Implementation and Repeatability Aspects Combined with Refactoring for a Reviews Manager System

Agorakis Bompotas[1], Aristidis Ilias[1], Andreas Kanavos[2], Panayiotis Kechagias[1], Panayiotis Arvanitakis[3], Nikos Zotos[3], Konstantinos Kovas[3] and Christos Makris[1]
[1]Computer Engineering and Informatics Department, University of Patras, Patras, Greece
[2]Department of Digital Media and Communication, Ionian University, Kefalonia, Greece
[3]Innovative Private Company, Patras, Greece

Keywords: Deep Learning, LSTM Neural Networks, Machine Learning, Natural Language Processing, Sentiment Analysis.

Abstract: With the advent of social media, there is a data abundance so that analytics can be reliably designed for ultimately providing valuable information towards a given product or service. Hotel customers express reviews for every accommodation service provided and/or for the accommodation as a whole. On the other hand, reviews are particularly interested for the tourism industry in order to extract customers' opinions and aspects, which will assist them to improve their provided services. In this paper, we delve into the detail of design and implementation of a system that initially utilizes some pre-processing techniques, as classic Natural Language Processing approaches, namely TF-IDF bag of words and word embeddings, are employed. These approaches can be further used as the input of various classifiers and Long Short Term Memory Neural Networks. The main aspects of this system have been described in (Bompotas et al., 2020a) and (Bompotas et al., 2020b). In the present article we essentially refactor the system that was described in and by embedding in the implementation the Latent Dirichlet Allocation (LDA) component and perform a repeatability study on the experimental findings that were reported in (Bompotas et al., 2020a) depicting that its experimental findings are valid.

1 INTRODUCTION

The inescapable utilization of social media websites has notably contributed to the prosper of the electronic word-of-mouth (eWOM) communication within the period of Web 2.0. Nowadays, due to the expanding readiness of customers to share and trade their individual encounters in social networking websites and platforms, online users' reviews, comments as well as reports have steadily picked up an outstanding interest from tourism businesses. The maintenance of a steady engagement with the customers with the goal of realizing and satisfying their requests consists a determining factor for businesses to sustain their competitive strength or advantage (Kumar and Pansari, 2016).

On the other hand, social networks websites and platforms have been regarded a vital factor in modifying the way of impacting the customers' engagement with tourism brands (Harrigan et al., 2017). As a result, businesses must create innovative marketing techniques and strategies oriented towards the customers' needs and fulfillment in order to adjust to the rapidly alternating environment. Thus, within the past few years, a surprising interest in recognizing and extracting valuable bits of knowledge on the customers' behavior together with sentiment by exploring online user-generated content, has been observed (He et al., 2016).

Online customer reviews are considered one of the foremost important sources of user-generated content with a vital affect on the tourism industry. This kind of reviews plays an important role within the decision-making procedure of the consumers as they provide them with the comfort to employ a comprehensive understanding of customers' past encounters. Specifically, these encounters can influence either in positive or negative way the future intentions and decisions of consumers. All things considered, the tremendous amount of user-generated content has emerged the problem of information overload. For confronting this issue, businesses must perceive inno-

Bompotas, A., Ilias, A., Kanavos, A., Kechagias, P., Arvanitakis, P., Zotos, N., Kovas, K. and Makris, C.
Implementation and Repeatability Aspects Combined with Refactoring for a Reviews Manager System.
DOI: 10.5220/0010727000003058
In *Proceedings of the 17th International Conference on Web Information Systems and Technologies (WEBIST 2021)*, pages 607-615
ISBN: 978-989-758-536-4; ISSN: 2184-3252

vative ways of successfully distinguishing and afterwards forecasting the value of online reviews (Gavilan et al., 2018).

Hence, a broad interest has been identified in the development of intelligent automated tools that will decrease the human resource prerequisites of businesses and will empower them to obtain fine-grained experiences on the customers' opinions and feelings (Zvarevashe and Olugbara, 2018). Subsequently, businesses will have a clear outline of the aspects that they should focus on and as a result, will be able to adjust to the customer requests by prioritizing and optimizing their marketing campaigns in time.

Generally, reviews often comprise of textual data together with a score rating mechanism, that unequivocally demonstrates the overall customer satisfaction. Despite the fact that the evaluation of customers have been proved to be highly correlated with the sentiment polarity of the particular textual content of the reviews (Geetha et al., 2017), there is still a solid interest in further checking and assessing the textual content under specific technical properties, which can impact customer ratings (Zhao et al., 2019). Thus, the reviews of the customers are considered a vital source of information for the tourism industry, as they empower businesses to have a crystal view of the foremost critical aspects inferring from them and hence, their marketing strategies can be better prioritized and optimized.

The development of advanced Natural Language Processing (NLP) techniques to effectively and productively extract profitable insights is caused because of the demanding need for identifying such underlying attributes and characteristics (Pablos et al., 2016). Especially, text and opinion mining systems have been proposed in the bibliography for analyzing and classifying online text reviews based on their sentiment content (Kasper and Vela, 2011; Sun et al., 2019). Additionally, since NLP consists a challenging and complex assignment, deep learning strategies have also been proposed in the bibliography with the aim of improving the granularity of aspect-based sentiment analysis procedures (Do et al., 2019). Consequently, it appears that the execution along with the accuracy of such NLP applications can be vital within the progression and vitality of tourism businesses and organizations.

Text summarization strategies have also been proposed in order to successfully distinguish the top-k most informative sentences of hotel reviews because of the huge volume of user-generated data (Hu et al., 2017). The present work in essence refactors the system that was described in (Bompotas et al., 2020a; Bompotas et al., 2020b) by embedding in the implementation the Latent Dirichlet Allocation (LDA) component and by performing a repeatibility study on the experimental findings that were reported in (Bompotas et al., 2020a). Moreover, it delves into the implementation details of a hotel review system that was described by the authors in (Bompotas et al., 2020a; Bompotas et al., 2020b).

In these articles a new approach was proposed for analyzing hotel reviews using Latent Dirichlet Allocation (LDA) for aspect mining and Neural Networks (NN) for sentiment analysis. A dynamic architecture, which receives the data stream, on-line or off-line in order not to overload the systems of the participating hotels or their service providers, is proposed. It extracts the aspects along with the sentiment of the hotel reviewers by applying LDA and NN modules accordingly, then stores the data and finally, attempts to correlate the data with the reviewers. The process is not obvious, given the anonymity of the reviewers, but the attempt to correlate them can be implemented with extensive training of the NN. The architecture proposes a novel platform utilizing the benefits of both algorithms, so that it can be used in an effective way in data forecasting. The design aspects of this system was presented in (Bompotas et al., 2020b) while experiments on the sentiments analysis module, without the presence of the LDA, were referred in (Bompotas et al., 2020a). In the present article, we prove that we are able to essentially reproduce the experimental findings of these above mentioned papers, and refactor the initial code by implementing and embedding in its workings an LDA component, depicting that our results can possibly be generalized to the space of topics.

The rest of this paper is organized as follows. Section 2 presents the related work. The main concepts of our system architecture, including the user interface, are covered in Section 3. The description of our dataset as well as the implementation details of the proposed sentiment analysis system and details concerning Long Short Term Memory Neural Networks (LSTMs) are presented in Section 4, while in Section 5, we present our experimental results on the performance of the classifiers based on a variety of metrics. Finally, in Section 6, we conclude the paper by outlining our findings and discussing future work.

2 RELATED WORK

Sentiment analysis procedure, which is also known as opinion extraction, possesses a pivotal role in the interpretation of natural languages and quantitative linguistics. Particularly, the investigation of sentiment is crucial to understanding user-generated text in social networks or product reviews and has attracted a lot of interest from both academia and industry (Pang and Lee, 2008). Academic literature has been given lots of attention to innovative strategies of handling valuable hotel data and to extrapolate important and relevant information that can be later utilized for sustainable economic development because of the increasing abundance of data from hotels across the world on a daily basis.

A summary of the rating monitoring strategies, where several hotel reviews have been assembled in order to address the viewpoints of guests as well as of the hotel, is provided in (Kasper and Vela, 2011). Around the same context, authors in (Hu and Liu, 2004) incorporate a more common and non-context specific strategy related to opinion mining, which is focused on feedback of consumers. Specifically, feedback and various opinions on person merchandise were analyzed and as a second step, a detailed percentage of polarity that reflects the thoughts of the consumers was derived. Altogether, both the evaluation and the overall impression of the growing number of hotel comments contribute to a positive intuition either by identifying challenges that management ought to settle or by empowering prospective customers to choose their next hotel (Liu and Zhang, 2012).

Social media consist of sites that accept a vast range of product and service feedback, which provides an immeasurable benefit over classic remarks under the company; more to the point, visual depiction and linkages between different values of feedback will have greater latent relations between opinion and rating. Authors in (Kanavos et al., 2017) illustrated the scalability of their methodologies, where large quantities of review data were analyzed utilizing distributed computing systems. Furthermore, a range of empirical research that focuses on the interpretation of substantive emotions across the lens of social networking has demonstrated in (Kanavos et al., 2018b).

Nevertheless, the strategy of assessing the emotional polarity of the feedback is not explicitly communicated within the raw data gathered. A variety of pre-processing layers was carried out so that this importance will be strongly focused within the research in (Haddi et al., 2013). Two fundamental layers exist until moving to the classification and performance assessment phases; data transformation and filtering. Data were initially prepared and unnecessary identifiers were removed, followed by stemming and lemmatization procedures. Amid the filtering phase, a statistical analysis was carried out with the aid of the Chi-square test in order to decide the association between the term and the group utilized in the phrase. All the metrics have been refined when considering the pre-processing method compared to completely skipping this stage in terms of the three essential evaluation metrics, namely Precision, Recall and F1-Measure, as it can be depicted in the performance evaluation phase.

Subsequently, the review management is naturally dependent on the essence of the comments implied to above and can be considered as nothing more than a set of texts. Therefore, the text mining concept, as a methodology to assist this phase, is considered to be a very important aspect (Blei, 2012; García et al., 2015). As a previous study on opinion clustering in comments, the configuration described in (Dave et al., 2003; Gourgaris et al., 2015), can be observed. Other recent studies related to consumer shopping patterns are described in (Domingos and Richardson, 2001; Iakovou et al., 2016; Kanavos et al., 2018a; Leskovec et al., 2007).

Target-dependent classification with respect to emotion is usually utilized in literature as a text classification problem. The majority of current studies develop emotion classifiers with a number of supervised approaches from machine learning, such as a feature-based Support Vector Machine (Jiang et al., 2011) or a Neural Network method (Dong et al., 2014; Vo and Zhang, 2015). Neural networks have delivered state-of-the-art efficiency in a number of Natural Language Processing activities, such as the automatic translation (Lample et al., 2016), the document summarization (Rush et al., 2015), the query addressing (He and Golub, 2016) and the paraphrase recognition (Yin et al., 2016). With respect to the recurrent layers of the schema, an explanatory evaluation and examination of different Recurrent Neural Networks (RNNs) such as Gated Recurrent Units (GRUs) and Large Short-Term Memory Units (LSTMs) is presented in (Chung et al., 2014). Within the same scope, LSTMs were also employed for sentiment classification in (Wang et al., 2012); however, the work at hand pertained to specific aspects and how they reflect particular sentiment.

The machine learning algorithms have the advantage of dealing with high dimensional and nonlinear relationships, which is especially suitable for establishing train dynamic model and train speed prediction on account of the dynamic and nonlinear nature (Savvopoulos et al., 2018). One of the most classic text mining techniques that composed the foundation for modern opinion mining is the Latent Dirichlet Allocation (LDA) (Blei et al., 2003; Griffiths, 2002; Griffiths and Steyvers, 2004). LDA is a probabilistic algorithm that can discover the latent topics that may exist within the reviews of the collection. More specifically, LDA extracts the top N topics that are most common in a review, based on the representations of the most frequent words with the input being a term document matrix, whereas two distributions are considered as output; one for document-topic relations and the other for topic-word ones.

In this article we try to validate the quality of the experimental results presented in (Bompotas et al., 2020a) and describe the implementation details of the design aspects of the system proposed in (Bompotas et al., 2020b). As described in Technology Networks[1], when measuring the quality of experiments, repeatability and reproducibility are key notions. Repeatability is ”a measure of the likelihood that, having produced one result from an experiment, you can try the same experiment, with the same setup, and produce that exact same result” while reproducibility is ”a measure of whether results in a paper can be attained by a different research team, using the same methods.” Our article is in essence a repeatability study on a refactored version of the system presented in (Bompotas et al., 2020a).

3 SYSTEM COMPONENTS

The system-starting point is the product ”BookOnCloud” which will offer to the customers-owners of its tourist accommodation various packages that will enable them to monitor the competition and their position in it at any time. The end result will be that this useful information will be displayed on the customer management screens.

Each customer of ”BookOnCloud” depending on the offer package he has purchased will receive the requested information during the period covered by the package he has purchased. This time period is translated into cron expressions and stored in a table of a postgres database along with the unique id of the client and the id of the target. In this way a complete customer request has been made.

In Figure 1, the main architecture of the scraping system is illustrated. The system is mainly developed in Node.js and uses various components. Apache Kafka has been used as a distributed event streaming platform, to handle the requests for scraping pages. The implemented Kafka instance uses three Topics. The Scrap Topic contains the requests for scraping pages. The Parse Topic contains the parsing results from scraping the pages. The Error Topic contains the error events produced during scraping and parsing.

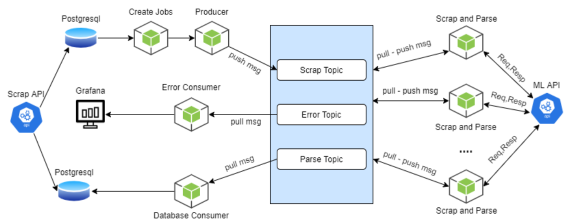

Figure 1: Architecture of the Scraping System.

Grafana has been used as a visualization platform to be able to visualize and query upon the error events that may occur. PostgreSQL is used as a database for storing both the configuration of each hotel and the data retrieved by scraping the review pages.

The ScrapAPI has endpoints that allow users to configure the scraping sources (the URLs from where Reviews are going to periodically be retrieved). These URL sources are stored in a database table along with configuration options (e.g. how often the page should be scrapped). A node service ”Create Jobs” constantly checks the URL sources and based on the configuration options, decides when and what page needs to be scrapped by creating new entries (scrap requests) in another database table. The node service ”Producer” constantly checks for scrap requests in the database and puts them in the Scrap Topic of the Kafka component. There are multiple ”Scrap and Parse” services running concurrently and the scrap requests in the Scrap Topic are distributed among them. Each ”Scrap and Parse” service pulls the request from the Scrap Topic and then pushes results to the Parse Topic (if successful) or the Error Topic (if an error occurred). The ”Error Consumer” service pulls from the Error Topic and feeds the data in the Grafana API, that allows monitoring the error cases through useful graphs. The ”Database Consumer” service pulls from the Parse Topic, manipulates the data and stores the results (scrapped data) in the database. Finally, the ScrapAPI is used to retrieve the scrapped data from the database.

Here is another subsystem, which is implemented with Apache Kafka. A node.js script undertakes to read the table from the database and execute a scheduled process that will run every now and then so that this is equal to the time specified by the cron expression for each client request. This process essentially activates an Apache Kafka producer, who places these requests in a queue. These applications are the first topic of Apache Kafka.

This topic is ”consumed” by a group of consumers. We use groups to take advantage of more consumers who will read the messages of the topic

[1] https://www.technologynetworks.com/informatics/articles/repeatability-vs-reproducibility-317157

as the queue grows. In this way the system becomes more efficient and faster since the processes are performed in parallel on the separate servers offered by Apache Kafka. This is why Apache Kafka was preferred because it can efficiently manage queues and transfer real-time data from sender to recipient.

An example application that uses the Scrap API to configure the scraping requests and retrieve and display the results. Its architecture is depicted in Figure 2. The application has been developed in Node.js and React. Users have to log in to the Extranet application. The React application receives a JSON Web Token (JWT) that uses throughout its communication with the Scrap API. The implemented React application is loaded in a single web page and has three separate tabs. The ”Information” tab, is where the user can view, add, and edit the URLs from which he wants to retrieve reviews. The ”Review” tab is where the user can view and filter the retrieved reviews across all the requested URLs. Apart from the information retrieved, additional information is displayed from the Machine Learning Analysis performed on the review text. This analysis can identify if the overall review is positive, negative, or neutral. It may also identify polarity regarding specific hotel aspects (e.g. cleanliness, amenities, price). Finally, the ”Global Score” tab displays information retrieved from the scraped URLs, regarding the score of the hotel, based on the ratings from the reviewers.

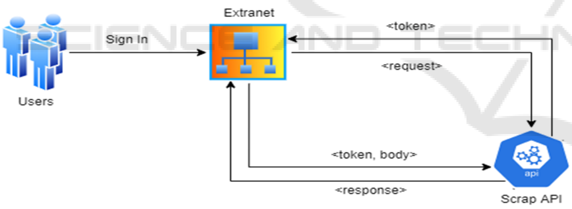

Figure 2: Architecture of the Example Application.

3.1 Review Manager User Interface

Concerning the component of the system architecture that will perform sentiment analysis, it is depicted in Figure 3. It consists of an Application Programming Interface (API) acting as gateway to an online hotel booking platform, a NoSQL database and the Sentiment Analysis Infrastructure.

Hotel reviews are inserted in the database through the corresponding API, the Natural Language Processing module initially parses the stored reviews, transforms them into the appropriate form and then passes them to the Aspect Mining and Sentiment Analysis modules, which produces the final outputs and stores them back to the database. Both the initial reviews and the results of the analysis are easily accessible through the API.

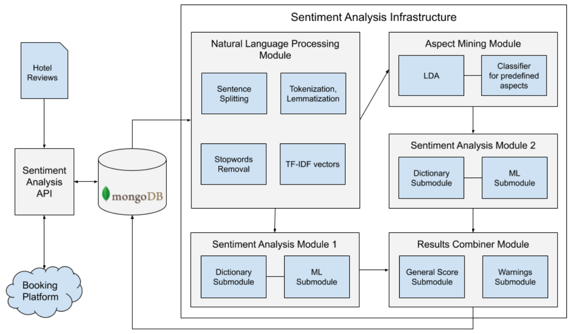

Figure 3: Hotel Reviews Sentiment Analysis Platform Architecture.

Table 1: Positive vs Negative reviews.

Sentiment	Number	Percentage
Positive	352.029	50%
Negative	352.029	50%
Total	704.058	100%

4 IMPLEMENTATION

In order to evaluate the system proposed we mainly focused on the sentiment analysis component accompanied with the aspect component. For the sentiment analysis component we performed a set of experiments that in essence reproduce the outcomes of the experiments of a previous publication of ours, while we performed also a set of experiments for the aspect component. Note that since there are common authors in the previous publication and the present study what we do is mainly a repeatability study.

4.1 Dataset

Since the paper is a repeatibility/reproducibility study, we have to use the same experimental setup with the previous work, but with different parameters concerning the input data used. In particular we use the same experimental setup (input data and partitioning procedure), however we split in a different way our data.

Hence the dataset consists of $515,000$ hotel reviews in Europe[2] was taken into consideration during the training and evaluation processes. The dataset contain almost one million rows including both a positive and a negative review along with additional

[2]https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe

metadata about the hotel, the reviewer, and the reviews themselves. In our case we mainly care or the plain text, and not for the extra information, and since the meta-information has almost no value for our proposed schema and was removed during the preprocessing stage. Additionally the dataset was post-processed in order to contain one labeled review per row, either as positive or negative and empty reviews were removed. The final dataset consisted of $704,058$ reviews, divided exactly in half with $352,029$ positive and $352,029$ negative reviews, and it is depicted in Table 1.

4.2 Sentiment Analysis Infrastructure

4.2.1 Sentiment Analysis Module by Utilizing LSTM Neural Networks

As described in (Bompotas et al., 2020a), we have tested various classifiers: Support Vector Machines, Random Forest, Logistic Regression, Ridge Classifier, Multilayer Perceptron, Passive Aggressive, AdaBoost, Gradient Boosting, Perceptron, Decision Tree, Nearest Centroid and k-Nearest Neighbors and an LSTM Neural Network. In all classifiers except LSTM we employed a classic TF-IDF (Term Frequency - Inverted Document Frequency) bag of words model, while for the LSTM Neural Network a word embedding model is employed.

Furthermore, as described in (Bompotas et al., 2020b), LSTMs' architecture is based on "*cell state*" and "*gates*" through which the input information is propagated. More accurately, there are three gates and two states in LSTMs: the *forget gate* (f_t) whose responsibility is to remove unnecessary information from the *cell state* taking as input the hidden state of the previous cell h_{t-1} and data record x_t. Next, the *input gate* (i_t) adds new information on the cell state by creating a vector of all possible values and multiplying them with the *tanh* function. Hence the flow of information starts by feeding word sequences to the embedding layer of the neural network. Since longer embeddings increase the complexity and reduce the accuracy of the sentiment analysis we have chosen the output length of the embedding layer was decided to be only 256.

The output of the embedding layer is then forced through a dropout layer with a drop probability equal to 0.2. The third step of our process consists of a layer of 256 LSTM units that correspond to the embedding layer's mappings and have a recurrent dropout equal to 0.2. The information after passing through a dropout layer, is processed by a fully connected layer of 256 Rectified Linear Units (ReLUs). Then the output of the layer is filtered by a dropout layer before it reaches the softmax classification layer, where the review's sentiment is decided.

4.2.2 Aspect Mining Module

Although extracting the overall sentiment of a review is a meaningful data mining task, a system that informs hoteliers that there is room for improvement would not be complete without revealing them what exactly left their customers dissatisfied. To address this, the proposed Review Manager System provides augmented functionality by deciding the polarity of the discrete aspects contained in each review.

To achieve this goal, during the preprocessing stage our system splits each review into sentences and the tokenized output is then parsed by the Aspect Mining Module which employs the powerful Latent Dirichlet Allocation method to cluster every sentence into a predefined number of topics. These topics are characterized by a set of words that are most likely to occur in the documents assigned to them and thus they have a semantic meaning that is useful to our analysis. Subsequently, a sentiment score is assigned to each sentence by the trained LTSM neural network of the Sentiment Analysis Module described above and these scores are aggregated to produce the final report of each review which consists of the list of the discovered aspects and their respective polarity additionally to the review's overall sentiment.

However, because LDA is an unsupervised technique, the produced results are arbitrary and may not align with the actual topics that customers consider important when reading or composing a hotel review. To further improve our system, the domain experts of our team concluded on the list of aspects showing in Table 2 as the ones that our system should be searching for. For converging to a set of predefined aspects the LDA had to be modified to create a semi-supervised method. Out of the techniques tested, the approach of SeededLDA (Jagarlamudi et al., 2012) was the one that stood out and suited our needs the best. SeededLDA is a variation of the original algorithm where prior to its execution some words can be influenced via an input weight or seed to lean towards a specific topic. As a consequence, by carefully selecting and boosting a list of words that are related to the predefined aspects the execution can be guided to produce the desired topics.

Table 2: Predefined Aspects.

Staff / Service
Comfort
Facilities / Amenities
Value for money
Cleanliness
Location

5 EVALUATION

In our work we mainly refactor the implementation of (Bompotas et al., 2020a) by restructuring carefully the various components of our code and by embedding the LDA machinery in it.

The new experiments ran on the same machinery, namely a Dell Precision 7520 mobile workstation with an Intel i7-6820HQ processor, 32GB of RAM and an NVIDIA Quadro M1200 graphics card with characteristics such as 4GB of dedicated memory, 640 CUDA cores and computation capability equal to 5.0 that enabled us to reduce training time for the LSTM Neural Network. The machine's operating system was Windows 10, and the implementation of the algorithms were developed in Python 3.7 with TensorFlow 2.2.0 and CUDA 10.1.

The dataset employed was the same as in the previous work however we have chosen a different separation of the datasets to 10 splits (in comparison to the previous work) in order to see if the attained results agree with that of the previous article. During the execution of each algorithm we again tried to determine the parameters, that could optimize the performance and then we split the datasets into training and test sets with a ratio of 75% to 25%.

The following Table 3 summarizes the result of the experiments for the set of the methods:

As was the case from our previous experiments and it is evident from Table 3 the LSTM Neural Network outperforms all the other algorithms that were evaluated against by a large margin in every metric score. In addition the LSTM seems to achieve nice performance even when considering the various aspects. Furthermore, the LSTM Neural Network proved to perform equally well for every metric and for both classes (positive and negative).

In order to validate the statistical significance of our results we employed t-test and we computed the p values for the null hypothesis testing of LSTM in comparison with the other algorithmic schemes. The desired value should be less than 5% and as we see in Table 4 this is achieved in the majority of the results.

In the Table of p values the values are depicted with an accuracy of 10 decimal points, and that is why a lot of cells are with 0 values.

Moreover, to fine tune and later evaluate the quality of the Aspect Mining Module, a series of tests were conducted using real data. The lack of a big dataset annotated with the topics provided by our domain experts meant that we had to construct our own. This is an ongoing task but we were able to test our model with a smaller dataset of approximately 70 records that was ready during the writing of this and the results were quite promising as the system was able to identify the correct aspects of each review and detect their polarity with an accuracy that matched our previous results. Further evaluation is needed and is left as future work.

6 CONCLUSIONS AND FUTURE WORK

In the present article we delved into the detail of design and implementation of a system that initially utilizes some pre-processing techniques, as classic Natural Language Processing approaches, namely TF-IDF, bag of words and word embeddings, in order to be used as the input of various classifiers and Long Short Term Memory Neural Networks, for testing the sentiment output of particular hotel reviews. A dynamic architecture, which receives the data stream in order not to overload the systems of the participating hotels or their service providers, is proposed.

The main aspects of this system have been described in (Bompotas et al., 2020a) and (Bompotas et al., 2020b). In the present article we essentially refactor the system that was described in these works and by embedding in the implementation the Latent Dirichlet Allocation (LDA) component, we perform a repeatability study on the experimental findings that were reported in (Bompotas et al., 2020a) depicting that its experimental findings remain the same. The outcome of the experiments verify the findings presented in (Bompotas et al., 2020b), while the embedding of the LDA component seems to work without problem providing to the expert another source of information.

For future work, it would be interesting to apply our methodology to a much larger sample of data. In addition, it is necessary to study the total execution times in order to magnify our methodology. Furthermore, another potential approach could be implemented concerning the complexity of the architecture. Specifically, as a model deepens in terms of layers as well as in the size of its graph, new ways for defining

Table 3: Aggregate Experimental evaluation.

Method	Precision		Recall		f1-score		Accuracy
	Negative	*Positive*	*Negative*	*Positive*	*Negative*	*Positive*	
AdaBoost	0.86	0.89	0.89	0.85	0.87	0.87	0.87
Decision Trees	0.84	0.85	0.86	0.84	0.85	0.85	0.85
Gradient Boosting	0.84	0.91	0.91	0.82	0.87	0.86	0.87
K-Nearest Neighbor (KNN)	0.65	0.80	0.87	0.52	0.74	0.63	0.70
Logistic Regression	0.88	0.93	0.93	0.88	0.91	0.90	0.90
Long Short Term Memory (LSTM)	0.92	0.93	0.94	0.91	0.93	0.92	0.93
Multilayer Perceptron	0.87	0.87	0.87	0.87	0.87	0.87	0.87
Nearest Centroid	0.78	0.94	0.96	0.73	0.86	0.83	0.85
Passive Aggressive	0.87	0.87	0.87	0.87	0.87	0.87	0.87
Perceptron	0.85	0.86	0.86	0.85	0.85	0.85	0.85
Random Forest	0.89	0.92	0.92	0.88	0.91	0.90	0.90
Ridge	0.88	0.92	0.92	0.87	0.90	0.89	0.90
Support Vector Machines (SVM)	0.89	0.94	0.94	0.88	0.91	0.91	0.91

Table 4: Experimental Evaluation for p.

Method	P-Value
AdaBoost	0.0000000000
Decision Trees	0.0000000000
Gradient Boosting	0.0000000000
K-Nearest Neighbor (KNN)	0.0000000000
Logistic Regression	0.0000078402
Multilayer Perceptron	0.0000000000
Nearest Centroid	0.0000000000
Passive Aggressive	0.0000000000
Perceptron	0.0000000000
Random Forest	1.5304732684
Ridge	0.0000012907
Support Vector Machines (SVM)	0.0074408588

the optimal connection within this stack of layers can be emerged.

ACKNOWLEDGEMENT

This work has been co-financed by the European Union and Greek national funds through the Regional Operational Program “Western Greece 2014-2020”, under the Call “Regional Research and Innovation Strategies for Smart Specialisation - RIS3 in Information and Communication Technologies” (project: 5038701 entitled “Reviews Manager: Hotel Reviews Intelligent Impact Assessment Platform”).

REFERENCES

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4):77–84.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Bompotas, A., Ilias, A., Adamopoulos, M., Kanavos, A., Makris, C., Rompolas, G., and Savvopoulos, A. (2020a). A sentiment-based hotel review summarization using LSTM neural networks. In *11th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pages 1–7.

Bompotas, A., Ilias, A., Kanavos, A., Makris, C., Rompolas, G., and Savvopoulos, A. (2020b). A sentiment-based hotel review summarization using machine learning techniques. In *16th International Conference on Artificial Intelligence Applications and Innovations (AIAI)*, volume 585, pages 155–164.

Chung, J., Gülçehre, Ç., Cho, K., and Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *CoRR*, abs/1412.3555.

Dave, K., Lawrence, S., and Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *12th International World Wide Web Conference (WWW)*, pages 519–528.

Do, H. H., Prasad, P. W. C., Maag, A., and Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: A comparative review. *Expert Systems with Applications*, 118:272–299.

Domingos, P. M. and Richardson, M. (2001). Mining the network value of customers. In *7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 57–66.

Dong, L., Wei, F., Tan, C., Tang, D., Zhou, M., and Xu, K. (2014). Adaptive recursive neural network for target-dependent twitter sentiment classification. In *52nd*

Annual Meeting of the Association for Computational Linguistics (ACL), pages 49–54.

García, S., Luengo, J., and Herrera, F. (2015). *Data Preprocessing in Data Mining*, volume 72 of *Intelligent Systems Reference Library*. Springer.

Gavilan, D., Avello, M., and Martinez-Navarro, G. (2018). The influence of online ratings and reviews on hotel booking consideration. *Tourism Management*, 66:53–61.

Geetha, M., Singha, P., and Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels - an empirical analysis. *Tourism Management*, 61:43–54.

Gourgaris, P., Kanavos, A., Makris, C., and Perrakis, G. (2015). Review-based entity-ranking refinement. In *11th International Conference on Web Information Systems and Technologies (WEBIST)*, pages 402–410.

Griffiths, T. L. (2002). Gibbs sampling in the generative model of latent dirichlet allocation.

Griffiths, T. L. and Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5228–5235.

Haddi, E., Liu, X., and Shi, Y. (2013). The role of text preprocessing in sentiment analysis. In *1st International Conference on Information Technology and Quantitative Management (ITQM)*, pages 26–32.

Harrigan, P., Evers, U., Miles, M., and Daly, T. (2017). Customer engagement with tourism social media brands. *Tourism Management*, 59:597–609.

He, W., Tian, X., Chen, Y., and Chong, D. (2016). Actionable social media competitive analytics for understanding customer experiences. *Journal of Computer Information Systems*, 56(2):145–155.

He, X. and Golub, D. (2016). Character-level question answering with attention. pages 1598–1607.

Hu, M. and Liu, B. (2004). Mining opinion features in customer reviews. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI)*, pages 755–760.

Hu, Y., Chen, Y., and Chou, H. (2017). Opinion mining from online hotel reviews - A text summarization approach. *Information Processing and Management*, 53(2):436–449.

Iakovou, S. A., Kanavos, A., and Tsakalidis, A. K. (2016). Customer behaviour analysis for recommendation of supermarket ware. In *12th IFIP International Conference and Workshops (AIAI)*, pages 471–480.

Jagarlamudi, J., Daumé III, H., and Udupa, R. (2012). Incorporating lexical priors into topic models. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 204–213, Avignon, France. Association for Computational Linguistics.

Jiang, L., Yu, M., Zhou, M., Liu, X., and Zhao, T. (2011). Target-dependent twitter sentiment classification. In *49th Annual Meeting of the Association for Computational Linguistics*, pages 151–160.

Kanavos, A., Iakovou, S. A., Sioutas, S., and Tampakas, V. (2018a). Large scale product recommendation of supermarket ware based on customer behaviour analysis. *Big Data and Cognitive Computing*, 2(2).

Kanavos, A., Nodarakis, N., Sioutas, S., Tsakalidis, A., Tsolis, D., and Tzimas, G. (2017). Large scale implementations for twitter sentiment classification. *Algorithms*, 10(1):33.

Kanavos, A., Perikos, I., Hatzilygeroudis, I., and Tsakalidis, A. (2018b). Emotional community detection in social networks. *Computers & Electrical Engineering*, 65:449–460.

Kasper, W. and Vela, M. (2011). Sentiment analysis for hotel reviews. In *Computational Linguistics-Applications Conference*, pages 45–52.

Kumar, V. and Pansari, A. (2016). Competitive advantage through engagement. *Journal of Marketing Research*, 53(4):497–514.

Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., and Dyer, C. (2016). Neural architectures for named entity recognition. pages 260–270.

Leskovec, J., Adamic, L. A., and Huberman, B. A. (2007). The dynamics of viral marketing. *ACM Transactions on the Web*, 1(1):5.

Liu, B. and Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In *Mining Text Data*, pages 415–463.

Pablos, A. G., Cuadros, M., and Linaza, M. T. (2016). Automatic analysis of textual hotel reviews. *Journal of Information Technology & Tourism*, 16(1):45–69.

Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.

Rush, A. M., Chopra, S., and Weston, J. (2015). A neural attention model for abstractive sentence summarization. pages 379–389.

Savvopoulos, A., Kanavos, A., Mylonas, P., and Sioutas, S. (2018). LSTM accelerator for convolutional object identification. *Algorithms*, 11(10):157.

Sun, Q., Niu, J., Yao, Z., and Yan, H. (2019). Exploring ewom in online customer reviews: Sentiment analysis at a fine-grained level. *Engineering Applications of Artificial Intelligence*, 81:68–78.

Vo, D. and Zhang, Y. (2015). Target-dependent twitter sentiment classification with rich automatic features. In *24th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1347–1353.

Wang, H., Can, D., Kazemzadeh, A., Bar, F., and Narayanan, S. (2012). A system for real-time twitter sentiment analysis of 2012 U.S. presidential election cycle. In *Annual Meeting of the Association for Computational Linguistics*, pages 115–120.

Yin, W., Schütze, H., Xiang, B., and Zhou, B. (2016). ABCNN: attention-based convolutional neural network for modeling sentence pairs. *Transactions of the Association for Computational Linguistics*, 4:259–272.

Zhao, Y., Xu, X., and Wang, M. (2019). Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *International Journal of Hospitality Management*, 76:111–121.

Zvarevashe, K. and Olugbara, O. O. (2018). A framework for sentiment analysis with opinion mining of hotel reviews. In *Conference on Information Communications Technology and Society (ICTAS)*, pages 1–4.