect certain features have in classification performance. For that purpose, we chose to examine classification performance by using two different combinations of features; only the text-based features and only the non-textual features (rating, additional ratings, travel type). As before, we followed the same procedure for selecting the parameters of each algorithm in order to perform classification using the aforementioned feature combinations for both languages.

Fig. 1 presents the F-score of the classification of Greek reviews for each algorithm, using all features, text-based features and non-textual features only. As we can see, classification using only the text-based features yields worse results for all classifiers compared to using the whole set of features, with the biggest difference observed for the DT classifier. On the contrary, when applying only the non-textual features performance is either on the same level (SVM, DT) or better (RF) compared to using all features.

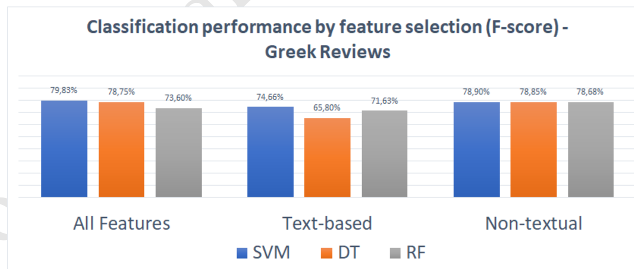


Figure 1: Classification performance of ML algorithms by feature selection - Greek reviews

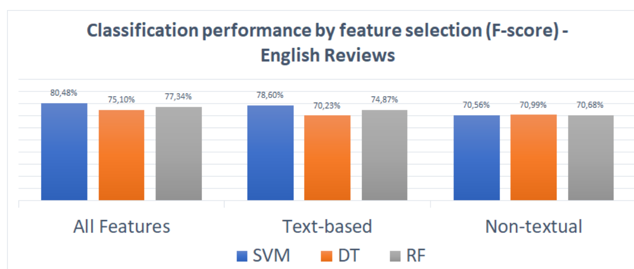


Figure 2: Classification performance of ML algorithms by feature selection - English reviews

The corresponding results for the classification of English reviews, are presented in Fig. 2. Results show that using only text-based or non-textual features has a diminishing effect on the classification performance. Contrary to results regarding Greek reviews, text-based features achieve better performance than non-textual for each classification algorithm. The combination of both text-based and non-textual features results to the best performance.

5.2 Experiment 2 - Neural classifiers

Further from these results, we also employed a more modern approach by exploiting the deep learning capabilities offered by the

Tensorflow framework (using Keras and Tensorflow 2.1 over RStudio). Our evaluation metric is *accuracy*, since we have a balanced dataset and do not directly compare with previous results.

We started with the evaluation using only the textual components of the review. This was done by analysing three scenarios in each language, using the main review text only, the review title only and finally a concatenation of the title, main text and additional tip text (where present). Text pre-processing involved tokenizing the text, removing Greek and English stopwords, symbols, numbers and punctuation, using the *quanteda* R package for NLP. Next, we created a dictionary of the *top N* words in the texts (by frequency) and generated integer-encoded tensors for each text through this dictionary. Our network consists of one embedding layer with the dictionary length as the input size and 16 output units, a 1-D global average pooling layer, a further dense layer with 16 units (*RELU* activation) and a 0.5 dropout to avoid overfitting. The output layer consists of 1 unit (since we are doing binary classification) with sigmoid activation. An ADAM optimiser is used, and binary cross-entropy is used as the loss function during training.

First we experimented with the size of the dictionary. As can be seen in the model training history in Fig. 5, a larger dictionary yields better accuracy for both languages, hence we proceeded with 500 words as the dictionary size.

Using a single training-test run with an 80-20 stratified split on the dataset (and splitting the training dataset again 80-20 to evaluate training performance), the best results in terms of accuracy are with either the main review text alone or all text (*All text* EL:69.87%, EN:75.11%; *Review only* EL:69.98%, EN:74.89%; *Title only* EL:50.00%, EN:59.99%). Performance is consistently better with English. Based on these results, we performed a k-fold cross validation ($k = 10$) on the "All text" scenario. The results are quite similar to the single run (EL: $\mu = 69.97\%, \sigma = 2.59\%$; EN: $\mu = 75.40\%, \sigma = 2.00\%$).

Next, we examine the classification performance using only numeric features. We selected the trip type, review rating, Guiraud's R and the aspect-based similarities of review text. Trip type and review rating were one-hot encoded due to their categorical nature. The network consists of a dense input layer with the same size as the number of features, a 32-unit dense hidden layer (*RELU activation*) and a single-unit output layer. An ADAM optimiser is used, and binary cross-entropy is used as the loss function.

The first run was a single training-test run with an 80-20 stratified split on the dataset (and splitting the training dataset again 80-20 to evaluate training performance). A k-fold cross validation ($k = 10$) revealed a considerable performance improvement over text-only for Greek, and comparable performance for English (EL: $\mu = 77.05\%, \sigma = 1.58\%$; EN: $\mu = 73.81\%, \sigma = 2.04\%$).

Next, we attempted to combine the text and numeric inputs using a complex model, which receives the two types of dataset as input, performs processing via straight DNN or word-embedding based DNN analysis and merges the results. The architecture is shown in Fig. 4. The results are more favourable for the English dataset (EL: $\mu = 75.60\%, \sigma = 2.53\%$; EN: $\mu = 79.16\%, \sigma = 1.25\%$). Overall, as can be seen in Fig. 3, the results are mixed depending on the language of each dataset. For Greek, best performance is captured using the numeric data only, while the treatment of English reviews benefits from the combination of numeric and text data. One plausible explanation for this is that English and Greek languages may benefit from different pre-processing steps (e.g. stemming, lemmatization) which we did not employ, and which may be hindering the effectiveness of the text analysis elements.

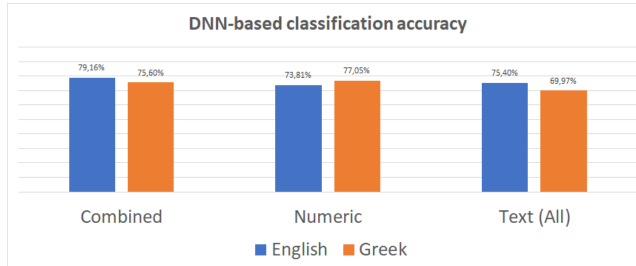


Figure 3: DNN model results

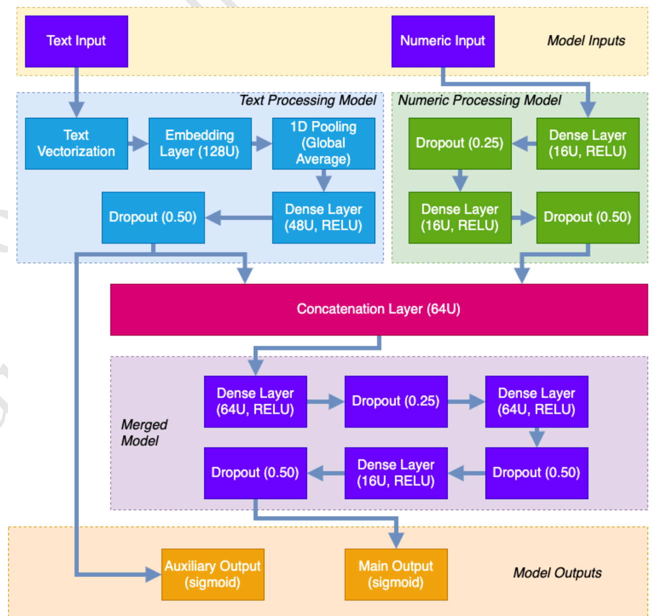
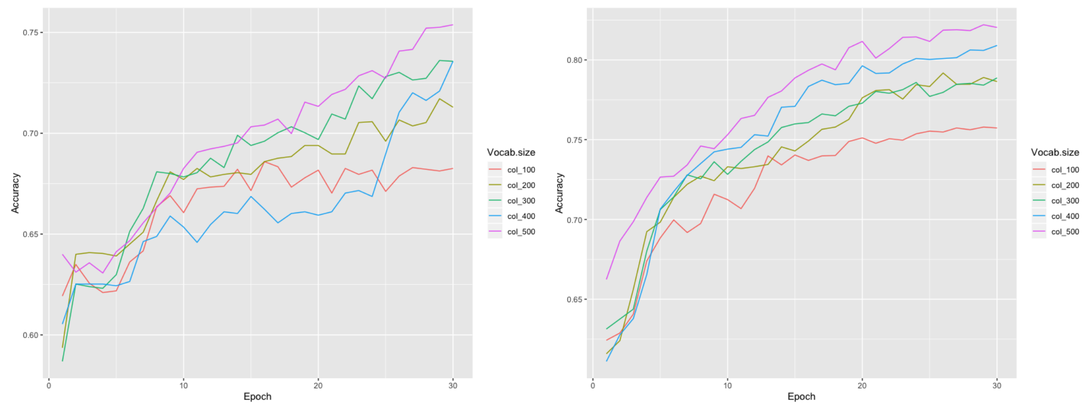


Figure 4: Complex DNN model architecture

6 DISCUSSION AND FUTURE WORK

In this paper, we presented a method for predicting the helpfulness of online hotel reviews using a range of text-based and non-textual features. We employed features that capture the textual content of the review as well as review metadata in order to automatically predict the helpfulness of reviews written both in Greek or English. We evaluated the performance of our approach in the review helpfulness classification task using ML and Neural classifiers and experimenting with different feature sets.

Results from ML classifiers, show that SVM performs better than RF and DT for both languages, with the best performance observed for English reviews, achieving promising F-Score and Accuracy values for review helpfulness classification. Text-based features have a bigger impact on the classification performance



(a) Greek

(b) English

Figure 5: Model training history with various vocabulary sizes

for English reviews, whereas non-textual features perform better for Greek reviews. In the majority of cases the effectiveness of combining both sets of features is apparent, suggesting that the word embeddings-based representation of the text vector captures accurately the semantic context of the text.

Experiments using Neural classifiers, confirm that features based on the textual components of reviews lead to better performance for English compared to Greek, while the numeric features we employed offer a considerable performance boost to Greek reviews yielding better results compared to combining both feature sets or using only textual features. The most effective feature selection for English reviews is the combination of text-based and numeric features. In the future, we plan to conduct further experiments with larger and more diverse datasets in order to further assess the effect application of different feature sets and classification algorithms have in the prediction of hotel reviews helpfulness.

ACKNOWLEDGMENTS

Research in this paper was funded by the Hellenic Government and the European Regional Development Fund under the NSRF 2014-2020 program (Filoxeno 2.0 project, T1EDK-00966)

REFERENCES

[1] [n.d.]. FastText: Pre-trained word vectors for 157 languages. https://fasttext.cc/docs/en/crawl-vectors

[2] [n.d.]. Php-ML: Machine Learning Library for PHP. https://github.com/jorgecasas/php-ml

[3] [n.d.]. TripAdvisor: Number of Reviews 2019. https://www.statista.com/statistics/684862/tripadvisor-number-of-reviews/

[4] Wasim Ahmad and Jin Sun. 2018. Modeling Consumer Distrust of Online Hotel Reviews. *International Journal of Hospitality Management* 71 (April 2018), 77–90. https://doi.org/10.1016/j.ijhm.2017.12.005

[5] Muhammad Bilal, Mohsen Marjani, Ibrahim Abaker Targio Hashem, Akibu Mahmoud Abdullahi, Muhammad Tayyab, and Abdullah Gani. 2019. Predicting Helpfulness of Crowd-Sourced Reviews: A Survey. In *2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)*. 1–8. https://doi.org/10.1109/MACS48846.2019.9024814

[6] Bin Fang, Qiang Ye, Deniz Kucukusta, and Rob Law. 2016. Analysis of the Perceived Value of Online Tourism Reviews: Influence of Readability and Reviewer Characteristics. *Tourism Management* 52 (Feb. 2016), 498–506. https://doi.org/10.1016/j.tourman.2015.07.018

[7] Hong Hong, Di Xu, G. Alan Wang, and Weiguo Fan. 2017. Understanding the Determinants of Online Review Helpfulness: A Meta-Analytic Investigation. *Decision Support Systems* 102 (Oct. 2017), 1–11. https://doi.org/10.1016/j.dss.2017.06.007

[8] Feng Hu. 2020. What Makes a Hotel Review Helpful? An Information Requirement Perspective. *Journal of Hospitality Marketing & Management* 29, 5 (July 2020), 571–591. https://doi.org/10.1080/19368623.2019.1661931

[9] Ya-Han Hu and Kuanchin Chen. 2016. Predicting Hotel Review Helpfulness: The Impact of Review Visibility, and Interaction between Hotel Stars and Review Ratings. *International Journal of Information Management* 36, 6, Part A (Dec. 2016), 929–944. https://doi.org/10.1016/j.ijinfomgt.2016.06.003

[10] Linchi Kwok and Karen L. Xie. 2016. Factors Contributing to the Helpfulness of Online Hotel Reviews: Does Manager Response Play a Role? *International Journal of Contemporary Hospitality Management* 28, 10 (Jan. 2016), 2156–2177. https://doi.org/10.1108/IJCHM-03-2015-0107

[11] Pei-Ju Lee, Ya-Han Hu, and Kuan-Ting Lu. 2018. Assessing the Helpfulness of Online Hotel Reviews: A Classification-Based Approach. *Telematics and Informatics* 35, 2 (May 2018), 436–445. https://doi.org/10.1016/j.tele.2018.01.001

[12] Sai Liang, Markus Schuckert, and Rob Law. 2019. How to Improve the Stated Helpfulness of Hotel Reviews? A Multilevel Approach. *International Journal of Contemporary Hospitality Management* 31, 2 (Jan. 2019), 953–977. https://doi.org/10.1108/IJCHM-02-2018-0134

[13] Zhiwei Liu and Sangwon Park. 2015. What Makes a Useful Online Review? Implication for Travel Product Websites. *Tourism Management* 47 (April 2015), 140–151. https://doi.org/10.1016/j.tourman.2014.09.020

[14] Yufeng Ma, Zheng Xiang, Qianzhou Du, and Weiguo Fan. 2018. Effects of User-Provided Photos on Hotel Review Helpfulness: An Analytical Approach with Deep Leaning. *International Journal of Hospitality Management* 71 (April 2018), 120–131. https://doi.org/10.1016/j.ijhm.2017.12.008

[15] Lionel Martin and Pearl Pu. 2014. Prediction of Helpful Reviews Using Emotions Extraction. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI'14)*. AAAI Press, Québec City, Québec, Canada, 1551–1557.

[16] Michael P. O'Mahony and Barry Smyth. 2009. Learning to Recommend Helpful Hotel Reviews. In *Proceedings of the Third ACM Conference on Recommender Systems (RecSys '09)*. Association for Computing Machinery, New York, NY, USA, 305–308. https://doi.org/10.1145/1639714.1639774

[17] Seunghun Shin, Qianzhou Du, and Zheng Xiang. 2019. What's Vs. How's in Online Hotel Reviews: Comparing Information Value of Content and Writing Style with Machine Learning. In *Information and Communication Technologies in Tourism 2019*, Juho Pesonen and Julia Neidhardt (Eds.). Springer International Publishing, Cham, 321–332. https://doi.org/10.1007/978-3-030-05940-8_25

[18] Konstantinos Tsamis, Andreas Komninos, and John Garofalakis. 2019. Leveraging Social Media Linguistic Features for Bilingual Microblog Sentiment Classification. In *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)*. 1–8. https://doi.org/10.1109/IISA.2019.8900674

[19] Zheng Xiang, Qianzhou Du, Yufeng Ma, and Weiguo Fan. 2017. A Comparative Analysis of Major Online Review Platforms: Implications for Social Media Analytics in Hospitality and Tourism. *Tourism Management* 58 (Feb. 2017), 51–65. https://doi.org/10.1016/j.tourman.2016.10.001

[20] Sung-Byung Yang, Seung-Hun Shin, Youhee Joun, and Chulmo Koo. 2017. Exploring the Comparative Importance of Online Hotel Reviews' Heuristic Attributes in Review Helpfulness: A Conjoint Analysis Approach. *Journal of Travel & Tourism Marketing* 34, 7 (Sept. 2017), 963–985. https://doi.org/10.1080/10548408.2016.1251872